

# Remote Sensing & GIS Applications



**Dr. K N Tiwari**

**Dr. C Chatterjee**

**Prof. N K Gontia**

**Dr. Susanta Kumar Jena**



## Remote Sensing and GIS Application

-: Course Content Developed By :-

**Dr. K N Tiwari**

Professor

Dept. of Agricultural and Food Engg., IIT, Kharagpur

**Dr. C Chatterjee**

Associate Professor

Dept. of Agricultural and Food Engg., IIT, Kharagpur

-:Content Reviewed by :-

**Prof. N K Gontia**

Dean

College of Agricultural Engineering and Technology, Junagadh Agricultural University,  
Junagadh

**Dr. Susanta Kumar Jena**

Principal Scientist (SWCE)

Directorate of Water Management (ICAR)

---



**AGRIMOON.COM**

**All About Agriculture...**

# INDEX

Lesson Name	Page No
<b>Module 1: Introduction and Concepts of Remote Sensing</b>	
Lesson 1 Introduction to Remote Sensing	5-14
Lesson 2 Energy Sources and Interaction	15-25
Lesson 3 Remote Sensing Applications	26-32
<b>Module 2: Sensors, Platforms and Tracking System</b>	
Lesson 4 Sensors Characteristics	33-37
Lesson 5 Sensors and Tracking System	38-43
Lesson 6 Remote Sensing Platforms	44-47
<b>Module 3: Fundamentals of Aerial Photography</b>	
Lesson 7 Basic Principles of Photogrammetry	48-56
Lesson 8 Stereoscopy and Photogrammetry	57-61
Lesson 9 Image Interpretation	62-67
<b>Module 4: Digital Image Processing</b>	
Lesson 10 Digital Image	68-77
Lesson 11 Pre Processing Techniques	78-90
Lesson 12 Image Classification	91-95
Lesson 13 Classification Accuracy Assessment	96-101
<b>Module 5: Microwave and Radar System</b>	
Lesson 14 Microwave Remote Sensing	102-107
Lesson 15 Radar Systems	108-115
<b>Module 6: Geographic Information Systems (GIS)</b>	
Lesson 16 Introduction to GIS	116-131
Lesson 17 Applications of GIS	132-153
<b>Module 7: Data Models and Structures</b>	
Lesson 18 Data Models	154-163
Lesson 19 Data Model Classification	164-172
Lesson 20 Data Structures	173-186
<b>Module 8: Map Projections and Datum</b>	
Lesson 21 Map Projections	187-197
Lesson 22 Coordinate and Datum System	198-207
Lesson 23 Projection System	208-220

<b>Module 9: Operations on Spatial Data</b>	
Lesson 24 Map Algebra	221-235
Lesson 25 Spatial Data	236-244
Lesson 26 Vector and Raster Map Algebra	245-260
<b>Module 10: Fundamentals of Global Positioning System</b>	
Lesson 27 GPS	261-274
Lesson- 28 GPS Application	275-295
<b>Module 11: Applications of Remote Sensing for Earth Resources Management</b>	
Lesson 29 Agriculture and Forestry	296-303
Lesson 30 Land Use and Land Cover Mapping and Change Detection	304-308
Lesson 31 Soil Mapping and Soil Erosion Estimation	309-316
Lesson 32 Water Resources and Flood Damage Assessment	317-322



## Module 1: Introduction and Concepts of Remote Sensing

### Lesson 1 Introduction to Remote Sensing

#### 1.1 Definition and Concepts of Remote Sensing

Remote sensing is the science and art of acquiring information (spectral, spatial, and temporal) about material objects, area, or phenomenon, without coming into physical contact with the objects, or phenomenon under investigation. Remote Sensing means sensing of the earth's surface from space by making use of the properties of electromagnetic wave emitted, reflected, or diffracted by the sensed objects for the purpose of improving natural resource management land and the protection of the environment. Without direct contact, some means of transferring information through space must be utilized. In remote sensing information transfer is accomplished by use of electromagnetic radiation (EMR).

Remote sensing in the broad sense, the measurement or acquisition of information of some property of an object or phenomenon, by a recording device that is not in physical or intimate contact with the object or phenomenon under study; e.g., the utilization at a distance (as from aircraft, spacecraft, or ship) of any device and its attendant display for gathering information pertinent to the environment, such as measurements of force fields, electromagnetic radiation, or acoustic energy. The technique employs such devices as the camera, lasers, and radio frequency receivers, radar systems, sonar, seismographs, gravimeters, magnetometers, and scintillation counters.

Remote sensing is a technique to observe the earth surface or the atmosphere using airborne or space borne platforms. It uses several parts of the electromagnetic spectrum. It records the electromagnetic energy reflected or emitted by the earth's surface.

#### Concept of Remote Sensing

Normally, if one comes across the term remote sensing, one wonders what does it mean? 'remote' means far away, and sensing means believing or observing or acquiring some information. Remote sensing means acquiring information of things from distance.

We use remote sensing when we use sense of sight for watching a cricket match from in stadium, sense smell freshly cooked curry, and sense of hearing from a telephone ring.

In the world of geospatial science, remote sensing, also known as the earth observation, means observing the earth with sensors from high above its surface. Sensors are like simple cameras except that they not only use visible light but also other bands of the electromagnetic spectrum such as infrared, microwaves, and ultraviolet regions. They are so high up that can make images of a very large area. Nowadays, remote sensing is mainly done from space using satellites, (*Bhatta, 2008*).

## 1.2 History of Remote Sensing and GIS

Remote sensing as a technology started with the first photographs in the early nineteenth century. Remote sensing began in 1858 when Gaspard-Felix Tournachon first took aerial photographs of Paris from a hot air balloon. Remote sensing continued to grow from there; one of the first planned uses of remote sensing occurred during the U.S. Civil War when messenger pigeons, kites, and unmanned balloons were flown over enemy territory with cameras attached to them.

The first governmental-organized air photography missions were developed for military surveillance during World Wars I and II but reached a climax during the Cold War. Cameras mounted on airplanes, or more commonly held by aviators, provided aerial views of fairly large surface areas that were invaluable for military reconnaissance. From then until the early 1960s, the aerial photograph remained the single standard tool for depicting the surface from a vertical or oblique perspective.

Aerial photographs were used in India as early as in 1920 for land survey. With the advent of space technology, images taken from space opened new vistas in their use for survey and monitoring natural resources. The whole technology was termed remote sensing. Remote sensing program in India took roots in 1970 with the conduct of the famous Coconut-wilt disease detection experiment. From then onwards, a systematic evolution of the RS program has resulted in the established of an integrated end-to-end system - having space, ground and application segments.

With active participation and financial support from both Central and State Governments, the technology, over the last few decades, has matured to cover diverse resources themes/ areas such as forestry, wasteland mapping, agricultural crop acreage and yield estimation, flood monitoring and damage assessment, landuse/land cover mapping, water resources management, ground water targeting, marine resources survey, urban planning, mineral targeting and environmental impact assessment etc. While, the remote sensing data integrated with Geographic Information System (GIS) is operationally being utilized for mapping various resources, now it is realized that the need is to step-ahead from mapping towards integrating these resources maps with other resource information and socio-economic data and provide a path for sustainable development. The ability of GIS to manage geospatial data establishes GIS as an important tool for a wide variety of applications. Since the 1970s GIS has been important in the management of natural resources. GIS provides an exceptional means for integrating timely remote sensing data with other spatial and thematic data types.

The rapid progress, and increased visibility, of remote sensing and GIS since the 1990s has been made possible by a paradigm shift in computer technology, computer science, and software engineering, as well as airborne and space observation technologies. As a result a new field of the study named geomatics engineering or geospatial technology or geoinformatics technology is now in its maturity. The term 'geomatics' is fairly young and is commonly used to define the tools and techniques used in land surveying, remote sensing, GIS, global navigation satellite systems (GNSS), and related forms of the earth mapping. Due to multidisciplinary applications and integration with other scientific and technological fields, in the recent years remote sensing and GIS has become a distinct field of study.

## Indian Space Program

Though ancient Indians were known to have knowledge about rocket science- it being used during wars- it was only after independence that the process of exploring space really accelerated. It was Dr. Vikram, who founded the Physical Research Laboratory (PRL) at Ahmedabad, Gujarat on November 11, 1947. This was the first step that India took towards becoming a space power. Our first biggest success was on April 19, 1975, when India launched its first satellite into space. It was launched by the Soviet Union from Kapustin Yar using a Cosmos-3M launch vehicle.

During the formative decade of 1960s, space research was conducted by India mainly with the help of sounding rockets. The Indian Space Research Organization (ISRO) was formed in 1969. Since India set its sights on space research and exploration in the 1970's it's had a number of successful missions including two satellite programs. The first satellite program, called INSAT, was to take care of the telecommunications needs of India. The second satellite program, called Indian Remote Sensing Satellites, is for natural resource management and economic planning. Space research activities were provided additional fillip with the formation of the Space Commission and the Department of Space by the Government of India in 1972. In the history of the Indian space program, 70s were the era of experimentation during which experimental satellite programs like Aryabhata, Bhaskara, Rohini and Apple were conducted. The success of these programs, led to era of operationalisation in 80s during which operational satellite programs like INSAT and IRS came into being. Today, INSAT and IRS are the major programs of ISRO.


Also, India has developed various launch vehicles that make a space program independent and are the most important technological measure of its advancement. Prominent among them are Satellite Launch Vehicle (SLV), Augmented Satellite Launch Vehicle (ASLV), Polar Satellite Launch Vehicle (PSLV) and Geosynchronous Satellite Launch Vehicle (GSLV).

## Indian Remote Sensing Satellite (IRS) System

Indian Remote Sensing (IRS) satellite system was commissioned with the launch of IRS-1A, in 1988. With eleven satellites in operation, IRS is the largest civilian remote sensing satellite constellation in the world providing imageries in a variety of spatial resolutions, spectral bands and swaths. The data is used for several applications covering agriculture, water resources, urban development, mineral prospecting, environment, forestry, drought and flood forecasting, ocean resources and disaster management. Table 1.1 provides brief description of IRS satellites launched by ISRO.

**Table 1.1 IRS satellites series**

Satellite	Launched Date	Payloads	Purpose
IRS-1A	17.03.1988	LISS-I (Resolution (R): 72.5 m) LISS-IIA and LISS-IIB (R: 36.25 m)	First operational remote sensing satellite for large scale mapping.

SROSS-2	13.07.1988		Failed
IRS-1B	29.08.1991	LISS-I (R: 72.5 m), LISS-2A and LISS-2B	Same as IRS-1A
IRS-1E	20.09.1993		Failed
IRS-P2	15.10.1994	LISS-IIA & LISS-IIB (R: 32.74 m)	Same as IRS-1B
IRS-1C	28.12.1995	PAN (R: <6 m ) LISS-III(R: 23.6 m) WideField Sensor (WiFS), (R: 189 m)	PAN data is of finer spatial resolution used to sharpen or increase the resolution of coarser resolution imagery. It is also used in various geological, biological, and engineering surveys and mapping. LISS-III had improved spatial resolution. Land and vegetation observation. WiFS data used for very large scale mapping; mainly used for ocean monitoring.
IRS-P3	21.03.1996	WiFS, Modular Opto - electronic Scanner (MOS), Indian X-ray Astronomy Experiment (IXAE), C-band transponder (CBT)	Remote sensing of earth's natural resources. Study of X-ray Astronomy. Periodic calibration of PSLV tracking radar located at tracking stations.
IRS-1D	27.09.1997	PAN (R: 6 m) LISS-III (R: 23.6 m) WiFS (R: 189 m)	For continuation of IRS-1C.
IRS-P4 / Oceansat	26.05.1999	Ocean Colour Monitor (OCM) Multi - frequency Scanning Microwave Radiometer (MSMR)	To study surface winds and ocean surface strata, observation of chlorophyll concentrations, monitoring of phytoplankton blooms, study of atmospheric aerosols and suspended sediments in the water.
TES	26.05.1999	PAN	TES is an experimental satellite to demonstrate and validate the technologies like attitude and orbit control system, high-torque reaction wheels, new reaction control system etc.
IRS P6 / Resourcesat-1	17.10.2003	LISS-IV (R: 5.8 m.), LISS-III (R: 23.5 m), AWiFS-A and AWiFS-B	Monitoring of vegetation dynamics, crop yield estimates, disaster management support etc.
CARTOSAT-1	05.05.2005	PAN FORE, PAN - AFT	The first IRS Satellite capable of providing in-orbit stereo images. Used for Cartographic

		(R: 2.5 m)	applications at cadastral level, urban and rural infrastructure development and management, as well as applications in Land Information System (LIS) and Geographical Information System (GIS). It provides stereo pairs required for generating DEM, Ortho Image products.
CARTOSAT-2	10.01.2007	PAN (R: <1 m)	Same as CARTOSAT-1.
CARTOSAT-2A	28.04.2008	PAN (R: <1 m)	Same as CARTOSAT-1 and 2.
RISAT-2	20.04.2009	Synthetic Aperture Radar (SAR)	It is useful in all weather remote sensing application. Disaster Management applications.
Oceansat-2	23.09.2009	Ocean Colour Monitor (OCM) Ku-band Pencil Beam scatterometer (SCAT) Radio Occultation Sounder for Atmosphere (ROSA)	It is envisaged to provide continuity of operational services of Oceansat-1 (IRS-P4) with enhanced application potential.
CARTOSAT-2B	12.07.2010	PAN (R: <1 m)	Continuation of previous CARTOSAT missions.
RESOURCESAT-2	20.04.2011	LISS-III, LISS-IV, AWiFS-A, AWiFS-B, AIS	It follows RESOURCESAT-1 mission with enhanced multispectral, spatial coverage and radiometric accuracy. AIS- for ship surveillance.
Megha-Tropiques	12.10.2011	Microwave Analysis and Detection of Rain and Atmospheric Structures (MADRAS), Sounder for Probing Vertical Profiles of Humidity (SAPHIR), Scanner for Radiation Budget (ScaRaB), Radio Occultation Sensor for Vertical Profiling of Temp. and Humidity (ROSA),	The main objective of this mission is to understand the life cycle of convective systems that influence the tropical weather and climate and their role in associated energy and moisture budget of the atmosphere in tropical regions.
RISAT-1	26.04.2012	Synthetic Aperture Radar (SAR)	Enables imaging of the surface features during both day and night under all weather conditions enable applications in agriculture,

			particularly paddy monitoring in kharif season and management of natural disasters like flood and cyclone.
--	--	--	--

(Source: [www.isro.org/satellites/earthobservationsatellites.aspx](http://www.isro.org/satellites/earthobservationsatellites.aspx); July 27, 2012)

### Indian National Satellite (INSAT) System

The Indian National Satellite (INSAT) systems which are placed in Geo-stationary orbits are one of the largest domestic communication satellite systems in Asia-Pacific region. Established in 1983 with commissioning of INSAT-1B, in a joint venture of Department of Space (DOS), Department of Telecommunications, India Meteorological Department, All India Radio and Doordarshan. It initiated a major revolution in India's communications sector and sustained the same later. INSAT space segment consists of 24 satellites out of which 10 are in service (INSAT-2E, INSAT-3A, INSAT-4B, INSAT-3C, INSAT-3E, KALPANA-1, INSAT-4A, INSAT-4CR, GSAT-8 and GSAT-12).

**Table 1.2. INSAT programs**

Satellite	Launched date
INSAT-1A	10 <sup>th</sup> April, 1982
INSAT-1B	30 <sup>th</sup> August, 1983
INSAT-1C	21 <sup>st</sup> July, 1988
INSAT-1D	12 <sup>th</sup> June, 1990
INSAT-2A	10 <sup>th</sup> July, 1992
INSAT-2B	23 <sup>rd</sup> July, 1993
INSAT-2C	7 <sup>th</sup> December, 1995
INSAT-2D	4 <sup>th</sup> June, 1997
INSAT-2DT acquired from ARABSAT	26 <sup>th</sup> February, 1992
INSAT-2E	3 <sup>rd</sup> April, 1999
INSAT-3B	22 <sup>nd</sup> March, 2000
GSAT-1	18 <sup>th</sup> April, 2001
INSAT-3C	24 <sup>th</sup> January, 2002
KALPANA	12 <sup>th</sup> September, 2002
INSAT-3A	10 <sup>th</sup> April, 2003



GSAT-2	8 <sup>th</sup> May, 2003
INSAT-3E	28 <sup>th</sup> September, 2003
EDUSAT	20 <sup>th</sup> September, 2004
HAMSAT	5 <sup>th</sup> May, 2005
INSAT-4A	22 <sup>nd</sup> December, 2005
INSAT-4B	12 <sup>th</sup> March, 2007
INSAT-4C-R	2 <sup>nd</sup> September, 2007
GSAT-4	15 <sup>th</sup> April, 2010
GSAT-5P	25 <sup>th</sup> December, 2010
GSAT-8	21 <sup>st</sup> May, 2011
GSAT-12	15 <sup>TH</sup> July, 2011

(Source: [www.isro.org/satellites/geostationary.aspx](http://www.isro.org/satellites/geostationary.aspx); July 30)

### Indian Regional Navigation Satellite System - GAGAN

For satellite based navigation in India two core constellations--Global Positioning System (GPS) of the United States and GLONASS of the Russian Federation are available. The position accuracies achievable with these core constellations are not good enough for precision approach and landing requirements of Civil Aviation. The Ministry of Civil Aviation has decided to implement an indigenous Satellite-Based Regional GPS Augmentation System also known as Space-Based Augmentation System (SBAS) named GAGAN (GPS and Geo Augmented Navigation), it will incorporate GPS augmented by a geosynchronous segment as part of the Satellite-Based Communications, Navigation and Surveillance (CNS)/Air Traffic Management (ATM) plan for civil aviation.

The next major milestone in GAGAN is the conduct of PSAT (Preliminary System Acceptance Testing) which has been successfully completed in Dec 2010. The first GAGAN navigation payload is slated on GSAT-8 which was launched on May 21, 2011. The second GAGAN payload is scheduled to be launched on GSAT-10 shortly.

### Satellite Launch Vehicle (SLV)

The Satellite Launch Vehicle, usually known by its abbreviation SLV or SLV-3 was a 4-stage solid-fuel light launcher. It was intended to reach a height of 500 km and carry a payload of 40 kg. Its first launch took place in 1979 with 2 more in each subsequent year, and the final launch in 1983. Only two of its four test flights were successful.

### **Augmented Satellite Launch Vehicle (ASLV)**

The Augmented Satellite Launch Vehicle, usually known by its abbreviation ASLV was a 5-stage solid propellant rocket with the capability of placing a 150 kg satellite into LEO. This project was started by the ISRO during the early 1980s to develop technologies needed for a payload to be placed into a geostationary orbit. Its design was based on Satellite Launch Vehicle. The first launch test was held in 1987, and after that 3 others followed in 1988, 1992 and 1994, out of which only 2 were successful, before it was decommissioned.

### **Polar Satellite Launch Vehicle (PSLV)**

The Polar Satellite Launch Vehicle, usually known by its abbreviation PSLV, is an expendable launch system developed to allow India to launch its Indian Remote Sensing (IRS) satellites into sun synchronous orbits, a service that was, until the advent of the PSLV, commercially viable only from Russia. PSLV can also launch small satellites into geostationary transfer orbit (GTO). The reliability and versatility of the PSLV is proven by the fact that it has launched 30 spacecraft (14 Indian and 16 from other countries) into a variety of orbits so far. In April 2008, it successfully launched 10 satellites at once, breaking a world record held by Russia.

On July 15, 2011 the PSLV flew its 18th consecutive successful mission orbiting satellites. Its only failure in 19 flights was its maiden voyage in September 1993, providing the rocket with a 94 percent success rate. It has also been used for secondary payload launches for several countries including Germany, Belgium, South Korea, Indonesia, Argentina, Israel, Canada, Denmark, Japan, and the Netherlands. Its first dedicated launch for a foreign client took place in April 2007 when it launched the Italian satellite AGILE.

### **Geosynchronous Satellite Launch Vehicle (GSLV)**

It was developed to enable India to launch its INSAT-type satellites into geostationary orbit and to make India less dependent on foreign rockets. The Geosynchronous Satellite Launch Vehicle (GSLV) project was initiated in 1990 with the objective of acquiring launch capability for Geosynchronous satellites. Until then, India depended on the former Soviet Union for the launch of heavy satellites.

ISRO is also developing the next generation GSLV that will be able to launch larger INSAT satellites in the 4 metric ton category. India's current generation of INSAT satellites weigh around 2 to 3 metric tons. GSLV Mk 3 will also be capable of placing 10 ton payloads into Low Earth Orbit (LEO).

### **The Ground Segment**

ISRO Telemetry, Tracking and Command Network (ISTRAC) provide mission support to Low-Earth Orbit (LEO) satellites as well as launch vehicle missions. ISTRAC has its headquarters and a multi-mission Spacecraft Control Centre at Bangalore. It has a network of ground stations at Bangalore, Lucknow, Sriharikota, Port Blair and Thiruvananthapuram in India besides stations at Mauritius, Bearslake (Russia), Brunei and Biak (Indonesia). ISTRAC activities are organized into network operations, network augmentation, mission operation

and spacecraft health monitoring, communications and computers and control centre facilities and development projects.

ISTRAC has also set up an Indian Deep Space Tracking Network station at Bangalore for India's mission to moon, Chandrayaan-1.

In addition to ISTRAC, there is the Master Control Facility (MCF) located in Hassan, Karnataka and Bhopal, Madhya Pradesh. The MCF monitors and controls all the geo-stationary satellites. MCF carries out operations related to initial orbit-raising for satellites, in-orbit payload testing, and in-orbit operations throughout the life of these satellites. The operations involve continuous tracking, telemetry and commanding, special operations like eclipse management, station-keeping maneuvers and recovery in case of contingencies. MCF interacts with the user agencies for effective utilization of the satellite payloads and to minimize service disturbances during special operations.

### **1.3 Modern Remote Sensing Technology Versus Conventional Aerial Photography**

The use of different and extended portions of the electromagnetic spectrum, development in sensor technology, different platforms for remote sensing (spacecraft, in addition to aircraft), emphasize on the use of spectral information as compared to spatial information, advancement in image processing and enhancement techniques, and automated image analysis in addition to manual interpretation are points for comparison of conventional aerial photography with modern remote sensing system.

During early half of twentieth century, aerial photos were used in military surveys and topographical mapping. Main advantage of aerial photos has been the high spatial resolution with fine details and therefore they are still used for mapping at large scale such as in route surveys, town planning, construction project surveying, cadastral mapping etc. Modern remote sensing system provide satellite images suitable for medium scale mapping used in natural resources surveys and monitoring such as forestry, geology, watershed management etc. However the future generation satellites will to provide much high-resolution images for more versatile applications. (Source: [wamis.org/agm/pubs/agm8/Paper-2.pdf](http://wamis.org/agm/pubs/agm8/Paper-2.pdf)).

### **1.4 Remote Sensing Processes**

Remote Sensing is a multi disciplinary activity which deals with the inventory, monitoring and assessment of natural resources through the analysis of data obtained by observations from a remote platform. When viewed in this context remote sensing covers various disciplines from astronomy to laboratory testing of materials. However, remote sensing is currently used more commonly to denote identification of earth features by detecting the characteristic of electromagnetic radiation that is reflected, emitted or scattered by the earth surface. The electromagnetic radiation extending from the ultraviolet to the far infra-red and microwave regions provides the greatest potential in the context of earth resources survey. For collection of remotely sensed data the basic requirements are a platform and a sensor.

In much of remote sensing, the process involves an interaction between incident radiation and the targets of interest.

This is exemplified by the use of imaging systems where the following seven elements are involved. Further remote sensing also involves the sensing of emitted energy and the use of non-imaging sensors.

**A) Energy source or illumination-** The first requirement for remote sensing is to have an energy source which illuminates or provides electromagnetic energy to the target of interest.

**B) Radiation and the atmosphere-** As the EMR travels from its source to the target, it will come in contact with and interact with the atmosphere it passes through. This interaction may take place a second time as the energy travels from the target to the sensor.

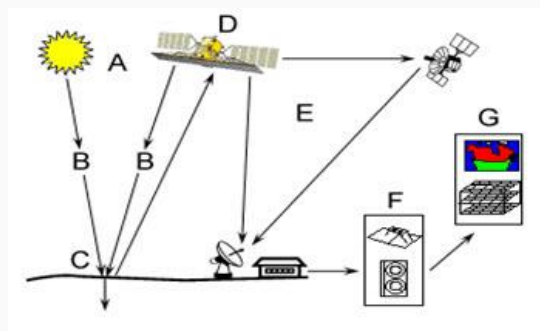


Fig. 1.2. Remote Sensing process.

(Source: [gisceu.net/tutorial/chap1/c1p1\\_i2e.html](http://gisceu.net/tutorial/chap1/c1p1_i2e.html); August: 10)

**C) Interaction with the target-** once the energy makes its way to the target through the atmosphere, it interacts with the target depending on the properties of both the target and the radiation.

**D) Recording of energy by the sensor-** after the energy has been scattered by, or emitted from the target, a sensor is required (remote- not in contact with the target) to collect and record the electromagnetic radiation.

**E) Transmission, reception and processing-** the energy recorded by the sensor has to be transmitted, often in electronic form, to a receiving and processing station where the data are processed into an image (hardcopy and/or digital).

**F) Interpretation and analysis-** the processed image is interpreted, visually and/or digitally or electronically, to extract information about the target which was illuminated.

**G) Application -** the final element of the remote sensing process is application i.e. after extracting the information from the image to solve a particular problem.

**Keywords:** Remote sensing, GIS, IRS, INSAT, PSLV, ASLV, GSLV.

## Lesson 2 Energy Sources and Interaction

### 2.1 Source of Energy

As noted earlier in lesson 1, the first requirement for remote sensing is to have an energy source to illuminate the target. Just as our eyes need objects to be illuminated by light so that we can see them, sensors also need a source of energy to illuminate the earth's surface. The sun is the natural source of energy. This energy is in the form of electromagnetic radiation (EMR). The following subsections explain the classification of remote sensing.

Depending on the predominant source of electromagnetic energy in the remote sensing system, the remote sensing can be passive or active.

**Passive Remote Sensing** depends on a natural source to provide energy. The sun is the most commonly used source of energy for passive remote sensing. The satellite sensor in this case records primarily the radiation that is reflected from the target. Remote sensing in the visible part of the electromagnetic spectrum is an example of passive (reflected) remote sensing.

A portion of the sun's radiation that is not reflected back to the sensor is absorbed by the target, raising the temperature of target material. The absorbed radiation is later emitted by the material at a different wavelength. Passive remote sensing can also be carried out in the absence of the sun. In this latter case, the source of energy is the target material itself and the sensor records primarily emitted radiation. Remote sensing in the thermal infrared portion of the electromagnetic spectrum is an example of passive (emitted) remote sensing.

**Active Remote Sensing** uses an artificial source for energy. For example the satellite itself can send a pulse of energy which can interact with the target. In active remote sensing, humans can control the nature (wavelength, power, duration) of the source energy. Remote sensing in the microwave region of the electromagnetic spectrum (radar remote sensing) is an example of active remote sensing. Active remote sensing can be carried out during day and night and in all weather conditions.

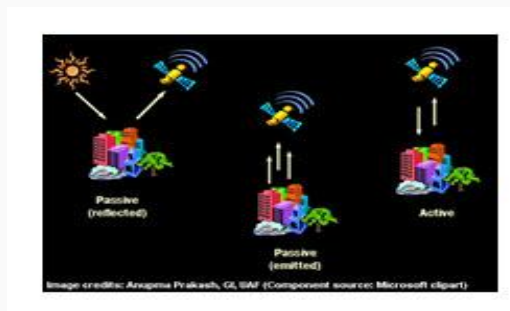


Fig. 2.1. Active and passive satellite sensors.

(Source: [essedesignguide.org/index.php?option=com\\_content&task=view&id=165&Itemid=9](http://essedesignguide.org/index.php?option=com_content&task=view&id=165&Itemid=9) ; Aug16,2012.)

## Remote Sensing Classification

Based on wavelength regions the remote sensing can be classified as:

1. Visible and reflective infrared remote sensing.
2. Thermal infrared or emitted remote sensing.
3. Microwave remote sensing.

The energy source used in the *visible and reflective infrared remote sensing* is the sun. The sun radiates EM energy with a peak wavelength of about 0.5  $\mu\text{m}$ . Remote sensing data obtained in the visible and reflective infrared regions mainly depends on the reflectance of objects on the ground surface. Therefore, information about objects can be obtained from the spectral reflectance. However laser radar is exceptional because it does not use the solar energy but the laser energy of the sensor.

The source of the radiant energy in the thermal infrared remote sensing is the object itself, because any object with a normal temperature of about 27°C will emit EM radiation with a peak at about 9.7  $\mu\text{m}$ .

In microwave region, there are two types of *microwave remote sensing*, passive microwave remote sensing and active microwave remote sensing. In passive microwave remote sensing, the microwave radiation emitted from an object is detected, while the back scattering coefficient is detected in active microwave sensing ([www.jars1974.net/pdf/02\\_Chapter01.pdf](http://www.jars1974.net/pdf/02_Chapter01.pdf)).

## 2.2 Energy Interaction with Atmosphere

Irrespective of source, all radiation detected by remote sensors passes through some distance, or path length of atmosphere. The path length involved can vary widely. For example, space photography results from sunlight that passes through the full thickness of the earth's atmosphere twice on its journey from source to sensor. On the other hand, an airborne thermal sensor detects energy emitted directly from objects on the earth, so a single, relatively short atmospheric path length is involved. The net effect of the atmosphere varies with these differences in path length and also varies with the magnitude of the energy signal being sensed, the atmospheric conditions present, and the wavelengths involved.

These effects are caused principally through the mechanisms of absorption, atmospheric scattering and reflection.

### Absorption

Absorption is the process by which radiant energy is absorbed and converted into other forms of energy. The absorption of the incident radiant energy may take place in the atmosphere and on the terrain. An absorption band is a range of wavelengths (or frequencies) in the electromagnetic spectrum within which radiant energy is absorbed by a substance. The cumulative effect of the absorption by the various constituents can cause the atmosphere to close down completely in certain regions of the spectrum. This is not desired for remote sensing as no energy is available to be sensed.



Ozone, carbon dioxide, and water vapour are the three main atmospheric constituents that absorb radiation. Ozone serves to absorb the harmful (to most living things) ultraviolet radiation from the sun. Without this protective layer in the atmosphere our skin would burn when exposed to sunlight.

Carbon dioxide referred to as a greenhouse gas. This is because it tends to absorb radiation strongly in the far infrared (thermal infrared) portion of the spectrum –area associated with thermal heating –which serves to trap this heat inside the atmosphere. The water vapour in the atmosphere absorbs much of the incoming long wave (thermal) infrared and shortwave microwave radiations. The presence of water vapour in the lower atmosphere varies greatly from location to location and at different times of the year.

## Scattering

Scattering is unpredictable diffusion of electromagnetic radiation by atmospheric particles. It occurs when particles or large gas molecules present in the atmosphere interact with and cause the EMR to be redirected from its original path. The amount of scattering takes place depends on several factors including the wavelength of the radiation, the diameter of particles or gases, and the distance the radiation travels through the atmosphere. There are two classes of scattering – selective and non- selective.

Selective scattering: There are three types of selective scattering Rayleigh scattering, Mie scattering, and Raman scattering:

- **Rayleigh scattering** (sometimes referred to as molecular scattering) occurs when the effective diameter of the matter (usually air molecules air molecules such as oxygen and nitrogen in the atmosphere) is many times (usually < 0.1 times) smaller than the wavelength of the incident EMR. The amount of scattering is inversely related to the fourth power of wavelength of radiation. For example, ultraviolet light at 0.3  $\mu\text{m}$  most Rayleigh scattering takes place in the upper 4.5 km of the atmosphere. It is responsible for the blue appearance of the sky. The shorter violet and blue wavelengths are more efficiently scattered than the longer green and red wavelengths. That is why most remote sensing systems avoid detecting and recording wavelengths in the ultraviolet and blue portions of the spectrum.
- **Mie Scattering** (also referred to as non – molecular scattering) takes place in the lower 4.5 km of the atmosphere, where there may be many essentially spherical particles present with diameters approximately equal to the size of the wavelength of the incident energy. The actual size of the particles may range from 0.1 to 10 times the wavelength of the incident energy. For visible light, the main scattering agents are dust and other particles ranging from a few tenths of the micrometer to several micrometers in diameter. Fig. 2.2 shows the scattering of electromagnetic radiation.
- **Raman Scattering** is caused by atmospheric particles, which are larger, smaller, or equal to that of the wavelength of the radiation being sensed. The atmospheric particles may be gaseous molecules, water droplets, fumes, or dust particles. These portions have an elastic collision with the atmospheric particles which result in either loss or gain of energy and thus an increase or decrease in wavelength.

- **Non- Selective Scattering** takes place in the lowest portions of the atmosphere where there are particles greater than 10 times the wavelength of the incident EMR. This type of scattering is non- selective, that is, all wavelength of light are scattered, not just blue, green, or red. Thus, that is water droplets and ice crystals that make up clouds and fog banks scatter all wavelengths of visible light equally well, causing the cloud to appear white, (blue + green + red light = white light).

## Refraction

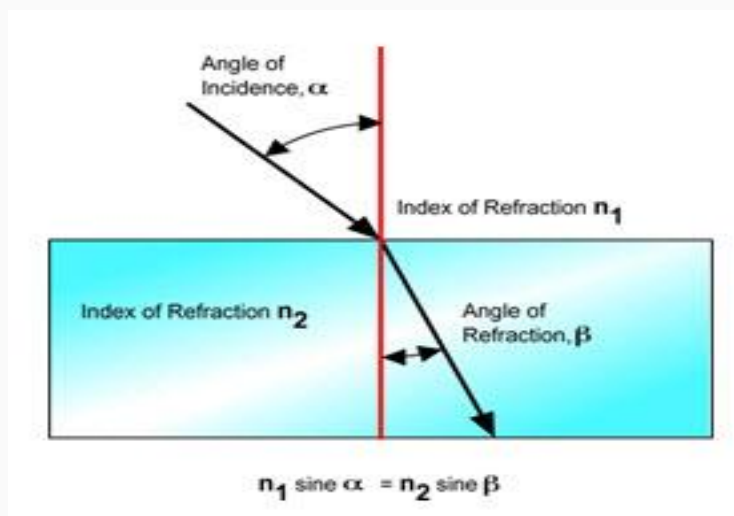
Refraction is the deflection of EM radiation as it passes from one medium with one refractive index to a medium with a different refractive index. Refractive index is defined as the ratio or the speed of light in a vacuum to the speed of light in the medium and is calculated by

$$\eta = \frac{C}{C_n} \quad (2.1)$$

$n$  = Refractive index.

$C$  = Speed of light in vacuum.

$C_n$  = Speed of light in medium.



**Fig. 2.3. Refraction of EM wave. (Source: [chemicalparadigms.wikispaces.com/Unit+2+Refraction+of+light](http://chemicalparadigms.wikispaces.com/Unit+2+Refraction+of+light))**

In the Earth's atmosphere temperature, compositions and humidity all affect the density which affects the refractive index. Serious errors in location due to refraction can occur in images formed from energy detected at high altitude or at acute angle. However, these location errors are predictable by Snell's law and can be removed.

The angle that the radiation (Fig. 2.3) will be bent is defined by Snells Law:

$$n_1 \sin \alpha = n_2 \sin \beta \quad (2.2)$$

$n_1$  = refractive index of first medium.

$n_2$  = refractive index of second medium.

$\sin \alpha$  = angle of incidence.

$\sin \beta$  = angle of refraction.

## Reflection

Reflection is the process whereby radiation 'bounces off' an object like the top of a cloud, a water body, or the terrestrial earth. Reflection differs from scattering in that the direction associated with scattering is unpredictable but in case of reflection it is predictable. Reflection exhibits fundamental characteristics that are important in the remote sensing. First, the incident radiation, the reflected radiation and a vertical to the surface from which the angle of incidence and reflection are measured all lie in the same plane. Second, the angle of incidence and the angle of reflection are approximately equal.

A considerable amount of incident radiant flux from the sun is reflected from the top of clouds and other material in the atmosphere. A substantial amount of this energy is reradiated back to space. The reflection principles that apply to clouds also apply to the terrain.

### 2.2.1 Energy Interactions with Earth Surface Features

When electromagnetic energy is incident on any given earth surface feature, three fundamental energy interactions with the feature are possible. This is illustrated in Fig. 2.4 for an element of the volume of a water body. Various fractions of the energy incident on the element are reflected, absorbed, and/ or transmitted. The amount of radiant energy onto, off of, or through a surface per unit time is called radiant flux ( $\Phi$ ) and is measured in watts (W).

Applying the principle of conservation of energy, the interrelationship between these three energy interactions can be expressed as

$$E_1(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \quad (2.3)$$

where,

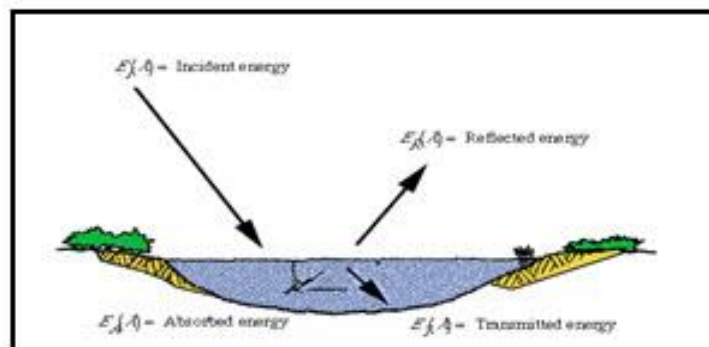
$E_1$  = incident energy

$E_R$  = reflected energy

$E_A$  = absorbed energy

$E_T$  = transmitted energy

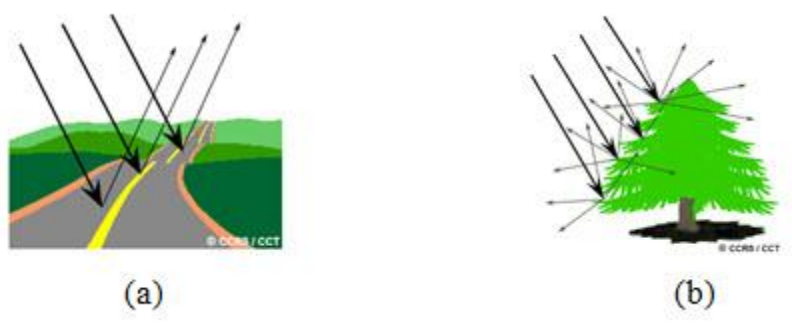
All energy components are function of wavelength  $\lambda$ .



**Fig. 2.4. Basic interactions between electromagnetic energy and an earth surface. (Source: Lillesand, Kiefer, 2002)**

In above equation two points concerning this relationship should be noted. First, the proportions of energy reflected, absorbed, and transmitted will vary for different earth features, depending on their material type and condition. These differences permit us to distinguish different features on an image. Second, the wavelength dependency means that, even within a given feature type, the proportion of reflected, absorbed, and transmitted energy will vary at different wavelengths. Thus, two features may be indistinguishable in one spectral range and be very different in another wavelength band.

In remote sensing the radiation reflected from targets. We refer to two types of reflection is measured reflection from a target: can be specular reflection and diffuse reflection. For a smooth surface specular or mirror-like reflection occurs where all (or almost all) of the energy is directed away from the surface in a single direction. Diffuse reflection occurs when the surface is rough and the energy is reflected almost uniformly in all directions. Most earth surface features lie somewhere between perfectly specular or perfectly diffuse reflectors. Fig's 2.5 and 2.6 illustrate specular and diffused reflection. Whether a particular target reflects specularly or diffusely, or somewhere in between, depends on the surface roughness of the feature in comparison to the wavelength of the incoming radiation. If the wavelengths are much smaller than the surface variations or the particle sizes that make up the surface, diffuse reflection will dominate. For example, fine-grained sand would appear fairly smooth to long wavelength microwaves but would appear quite rough to the visible wavelengths.



**Fig. 2.5. (a) Specular Reflection and (b) Diffuse Reflection. (Source: [www.cps-amu.org/sf/notes/mw4-2-4.htm](http://www.cps-amu.org/sf/notes/mw4-2-4.htm))**

Diffuse reflections contain spectral information on the “color” of the reflecting surface, whereas specular reflections do not. Hence, in remote sensing, we are most often interested in measuring the diffuse reflectance properties of terrain features.

This is measured as a function of wavelength and is called spectral reflectance,  $\rho(\lambda)$ .

It is can be expressed as

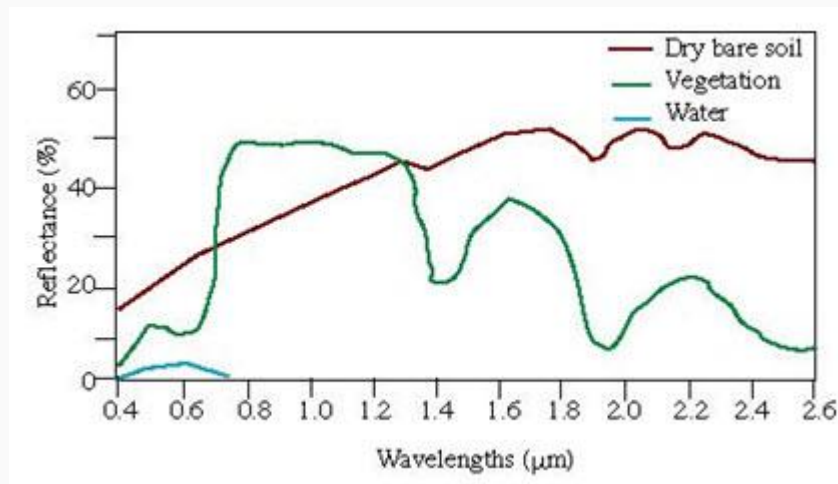
$$\rho_{\lambda} = \frac{E_R(\lambda)}{E_I(\lambda)} \quad (2.4)$$

$$= \frac{\text{Energy of wavelength } \lambda \text{ reflected from the object}}{\text{Energy of wavelength } \lambda \text{ incident upon the object}} \times 100$$

Where  $\rho(\lambda)$  is expressed as a percentage.

The spectral response of objects can be built from the measured reflected energy for different wavelengths. The spectral response of a material to different wavelengths of EMR can be represented graphically as a Spectral Reflectance Curve.

The comparison of spectral reflectance curves of different objects one can distinguish between them. For example, water and vegetation may reflect somewhat similarly in the visible wavelengths but are almost always separable in the infrared. Fig. 2.7 shows typical spectral reflectance curves for three basic types of earth features: healthy green vegetation, dry bare soil, and clear lake water. These curves indicate how much incident energy would be reflected from the surface, and subsequently recorded by a remote sensing instrument. At a given wavelength, the higher the reflectance, the brighter the object appears in an image.



**Fig. 2.7. Spectral reflectance curve.**

(Source: [www.ucalgary.ca/GEOG/Virtual/remotaintro.html](http://www.ucalgary.ca/GEOG/Virtual/remotaintro.html))

Note that vegetation reflects much more energy in the near-infrared (0.8 to 1.4 microns) than it does in visible light (0.4 to 0.7 microns). The amount of energy that vegetation reflects is related to the internal structure of the plant, and the amount of moisture in the plant. A surface like astro-turf, which is colored green, will appear dark in the near infrared, because it doesn't have the internal structure of living vegetation. Another feature to notice is that clear water reflects visible light only, so it will appear dark in infrared images.

Thus, to utilize remote sensing data effectively, one must know and understand the spectral characteristics of the particular features under investigation in any given application.

2.3 Data Acquisition and Interpretation

An image generated from an electronic sensor array is of a two-dimensional rectangular grid of numerical values that represent differing brightness levels. Each value represents the average brightness for a portion of the surface, represented by the square unit areas or grid in the image. The grid is a group of cells or pixels. When displayed on your computer, the brightness values in the image raster are translated into display brightness on the screen.

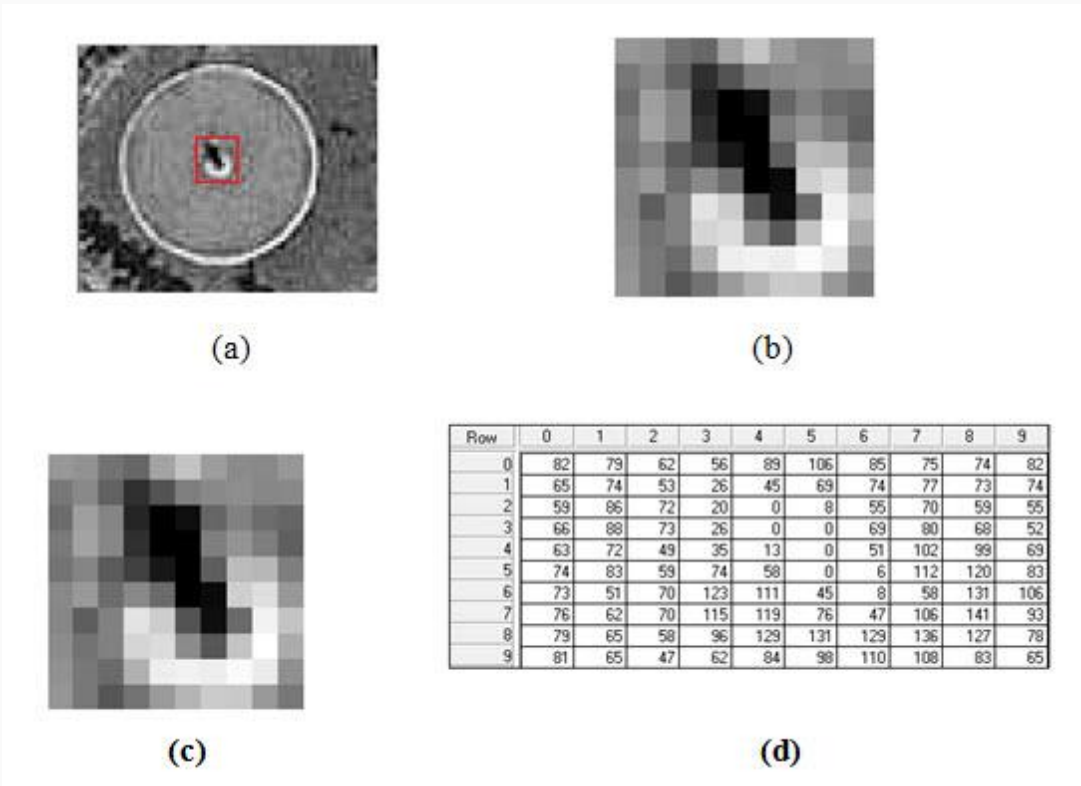
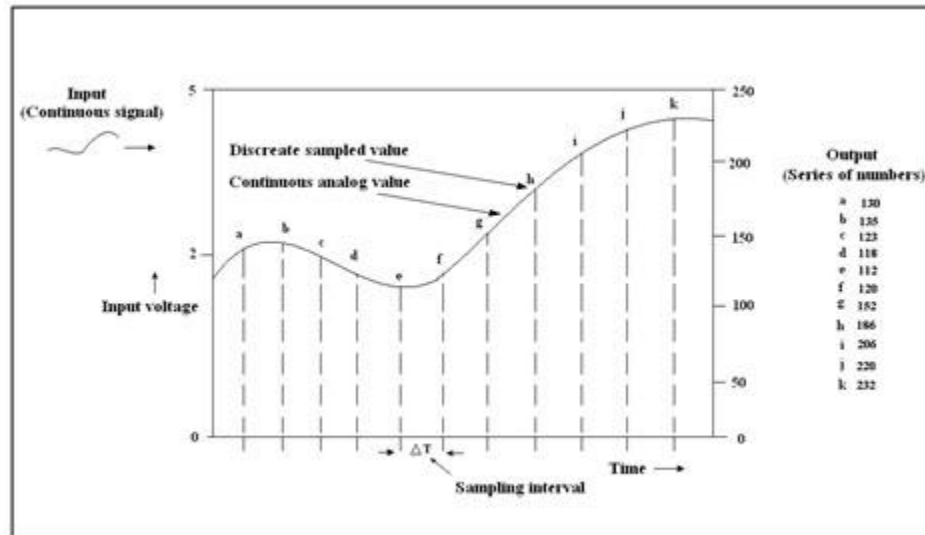


Fig. 2.8. Basic character of digital image data. (a) Original 563 rows x 468 columns digital image. (b) Enlargement showing 9 rows x 9 columns area of pixels showing in a red box in (a). (c) 9 rows x 9 columns enlargement. (d) Digital numbers corresponding to the radiance of each pixel shown in (c).

The basic characteristics of digital image data is illustrated in Fig. 2.8, though the image shown in (a) appears to be a continuous tone photograph, it is actually composed of a two dimensional array of discrete picture elements, or pixels. The intensity of each pixel corresponds to the average brightness, or radiance, measured electronically over the ground area corresponding to each pixel. A total of 320 rows and 480 columns of pixels are shown in



Fig. 2.8 (a). Whereas the individual pixels are virtually impossible to discern in (a), they are readily observable in the enlargements shown in (b) and (c). Corresponding to the average radiance measured in each pixel shown in Fig. 2.8 (c). These values are simply positive integers that result from quantizing the original electrical signal from the sensor into positive integer values using a process called analog-to-digital (A-to-D) signal conversion.



**Fig. 2.9. Analog-to-digital conversion process.**

The original electrical signal from the sensor is a continuous analog signal (shown by the continuous line plotted in the Fig. 2.9). This continuous signal is sampled at a set time interval ( $\Delta T$ ) and recorded numerically at each sample point (a, b... j, k). The sampling rate for a particular signal is determined by the highest frequency of change in the signal. The sampling rate must be at least twice as high as the highest frequency present in the original signal in order to adequately represent the variation in the signal. Fig. 2.9 is a graphical representation of the A-to-D conversion process.

Typically, the DNs constituting a digital image are recorded over such numerical ranges as 0 to 255, 0 to 511, 0 to 1023, or higher. These ranges represent the set of integers that can be recorded using 8-, 9-, and 10-bit binary computer coding scales, respectively. (That is  $2^8 = 256$ ,  $2^9 = 512$ , and  $2^{10} = 1024$ .) In such numerical formats, the image data can be readily analyzed with the aid of a computer.

Data alone cannot be used for decision making. It must be interpreted or analyzed before one can extract information. The analysis of remotely sensed data is performed using a variety of image interpretation and processing techniques which fall into two broad categories: Analog (visual) image interpretation and Digital image interpretation (DIP).

Visual interpretation techniques have certain disadvantages, however, in that they may require extensive training and are labor intensive. In addition, spectral characteristics are not always fully evaluated in visual interpretation effort. This is partly because of the limited ability of the eye to discern tonal analyze numerous spectral images. In applications where

spectral patterns are highly informative, it is therefore preferable to analyze digital, rather than pictorial, image data.

The use of computer- assisted analysis techniques permits the spectral patterns in remote sensing data to be more fully examined. It also permits the data analysis process to be largely automated, providing cost advantages over visual interpretation techniques. However, just as humans are somewhat limited in their ability to interpret spectral patterns, computers are somewhat limited in their ability to interpret spectral patterns. Therefore, visual and numerical techniques are complementary in nature, hence combination of both approaches are generally adopted.

## 2.4 Advantages of Remote Sensing

Remote sensing has several unique advantages as well as some limitations of remote sensing, to use it more effectively. Remote sensing is unobtrusive if the sensor is passively recording the EM energy reflected from or emitted by the phenomenon of interest. This is a very important consideration, as passive remote sensing (using natural source of energy, e.g., the sun) does not disturb the object or area of interest.

Under carefully controlled conditions, remote sensing can provide fundamental biophysical data, including: x,y location, z elevation or depth, biomass, temperature, moisture content, etc., in this sense, it is much like surveying, providing fundamental data that other sciences can use when conducting scientific investigations. However, unlike much of surveying, the remotely sensed data may be obtained systematically over very large geographic areas rather than just single point observation. **The major advantages are stated below.**

- **Synoptic View**

Remote sensing process facilitates the study of various earths' surface features in their spatial relation to each other and helps to delineate the required features and phenomena. Data about the entire earth is obtained in a short period of time can be used for different purposes without taking permission.

- **Repeativity**

The remote sensing satellites provide repetitive coverage of the earth and this temporal information is very useful for studying landscape dynamics, phenological variations of vegetation and other land features and change detection analysis.

- **Accessibility**

Remote sensing process made it possible to gather information about the area when it is not possible to do ground survey like in mountainous areas and foreign areas. Passive remote sensing can be used in all weather and all time of a day.

- **Time saving**

Since information about a large area can be gathered quickly, the techniques save time and efforts of human. It also saves the time of fieldwork.

- **Cost Effective**

Remote sensing especially when conducted from space, is an intrinsically expensive activity. Nevertheless, cost-benefit analysis demonstrates its financial effectiveness, and much

speculation or developmental remote sensing activity can be justified in this way. It is a cost-effective technique as again and again fieldwork is not required and also a large number of users of different disciplines can share and use the same data.

## 2.5 Limitations of Remote Sensing

Remote sensing science has various limitations. Perhaps the greatest limitation is that its utility is often oversold (Jensen 2004). It is not a panacea that will provide all the information needed for conducting physical, biological, or social science. It simply provides some spatial, spectral, and temporal information of value.

Human beings select the most appropriate sensor to collect the data, specify the resolution of the data, calibrate the sensor, select the platform that carry the sensor determine when the data will be collected, and specify how the data are processed. Thus, human method produced error may be introduced, as the various remote sensing instrument and mission parameters are specified (Jensen 2004).

Powerful active remote sensor system, such as lasers or radars that emit their own EMR, can be intrusive and affect that phenomenon being investigated.

Remote sensing instruments like in situ instruments often become uncalibrated, resulting in uncalibrated remote sensing data. Finally, remote sensors data may be expensive to collect, interpret, or analyze. But the information derived from the remote sensor data is so expensive that the expense is warranted.

The limitations of satellite remote sensing also include the inability of many sensors to obtain data and information through cloud cover (although microwave sensors can image earth through clouds) and the relatively low spatial resolution achievable with many satellite-borne earth remote sensing instruments. In addition, the need to correct for atmospheric absorption and scattering and for the absorption of radiation through water on the ground can make it difficult to obtain desired data and information on particular variables.

**Keywords:** Energy sources, absorption, scattering, specular, diffused reflector, spectral reflectance, Analogue to digital conversion.

## Lesson 3 Remote Sensing Applications

### 3.1 Introduction

Remote sensing may be used for numerous applications including weapon guidance system (e.g., the cruise missile), medical image analysis (e.g., X-raying a broken arm), non-destructive evaluation of machinery and products (e.g., of the assembly line), analysis of the earth's resources, etc. Earth resource information is defined as any information concerning terrestrial vegetation, soils, minerals, water, ocean, urban infrastructure etc. as well as certain atmospheric characteristics.

Remote sensing system is able to provide a synoptic view of a wide area in a single frame. The width of a single frame, or swath width, could be 60 km x 60 km in the case of the European SPOT satellite, or as wide as 185 km x 185 km in the case of Landsat, or 23 km x 23 km in the case of LISS-IV of IRS P6. Remote sensing systems can provide data and information in areas where access is difficult as rendered by terrain, weather, or military security. The towering Himalayas and the bitterly cold Antarctic regions provide good examples of these harsh environments. Active remote sensing systems provide cloud-free images that are available in all weather conditions, day or night. Such systems are particularly useful in tropical countries where constant cloud cover may obscure the target area. In 2002, the United States military initiatives in Afghanistan used remote sensing systems to monitor troops and vehicle convoy movements at spatial resolutions of less than one meter to a few meters.

Remote sensing can provide information in two different classes of variables: biophysical and hybrid. Biophysical variables can provide fundamental biological and/or physical biophysical information directly, without having to use other surrogate or ancillary data. For example, a thermal infrared sensor can record the apparent temperature of a rock by measuring the radiant flux emitted from its surface. Similarly, it is possible to conduct remote sensing in a very specific region of the electromagnetic spectrum and identify the amount of water vapor in the atmosphere. It is also possible to measure soil moisture content directly using microwave remote sensing technique.

The second general group hybrid variables, created by systematically analyzing more than biophysical variable. For example, by remotely sensing a plant's chlorophyll absorption characteristics, temperature, and moisture content, it may be possible to model these data to detect vegetation stress. The variety of hybrid variables is large. Basically, successful remote sensing modeling predicts how much radiant flux in certain wavelengths should exit from a particular object.

At present applications of remote sensing are numerous and varied. They include land cover mapping and analysis, land use mapping, agricultural plant health monitoring and harvest forecast, water resources, wildlife ecology, archeological investigations, snow and ice

monitoring, disaster management, geologic and soil mapping, mineral exploration, coastal resource management, military surveillance, and many more.

### **3.2 Applications of Remote Sensing Technology**

Each sensor in remote sensing system is designed for a specific purpose. With optical sensors, the design focuses on the spectral bands to be used. With radar imaging, the incidence angle and microwave band used plays an important role in defining which applications the sensor is best suited for. Each application itself has specific demands, for spectral resolution, spatial resolution, radiometric resolution and temporal resolution.

In the last four decades it has grown as a major tool for collecting information on almost every aspect on the earth. With the availability of very high spatial resolution satellites in the recent years, the applications have multiplied. In India remote sensing has been used for various applications during the last four decades and has contributed significantly towards development.

Some of the important projects carried out in the country include groundwater prospects mapping under drinking water mission, Forecasting Agricultural Output Using Space, Agrometeorology And Land Based Observations (FASAL), forest cover/type mapping, grassland mapping, biodiversity characterization, snow & glacier studies, land use/cover mapping, coastal studies, coral and mangroves studies, wasteland mapping etc. The information generated by large number of projects have been used by various departments, industries and others for different purposes like development planning, monitoring, conservation etc. There can be many applications for remote sensing, in different fields, as described below.

#### **3.2.1 Agriculture**

Agriculture plays a dominant role in economies of both developed and undeveloped countries. Satellite and airborne images are used as mapping tools to classify crops, examine their health and viability, and monitor farming practices. Agricultural applications of remote sensing include the following:

- crop type classification
- crop condition assessment
- crop yield estimation
- mapping of soil characteristics
- mapping of soil management practices
- compliance monitoring (farming practices)
- Monitoring of pests and Diseases

### 3.2.2 Forestry

Forests are a valuable resource providing food, shelter, wildlife habitat, fuel, and daily supplies such as medicinal ingredients and paper. Forests play an important role in balancing the Earth's CO<sub>2</sub> supply and exchange, acting as a key link between the atmosphere, geosphere, and hydrosphere. Forestry applications of remote sensing include the following:

- **Reconnaissance mapping:** objectives to be met by national forest/environment agencies include forest cover updating, depletion monitoring, and measuring biophysical properties of forest stands.
- **Commercial forestry:** of importance to commercial forestry companies and to resource management agencies are inventory and mapping applications; collecting harvest information, updating of inventory information for timber supply, broad forest type, vegetation density and biomass measurements.
- **Environmental monitoring:** conservation authorities are concerned with monitoring the quantity, health, and diversity of the Earth's forests.

### 3.2.3 Geology

Geology involves the study of landforms, structures, and the subsurface, to understand physical processes creating and modifying the earth's crust. It is most commonly understood as the exploration and exploitation of mineral and hydrocarbon resources, generally to improve the conditions and standard of living in society.

Geological applications of remote sensing include the following:

- Surficial deposit / bedrock mapping
- Lithological mapping
- Structural mapping
- Sand and gravel (aggregate) exploration/ exploitation
- Mineral exploration
- Hydrocarbon exploration
- Environmental geology
- Geobotany
- Baseline infrastructure
- Sedimentation mapping and monitoring
- Event mapping and monitoring
- Geo-hazard mapping

- Planetary mapping

### 3.2.4 Hydrology

Hydrology is the study of water on the earth's surface, whether flowing above ground, frozen in ice or snow, or retained by soil.

Examples of hydrological applications include:

- Wetlands mapping and monitoring,
- Soil moisture estimation,
- Snow pack monitoring / delineation of extent,
- Measuring snow thickness,
- Determining snow-water equivalent,
- River and lake ice monitoring,
- Flood mapping and monitoring,
- Glacier dynamics monitoring (surges, ablation)
- River / delta change detection
- Drainage basin mapping and watershed modeling,
- Irrigation canal leakage detection
- Irrigation scheduling
- Crop water estimation
- Monitoring of crop water stresses
- Canal water distribution and management

### 3.2.5 Sea Ice

Ice covers a substantial part of the earth's surface and is a major factor in commercial shipping and fishing industries, coast guard and construction operations, and global climate change studies.

Examples of sea ice information and applications include:

- Ice concentration
- Ice type / age / motion
- Iceberg detection and tracking

- Surface topography
- Tactical identification of leads: navigation: safe shipping routes/rescue
- Ice condition (state of decay)
- Historical ice and iceberg conditions and dynamics for planning purposes
- Wildlife habitat
- Pollution monitoring
- Meteorological / global change research

### 3.2.6 Land Cover & Land Use

Although the terms land cover and land uses are often used interchangeably, their actual meanings are quite distinct. Land cover refers to the surface cover on the ground, while Land use refers to the purpose the land serves. The properties measured with remote sensing techniques relate to land cover, from which land use can be inferred, particularly with ancillary data or a priori knowledge.

Land use applications of remote sensing include the following:

- Natural resource management
- Wildlife habitat protection
- Baseline mapping for GIS input
- Urban expansion / encroachment
- Routing and logistics planning for seismic / exploration / resource extraction activities
- Damage delineation (tornadoes, flooding, volcanic, seismic, fire)
- Legal boundaries for tax and property evaluation
- Target detection - identification of landing strips, roads, clearings, bridges, land/water interface

### 3.2.7 Mapping

Mapping constitutes an integral component of the process of managing land resources, and mapped information is the common product of analysis of remotely sensed data.

Mapping applications of remote sensing include the following:



- **Planimetry**

Land surveying techniques accompanied by the use of a GPS can be used to meet high accuracy requirements, but limitations include cost effectiveness, and difficulties in attempting to map large, or remote areas. Remote sensing provides a means of identifying and presenting planimetric data in convenient media and efficient manner. Imagery is available in varying scales to meet the requirements of different users. Defence applications typify the scope of planimetry applications - extracting transportation route information, building and facilities locations, urban infrastructure, and general land cover.

- **Digital elevation models (DEM's)**

Generating DEMs from remotely sensed data can be cost effective and efficient. A variety of sensors and methodologies to generate such models are available and proven for mapping applications. Two primary methods for generating elevation data are

1. Stereogrammetry techniques using airphotos (photogrammetry), VIR imagery, or radar data (radargrammetry), and
2. Radar interferometry.

- **Baseline thematic mapping/ topographic mapping**

As a base map, imagery provides ancillary information to the extracted planimetric or thematic detail. Sensitivity to surface expression makes radar a useful tool for creating base maps and providing reconnaissance abilities for hydrocarbon and mineralogical companies involved in exploration activities. This is particularly true in remote northern regions, where vegetation cover does not mask the microtopography and generally, information may be sparse. Multispectral imagery is excellent for providing ancillary land cover information, such as forest cover. Supplementing the optical data with the topographic relief and textural nuance inherent in radar imagery can create an extremely useful image composite product for interpretation.

### 3.2.8 Oceans & Coastal Monitoring

The oceans not only provide valuable food and biophysical resources, they also serve as transportation routes, are crucially important in weather system formation and CO<sub>2</sub> storage, and are an important link in the earth's hydrological balance. Coastlines are environmentally sensitive interfaces between the ocean and land and respond to changes brought about by economic development and changing land-use patterns. Often coastlines are also biologically diverse inter-tidal zones, and can also be highly urbanized.

Ocean applications of remote sensing include the following:

- Ocean pattern identification:
- Currents, regional circulation patterns, shears
- Frontal zones, internal waves, gravity waves, eddies, upwelling zones, shallow water bathymetry,
- Storm forecasting

- Wind and wave retrieval
- Fish stock and marine mammal assessment
- Water temperature monitoring
- Water quality
- Ocean productivity, phytoplankton concentration and drift
- Aquaculture inventory and monitoring
- Oil spill
- Mapping and predicting oil spill extent and drift
- Strategic support for oil spill emergency response decisions
- Identification of natural oil seepage areas for exploration
- Shipping
- Navigation routing
- Traffic density studies
- Operational fisheries surveillance
- Near-shore bathymetry mapping
- Intertidal zone
- Tidal and storm effects
- Delineation of the land / water interface
- Mapping shoreline features / beach dynamics
- Coastal vegetation mapping
- Human activity / impact

(Source: [geography.huji.ac.il/personal/Noam%20Levin/1999-fundamentals-of-remote-sensing.pdf](http://geography.huji.ac.il/personal/Noam%20Levin/1999-fundamentals-of-remote-sensing.pdf))

**Keywords:** Remote sensing applications, Land use / Land cover, mapping, DEM, Planimetry, Hydrology, Agriculture.



## ***Module 2: Sensors, Platforms and Tracking System***

### **Lesson 4 Sensors Characteristics**

#### **4.1 Remote Sensing Sensors**

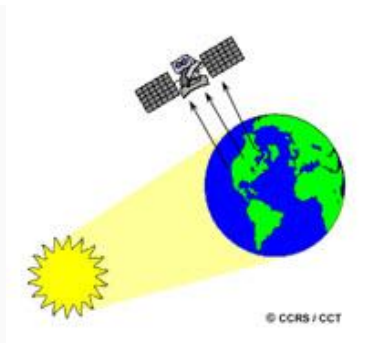
Sensor is an electronic circuit which can record the electromagnetic radiation incident upon it. A sensor is a device comprising of optical component or system and a detector with electronic circuit. It senses a variation in input energy to produce a variation in another or same form of energy. All sensors employed on earth observation platforms use electromagnetic radiation to observe terrain features. Various types of sensor are employed based on requirements and purpose. Remote sensing sensors measure radiance of objects under study in a given wavelength. They comprise of several components such as-

- System to receive radiation from the pixel and a telescope (objective),
- Calibration source and spectrometer,
- Amplifier and recording system,

The sensor thus represents an imaging radiometer, often called scanner, essentially consisting of a radiometer supported by a system of image acquisition, pixel by pixel.

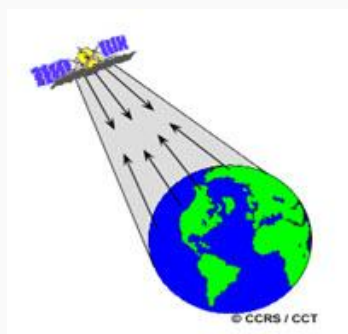
#### **4.2 Classification of Remote Sensing Sensors**

Remote sensors can be broadly classified as passive sensors and active sensors. Sensors which sense natural radiation, either emitted or reflected from the earth are called passive sensors. This process is known as passive remote sensing (Fig. 4.1.). It is also possible to produce electromagnetic radiation of a specific wavelength or band of wavelengths as a part of the sensor system. The interaction of this radiation with the target could then be studied by sensing the scattered radiation from the targets. Such sensors, which produce their own electromagnetic radiation, are called active sensors. This process of remote sensing is active remote sensing (Fig. 4.2.). Since the technology for developing microwave sensors is quite different from that of optical infrared (OIR) sensors, from the standpoint of understanding the design and realization of the sensors, it is convenient to classify the sensors (both passive and active) as those operating in the OIR region and those operating in the microwave region. The OIR and microwave sensors could be either imaging or non imaging sensors. Imaging sensors give a two dimensional spatial distribution of the emitted or reflected intensity of the electromagnetic radiation (as in a photographic camera), while the non imaging sensors measure the intensity of radiation, within the field of view, and in some cases as a function of distance along the line of sight of the instrument (for example, vertical temperature profiling radiometer -VTPR). Fig. 4.3. shows the remote sensors classification.



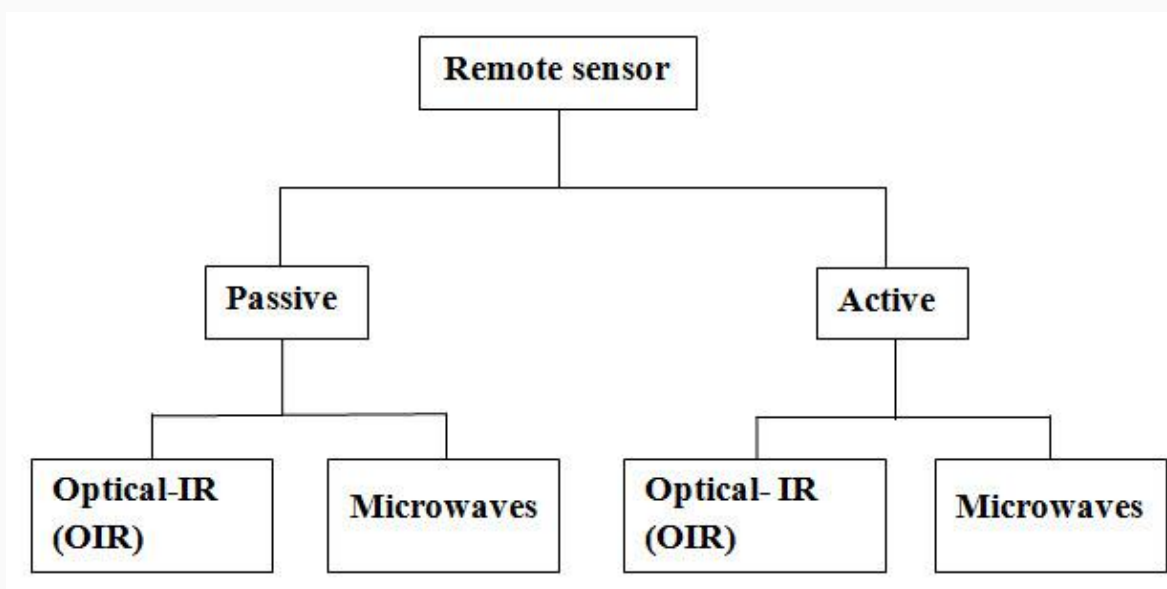
**Fig. 4.1. Passive remote sensing.**

(Sources: [www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1212](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1212), August 17, 2012.)



**Fig. 4.2. Active remote sensing.**

(Sources: [www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1212](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1212), August 17, 2012.)



**Fig. 4.3. Remote sensing sensors classification. (Source: Joseph, 2005)**

The sensors may be broadly be divided classified based on their working principles and recording medicines.

- 1) Photographic camera
- 2) The vidicon using detecting media as television camera
- 3) The optical scanner
- 4) Microwave radiometer
- 5) Microwave radar

### 4.3 Sensors Parameters and Resolution

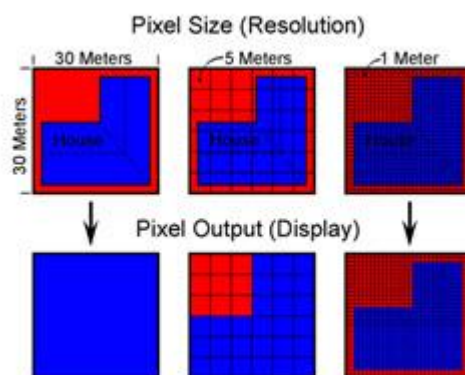
The information collected by the remote sensors is meant to identify and map various earth surface objects. Therefore, we may say that the performance of the sensor is evaluated based on its classification as well as its mapping accuracy requirements. It is reasonable to assume that this will depend on the instruments ability to detect small differences in the emittance/reflectance of the earth's surface in a number of spectral bands for as small an object as possible and as often as possible.

Then the important question is what is the optimum set of specifications for a remote sensor? Unfortunately, there is no unique answer, since the choice of the optimum parameters depends on the theme under study. Even if one identifies an ideal set of parameters, the realization of a combination of these parametric values (i.e., spatial resolution, number of spectral bands, spectral bandwidth, signal to noise ratio, etc.) in a sensor system is a complex problem due to the strong interrelationship among these parameters and requirements of the sensor. We may consider the sensor parameters under four domains: (1) spatial (2) spectral (3) radiometric (4) temporal.

#### Resolution

In remote sensing, a given area of earth surface is observed by the sensor with each measurement corresponding to an elemental area on the surface over number of spectral bands. Then one can think of resolution. Resolution is defined as a measure of sensor ability to distinguish between signals. All remote sensing information is resolution dependent. The various parameters which characterize theses different kinds of sensor systems are described by resolution. The quality of remote sensing data depends on its spatial, spectral, radiometric and temporal resolution.

**Spatial resolution** refers to the size of the smallest object that can be detected in an image. The basic unit in an image is called a pixel. One-meter spatial resolution means each pixel image represents an area of one square meter in the ground. Smaller an area represented by one pixel, higher the spatial resolution of the image. Fig. 4.4. explains the details of spatial resolution. For example spatial resolution of Landsat 7, ETM+ sensor 30 m for band 1-5, 7; 60 m for band 6, and 15 m for band 8. It means that one image pixel corresponds to 30 m x 30 m area in the terrain for band 1 – 5; 60 m x 60 m area in the terrain for band 6 and 15 m x 15 m area in the terrain for band 8.



**Fig. 4.4. Pixel resolution.** (Source:[www.satimagingcorp.com/characterization-of-satellite-remote-sensing-systems.html](http://www.satimagingcorp.com/characterization-of-satellite-remote-sensing-systems.html))

**Spectral resolution** refers to the number of bands and the wavelength width of each band. A band is a narrow portion of the electromagnetic spectrum. Shorter wavelength widths can be distinguished in higher spectral resolution images. Multi-spectral imagery can measure several wavelength bands such as visible green or NIR. Landsat, Quickbird and Spot satellites use multi-spectral sensors. Hyperspectral imagery measures energy in narrower and more numerous bands than multi-spectral imagery. The narrow bands of hyperspectral imagery are more sensitive to variations in energy wavelengths and therefore have a greater potential to detect crop stress than multi-spectral imagery. Multi-spectral and hyperspectral imagery are used together to provide a more complete picture of crop conditions.

**Table 4.1 Specification of LISS-III and LISS-IV sensors of IRS-P6**

		LISS-IV	LISS-III
Spatial Resolution	Band 2 (green)	5.8 m	23.5 m
	Band 3 (red)	5.8 m	23.5 m
	Band 4 (NIR)	5.8 m	23.5 m
	Band 5 (SWIR)	5.8 m	23.5 m
Radiometric resolution,	all Bands	7 bit	7 bit
Spectral Coverage	Band 2 (green)	520-590 nm	520-590 nm
	Band 3 (red)	620-680 nm	620-680 nm
	Band 4 (NIR)	770-860 nm	770-860 nm
	Band 5 (SWIR)	1550-1700 nm	1550-1700 nm

(Source: [www.euromap.de/docs/doc\\_005.html](http://www.euromap.de/docs/doc_005.html))

**Radiometric resolution** refers to the sensitivity of a remote sensor to variations in the reflectance levels. The higher the radiometric resolution of a remote sensor, the more sensitive it is to detecting small differences in reflectance values. Higher radiometric resolution allows a remote sensor to provide a more precise picture of a specific portion of the electromagnetic spectrum. Thus it refers to quantization levels of DV values in an image. Radiometric resolution is expressed in bits such 8-bit, 9-bit and 10-bit image has radiometric

resolution  $256(=2^8)$ ,  $512(=2^9)$  and  $1024(=2^{10})$  respectively. Resourcesat-2 has LISS-III and LISS-IV sensor with 10-bit radiometric resolution and AWiFS sensor with 12-bit radiometric resolution.

**Temporal resolution** refers to how often a remote sensing platform can provide coverage of an area. Geo-stationary satellites can provide continuous sensing while normal orbiting satellites can only provide data each time they pass over an area. Remote sensing taken from cameras mounted on airplanes is often used to provide data for applications requiring more frequent sensing. Cloud cover can interfere with the data from a scheduled remotely sensed data system. Remote sensors located in fields or attached to agricultural equipment can provide the most frequent temporal resolution.

Temporal resolution or the revisit time period is 5 days for LISS-IV and 24 days for LISS-III of IRS-P6, 16 days of Landsat ETM+, 2-3 days (latitude) of SPOT 5.

**Keywords:** Remote sensor, Active sensor, Passive sensor, Resolution, Radiometer, Optical scanner, Pixel.



## Lesson 5 Sensors and Tracking System

### 5.1 Optical-Infrared Sensors

Optical infrared remote sensors are used to record reflected/emitted radiation of visible, near-middle and far infrared regions of electromagnetic radiation. They can observe for wavelength extend from 400-2000 nm. Sun is the source of optical remote sensing. There are two kinds of observation methods using optical sensors: visible/near infrared remote sensing and thermal infrared remote sensing.

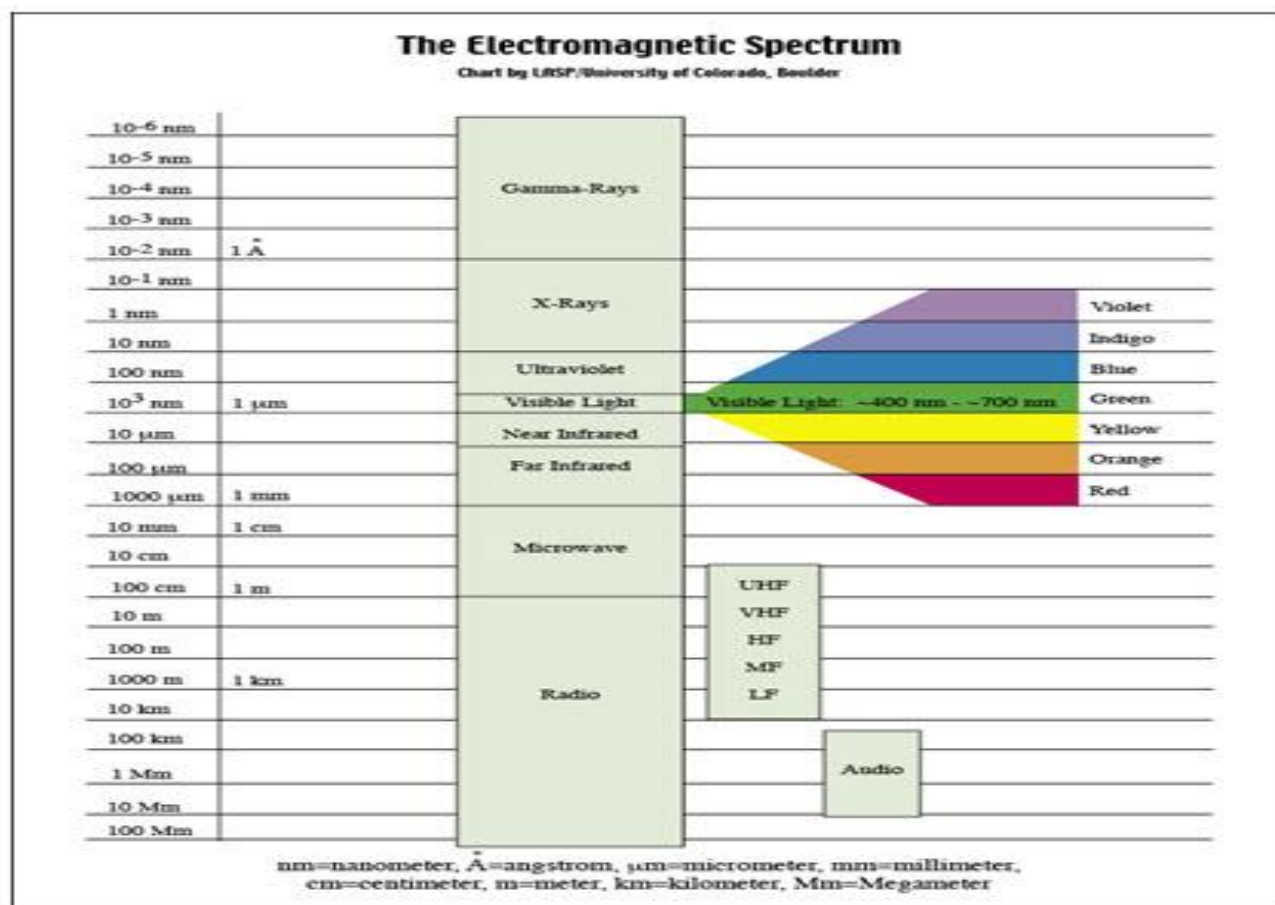
#### a) Visible/Near Infrared Remote Sensing

In this observation method visible light and near infrared rays of sunlight reflected by objects on the ground is observed. The magnitude of reflection infer the conditions of land surface, e.g., plant species and their distribution, forest farm fields, rivers, lakes, urban areas etc. In the absence of sunlight or darkness, this method cannot be used.

#### b) Thermal Infrared Remote Sensing

In thermal infrared remote sensing, the land surface radiate heat due to interaction of earth surface with solar radiation. Also this is used to observe the high temperature areas, such as volcanic activities and forest fires. Based on the strength of radiation, one can surface temperatures of land and sea, and status of volcanic activities and forest fires. This method can observe at night when there is no cloud.





**Fig. 5.1. Electromagnetic spectrum. (Source: [lasp.colorado.edu/cassini/education/Electromagnetic%20Spectrum.htm](http://lasp.colorado.edu/cassini/education/Electromagnetic%20Spectrum.htm))**

The optical remote sensing can be classified into panchromatic imaging system, multispectral imaging system and hyperspectral imaging system.

### Panchromatic Imaging System

A single channel sensor with broad wavelength range is used to detect radiation within a broad wavelength range. In panchromatic band, visible and near infrared are included. The imagery appears as a black and white photograph. The color of the target is not available. Examples of panchromatic imaging system are Landsat ETM+ PAN, SPOT HRV-PAN and IKONOS PAN, IRS-1C, IRS-1D and CARTOSAT-2A. Spectral range of Panchromatic band of ETM+ is 0.52 μm to 0.9 μm, CARTOSAT-2B is 0.45-0.85 μm, SPOT is 0.45- 0.745 μm.

### Multispectral imaging system

The multispectral imaging system uses a multichannel detectors and records radiation within a narrow range of wavelength. Both brightness and color informations are available on the image. LANDSAT, LANDSAT TM, SPOT HRV-XS and LISS etc are the examples.

## Hyperspectral imaging system

Hyperspectral imaging system records the radiation of terrain in 100s of narrow spectral bands. Therefore the spectral signature of an object can be achieved accurately, helps in object identification more precisely. Example, Hyperion data is recorded in 242 spectral bands, and AVIRIS data is recorded in 224 spectral bands.

## 5.2 Microwave Sensors

The region Microwave sensors receive microwaves, which are longer wavelength than visible light and infrared rays, and observation is not affected by day, night or weather. Microwave portion of the spectrum includes wavelengths within the approximate range of 1 mm to 1m. Thus, the longest microwaves are about 2,500,000 times longer than the shortest light waves. There are two types of observation methods using microwave sensor: active and passive.

**a) Active sensor:** The sensor emits microwaves and observes microwaves reflected by land surface features. It is used to observe mountains, valleys, surface of oceans wind, wave and ice conditions.

**b) Passive sensor:** This type of sensor records microwaves that naturally radiated from earth surface features. It is suitable to observe sea surface temperature, snow accumulation, thickness of ice, soil moisture and hydrological applications etc.

RISAT is an Indian remote sensing satellite provides microwave data.

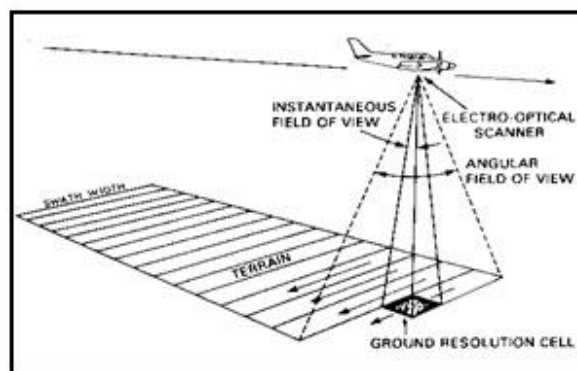
## 5.3 Scanning Mechanism

Multispectral scanner images are acquired by means of along track or across track scanning system. Depending on the way of scanning, remote sensing scanner can be categorized as: across track (whiskbroom) and along track (push broom) scanning.

### 5.3.1 Across-track scanning

Multispectral scanning systems make two dimensional images of terrain for swath beneath the platform. Across track scanners scan the earth terrain in a series of lines. The lines are oriented in the direction perpendicular to the motion of the sensor platform (i.e. across the swath). Hence in across track scanning an optical-mechanical scanner (also known as whiskbroom scanner) scans along the swath from one side to another. Swath is width of the strip of a scene along the across track direction.

The MSS and thematic mapper (TM) of Landsat, and AVHRR of NOAA are the examples of optical-mechanical scanners.



**Fig. 5.1. Across track scanner.**

(Source: [www.fao.org/docrep/003/T0355E/T0355E04.htm](http://www.fao.org/docrep/003/T0355E/T0355E04.htm))

Fig. 5.1. illustrates the operation of an across-track, or whiskbroom scanner. Such systems scan the terrain using a single or a set of detectors per band for recording the PAN and the multi-spectral signals respectively. Using a rotating or oscillating mirror, the detector starts looking at the adjacent element on the ground. Thus, information is collected pixel by pixel by the detector produce a strip or a scan line. Successive scan lines are covered as the aircraft moves forward, yielding a series of contiguous, or just touching, narrow strips of observation comprise a two dimensional image of rows (scan lines) and columns.

The length of time, the scanner “sees” the ground resolution cell is called dwell time within the system’s instantaneous field of view (IFOV). The total angle along a scan line or swath is termed as field of view (FOV). The IFOV is normally expressed as the cone angle ( $\beta$ ) within which incident energy is focused on the detector. Radiance of all the land cover types or features included in the IFOV at any given instant will be recorded by the detector. Thus, a pixel typically contains the radiance of a single land cover type and multiple land cover types termed as pure pixel and mixed pixels respectively.

If  $H$  be the sensor platform height and  $\beta$  be the IFOV, the footprint ( $D$ ) of the detector on the ground can be expressed as

$$D = H\beta \quad (5.1)$$

The ground segment sensed at any instant is called the ground resolution element or ground resolution cell. The ground resolution cell is referred to as the system’s spatial resolution. For example, the spatial resolution ( $D$ ) of a scanner having a 2.8588 -milliradian (mrad) IFOV and being operated from 822 km above the terrain will be  $822 \times 1000\text{m} \times (2.8588 \times 10^{-5} \text{ rad}) = 23.5\text{m}$ .

A larger IFOV means a greater quantity of total energy is focused on a detector as the scanners mirror sweeps across a ground resolution cell. This permits more sensitive scene radiance measurements due to higher signal levels. The result is an improvement in the radiometric resolution, or the ability to discriminate very slight energy differences. Thus, there is a trade-off between high spatial resolution and high radiometric resolution in the design of multispectral scanner systems. A large IFOV yields a signal that is much greater

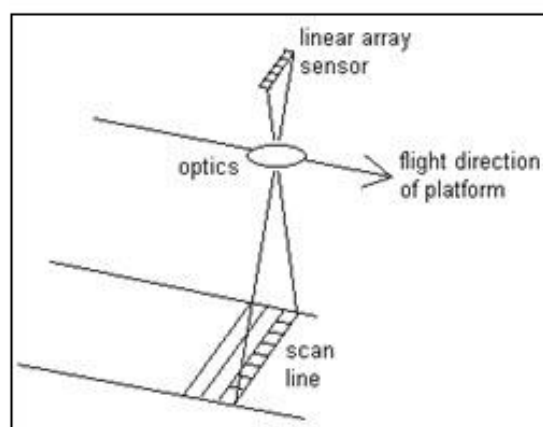
than the background electronic noise (extraneous, unwanted responses) associated with any given system. Thus, other things being equal, a system with a large IFOV will have a higher *signal-to-noise* ratio than with a small IFOV. Again, this results from the fact that a large IFOV results in a longer dwell time. What is sacrificed for these higher signals levels is spatial resolution. In a similar vein, the *signal-to-noise* ratio can be increased by broadening the wavelength band over which a given detector operates.

### 5.3.2 Along-track scanning

To achieve better spatial / spectral resolution than those provided by opto-mechanical scanners, push broom imaging systems are currently being used, referred as along track scanning. In along track scanning the detector scans the terrain of earth in the flight direction directly beneath the platform. The detector scans the whole terrain equal to width of swath in a strip or line. To build up a two dimensional image by recording successive scan lines, the detectors are oriented at right angles to the flight direction.

However, there is a distinct difference between along track and across track systems in the manner in which each scan line is recorded. In an along-track system there is a mechanical scanning mirror for pixel by pixel scanning along the swath direction. Instead a mechanical scanning mirror, a linear array of detectors is used in along track scanning system scan the whole swath. Linear arrays typically consist of numerous charge coupled devices (CCDs) positioned end to end. Each detector element of an array measures the radiance for a single ground resolution cell. Thus the size and IFOV of a single detector determine spatial resolution of the system.

Linear array CCDs are designed to be very small and a single array may contain over 10,000 individual detectors. Each spectral band of sensing requires its own linear array. Normally, the arrays are located in the focal plane of the scanner such that each scan line is viewed by all arrays simultaneously.



**Fig. 5.2: Push broom scanner.**

(Source: [www.nln.geos.ed.ac.uk/courses/english/frs/f5300/f5300008.htm](http://www.nln.geos.ed.ac.uk/courses/english/frs/f5300/f5300008.htm))

Linear array systems afford a number of advantages over across-track mirror scanning systems. First, linear arrays provide the opportunity for each detector to have a longer dwell time over which to measure the energy from each ground resolution cell. This enables a stronger signal to be recorded (and, thus, a higher signal to noise ratio) and a greater range in the signal levels that can be sensed, which leads to better radiometric resolution. In addition, the geometric integrity of linear array systems is better because of the fixed relationship among detector elements recording each scan line. The geometry along each row of data (scan line) is similar to an individual photo taken by an aerial camera. The geometric errors introduced into the sensing process by variations in the scan mirror velocity of across track scanners are not present in along track scanners. Because linear arrays are solid-state microelectronic devices, along track scanners are generally smaller in size and weight and require less power for their operation than across track scanners. Also, having no moving parts, a linear array system has higher reliability and longer life expectancy.

One disadvantage of linear array systems is the need to calibrate many more detectors. Another limitation is limited spectral range of detectors.

The first use of CCD-based pushbroom scanners on an unmanned earth-observing spacecraft was on the French SPOT-1 launched in 1986. All the LISS (Linear Imaging Self Scanning) sensors are equipped with linear array CCDs in IRS satellite sensor (IRS 1A, 1B, 1C & D).

**Keywords:** Sensor, Thermal Infra red, Microwave, Along track scanning, Across track scanning.

AgriMoon



AGRIMOON.COM  
All About Agriculture...

## Lesson 6 Remote Sensing Platforms

### 6.1 What is a Platform?

For remote sensing applications, sensors should be mounted on suitable stable platforms. These platforms can be ground based air borne or space borne based. As the platform height increases the spatial resolution and observational area increases. Thus, higher the sensor is mounted; larger the spatial resolution and synoptic view is obtained. The types or characteristics of platform depend on the type of sensor to be attached and its application. Depending on task, platform can vary from ladder to satellite. For some task sensors are also placed on ground platforms. Though aircrafts and satellites are commonly used platforms, balloons and rockets are also used.

Brief descriptions of these platforms are given in subsequent sections.

### 6.2 Type of Platforms

Three types of platforms are used to mount the remote sensors –

1. Ground Observation Platform
2. Airborne Observation Platform, and
3. Space-Borne Observation Platform

#### 6.2.1 Ground Observation Platform

Ground observation platforms are used to record detailed information about the objects or features of the earth's surface. These are developed for the scientific understanding on the signal-object and signal-sensor interactions. Ground observation includes both the laboratory and field study, used for both in designing sensors and identification and characterization of land features. Ground observation platforms include – handheld platform, cherry picker, towers, portable masts and vehicles etc. Portable handheld photographic cameras and spectroradiometers are largely used in laboratory and field experiments as a reference data and ground truth verification.

#### 6.2.2 Air Borne Based Platform

Airborne platforms were the sole non-ground-based platforms for early remote sensing work. Aircraft remote sensing system may also be referred to as sub-orbital or airborne, or aerial remote sensing system. At present, airplanes are the most common airborne platform. Other airborne observation platforms include balloons, drones (short sky spy) and high altitude sounding rockets. Helicopters are occasionally used.

### 6.2.2.1 Balloon Platform

Balloons are used for remote sensing observation (aerial photography) and nature conservation studies. The first aerial images were acquired with a camera carried aloft by a balloon in 1859. Balloon floats at a constant height of about 30 km. It consists of a rigid circular base plate for supporting the entire sensor system which is protected by an insulating and shock proof light casing. The payload used for Indian balloon experiment of three Hasselblad cameras with different film filter combinations, to provide PAN, infra red black and white and infra red false color images. Flight altitude being high compared to normal aircraft height used for aerial survey, balloon imagery gives larger synoptic views. The balloon is governed by the wind at the floating altitude. Balloons are rarely used today because they are not very stable and the course of flight is not always predictable, although small balloons carrying expendable probes are still used for some meteorological research.

### 6.2.2.2 Drone

Drone is a miniature remotely piloted aircraft. It is designed to fulfill requirements for a low cost platform, with long endurance, moderate payload capacity and capability to operate without a runway or small runway. Drone includes equipment of photography, infrared detection, radar observation and TV surveillance. It uses satellite communication link. An onboard computer controls the payload and stores data from different sensors and instruments. The payload computer utilizes a GSM/GPRS (where available) or independent satellite downlink, and can be monitored its position and payload status from anywhere in the world connected to the internet.

Drone was developed in Britain during World War-II, is the short sky spy which was originally conceived as a military reconnaissance. Now it plays important role in remote sensing. The unique advantage is that it could be accurately located above the area for which data was required and capable to provide both night and day data.

### 6.2.2.3 Aircraft

Special aircraft with cameras and sensors on vibration less platforms are traditionally used to acquire aerial photographs and images of land surface features. While low altitude aerial photography results in large scale images providing detailed information on the terrain, the high altitude smaller scale images offer advantage to cover a larger study area with low spatial resolution.

The National High Altitude Photography (NHAP) program (1978), coordinated by the US Geological Survey, started to acquire coverage of the United States with a uniform scale and format. Beside aerial photography multi spectral, hyperspectral and microwave imaging is also carried out by aircraft; thereafter multi spectral, hyperspectral and microwave imaging were also initiated.

Aircraft platforms offer an economical method of remote sensing data collection for small to large study areas with cameras, electronic imagers, across-track and along-track scanners, and radar and microwave scanners. AVIRIS hyperspectral imaging is famous aircraft aerial photographic operation of USGS.

#### 6.2.2.4 High Altitude Sounding Rockets

High altitude sounding rocket platforms are useful in assessing the reliability of the remote sensing techniques as regards their dependence on the distance from the target is concerned. Balloons have a maximum altitude of approximately 37 km, while satellites cannot orbit below 120 km. High altitude sounding rockets can be used to a moderate altitude above terrain. Imageries with moderate synoptic view can be obtained from such rockets for areas of some 500,000 square kilometers per frame. The high altitude sounding rocket is fired from a mobile launcher. During the flight its scanning work is done from a stable altitude, the payload and the spent motor are returned to the ground gently by parachute enabling the recovery of the data. One most important limitations of this system is to ensure that the descending rocket not going to cause damage.

#### 6.2.3 Space-borne Observation Platforms

In spaceborne remote sensing, sensors are mounted on-board a spacecraft (space shuttle or satellite) orbiting the earth. Space-borne or satellite platform are onetime cost effected but relatively lower cost per unit area of coverage, can acquire imagery of entire earth without taking permission. Space borne imaging ranges from altitude 250 km to 36000 km.

Spaceborne remote sensing provides the following advantages:

- Large area coverage;
- Frequent and repetitive coverage of an area of interest;
- Quantitative measurement of ground features using radiometrically calibrated sensors;
- Semi-automated computerised processing and analysis;
- Relatively lower cost per unit area of coverage.

There are two types of well recognized satellite platforms- manned satellite platform and unmanned satellite platform.

**Manned Satellite Platforms:** Manned satellite platforms are used as the last step, for rigorous testing of the remote sensors on board so that they can be finally incorporated in the unmanned satellites. This multi- level remote sensing concept is already presented. Crew in the manned satellites operates the sensors as per the program schedule. Information on a series of NASA's manned satellite programs are given in table 6.1.

**Table 6.1: Manned satellite programs of NASA**

Program	Year	Crew	Sensors used
Mercury	1962-	One	Head-heal Camera
Gemini	1963	Two	Head- held Camera
Apollo	1964-	Three	Multispectral Camera



Skylab	1965	Three	Hand-held Camera
	1968-		Multispectral Scanner
Space Shuttle	1972	Three to	Head-head Camera, LFC, Sir , MOMS
International	1973-	Seven	
Space Station	1974		
	1981	Variable	Multiple sensors for remote sensing and a range of laboratory equipments for conducting physics-chemical and biological experiments. It is planned to serve as the base for launching smaller unmanned satellites into polar orbits from which remote sensing data can be relayed to earth stations. Crew from the Space station can also go to these polar satellites to repair and refuel them. Space Shuttle to provide transportation of Astronauts and necessary cargo between earth and the Space station.
	2000	1 <sup>st</sup> Station	
	Nov 02,	Crew	
	2000	Arrived	
	Mar 10,	2 <sup>nd</sup> Station	
	2001	Crew	
	Aug 12,	Arrived	
	2001	3 <sup>rd</sup> Station	
	Dec	Crew	
	07,2002	Arrived	
	June 07,	5 <sup>th</sup> Station	
	2002	Crew	
	Nov 25,	Arrived	
	2002	6 <sup>th</sup> Station	
	Feb 01,	Crew	
	2003	Arrived	
	Apr 18,	Space	
	2003	Shuttle	
	Oct 20,	Columbia	
	2003	Disaster	
	Apr 17,	7 <sup>th</sup> Station	
	2004	Crew	
		Arrived	
		8 <sup>th</sup> Station	
		Crew	
		Arrived	
		9 <sup>th</sup> Station	
		Crew	
		Arrived	

(Source: Panda, 2005)

### Unmanned Satellite Platforms

Landsat series, SPOT series and IRS series of remote sensing satellite, NOAA series of meteorological satellites, the entire constellation of the GPS satellites and the GOES and INSAT series of geostationary environmental, communication, television broadcast, weather and earth observation satellites etc are examples of unmanned satellite category.

**Keywords:** Platform, Air borne, Space-borne, Manned Satellite, Unmanned Satellite

### *Module 3: Fundamentals of Aerial Photography*

#### **Lesson 7 Basic Principles of Photogrammetry**

##### **7.1 What is Aerial Photograph?**

Aerial photography refers to taking photograph of earth surface from space. Platform of aerial photography includes aircraft, helicopter, balloon, parachute etc. Aerial photography was first practiced by the French photographer and balloonist Gaspard-Félix Tournachon, known as "Nadar", in 1858 over Paris, France. It was the first means of remote sensing with immense application potentiality even uses now-a-days in the age of satellites with sophisticated electronic devices.

The characteristics of aerial photography that make it widely popular are:

1. **Synoptic view point:** Aerial photograph gives bird's eye view enabling to see surface features of large area and their spatial relationships.
2. **Time freezing ability:** Aerial photographs provide a permanent and objective records of the existing conditions of the earth's surface at a point of time, thus can be used for historical records.
3. **Capability to stop action:** They provide a stop action view of the dynamic conditions of earth' surface features, thus useful in studying dynamic phenomenon such as flood, forest fire, agriculture etc.
4. **Spectral resolution and spatial resolution:** Aerial photograph can be achieved sensitive to the electromagnetic (EM) wave outside the spectral sensitivity of human eye with very high spatial resolution.
5. **Three dimensional perspectives:** Stereo-scopic view can be obtained from aerial photographs enabling for both vertical and horizontal measurements.
6. **Availability:** Aerial photographs of different scales are available in websites approved by agencies involved in Aerial photography mission.
7. **Economy:** They are much cheaper than that of field survey and more accurate than maps.

The aerial photographs that have been geometrically "corrected" using ground elevations data to correct displacements caused by differences in terrain relief and camera properties are known as Orthophotos.

## 7.2 Types of Aerial Photos

Aerial photos can be distinguished depending on the position of camera axis with respect to the vertical and motion of the aircraft. Aerial photographs are divided into two major groups, vertical and oblique photos.

**i) Vertical photos:** The optical axis of the camera or camera axis is directed vertically as straight down as possible (Fig 7.1). The nadir and central point of the photograph are coincident. But in real a truly vertical aerial photograph is rarely obtainable because of unavoidable angular rotation or tilts of aircraft. The allowable tolerance is usually  $+3^\circ$  from the perpendicular (plumb) line to the camera axis. Vertical photographs are taken for most common use in remote sensing and mapping purposes.

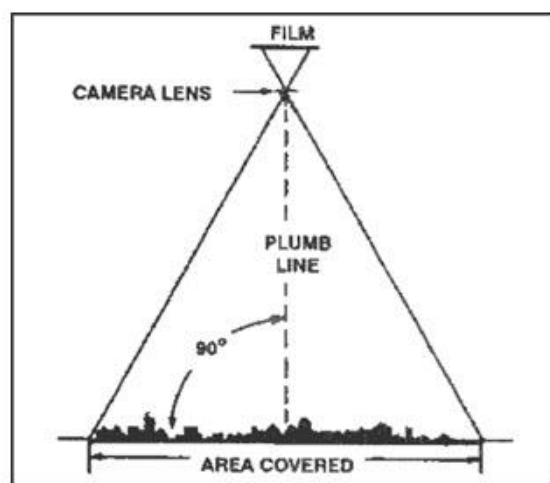


Fig. 7.1. Schematic diagram of taking a vertical photograph.

(Source: [www.map-reading.com/aptypes.php](http://www.map-reading.com/aptypes.php))

A vertical photograph has the following characteristics:

- (1) The camera axis is perpendicular to the surface of the earth.
- (2) It covers relatively small area than oblique photographs.
- (3) The shape of the ground covered on a single vertical photo closely approximates a square or rectangle.
- (4) Being a view from above, it gives an unfamiliar view of the ground.
- (5) Distance and directions may approach the accuracy of maps if taken over flat terrain.
- (6) Relief is not readily apparent.

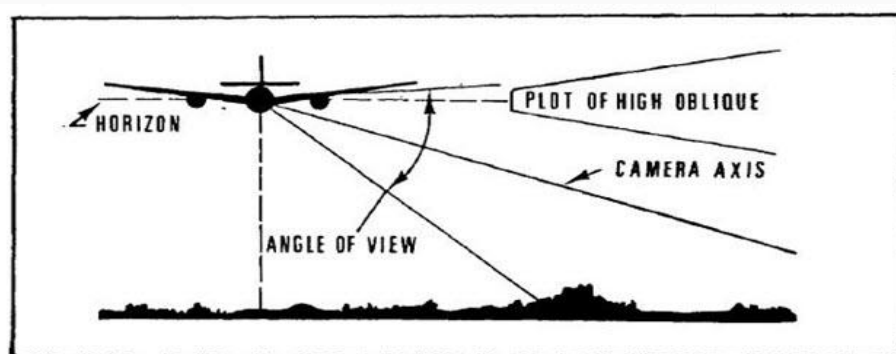
**ii) Oblique photos:** When the optical of the camera forms an angle of more than  $5^\circ$  with the vertical, oblique photographs are obtained (Fig. 7.2). The nadir and central point of the photograph are not coincident.



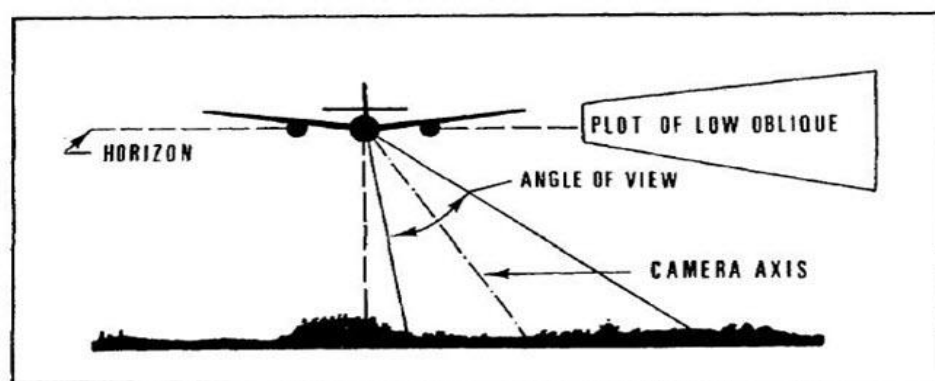
Fig. 7.2. Vertical and oblique photography.

(Source: fhf1.com/Aerial\_Photography.html)

There are two types of oblique aerial photography – high angle and low angle. In high angle oblique aerial photography the resulting images shows the apparent horizon and in low angle oblique photograph does not. Oblique photographs can be useful for covering very large areas in a single image and for depicting terrain relief and scale.



(a)



(b)

Fig. 7.3. (a) High oblique and (b) low oblique photographs.

(Source: armyintelligence.tpub.com/IT0649/IT06490010.htm)

A square outline on the ground appears as a trapezium in oblique aerial photo. These photographs can be distinguished as high oblique and low oblique (Fig.7.3). But these are not widely used for mapping as distortions in scale from foreground to the background preclude easy measurements of distance, area, and elevation.

An oblique photograph has the following characteristics:

1. Low oblique photograph covers relatively small area than high oblique photographs.
2. The ground area covered is trapezoid, but the photograph is square or rectangular. Hence scale is not applicable and direction (azimuth) also cannot be measured.
3. The relief is discernible but distorted.

### 7.3 Basic Geometric Characteristics of Aerial Photographs

Aerial photographs are taken using a camera fitted at the bottom of a aircraft along a line is termed as *flight line* or *flight strips* and the line traced on ground directly beneath the camera is called nadir line. The point on photograph where the nadir line meets the ground is termed as *principal point*. Lines drawn to connect marks located along opposite sides of the photo (*fiducial marks*) intersect precisely at the principal point. The point on the photo that falls on a line half- way between the principal point and the Nadir point is known as *isocenter*. The ground distance between the photo centers (principal points) is called *air base*.

In aerial photography, the aircraft acquires a series of exposures along each strip of multiple flight lines. Successive photographs are generally taken with some degree of overlap, which is known as endlap (Fig. 7.4). Standard endlap is 60%, which may be 80-90% in special cases such as in mountainous terrain. It ensures that each point of the ground appears in at least two successive photographs essential for stereoscopic coverage. Stereoscopic coverage consists of adjacent pairs of overlapping vertical photographs called stereopairs. Beside endlap the photograph is taken with some overlap of photographs of a strip with the adjacent strip, known as sidelap (Fig 7.5). It varies from 20% to 30% to ensure that no area of the ground is missing out to be photograph.

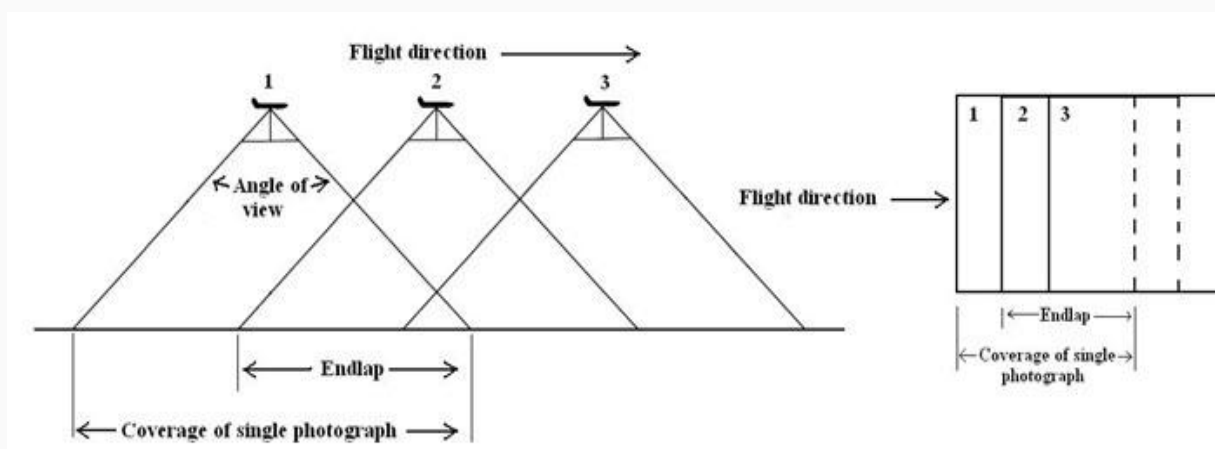
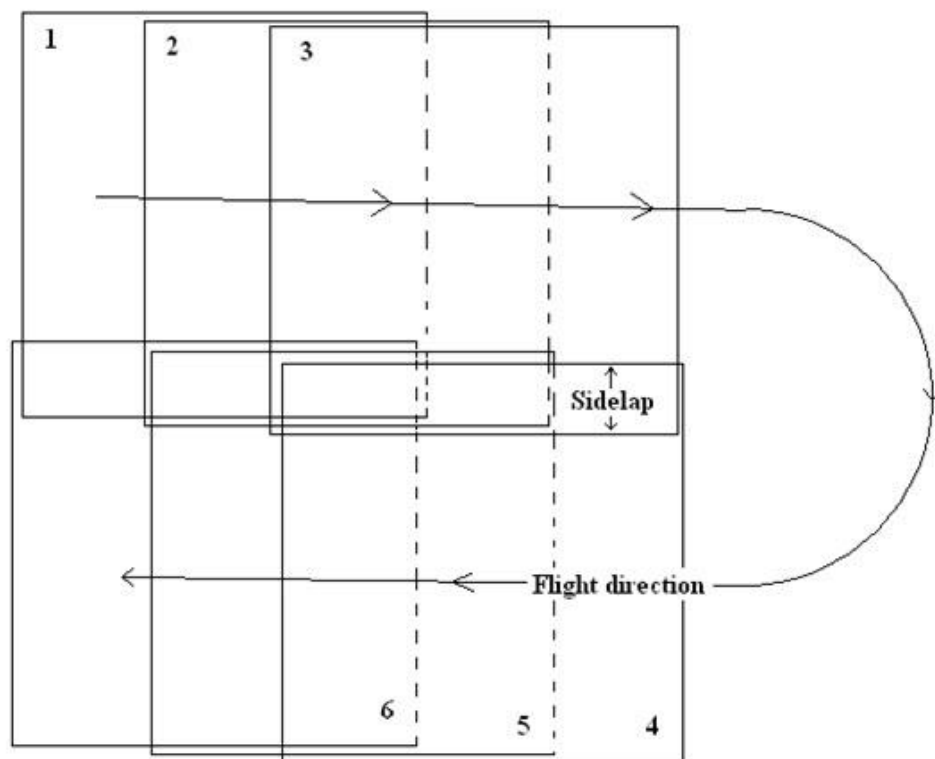


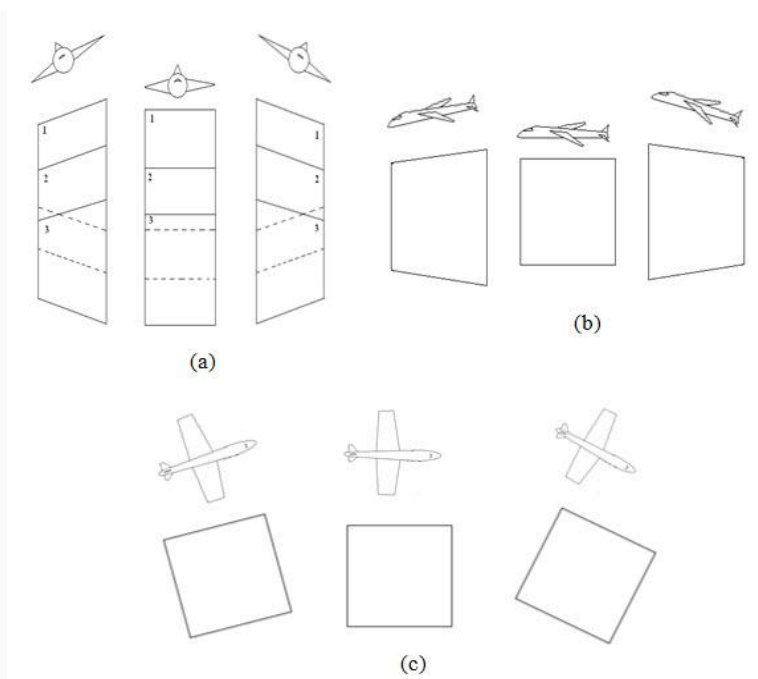
Fig. 7.4. Photographic coverage along flight line: endlap.



**Fig. 7.5. Positions of aerial photos: sidelap.**

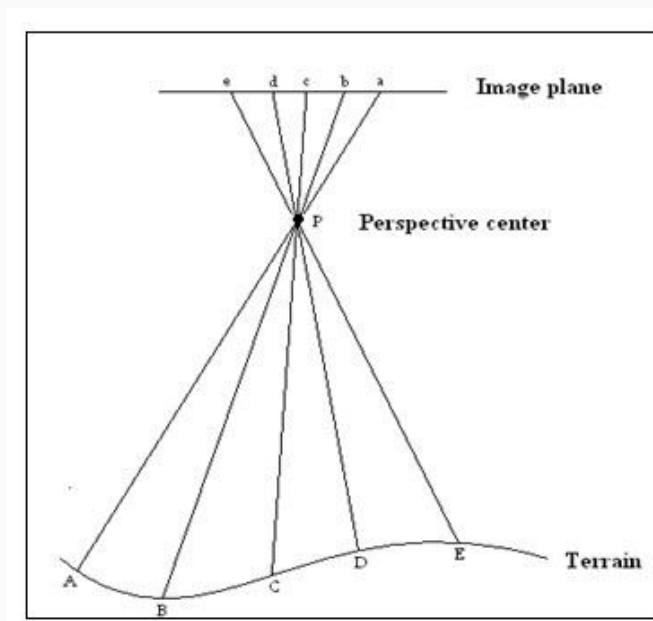
A truly vertical photograph is rarely obtained because of unavoidable angular rotations or tilts, caused by the atmospheric conditions (air pockets or currents), human error of the pilot fails to maintain a steady flight and imperfections in the camera mounting. There are three kinds of tilts as it can be seen from Fig 7.6.

1. Tilting forward and backwards (pitch)
2. Tilting sideways (roll)
3. Tilting in the direction of flight (yaw)



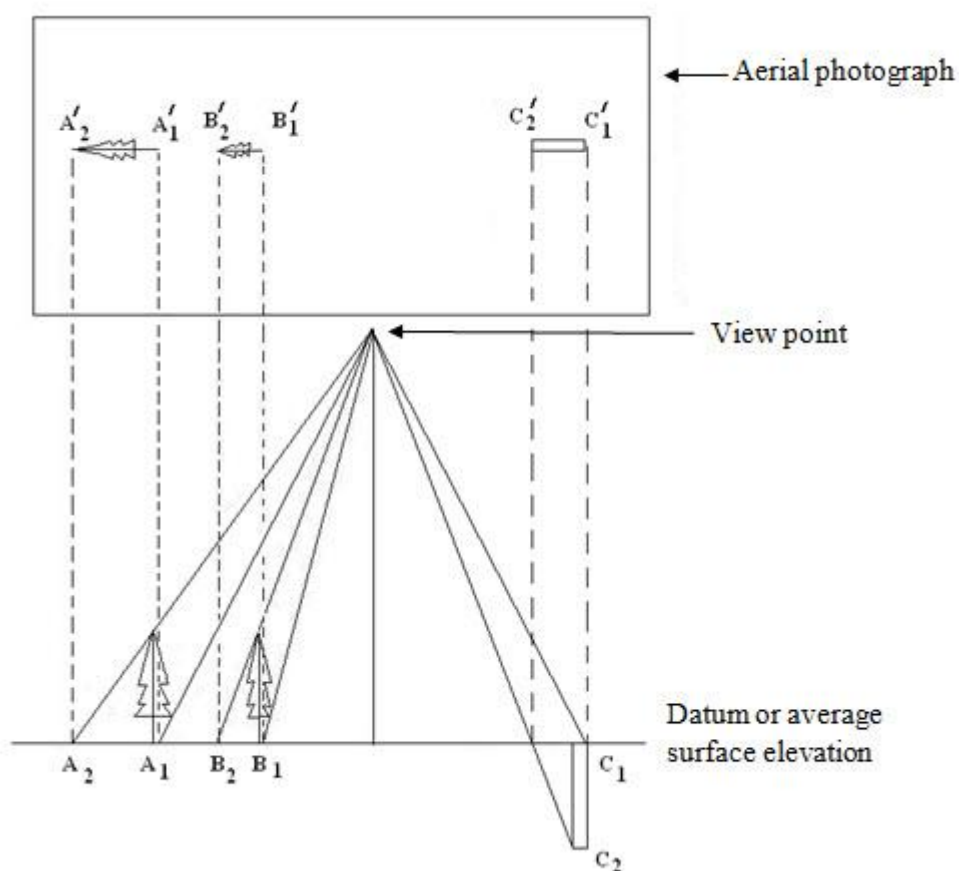
**Fig. 7.6. (a) Roll, (b) Pitch, (c) Yaw tilting of aircraft and corresponding ground coverage of aerial photograph.**

In order to understand the geometric characteristics of an aerial photograph it is also necessary to understand the viewing perspective and projection. In case of viewing perspective on a map the objects and features are both planimetrically and geometrically accurate, because the objects and features are located in the same position relative to each other as they are on the ground or on the surface of earth. But there is a change in scale. On the other hand in aerial photography, central or perspective projection system is used as it can be seen from Fig 7.7. Therefore, there are not only changes in scale but also change in relative position and geometry between the objects depending on the location of the camera.



**Fig. 7.7. Central or perspective projection.**

**Relief displacement:** Relief is the difference in elevation between the high and low points of a feature or object. Due to central perspective projection system used in aerial photography, vertical objects standing above datum (average elevation) other than principal point lean outward and objects standing below the datum (average elevation) lean inward in an aerial photograph. This distortion is called relief displacement (Fig 7.8). An aerial photograph is a three-dimensional scene transferred onto a two-dimensional plane. Thus three-dimensional squashes literally due to lack of vertical dimension. Therefore image objects above and below mean ground level or datum are displaced from their true horizontal location or relief displacement takes place. Camera tilts, earth curvature, terrain relief, object height and object position in respect to principal point are the main causes of relief displacement. Relief displacement allows the measurement of objects height from a single photo or stereopair.



**Fig. 7.8. Relief displacement in aerial photography.**

**Parallax:** The term parallax refers to the apparent change in relative position due to change in viewing position. Objects near to the observer will move faster than the objects of far. When one views objects from the window of a moving train, objects nearer to the train will move faster than objects are at far. More details are available in lesson no. 8.

**Orthophotos:** (These are also referred as Orthoimages, which are the digital version of orthophotos) that has been orthorectified. They do not contain relief displacement, tilt, scale variation caused by different errors includes aircraft movement, camera tilt, curvature of earth, elevation changes in topography etc. Therefore like photos, orthophotos show actual detail and like maps they have only one scale.



## 7.4 Photographic Scale

The scale of a map or photograph is defined as the ratio of distance measured on the map to the same distance on the ground. The amount of detail in an aerial photograph depends on the scale of the photograph. Scales may be expressed as unit equivalents or dimensionless representative fractions and ratios. For example, if 1 mm on a photograph represents 25 m on the ground, the scale of the photograph can be expressed as 1mm =25m (Unit equivalents), or 1/25,000(representative fraction) or 1:25,000 (ratio).

A convenient way to clear confusion between large scale and small scale photograph is that the same objects are smaller on a small scale photograph than on a large scale photograph. For example, two photographs having scale 1:50,000 and 1:10,000. Aerial photo with scale 1:10,000 images shows ground features at a larger, more detailed size but less ground coverage than 1:50,000 scale photo. Hence, in spite of its smaller ground coverage, the 1:10,000 photos would be termed the large scale photo.

The most straightforward method for determining photo scale is to measure the corresponding photo and ground distances between any two points (Fig 7.9). The scale S is then computed as the ratio of the photo distance d to the ground distance D.

In the Figure 7.8 the triangle  $\Delta Lab$  and  $\Delta LAB$  are similar.

Hence,  $ao/AO=Lo/LO$

or,  $d/D= f/H$

$S = d/D= \text{focal length}/\text{flying height}$

where,

S= scale of photograph

d= distance in photograph

D= distance in ground

f= focal length

H= flying height

Hence Scale of a photo  $\propto$  focal length of camera (f)

$\propto 1/\text{flying height (H)}$

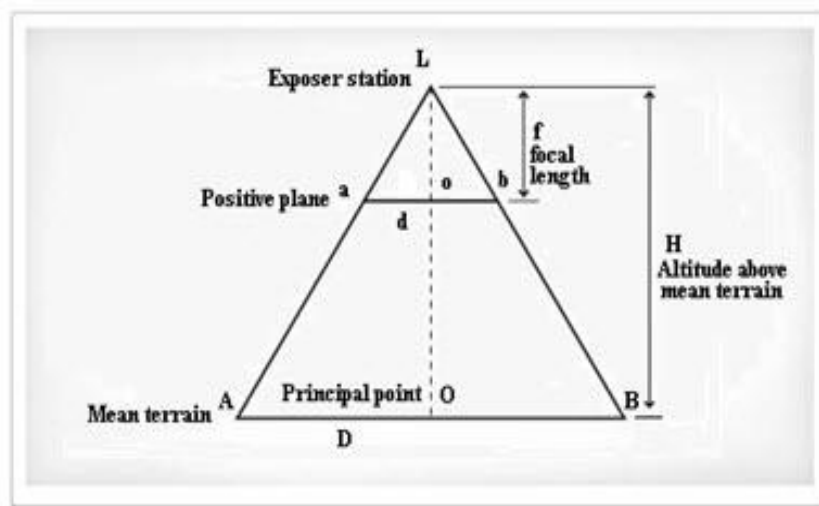


Fig. 7.9. Measurement of scale of an aerial photograph.

### 7.5 Ground Coverage of Aerial Photograph

The ground coverage of a photograph is, among other things, a function of camera format size. For example, an image taken with a camera having a 230×230 mm format (on 240 mm film) has about 17.5 times the ground area coverage of an image of equal scale taken with a camera having a 55×55-mm format (on 70-mm film) and about 61 times the ground area coverage of an image of equal scale taken with a camera having a 24×36-mm format (on 35-mm film). The ground coverage of a photo obtained with any given format is a function of focal length ( $f$ ) and flying height above ground ( $H$ ). The area covered in a photograph or ground coverage is inversely proportional to the scale of the photograph. Hence, for a constant flying height, the width of the ground area covered by a photo varies inversely with focal length and directly with flying height above terrain. Photos taken with shorter focal length lenses have larger areas of converge than taken with longer focal length lenses. On the other hand the ground coverage increases with increase in flying height.

**Keywords:** Aerial photograph, Vertical photograph, Oblique photograph, Photographic scale, Side lap, End lap.

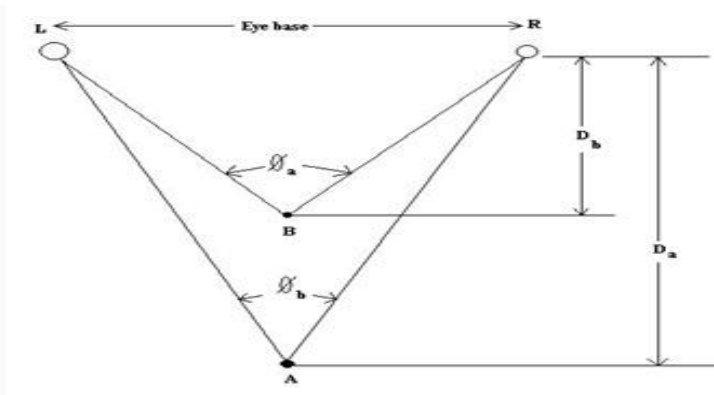
## Lesson 8 Stereoscopy and Photogrammetry

### 8.1 Definition of Stereoscopy

Stereoscopy, sometimes called stereoscopic imaging, is a technique used to enable a three dimensional effect, adding an illusion of depth to a flat image. In aerial photography, when two photographs overlap or the same ground area is photographed from two separate position forms a stereo-pair, used for three dimension viewing. Thus obtained a pair of stereoscopic photographs or images can be viewed stereoscopically. A stereoscope facilitates the stereoviewing process by looking at the left image with the left eye and the right image with the right eye. It is based on Porro-Koppe's Principle that the same light path can be generated in an optical system if a light source is projected onto the image taken by an optical system. Stereoscopic vision is constructed with a stereopair images using the relative orientation or tilt at the time of photography. Stereo viewing allows the human brain to judge and perceive in depth and volume. 3D representation of the earth's surface resulting in the collection of the geographic information with a greater accuracy compared to the monoscopic techniques.

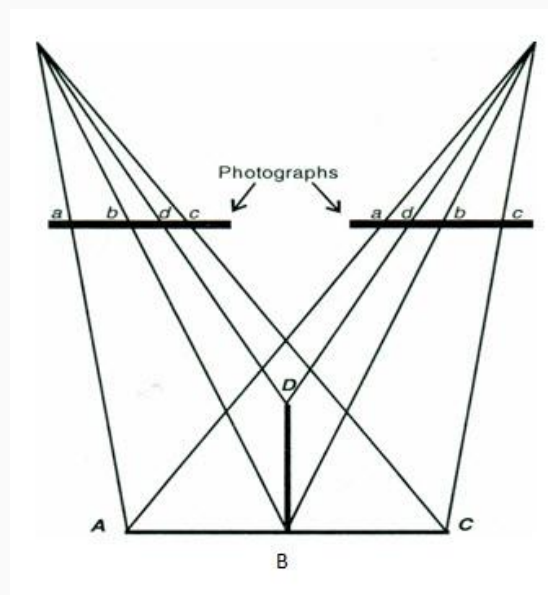
### 8.2 Stereoscopic Vision

On our daily life we unconsciously perceive and measure depth using our eyes. This stereo effect is possible because we have two eyes or binocular vision. The perception of depth through binocular vision is referred to as stereoscopic viewing, which means viewing an object from two different locations. Monoscopic or monocular vision refers to viewing surrounding objects with only one eye. Depth is perceived primarily based on the relative sizes of objects, shadow; distant objects appear smaller and behind closer objects. In stereoscopic vision, objects are viewed with both eyes a little distant from each other (approximately 65 mm) helps in viewing objects from two different positions and angles, thus a stereoscopic vision is obtained. The angle between the lines of sight of two eyes with each object known as parallax angle helps our brain in determining the relative distances between objects. Lesser the parallax angle higher the objects depth. Figure 8.1 shows the human stereoscopic vision, parallax angle  $\theta_a > \theta_b$ , helps the brain automatically to estimate the differences ( $D_a - D_b$ ) in depths between the objects A and B. This concept of distance estimation in stereoscopic vision is applied to view a pair of overlapping aerial photograph.



**Fig. 8.1. Human stereoscopic vision.**

As an example, in two photographs overlap the same region, in which objects *A*, *B* and *C* are situated at the same altitude and object *D* at a different altitude, the four objects will be observed in a different sequence in the two photographs *a*, *b*, *d*, *c* in the left photograph and *a*, *d*, *b*, *c* in the right (Fig. 8.2). In the same photograph, segments *ab* and *bc* are equal since they are at the same altitude, but segments *ad* and *dc* are not (source: Girard, 2003).



**Fig. 8.2. Perception of relief from two aerial photographs.**

(Source: Girard, 2003)

### 8.3 Stereoscopes

A stereoscope is used in conjunction with two aerial photographs taken from two different positions of the same area, (known as a stereo-pair) to produce a 3-D image. There are two types of stereoscopes: lens (or pocket) stereoscope and mirror stereoscope. Lens (or pocket) stereoscope has a limited view and therefore restricts the area that can be inspected where as in mirror stereoscope has wide view and enables a much larger area to be viewed on the stereo-pair. The most obvious feature when using a stereoscope is the enhanced vertical relief. This occurs because our eyes are only 65mm apart, but the air photos may be taken at

100s of meters apart, hence the difference in exposures is far greater than the difference between our eyes. Such an exaggeration also enables small features to become quite apparent and easily viewed.

A stereoscope (Fig. 8.3) consists of a double optical system (lenses, mirrors, prisms, etc.) mounted on a rigid frame supported on legs. In this way distance  $d$  is fixed and kept the focal distance. Thus the optical system creates a virtual image at infinity and consequently stereoscopic vision is obtained without eyestrain.

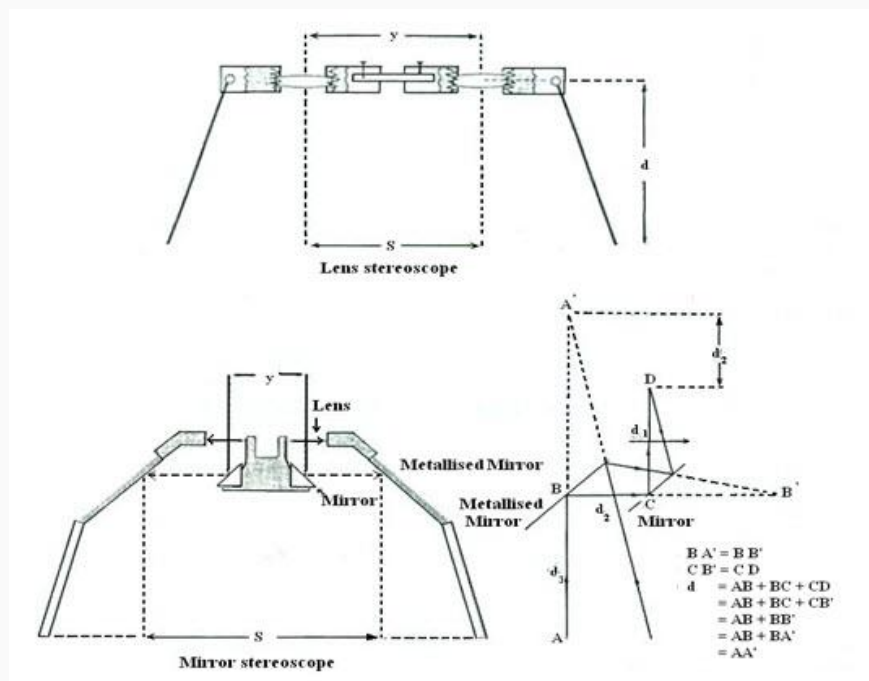


Fig. 8.3. Lens and mirror stereoscopes. (Source: Girard, 2003)

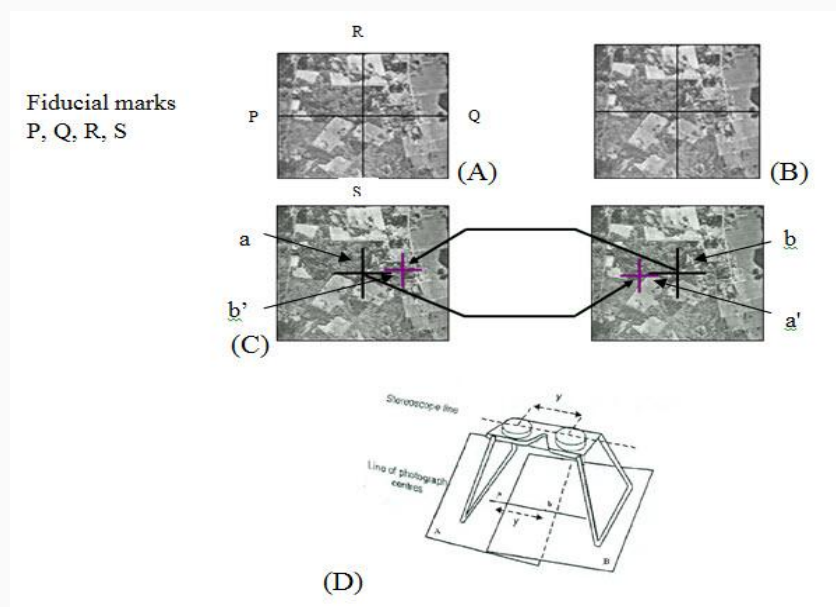
A simple lens stereoscope is made up of two achromatic convex lenses. The focal length is equal to  $d$  corresponding to the height of the stereoscope above the plane on which the stereo pair is placed. The lens spacing ( $y$ ) can allowed varying within 45 to 75 mm to accommodate individual eye spacing. The disadvantage of lens stereoscope is that the features just underneath the lens only are viewable but it has some magnification power. A mirror stereoscope comprises two metallised mirrors, two prisms, two lenses and two eyepieces having little or no magnification power. It enables viewing the optical part fixed on an arm and the photographic pairs are arranged on two different planes. They facilitate analyses of several stereo pairs consecutively without changing the arrangement in the whole overlap region compared to the lens stereoscope.

#### 8.4 Orientation of Photographs and use of Stereo Pairs

A vertical photograph is usually oriented in the same way as a map, i.e., when it is on a horizontal plane- which is normally farthest position from the observer and when it is in a vertical plane, north upwards. For quick appraisal, it is considered that the Northern hemispheres shadows are more or less directed northwards. Otherwise, points identifiable in the photo and a topographic map should be taken for identifying directions. Segments are drawn to connect various points and the photograph arranged in such a way that the

directions of the corresponding segments are parallel. It will be noticed that the directions of the segments do not correspond exactly when high relief is present (*source: Girard, 2003*).

**Use of a stereo pair:** A stereo pair consists of two consecutive photographs (A, and B) having a certain percentage of overlap, and these are placed in the manner in which they were taken during the mission; otherwise a pseudo- stereoscopy is obtained. Fiducial mark is the center of the each side of a photograph. Connecting the fiducial mark of either side, the principal point which is the geometric center of the photograph is obtained. In a stereopair the principal point of a photograph can be found on the either photograph, which is known as conjugate principal point on that photograph (marked as pink in Fig. 8.4 (C)). On each photograph a principal point and conjugate principal point are connected by a straight line. These two photographs are adjusted in the stereoscope so that that these two straight lines are coincident. To achieve good stereoscopic vision the distance of the two straight lines should be equal to the intraocular distance  $y$ .



**Fig. 8.4. (A), (B) are photographs makes a stereo pair, (C) principal points (a, b) are marked as black and conjugate principal points (a', b') are marked as pink. (D) Stereoscope. (Source: [www.cof.orst.edu/cof/teach/for220/lecture/Lecture11.ppt](http://www.cof.orst.edu/cof/teach/for220/lecture/Lecture11.ppt), and Girard, 2003)**

### 8.5 Parallax and Altitude Determination

In two successive photos taken in position  $L_1$  and  $L_2$  separated by a distance equal to the air base  $B$ , and the object height is  $h$ ;  $a$  and  $b$  are the top and bottom of the object (Fig. 8.5).  $B$  is the photo-base i.e. the distance between two successive exposures. For the convenience of the calculation, Figure 8.5 is transferred to Figure 8.6. Point  $a$  and  $b$  are projected on the joint photograph as  $x_a$ ,  $x_a'$  and  $x_b$ ,  $x_b'$  correspondingly (Fig. 8.5). The parallax of point  $a$  and  $b$  are  $P_a$  and  $P_b$  respectively. To represent this parallax, a pseudo-image is drawn at the focal distance of the camera. The parallax of point  $a$  and  $b$  is  $P_a$  and  $P_b$ .

$$P_a = x_a + x_a' \quad \text{and} \quad P_b = x_b + x_b'$$

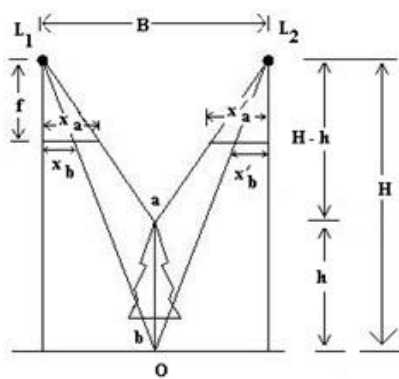


Fig. 8.5

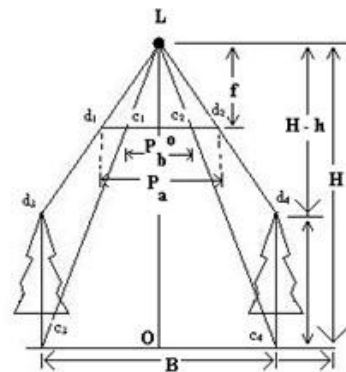


Fig. 8.6

$\Delta Lc_1c_2$  and  $\Delta Lc_3c_4$  are similar:

therefore,  $\frac{Lo}{LO} = \frac{C_1C_2}{C_3C_4}$  or,  $\frac{f}{H} = \frac{P_b}{B}$

$$\text{or, } P_b = \frac{Bf}{H}$$

Similarly, from  $\Delta Ld_1d_2$  and  $\Delta Ld_3d_4$ :  $P_a = \frac{Bf}{H-h}$

Now parallax difference ( $\Delta P$ ) =  $P_a - P_b$

$$\Delta P = P_a - P_b = \frac{Bf}{H-h} - \frac{Bf}{H}$$

$$\text{or, } \Delta P = \frac{Bf(H-H+h)}{(H-h)H}$$

$$\text{or, } \Delta P = \frac{Bf \cdot h}{(H-h)H}$$

$$\text{or, } \Delta P = \frac{P_a \cdot h}{H}$$

$$\text{or, } h = \frac{\Delta P \cdot H}{P_a}$$

Here, the object height measurement depends on the flight height and parallax of top and bottom of the object.

**Keywords:** Stereoscopy, Stereoscopic vision, Stereopair, Parallax, Stereoscope.



## Lesson 9 Image Interpretation

### 9.1 Visual Image Interpretation of Photographs and Images

Image interpretation of remote sensing data is to extract qualitative and quantitative information from the photograph or imagery. It involves identification of various objects on the terrain which may be natural or artificial consists of points, lines, or polygons. It depends on the way how different features reflect or emits the incident electromagnetic radiation and their recording by a camera or sensor. In the very beginning, when digital images and computerised classification were not available, the aerial photographs were analyzed only by visual interpretation. Accuracy of the interpretation depends on the training, experience, scale of photograph, geographic location of the study area, associated map, ground observation data etc. After the availability of satellite images, the data were categorized in two processing methods: analogue aerial photographs and digital satellite images. Though satellite images can be visually interpreted and aerial photographs can be processed by computers.

In image or photograph, some objects may be readily identifiable while other may not. It depends on individual perceptions and experience. The detail to which an image or photograph can be analyzed depends on the resolution of the image and scale of the photograph. Satellite images are generally have small scale than aerial photographs and cannot be analyzed stereoscopically.

### 9.2 Elements of Visual Interpretation

In our daily life we interpret many photos and images, but interpretation of aerial photographs and images are different because of three important aspects: (1) the portrayal of features from an overhead, often unfamiliar, perspective; (2) the frequent use of wavelengths outside of the visible portion of the spectrum; and (3) the depiction of the earth's surface at unfamiliar scales and. Eight fundamental parameters or elements are used in the interpretation of remote sensing images or photographs. These are tone or color, texture, size, shape, pattern, shadow, site and association. In some cases, a single such element is alone sufficient for successful identification; in others, the use of several elements will be required.

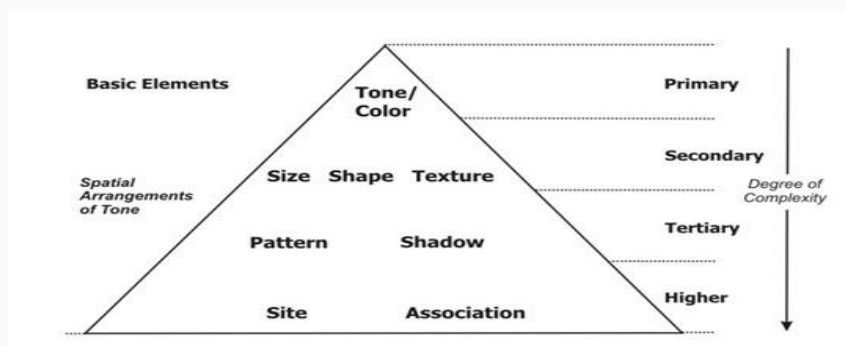
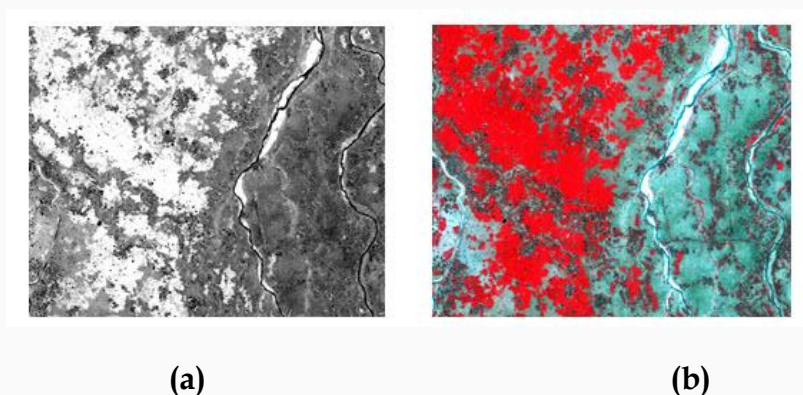


Fig. 9.1. Ordering of image elements in image interpretation.



**i) Tone or color:** Tone is the relative brightness of grey level on black and white image or color/F.C.C image. Tone is the measure of the intensity of the reflected or emitted radiation of the objects of the terrain. Lower reflected objects appear relatively dark and higher reflected objects appear bright. Figure 9.1a represents a band imaged in NIR region of the electromagnetic spectrum. Rivers does not reflect in NIR region thus appear black and the vegetation reflects much thus appears bright. Our eyes can discriminate only 16-20 grey levels in the black and white photograph, while more than hundreds of color can be distinguished in a color photograph. In multispectral imaging, optimal three bands are used to generate color composite image. False Color Composite (FCC) using NIR, red and green are most preferred combination for visual interpretation. In a standard FCC, NIR band passes through red channel, red band passes through green channel and green band passes through blue channel. Vegetation reflects much in NIR region of the electromagnetic spectrum therefore in standard FCC vegetation appears red (Fig. 9.1b), which is more suitable in vegetation identification.



**Fig. 9.2. Satellite image of area in (a) grey scale and in (b) standard FCC.**

**ii) Texture:** Texture refers to the frequency of tonal variation in an image. Texture is produced by an aggregate unit of features which may be too small to be clearly discerned individually on the image. It depends on shape, size, pattern and shadow of terrain features. Texture is always scale or resolution dependent. Same reflected objects may have difference in texture helps in their identification. As an example in a high resolution image grassland and tree crowns have similar tone, but grassland will have smooth texture compared to tree. Smooth texture refers to less tonal variation and rough texture refers to abrupt tonal variation in an imagery or photograph.



In the Fig. 9.3, the similar kinds of reflectance/color (green) appear throughout the image, three distinct land cover types can be seen from the image texture. The triangular patch at the bottom left corner is the plantation which has rough texture where individual trees can be seen. Right side on the top of the image, the trees are closer together, and the tree canopies merge together, forming medium textural pattern and at the right bottom corner with smooth texture indicating that it is probably an open field with short grass.

**Fig. 9.3. High resolution image showing different textures.**

(Source: [www.crisp.nus.edu.sg/~research/tutorial/opt\\_int.htm](http://www.crisp.nus.edu.sg/~research/tutorial/opt_int.htm))

**iii) Pattern:** Pattern refers to the spatial arrangement of the objects. Objects both natural and manmade have a pattern which aids in their recognition. The repetition of certain general form or relationship in tones and texture creates a pattern, which is characteristic of this element in image interpretation. In the Fig. 9.3 it could be easily understood that at the left bottom corner of the image, it is plantation, where the trees are nearly equally spaced. Whereas at the upper right and bottom right corners show natural vegetation.

**iv) Size:** Size of objects on images must be considered in the context of the image scale or resolution. It is important to assess the size of a target relative to other objects in the scene, as well as the absolute size, to aid in the interpretation of that target. A quick approximation of target size can make direct interpretation to an appropriate result more quickly. The most measured parameters are length, width, perimeter, area, and occasionally volume. For example, if an interpreter had to distinguish zones of land use, and had identified an area with a number of buildings in it, large buildings such as factories or warehouses would suggest commercial property, whereas small buildings would indicate residential use.



**Fig. 9.4. Satellite view of a part of a city.**

**(Source: [parallelsirals.blogspot.in/2010\\_05\\_01\\_archive.html](http://parallelsirals.blogspot.in/2010_05_01_archive.html))**

**v) Shape:** Shape refers to the general form, configuration or outline of an individual object. Shape is one of the most important single factors for recognizing object from an image. Generally regular shapes, squares, rectangles, circles are signs of man-made objects, e.g., buildings, roads, and cultivated fields, whereas irregular shapes, with no distinct geometrical pattern are signs of a natural environment, e.g., a river, forest. In a general case of misinterpretation in between roads and train line: roads can have sharp turns, joints perpendicularly, but rails line does not. From the shape of the following image, it can be easily said that the dark-blue colored object is a river.



**Fig. 9.5. Satellite image of an area**

**(Source: [geology.com/satellite/landsat-images-water.shtml](http://geology.com/satellite/landsat-images-water.shtml))**

**vi) Shadow:** Shadow is a helpful element in image interpretation. It also creates difficulties for some objects in their identification in the image. Knowing the time of photography, we can estimate the solar elevation/illumination, which helps in height estimation of objects. The outline or shape of a shadow affords an impression of the profile view of objects. But objects within shadow become difficult to interpret. Shadow is also useful for enhancing or identifying topography and landforms, particularly in radar imagery.



**Fig. 9.6. Shadow of objects used for interpretation. (Source: [wiki.landscapetoolbox.org/doku.php/remote\\_sensing\\_methods:image\\_interpretation](http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:image_interpretation))**

**vii) Association:** Association refers to the occurrence of certain features in relation to others objects in the imagery. In urban area a smooth vegetation pattern generally refers to a play ground or grass land not agricultural land (Fig 9.7).



**Fig. 9.7. Satellite image of an urban area.**

**(Source: [news.discovery.com/earth/haiti-satellite-earthquake-damage.html](http://news.discovery.com/earth/haiti-satellite-earthquake-damage.html))**



**viii) Site:** Site refers to topographic or geographic location. It is also an important element in image interpretation when objects are not clearly identified using the previous the elements. A very high reflectance feature in the Himalayan valley may be snow or cloud, but in Kerala one cannot say it as snow.

### 9.3 Interpretation Keys

The criterion for identification of an object with interpretation elements is called an interpretation key. The image interpretation depends on the interpretation key which an experienced interpreter has established from prior knowledge and utilizes in the interpretation of the current images. It provides guidance about the correct identification of features or conditions on the images. Generally, eight standardized keys are established to eliminate the difference between different interpreters. The eight interpretation elements are: size, shape, shadow, tone, colour, texture, pattern, and association. For agricultural and tree species identification a number of keys have been successfully employed used on a region-by-region and season-by-season basis, as vegetation can widely vary depending on location and season. Besides these, the time the photograph is taken, film type, and photo-scale should be carefully considered while developing interpretation keys.

The Table 9.1 shows an example of interpretation keys for forestry mapping. The keys are specified with respect to the crown's shape, rim shape of the crown, tone, shadow, projected, tree shape, pattern, texture, and other factors.

**Table 9.1 Interpretation keys for forestry mapping**

Species	Crown shape	Edge of Crown	Tone	Pattern	Texture
Cedar	Conical with sharp spear	Circular and sharp	Dark	Spotted grain	Hard and coarse
Cypress	Conical with round crown	Circular but not sharp	Dark but lighter than cedar	Spotted	Hard and fine
Pine	Cylindrical with shapeless crown	Circular but unclear	Light and unclear	Irregularly spotted	Soft but coarse
Larch	Conical with unclear crown	Circular with unclear edge	Lighter than cypress	Spotted	Soft and fine
Fir/spruce	Conical with wide crown	Circular with zig-zag edge	Dark and clear	Irregular	Coarse
Deciduous	Irregular shapes	Unclear	Lighter	Irregular	Coarse

(Source: Bhatta, 2008)

## 9.4 Generation of Thematic Maps

An image interpretation map is usually produced by transferring the interpreted information to a base map which has been prepared in advance. The requirements of the base map should be as follows:

1. Proper map scale to enable appropriate presentation of interpreted information.
2. Geographic coordinate system to establish the geographic reference.
3. Basic map information to be printed in light tones as background which results in enhancement of interpreted information

Normally, a topographic map (Toposheet), plan map, or orthophotomap is used as a base map.

A topographic map with a scale of 1:50,000 to 1:250,000 is usually the preferable base map for higher resolution satellite image interpretation. For oceanographic purposes or marine science, a scale of 1:50,000 to 1:500,000 may be used as the base map.

Orthophotomap are more easily used by cartographers for the transfer of interpreted information. For example, in case of forest classification.

The methods of transfer of information to a base map are as follows:

- The interpreted image is traced on to a base map by overlaying on a light table.
- The interpreted image is projected via a lens and a mirror onto a base map optical projection. The optical zoom transfer scope or mirror projector is very useful for image interpretation.
- Using grid system grid lines are drawn on both image and base map. Then the interpreted information in a grid on the image is transferred to the corresponding grid on the map.
- Photogrammetric plotting aerial photographs are interpreted into a thematic map using a photogrammetric plotter.

**Keywords:** Visual image interpretation, Elements, FCC, Interpretation keys, Thematic maps.



Module 4: Digital Image Processing

Lesson 10 Digital Image

10.1 Introduction to Digital Image

A digital image is a two dimensional array of discrete image elements or pixels representing spatial distribution of different parameters such as electromagnetic radiation, emission, temperature, or some geophysical or topographical elevation etc. Each pixel represented by a specific row ( $i$ ) and column ( $j$ ) value in an array or matrix. Each pixel associated with a pixel value is also called brightness value ( $BV_{ij}$ ), represented as Digital Number (DN), which are stored as binary digits in a computer. This  $BV_{ij}$ 's represented in a gray level or gray scale is in certain ranges such as 0-255 (8-bit image:  $2^8=256$ ) in a black and white image. 0 and 255 represented as completely black and white respectively. For colour pictures three image matrices (as parallel layers) with same ranges are required. A single image can be represented as a black and white image in a two dimensional (2D) array (matrix) of pixels having rows ( $i$ ) and columns ( $j$ ). A multi-spectral or hyper- spectral image can be represented in a 3D array of pixels having rows ( $i$ ), columns ( $j$ ) and bands ( $k$ ). Fig. 10.1. shows a digital image configuration of a single band.

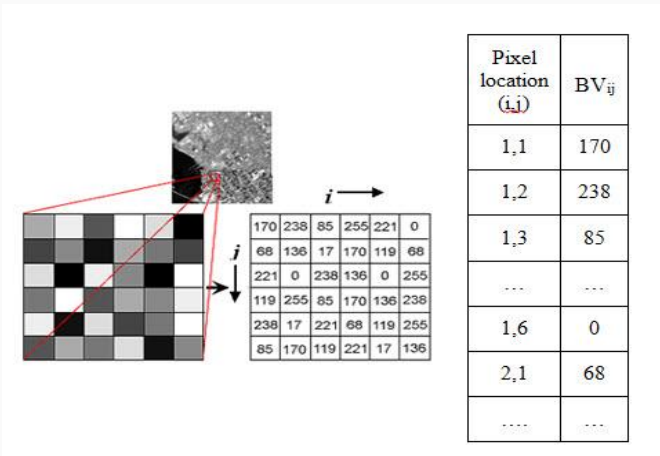


Fig. 10.1. Digital image configuration.

The features of the earth's surface are imaged simultaneously by different sensors with filters corresponding to different wavelength bandwidth recorded as analog signal. This analog signal is converted in digital format using Analog to Digital converter and scale into different ranges like 8-bit ( $2^8=256$ ), 9-bit ( $2^9=512$ ), 10-bit ( $2^{10}=1024$ ). A digital image is a data matrix

having integer values representing the average of energy reflected/emitted from a square area on the surface of earth.

## 10.2 Image Rectification and Restoration

Remotely sensed images are taken from a great distance from the surface of the earth affected by various parameters such as atmosphere, solar illumination, earth rotation, sensor motion with its optical, mechanical, electrical components etc. which causes distortion in the imagery. The intent of image rectification and restoration is to correct these distortions arises in image acquisition process both geometrically and radiometrically. Obviously, the nature of such procedures varies considerably with factors such as the digital image acquisition type (digital camera, along-track scanner, across-track scanner), platform (airborne, satellite), atmosphere and total field of view.

## 10.3 Geometric Correction

Raw digital images acquired by earth observation systems usually contain geometric distortions so they do not reproduce the image of a grid on the surface faithfully. Geometric distortions are the errors in the position of a pixel relative to other pixels in the scene and with respect to their absolute position in a particular projection system.

The geometric distortions are normally two types: (i) systematic or predictable, and (ii) random or unpredictable.

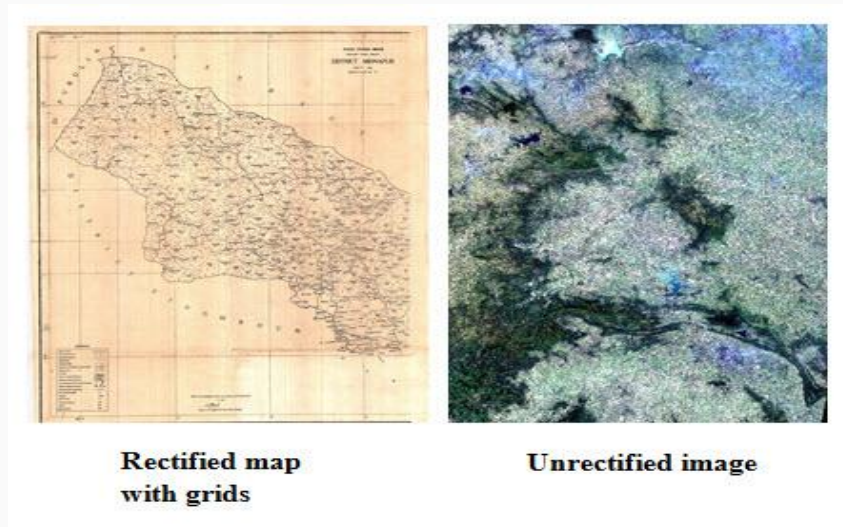
Due to rotation of earth and platform's forward motion the scan lines are not perpendicular to the direction of ground track, which makes the scanned area skewed and scale variation due to earth's curvature found in both push-broom and whisk-broom scanner. But in case of whisk-broom scanner, there are additional two errors founds (a) scale variation along the scan direction due to its scanning process, (b) error due to nonlinearity in scan mirror velocity.

Random distortions and residual unknown systematic distortions are difficult to account mathematically. These are corrected using a procedure called rectification.

Rectification is process of correcting an image so that it can be represented on a plain surface. It includes the process known as rubber sheeting, which involves stretching and warping image to georegister using control points shown in the image and known control points on the ground called Ground Control Point (GCP). Rectification is also called georeferencing.

The image analyst must be able to obtain both the distinct sets of co-ordinates: image co-ordinates and map co-ordinates (x,y) associated with GCPs. These values are used in a least squares regression analysis to determine geometric transformation coefficients that can be used to interrelate the geometrically correct (map) coordinates and the distorted- image coordinates. The GCPs are chosen such that these can be easily identified on the image and whose geographic co-ordinates can be obtained from ground (using GPS) or from existing geo-referenced map or image. The two common geometric correction procedures generally used are: image-to-map rectification and image-to-image registration, which uses an existing

georeferenced map and image correspondingly. Fig 10.2 shows rectified map which can be used to rectify an unrectified image.



**Fig. 10.2. Image to map rectification.**

Two basic operations are performed for geometric correction: (1) Spatial interpolation and (2) Intensity interpolation. Spatial interpolation accounts for the pixel position correction and intensity interpolation accounts for the DN value assignment from input image to its corresponding output pixel position.

### **(1) Spatial interpolation**

The first process of geometric correction involves the pixel position correction using rubber sheeting procedure. The input and output image co-ordinate systems can be related as:

$$x = f_1(X, Y) \text{ and } y = f_2(X, Y) \quad (10.1)$$

where,

$(x, y)$  = distorted input image coordinates (column, row)

$(X, Y)$  = correct image coordinates

$f_1, f_2$  = transformation functions

The mapping function could be polynomial. A first degree polynomial can model six kinds of distortions: translation and scale changes in  $x$  and  $y$ , skew and rotation.

The 1<sup>st</sup> degree polynomial (i.e. linear equation) can be mathematically expressed as

$$x = a_0 + a_1 X + a_2 Y \quad (10.2)$$

$$y = b_0 + b_1 X + b_2 Y \quad (10.3)$$

The co-efficients  $a_i$  and  $b_i$  are evaluated using GCPs.



## (2) Intensity interpolation

Intensity interpolation involves the extraction of DN value from input image to locate it in the output image to its appropriate location. Fig 10.3 used as an example, for an output pixel (3, 5) grey shaded, the corresponding input pixel is (2.4, 4.8). This means the output image may not have one-to-one match with the input image. Therefore an interpolation process is applied which is known as resampling. This process accounts the position of output pixel to its corresponding position of input pixel and the DN values of input pixel and their surroundings.

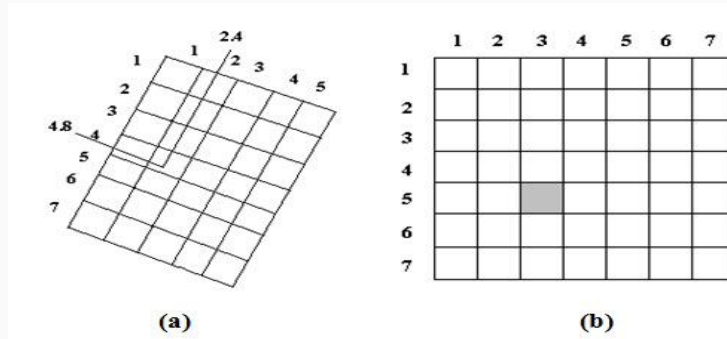


Fig. 10.3. a) Distorted image; b) Corrected image pixels matrix.

There are generally three kinds of resampling process-

### a) Nearest-neighbour interpolation

Let for an output pixel (3, 5), in Fig 10.3 the input pixel location is (2.4, 4.8). In nearest-neighbor interpolation, the output pixel (3, 5) will be filled up with DN value nearest to the input pixel location (2.4, 4.8) in the present case is (2, 5). In this process the radiometric values or DN values do not alter and it is computationally efficient. But features in the output matrix may be offset spatially by  $\pm 0.5$  pixel, thus a 'patchy' appearance arise at the edges and in linear features.

### b) Bilinear interpolation

Bilinear interpolation finds brightness values in two orthogonal direction of the desired input location (2.4, 4.8) in the input image. Distance-weighted average of DN values of four nearest pixels ((1, 4), (1, 5), (2, 4), (2, 5)) is computed in the input distorted image. Closer to the input location will have more weight; DN value of highest weighted pixel will be assigned to the output pixel.

$$BV_{wt} = \frac{\sum_{k=1}^4 \frac{R_k}{D_k^2}}{\sum_{k=1}^4 \frac{1}{D_k^2}} \quad (10.4)$$

where,  $R_k$  are the surrounding four DN values,  $D_k^2$  are the distance squared from the desired input location (2.4, 3.8).

This technique alters the DN value and generates smoother appearance in the resampled image.

### c) Cubic convolution interpolation

An improved restoration of the image is obtained in cubic convolution interpolation process. Similar to bilinear interpolation this technique is uses 16 neighbor pixel in the input image. It provides smoother appearance and sharpens edge and also changes the DN values but requires more computational time than previous processes.

Any interpolation is a low pass filtering operation and it produces a loss of higher frequency information depending on the method used.

## 10.4 Radiometric Correction and Noise Removal

Radiance value recorded for each pixel represents the reflectance/emittance property of the surface features. But the recorded value does not coincide with the actual reflectance/emittance of the objects due to some factors like: sun's azimuth and elevation, viewing geometry, atmospheric conditions, sensor characteristics, etc. To obtain the actual reflectance/emittance of surface features, radiometric distortions must be compensated. The main purpose of radiometric correction is to reduce the influence of errors. An error due to sensor's characteristic is also referred as noise.

Radiometric correction is classified into the following three types:

- 1) Sun angle and topographic correction
- 2) Atmospheric correction
- 3) Detector response calibration
  - a) De-stripping
  - b) Removal of missing scan line
  - c) Random noise removal
  - d) Vignetting removal

**1) Sun angle and topographic correction:** The sun elevation correction accounts for the seasonal position of the sun relative to the earth (Fig. 10.4). The correction is usually applied by dividing each pixel value in a scene by the sine of the solar elevation angle or the cosine of zenith angle (which is simple  $90^\circ$  minus the solar elevation angle).

$$DN_{\text{correction}} = DN / \sin\alpha \quad (10.5)$$

$$DN_{\text{correction}} = DN / \cos\theta \quad (10.6)$$

$\alpha$  = solar elevation angle,  $\theta$  = zenith angle ( $90^\circ$ - elevation angle)

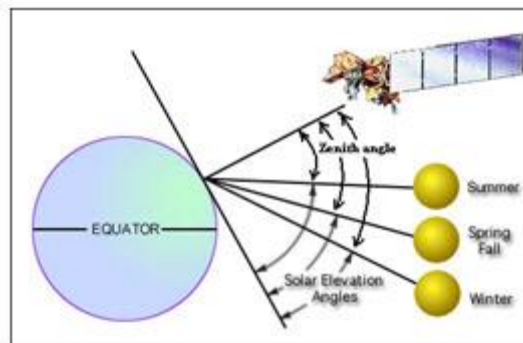


Fig. 10.4. Effects of seasonal changes on solar elevation angle.

(Source: [landsathandbook.gsfc.nasa.gov/data\\_properties/prog\\_sect6\\_3.html](https://landsathandbook.gsfc.nasa.gov/data_properties/prog_sect6_3.html); and modified )

The earth- sun distance correction is applied to normalize for the seasonal changes in the distance between the earth and the sun. The earth- sun distance is usually expressed in astronomical units. (An astronomical unit is equivalent to the mean distance the earth and the sun, approximately  $149.6 \times 10^6 \text{ km}$ ). The irradiance from the sun decreases as the square of the earth-sun distance (*Lillesand and Kiefer, 2003*).

**2) Atmospheric correction:** During transmission of EM wave from sun to objects and from objects to sensor through atmosphere, various kinds of absorption and scattering takes place. Objects receive not only the direct solar illumination, but also the scattered radiation from atmosphere and neighbouring objects and similarly the sensor receives reflection or radiation from target, its neighbouring objects, and scattered radiation from atmosphere (which is called path radiance or haze).

Therefore the sensor receives a composite reflection/ radiation which can be expressed as-

$$L_a = \frac{\rho ET}{\pi} + L_p \quad (10.7)$$

where,

$L_a$  = apparent spectral radiance

measured by sensor

$\rho$  = reflectance of objects

$E$  = irradiance on object

$T$  = transmission of atmosphere

$L_p$  = path radiance / haze

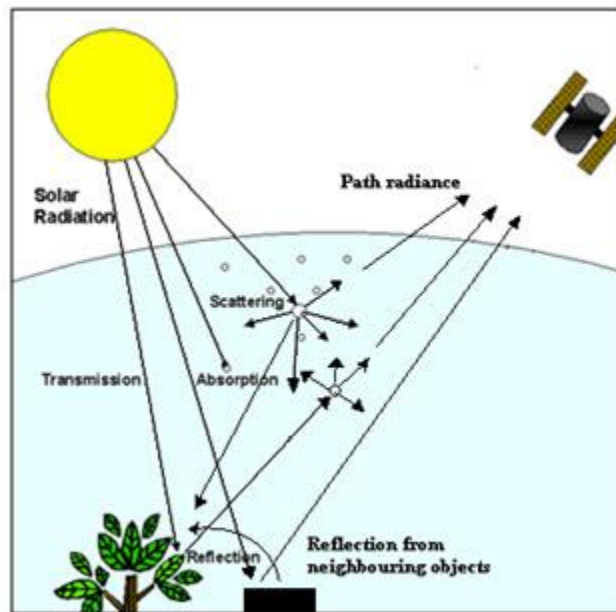


Fig. 10.5. Effect of solar illumination in atmosphere

For multispectral and hyperspectral data, the approaches of atmospheric correction is different.

The errors due to solar elevation, sun-earth distance and atmospheric effects can be reduced in two procedures: (i) absolute correction and (ii) relative correction.

### Absolute correction

The absolute correction procedure of Landsat data is stated below:

The radiance recorded by a sensor is converted in DN values in a range 0 – 255 (in 8 bit data). Therefore to obtain reflectance value of terrain objects it is required to convert DN values back to its equivalent the radiance value, and radiance value to reflectance value.

$$L_{\lambda} = \left( \frac{LMAX_{\lambda} - LMIN_{\lambda}}{Q_{cal\ max} - Q_{cal\ min}} \right) (Q_{cal} - Q_{cal\ min}) + LMIN_{\lambda} \dots\dots\dots (10.8)$$

where,

$L_{\lambda}$  = Spectral radiance at the sensor's aperture

$LMAX_{\lambda}$  = Spectral at-sensor radiance that is scaled to  $Q_{cal\ min}$ .

$LMIN_{\lambda}$  = Spectral at-sensor radiance that is scaled to  $Q_{cal\ max}$ .

$Q_{cal}$  = Quantized calibrated pixel value (DN).

$Q_{cal\ max}$  = Maximum quantized calibrated pixel value corresponding to  $LMIN_{\lambda}$ .

$Q_{cal\ min}$  = Maximum quantized calibrated pixel value corresponding to  $LMAX_{\lambda}$ .

$$\rho_{\lambda} = \frac{\pi L_{\lambda} d^2}{ESUN_{\lambda} \cos \theta_s} \dots\dots(10.9)$$

where,

$\rho_{\lambda}$  = TOA reflectance

$L_{\lambda}$  = Spectral radiance at the sensor's aperture

$d$  = Earth-Sun distance

$ESUN_{\lambda}$  = Mean exoatmospheric solar irradiance

$\theta_s$  = Solar zenith angle

The values of  $L_{\lambda}$ ,  $LMAX_{\lambda}$ ,  $LMIN_{\lambda}$ ,  $Q_{cal}$ ,  $Q_{cal\ max}$ ,  $Q_{cal\ min}$ ,  $\theta_s$  are obtained from the sensor parameters and satellite position.

**Table 10.1 Solar spectral irradiances for bands of ETM+ sensor**

Band	watts/(meter squared * $\mu m$ )
1	1969.000
2	1840.000
3	1551.000
4	1044.000
5	225.700
7	82.07
8	1368.000

(Source: [landsathandbook.gsfc.nasa.gov/data\\_prod/prog\\_sect11\\_3.html](https://landsathandbook.gsfc.nasa.gov/data_prod/prog_sect11_3.html))

**Table 10.2 Earth-Sun distance in astronomical units**

Julian Day	Distance	Julian Day	Distance	Julian Day	Distance	Julian Day	Distance	Julian Day	Distance
1	0.9832	74	0.9945	152	1.0140	227	1.0128	305	0.9925
15	0.9836	91	0.9993	166	1.0158	242	1.0092	319	0.9892

32	0.9853	106	1.0033	182	1.0167	258	1.0057	335	0.9860
46	0.9878	121	1.0076	196	1.0165	274	1.0011	349	0.9843
60	0.9909	135	1.0109	213	1.0149	288	0.9972	365	0.9833

(Source: [landsathandbook.gsfc.nasa.gov/data\\_prod/prog\\_sect11\\_3.html](http://landsathandbook.gsfc.nasa.gov/data_prod/prog_sect11_3.html))

**Noise Removal:** Image noise is any unwanted disturbance in image data that is due to limitations in the sensing, signal digitization, or data recording process. Radiometric errors due to detector response calibration are also termed as noise.

### 3) Detector response calibration

**Destriping:** Striping or banding in imagery occurs due to error in sensor adjustment or inability in sensing the radiation by sensors. This error was common in Landsat MSS sensor, where six detectors were used for each band drifted over time, resulting in relatively higher or lower values along every sixth line in the image. One common method of striping error correction is done by comparing the histogram between a set of data. A histogram is evaluated from scan lines 1, 7, 13, and so on; for a second lines 2, 8, 14, and so on; and so forth. These histograms are then compared in terms of their mean and standard deviations of six detectors to identify the problem detector(s) for the problem lines to resemble those for the normal data lines. Fig. 10.6(a) shows a stripped image and Fig. 10.6(b) is the output result after de-striping algorithm is applied.



Fig. 10.6. Destriping: (a) striped image, (b) destriped image.

(Source: [blamannen.wordpress.com/2011/07/12/destripe-landsat-7-etm-some-thoughts/](http://blamannen.wordpress.com/2011/07/12/destripe-landsat-7-etm-some-thoughts/))

**Line drop:** Line drop correction is similar to strip error. Line drop error is found when a sensor completely fails to scan. This can be minimized by normalizing with respect to their neighboring features or image values i.e. the data for missing lines is replaced by the average image value of neighbors. Fig. 10.7 shows the line drop correction.



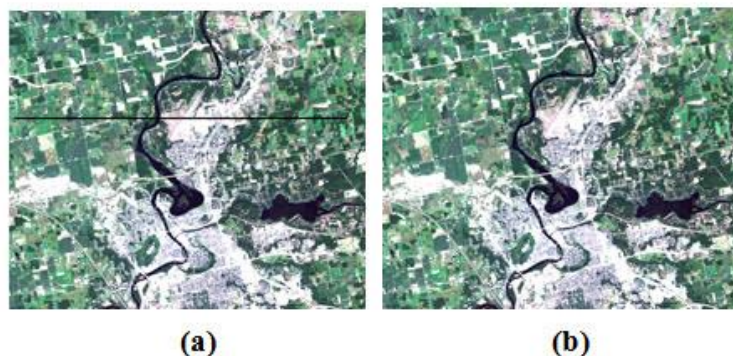


Fig. 10.7. Line drop correction: (a) original image (b) restored image.

**Random noise removal:** Random noise is not systematic. The noise in a pixel does not depend on the spatial location of the pixel in the imagery. It causes images to have a “salt and pepper” or “snowy” appearance. Moving window of  $3 \times 3$  or  $5 \times 5$  pixels is used for removing these errors.

**Vignetting removal:** Vignetting is found when the central portion image is more brighter than the periphery or corners. This error is found in images obtained from optical sensor, which uses lens. Images taken using wide-angle lenses often vignette due to refraction property of light. Proper calibration between the irradiance and sensor output signal is generally used for this correction. Fig. 10.8. shows the vignetting effect in left image and the right image is the output after vignetting removal is applied.

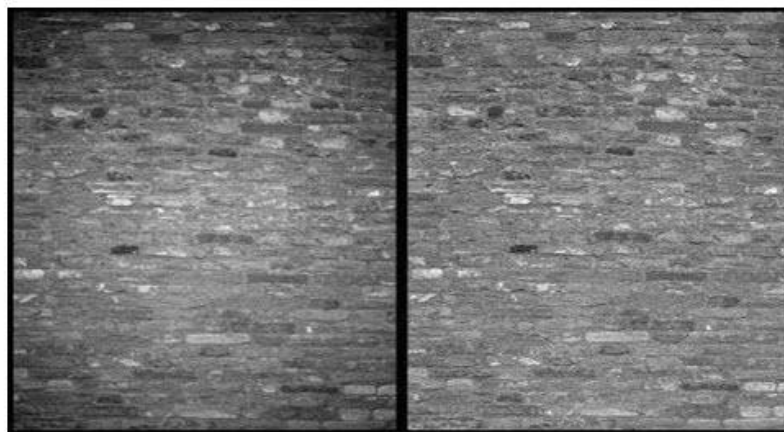


Fig. 10.8. Left image have vignetting effect in contrast to the uniform image.

(Source: [www.rbc.ufrj.br/\\_pdf\\_56\\_2004/56\\_1\\_06.pdf](http://www.rbc.ufrj.br/_pdf_56_2004/56_1_06.pdf))

**Keywords:** Digital image, Rectification, Geometric correction, Radiometric correction, Noise.

## Lesson 11 Pre Processing Techniques

Many image processing and analysis techniques have been developed to aid the interpretation of remote sensing images and to extract as much information as possible from the images. The choice of specific techniques or algorithms to use depends on the goals of each individual project. Image processing procedures can be divided into three broad categories: Image restoration, Image enhancement, and Information extraction. For better visualization and information extraction, enhancement of image is required are discussed as below.

### 11.1 Image Reduction and Image Magnification

#### 11.1.1 Image Reduction

Sometimes it is necessary to view the entire image in more detail e.g. for visual interpretation. Many image processing systems are not able to display the full image containing rows and column greater than 3000 and 3000 respectively. Let a study area be covered by 5028 x 4056 pixels. In such circumstances, image reduction allows the viewers to display the full image, which also reduces the image dataset to a smaller dataset. This technique is useful for orientation purposes as well as delineating the exact row and column coordinates of an area of interest. To reduce a digital image to just  $1/m^2$  of the original data, every  $m$ th row and  $m$ th column of the image are systematically selected and displayed. If  $m=2$  is used; this will create a 2x reduced image containing  $2514 \times 2028$  ( $= (5028/2) \times (4056/2)$ ) rows and columns respectively. This reduced image would contain only 25% of the pixels found in the original scene. The procedure of 2x reduction is shown in Fig. 11.1.

56	85	52	63	74	120
201	244	142	162	180	192
75	65	21	15	89	56
78	23	95	176	149	185
206	226	110	164	124	189
190	219	42	58	63	49

56	52	74
75	21	89
206	110	124

Fig. 11.1. Image reduction (2x).

In Fig. 11.1, every second row and second column is sampled. Thus a 2x reduction means every second row and second column of the original image will be sampled. A sample image reduction in 2x and 4x is shown in Fig. 11.2.



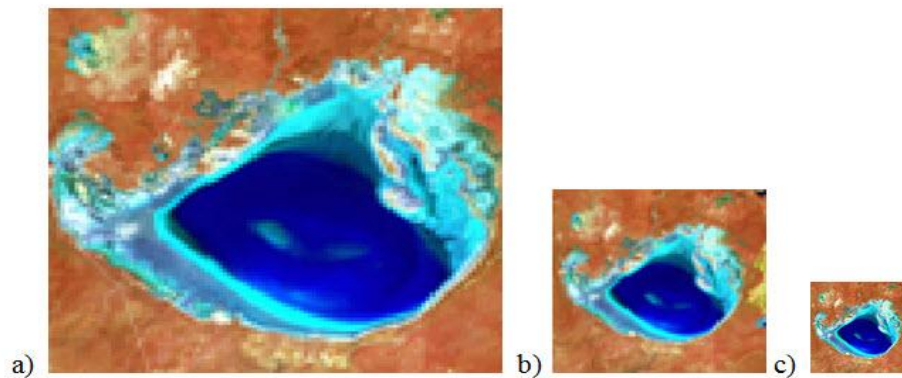


Fig. 11.2. a) Original image, b) 2x reduced image, c) 4x reduced image.

### 11.1.2 Image Magnification

Digital image magnification is often referred to as zooming. It improves the scale of the image for enhanced visual interpretation. In image reduction pixels are deleted depending on the order of reduction and image magnification is done by replication of rows and columns.

To magnify an image by an integer factor  $n$  squared, each pixel in the original image is usually replaced by an  $n \times n$  block of pixels all of which have the same pixel values as the original input pixel. The logic of a 2x magnification is shown in Fig. 11.3. This form of magnification doubles the size of each of the original pixel values. An example of magnification is shown in Fig. 11.4

26	30	86	
142	130	42	
65	210	180	

(a)

26	26	30	30	86	86	
26	26	30	30	86	86	
142	142	130	130	42	42	
142	142	130	130	42	42	
65	65	210	210	180	180	
65	65	210	210	180	180	

(b)

Fig. 11.3. a) Original image, b) 2x magnified image.

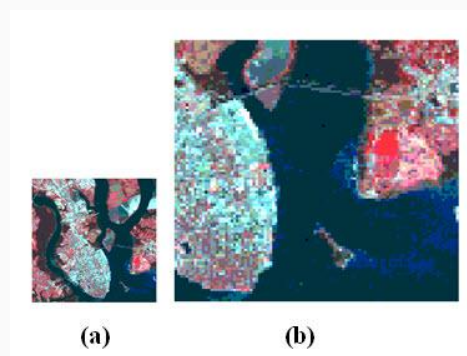


Fig. 11.4. a) Original image, b) 2x magnified image.

(Source: [www.uriit.ru/japan/Our\\_Resources/Books/RSTutor/Volume3/mod7/7-1/7-1.html](http://www.uriit.ru/japan/Our_Resources/Books/RSTutor/Volume3/mod7/7-1/7-1.html))

## 11.2 Color Compositing

In displaying a color composite image three primary colors (red, green, and blue) are used. Thus to create a colour image three bands of multi-band image is required. These three bands are passed through the primary color gun, creates a color image. Different color images may be obtained depending on the selection of three-band images and the assignment of the three primary colors. Color intensity of a feature in an image depends on the spectral reflectance of that feature in the three selected bands.

True colored image is said to be the image, which shows the features in their actual color. Multispectral image dataset contains spectral bands in the red, green, and blue spectrum region of the electromagnetic (EM) spectrum, may be combined to produce a true color image. True colored image can be obtained by passing the red band through red color gun, green band through green color gun, and blue band through blue color gun of a display system. Landsat ETM+, IKONOS image dataset are example of multispectral image dataset contains red, green, and blue bands.

Human eye is able to see the radiation having wavelength within the visual portion of the electromagnetic spectrum. Thus to see the radiations having wavelength lying out of the visible portion of the electromagnetic spectrum, False Color Composite (FCC) is required. In this case, spectral bands like Infrared (IR), Near-Infrared (N-IR), Thermal Infrared (T-IR) etc. are pass through the visual color gun to see the reflection/ radiation of features in these bands. Thus obtained image is known as FCC image. If the N-IR band radiation passes through the red color gun, red through green, and green through blue; thus obtained image is known as Standard FCC image. Compared to true color image, in standard FCC the vegetation appears better depending on the types and conditions, since vegetation reflects much in N-IR region of the EM spectrum than the green. As N-IR band passes through red color gun, vegetation appears red in Standard FCC image (Fig. 11.5).

Spectral bands of image datasets which do not contain three visible bands (IRS LISS-III/ LISS-IV or SPOT HRV) are mathematically combined resembles to a true color image, can be produced as follows:

RED colour gun = Red

GREEN colour gun =  $(3 * \text{Green} + \text{NIR}) / 4 = 0.75 * \text{Green} + 0.25 * \text{NIR}$

BLUE colour gun =  $(3 * \text{Green} - \text{NIR}) / 4 = 0.75 * \text{Green} - 0.25 * \text{NIR}$

In practice, we use various color combinations to facilitate the visual interpretation of an image depending on the nature of application and reflection/ radiation property of surface features.



**Fig. 11.5. False colour composite multispectral Landsat TM image.**

### **11.3 Image Enhancement**

Image enhancement procedures are applied to original image data to improve the quality of image, which would be more suitable in image interpretation and feature extraction depending on the nature of application. Although radiometric corrections are applied, the image may still not be optimized for highest level of visual interpretation. Thus it increases the amount of information that can be extracted visually. The image enhancement techniques are applied either to single-band images or separately to the individual bands of a multi-band image set. There are numerous procedures that can be performed to enhance an image. However, they can be achieved using two major functions: global operations (spectral or radiometric enhancement) and local operations (or spatial enhancement). Global operations change the value of each individual pixel independent of neighbor pixels, while local operations change the value of individual pixels in the context of the values of neighboring pixels.

#### **Global Operation or Radiometric Enhancement**

Contrast is the relative brightness in DN values or radiometric values. Object distinction in high contrast image is more suitable than low contrast image, as our eyes are good in discriminating higher radiometric differences. Thus the contrast or radiometric enhancement is used to improve the radiometric differences between the objects for better visual interpretability. It can normally be categorized as

- 1) Grey level thresholding
- 2) Density slicing
- 3) Contrast stretch

#### **1) Gray-Level Thresholding**

Gray-level thresholding is used to segment an input image into two classes: one for those pixels having DN values below an analyst-defined value and another for those above this value. Thus thresholding on an image creates a binary image containing pixels of only two DN values, 0 and 255. In the following example (Fig. 11.6.), let threshold value 185 be applied. In the

input image, all the pixels having DN values less than 185 will appear pure black (DN value=0) and all the pixels having DN values greater than 185 will appear white (DN value=255) in the output image.

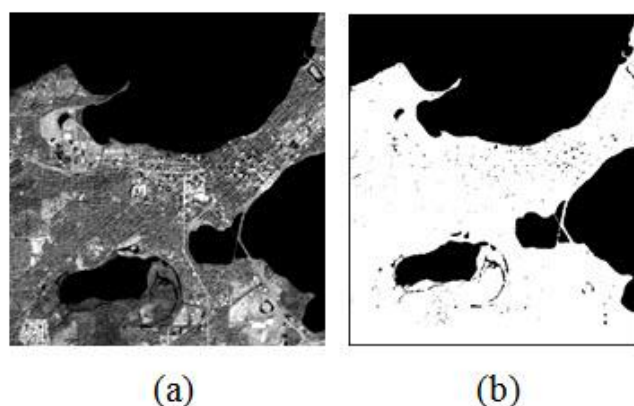


Fig. 11.6. a) Input Image b) Output image after thresholding.

## 2) Density or Level Slicing

Level slicing is similar to grey level thresholding. In grey level thresholding, the input image pixels are segmented in two groups lower and higher of a defined DN value or level, in case of level slicing a number of levels ( $m$ ) are used, which segmented the image in a numbers of groups ( $m+1$ ). Thus in density slicing or level slicing, the image histogram is divided into an analyst specified intervals or slices. All of the DN values in the input image falling within a specified are then displayed at a single DN in the output image. Consequently, if four slices are established, the output image will contain four different gray levels (Fig 11.7).

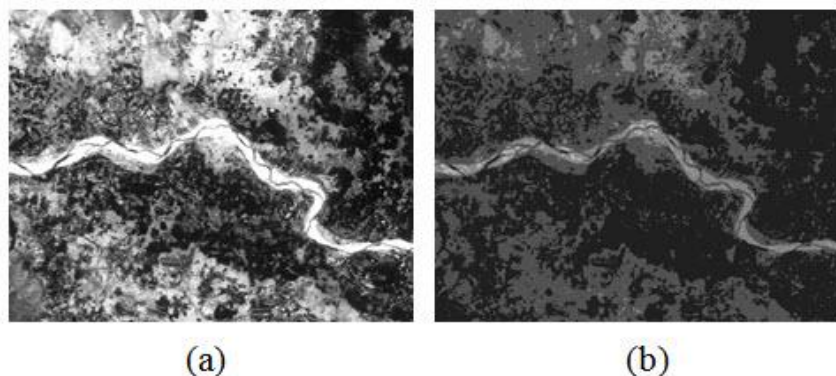
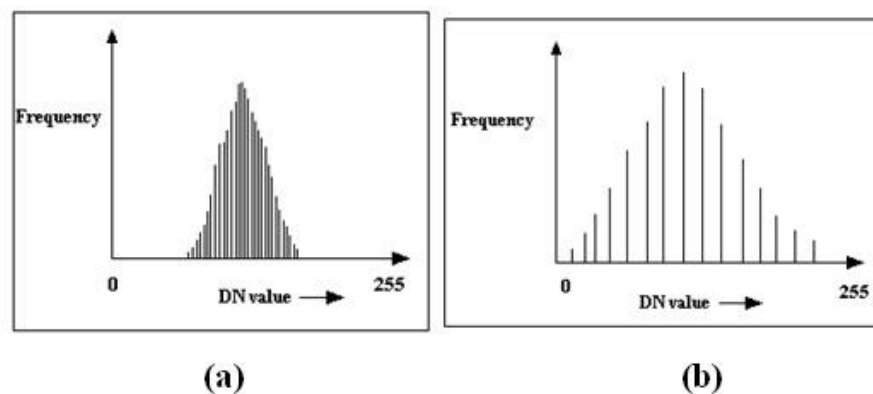


Fig. 11.7. a) Input image b) Output image (having 4 slices).

Application of density slice reduces the details of an image, but it also reduces the noise in the image. This is useful in highlighting the features in the imagery.

**Histogram:** Histogram is a graphical presentation of DN value of a pixel and occurrence of pixels having same DN value (frequency) in an image. From the histogram of an image the contrast of an image can easily be understood. In low contrast image the radiometric differences or differences in DN values are less, and in high contrast image the radiometric differences or differences in DN values are high (Fig. 11.8).



**Fig. 11.8. a) Low contrast histogram b) high contrast histogram**

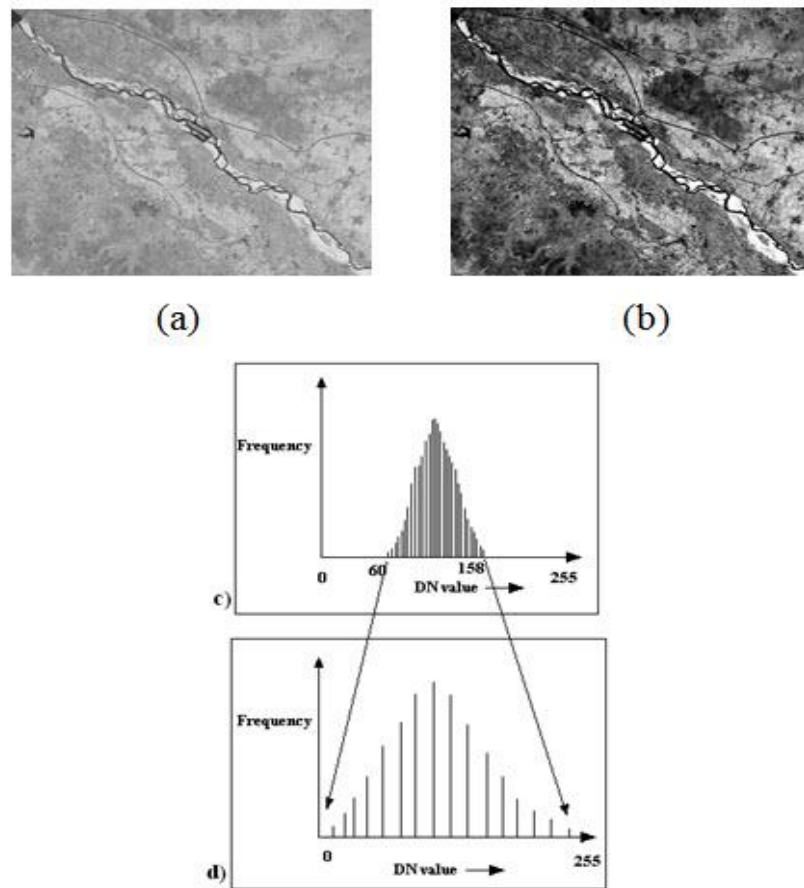
### 3) Contrast Stretching

One of the most useful method of contrast enhancement is contrast stretching. 8 bit image data contains 256 ( $=2^8$ ) levels in grey scale of DN values. Often remote sensing sensors fail to collect reflectance/radiation covering the entire 256 levels, normally confined to a narrow range of DN values, which produce a low contrast image. Let an input image have brightness values between 60 and 158. Contrast stretching operation expands the narrow range (60 to 158) of DN values of the input image to the full range (0 to 255) of DN values in the output image.

Contrast stretch is categorized as a) linear and b) non-linear.

#### a) Linear Contrast Stretch

In linear contrast stretch the input histogram is stretched uniformly. Thus it drags the minimum DN value to the zero (0; the lowest grey level of display system) and maximum DN value to the 255 (the highest grey level of the display system) and stretch all the interlaying DN values uniformly. This uniform expansion is called a linear stretch. This increase the contrast of an image, therefore subtle variations in input image data values would now be displayed in output tones that would be more readily distinguishable by an interpreter. Light tonal areas would appear lighter and dark areas would appear darker. An example of linear contrast stretch is shown in Fig. 11.9.



**Fig. 11.9. a) Original image, b) after linear contrast stretch, c) input histogram, d) output histogram.**

The linear contrast stretch would be applied to each pixel in the image uses the algorithm

$$DN' = \left( \frac{DN - MIN}{MAX - MIN} \right) \times 255$$

where,

DN' = digital number assigned to pixel in output image

DN = original digital number of pixel in input image

MIN = minimum value of input image, to be assigned a value of 0 in the output image (60 in the above example)

Max = maximum value of input image, to be assigned a value of 255 in the output image (158 in the above example).

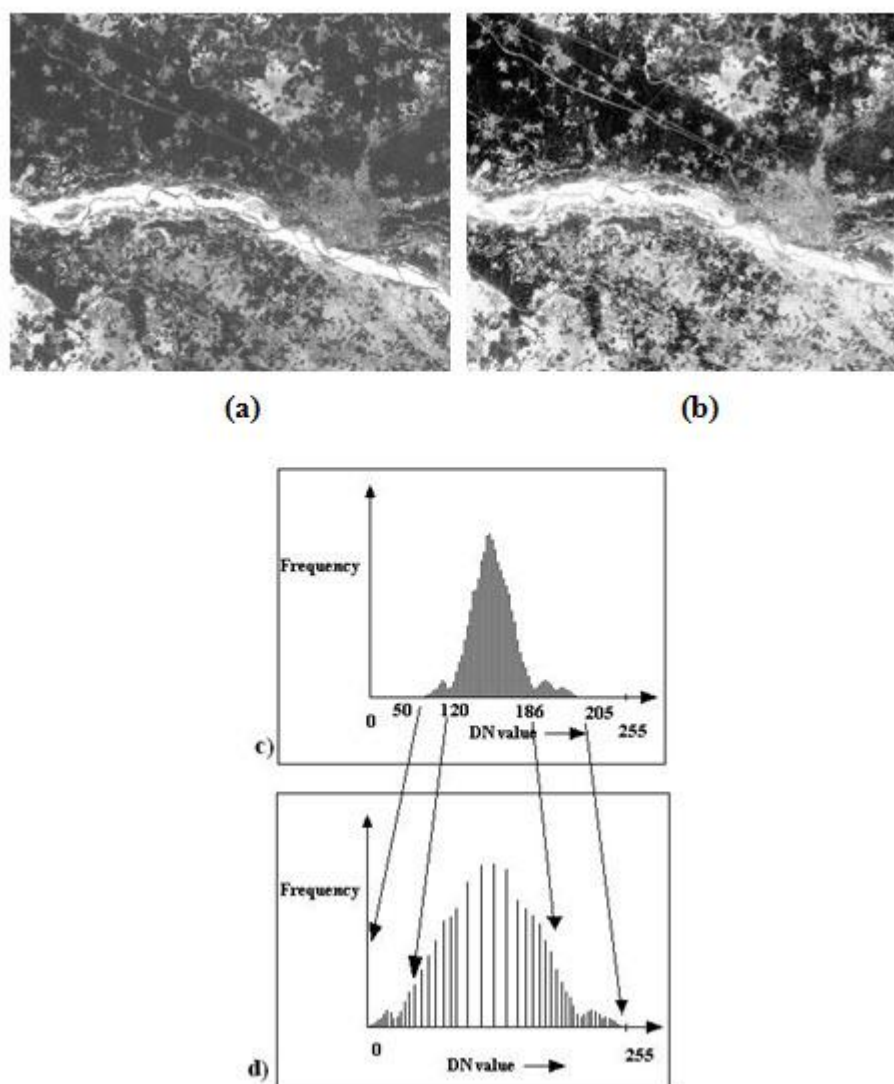
### **b) Non-linear Contrast Stretch**

One drawback of the linear stretch is that it assigns as many display levels to the rarely occurring DN values as it assigns as many occurring values. Due to uniform radiometric



enhancement, the stretch between DN values in the both tails (zones of low frequency DN values) is equal to the middle of the histogram (zones of high frequency DN values). The low frequency means less number of pixels, which is less important for information extraction, gets equal importance in the linear contrast enhancement operation. To avoid this, non-linear contrast stretch is applied to the image data.

**Histogram equalization-** One method of non-linear contrast stretch is histogram equalization (shown in Fig. 11.10.), in which the contrast stretching depends on the histogram of the image data. The contrast stretch will be higher between the DN values having higher frequency than the lower frequency DN values. But it suppresses the lower frequency pixels, which have less subject to interest.



**Fig. 11.10. a) Original image, b) output image after histogram equalization, c) input image histogram, d) output image histogram.**

As shown in Fig. 11.10., in the input histogram greater display values (and hence more radiometric detail) are assigned to the frequently occurring portion of the histogram. The image value range of 120 to 186 (high frequency) is stretched more than lower value range 50 to 120 (low frequency) and 186 to 205 (low frequency).

## Local Operations or Spatial Enhancement

Spatial image enhancement is used to increase the image quality through filtering. Spatial frequency is the rate of change of image values or DN values in the scene in a unit distance. Filtering operation emphasizes or de-emphasizes this spatial frequency for better image interpretation. Thus filters highlight or suppress specific features based on their spatial frequency hence it depends on the image value of neighbors. Different kinds of filters are explained below.

**Convolution filter:** In convolution filtering a window also known as operators or kernels having odd numbers of pixels (e.g., 3x3, 5x5, 7x7) used to move throughout the input image, the DN value of the output image is obtained by a mathematical manipulation. A convolution filter is classified as low-pass filter and high-pass filter.

**Low-pass filter:** A low-pass filter emphasizes low frequency or homogeneous areas and reduces the smaller details in the image. It is one of the methods of smoothing operation. It reduces the noise in the imagery. Low-pass filter normally sub-divided as average (mean), mode and median filter.

**Average filter:** In case of average or mean filter the output DN value is obtained by multiplying each coefficient of the kernel with the corresponding DN in the input image and adding all them all and dividing by the sum of kernels value. The coefficients of the kernel range from equal value to different weighted value.

**Mode filter:** A kernel is fitted in the input image, and computes the mode value within this kernel. The resultant value is the final value of the center pixel of the kernel.

**Median filter:** Median filter is similar to mode filter. In this case the median value of the kernel is computed which replaces the center pixel DN value. Median filter is very useful for removing random noise, salt-pepper pattern noise, speckle noise in RADAR imagery.

**High-pass filter:** High-pass filter is the opposite of low-pass filter. In contrast to low-pass filter, high-pass filter enhance the variations in the imagery. It increases the spatial frequency, thus sharpens the details in the imagery. High frequency filter works in the similar way of average filter, uses a kernel may be normal or weighted. Another way of getting the output of high-pass filtering is subtraction of the low-pass filtered image from the original image. The following 5x5 image explains the operation of low pass and high pass filters.

56	68	45	63	89
28	24	36	44	75
34	20	62	18	58
38	40	42	46	60
35	55	16	29	31

1	1	1
1	1	1
1	1	1

**Low-pass average  
filter**

-1	-1	-1
-1	16	-1
-1	-1	-1

**High-pass filter**

For, low-pass average filter, the central value of the of window in the output image will be



$$= \frac{(56 \times 1) + (68 \times 1) + (45 \times 1) + (28 \times 1) + (24 \times 1) + (36 \times 1) + (34 \times 1) + (20 \times 1) + (62 \times 1)}{9}$$

$$= 41$$

Similarly, if the high filter is applied, the calculated value will be 39, which will replace the value 24.

In the above example, the image values of the 3x3 window of the input image can be written in increasing order as 20, 24, 28, 34, 36, 45, 56, 62, 68. In this set 45 is the median value, this will replace the central kernel value 24. Similarly if a set of values of a 3x3 kernel is, as example 50, 62, 58, 20, 145, 19, 20, 96, 204, then the central kernel of the output image will be the mode value of the set, i.e. 20.

**Edge Enhancement filter:** Edge enhancement is quite similar to high frequency filter. High frequency filter exaggerate local brightness variation or contrast and de-emphasize the low frequency areas in the imagery. In contrast to this, edge enhancement filter emphasize both the high frequency areas along with low frequency areas. It enhances the boundary of features. Different kinds of kernels are used depending on the roughness or tonal variation of the image.

**Edge Detection filter:** There may arise confusion with edge enhancement filter with edge detection filter, which highlights the boundary of the features and de-emphasizes the low contrast areas totally. It highlights the linear features such as road, rail line, canal, or feature boundary etc. Sobel, Prewitt and Laplacian etc filters are the example of edge detection filter. Fig. 11.11 shows image filtering operations using low pass, high pass, edge enhancement and edge detection filters.

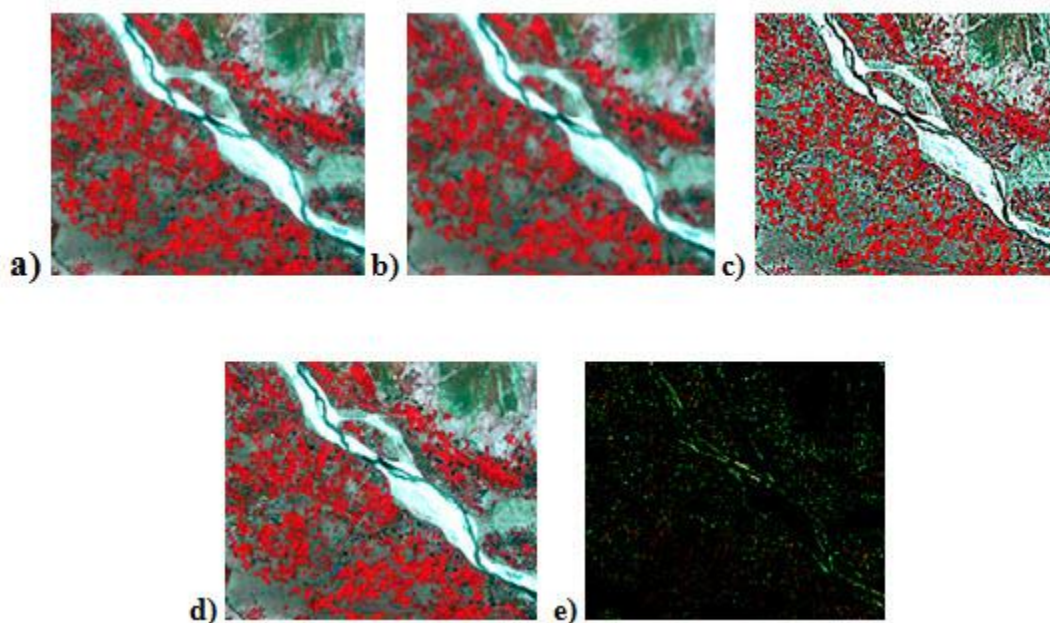
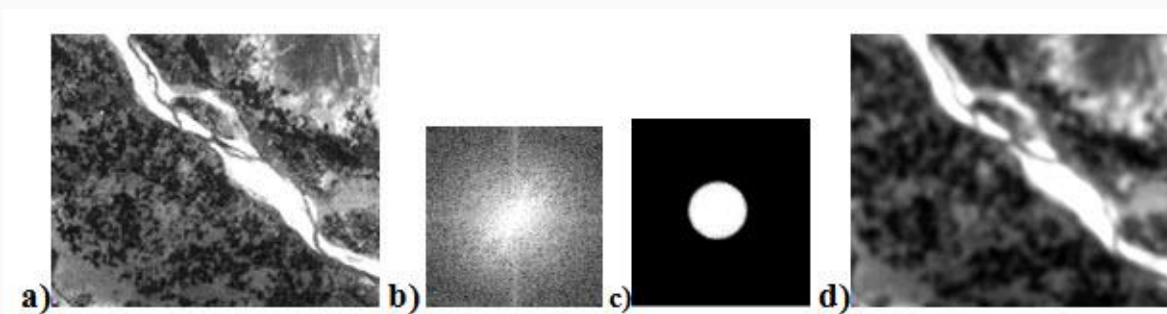


Fig. 11.11. a) Original image, output image applying: b) low pass filter, c) high pass filter, d) edge enhancement filter, e) edge detection filter.

**Fourier Transform:** The filters discussed above are spatial filters. Fourier analysis or transform the spatial domain is transferred to frequency domain. The frequency domain is represented as a 2D scatter plot, known as Fourier domain, where low frequencies falls at the center and high frequencies are progressively outward. Thus Fourier spectrum of an image can be used to enhance the quality of an image with the help of the low and high frequency block filters. After filtering the inverse Fourier transformation gives the output of the Fourier transformation. Fig. 11.12 shows the resulting images of application of Fourier spectrum and inverse Fourier spectrum.



**Fig. 11.12. a) Original image, b) Fourier spectrum of input image, c) Fourier spectrum after application of low pass filter, d) Output image after inverse Fourier transformation.**

#### 11.4 Image Manipulation

Each of the images acquired were manipulated in different ways to highlight various aspects of them.

These manipulations include mainly:

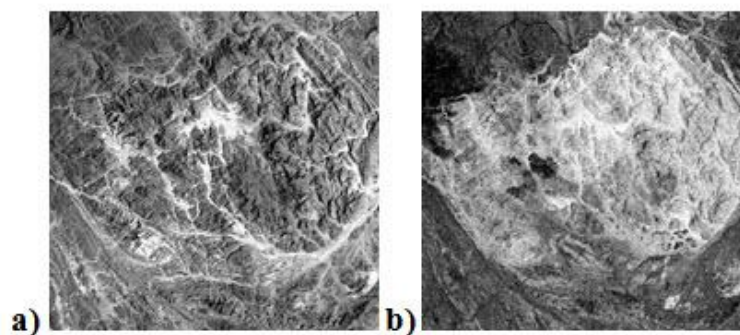
Band ratioing

Indexing

Principal Component Analysis

##### Spectral Ratioing

Radiance or reflectance of surface features differs depending on the topography, shadows, seasonal changes in solar illumination angle and intensity. The band ratio of two bands removes much of the effect of illumination in the analysis of spectral differences. Ratio image resulting from the division of DN values in one spectral band by the corresponding values in another band. In the scene, there are mainly two types of rocks which are discernable in Fig. 11.13. b) but not in Fig. 11.13.a).



**Fig. 11.13. a) Blue band shows topography due to illumination difference, b) ratio of band3/band2 removes illumination and yield different rock types.**

(Source: [geology.wlu.edu/harbor/geol260/lecture\\_notes/Notes\\_rs\\_ratios.html](http://geology.wlu.edu/harbor/geol260/lecture_notes/Notes_rs_ratios.html))

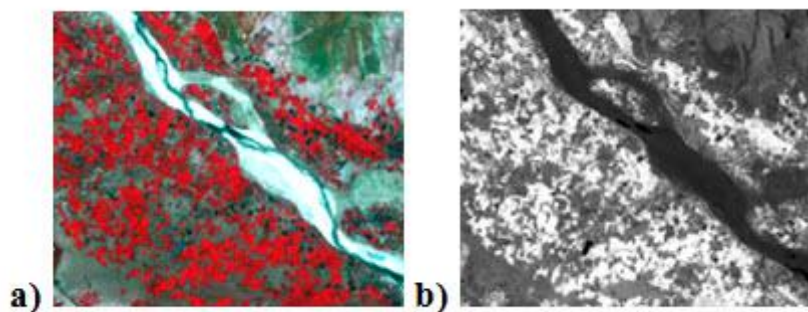
It should be kept in mind that different materials with different absolute radiances but having similar slopes may appear identical in band ratio image. One very useful band ratio image in vegetation monitoring is Ratio Vegetation Index (RVI) can be formulated as follows:

$$RVI = \frac{DN_{NIR}}{DN_{Red}}$$

## Indexing

Indexing is similar to band ratio. In contrast to band ratio or simple ratio, indexing is useful for nonlinear variation, thus overcome the discrepancy of misinterpretation sometime occurs in band ratio. As an example, Normalized Difference Vegetation Index (NDVI) uses the NIR and red bands; where the vegetation areas yield very high values because of high reflectance in NIR and low reflectance in visible band; whereas, clouds, water, snow yield negative index value and rock, bare soils etc. have very low or near zero index value (Fig. 11.14). Thus NDVI is very commonly used index in vegetation monitoring.

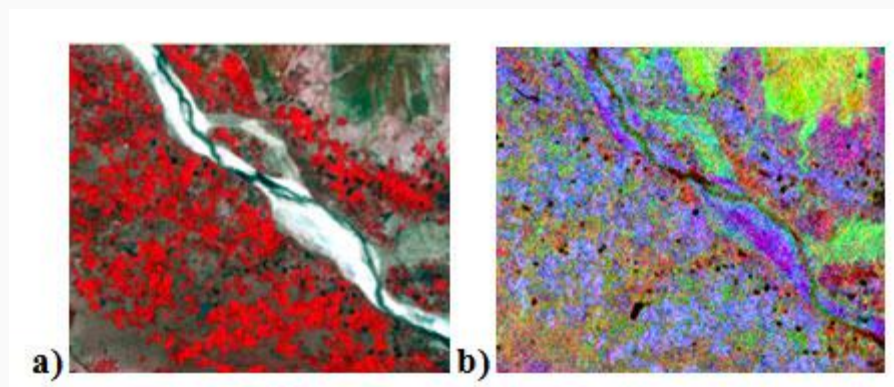
$$NDVI = \frac{DN_{NIR} - DN_{Red}}{DN_{NIR} + DN_{Red}}$$



**Fig. 11.14. a) FCC image of an area, b) NDVI output image; vegetation appears bright, river or wet land appears dark.**

## Principal Component Analysis

Spectral bands of multispectral data are highly correlated. Due to this high correlation in bands, the analysis of the data sometimes becomes quite difficult; thus images obtained by combining different spectral bands looks similar. To decorrelate the spectral bands, Principal Component Analysis (PCA) or Principal Component Transformation (PCT) is applied. PCA also have the capability of dimension reduction, thus PCA of a data may be either an enhancement operation prior to visual interpretation or a preprocessing procedure for further processing. Principal components analysis is a special case of transforming the original data into a new coordinate system. It enhances the subtle variation in the image data, thus many features will be identified which cannot be identifiable in raw image (Fig 11.15).



**Fig. 11.15. a) Input image (RGB:432) b) output image (RGB: 432)**

**Keywords:** Image reduction, Image magnification, Color composing, Image enhancement, Histogram equalization, Filtering, Fourier transform, PCA.

## Lesson 12 Image Classification

Different landuse landcover types in an image can be discriminated automatically using some image classification algorithms using spectral features, i.e. the brightness or image value (DN value) and "colour" information contained in each pixel. Thus, the purpose of image classification is to categorize pixels into several classes to create thematic maps. It is one of the most popular used information extraction techniques in remote sensing. It is also known as landuse-landcover classification. The usual categorization of image classification is discussed below-

### 12.1 Supervised Classification

In this technique, the image is classified on the priori knowledge of the analyst. The image is classified on the basis of predefined landuse-landcover classes and algorithm by the analyst. At first, the analyst must have some knowledge about the landuse-landcover classes of the study area; on this basis the landuse-landcover classes will be defined. There are three basic steps in supervised classification process. First is the training stage, in which representative sample sites of known feature types are collected, which is termed as training sites or signature class depending on the spectral response patterns of the features which we can identify from image or from other sources like aerial photography, ground truth data, personal experience, previous studies, or maps. In the second or classification stage, the unknown pixels of image are compared to the spectral patterns with the training sites and they are assigned to a class as defined by some algorithm. The third is the output presentation in the form of maps, table of area data and digital data files.

After the signatures are collected, the pixels are sorted into classes based on the algorithms used. The algorithms used in supervised classification are: a) Minimum Distance to Mean, b) Parallelepiped, c) Gaussian Maximum Likelihood.

**a) Minimum Distance to Mean Classifier:** The minimum distance to mean classifier is simplest mathematically and very efficient in computation. In this procedure the DN value of the training sets are plotted in a scatteromgram. Here a 2D scatteromgram is drawn for an example shown through Fig. 12.1. DN values of five training sets are plotted and their means are computed (shown with a subscript). Now a unknown pixel A will be classified or be assigned to a class by a distance calculation from the mean of each class to the pixel A. This is why this algorithm is named as minimum distance to mean classifier. Therefore the pixel A will be assigned to the class whose mean value is nearest to A. In this example, the unknown pixel A will be assigned to the sand class. This way the pixels of entire image are assigned to different landuse landcover classes. Thus for an n-Dimensional multispectral data, n-D scatter diagram is plotted; the mean of each classes are calculated and the image is classified according to the shortest distance class. The most used distance calculation method is Euclidean distance formulated as below:



$$d_e(m_{ik}) = \sqrt{\sum_{j=1}^{nb} (m_{ij} - k_j)^2} \quad \text{Numerical form}$$

$$= [(C_i - k)^T (C_i - k)]^{1/2} \quad \text{Matrix form}$$

where,

nb= number of bands

j= a particular band

i= a particular class

$k_j$  = DN value of pixel at band j

$C_{ij}$  = mean of DN values in band j for the sample for class i

$d_e(C_{ik})$  = Euclidean distance from the mean of a class to any unknown pixel

T = transposition function

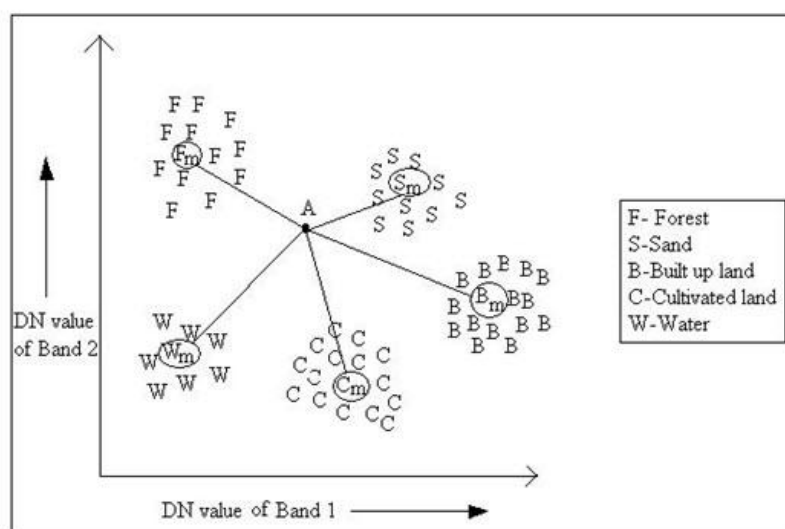
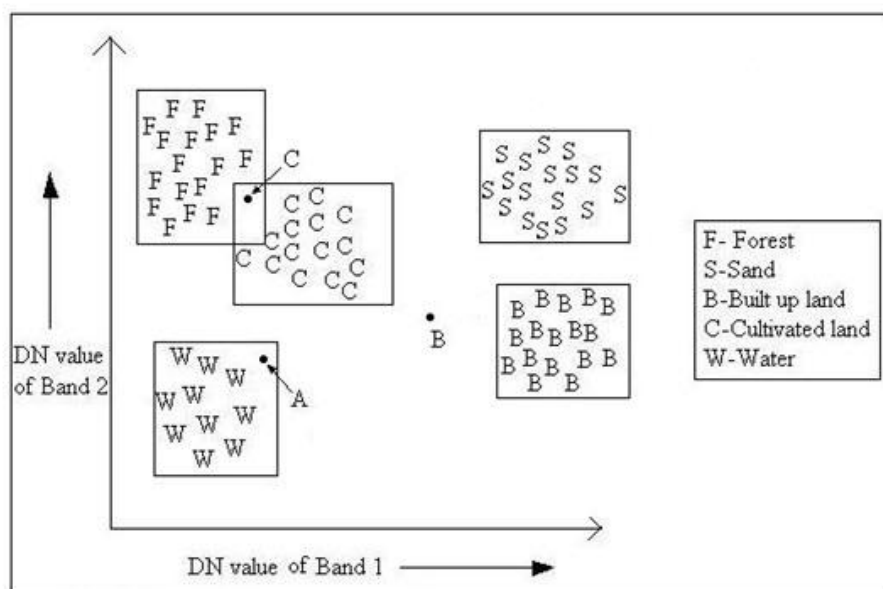


Fig. 12.1. Minimum distance to means classification strategy.

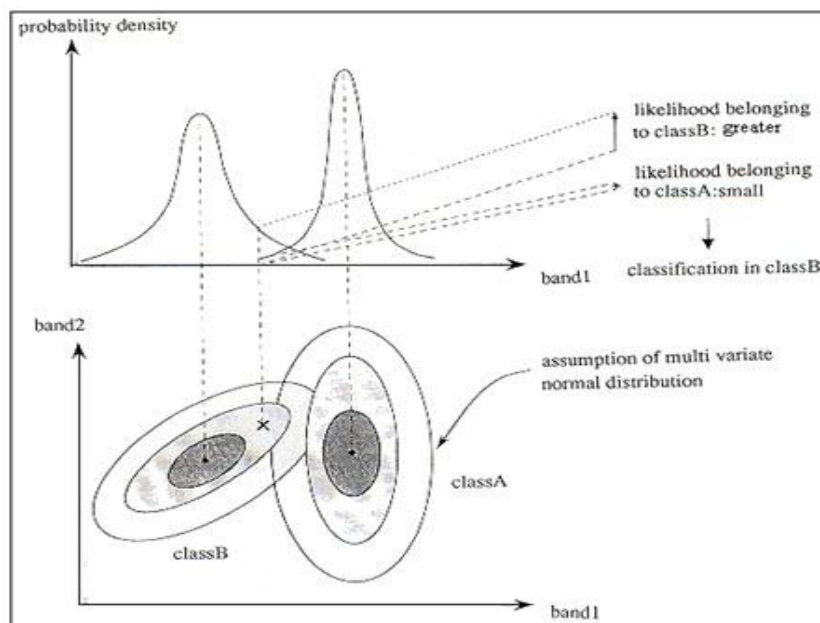
**b) Parallelepiped Classifier:** The parallelepiped classification strategy is also computationally simple and efficient. As an example, the DN values of two bands are plotted in a scatter diagram in the similar way to minimum distance to mean classifier. In this procedure a rectangular box is fitted for each class defined by the maximum and minimum values of each bands shown in Fig. 12.2. Now the classification of the pixels depends on whether any pixel falls inside any rectangular box (or parallelepiped decision region) or not. If the pixel falls inside the parallelepiped, it is assigned to the class. However, if the pixel falls within more than one class, it is put in the overlap class. If the pixel does not fall inside any class, it is assigned to the null class. In the given example, three unknown pixels are taken (A, B, and C). The pixel A will assigned to the water class, as it falls in parallelepiped of water

class; whereas pixel B will be labeled as unknown class and the pixel C will be labeled as overlap class. Overlap is caused due to high correlation or covariance between bands. Covariance is the tendency of spectral values to vary similarly in other bands. This method is poor as the spectral response patterns are frequently highly correlated and high covariance.



**Fig. 12.2. Parallelepiped classification strategy.**

**c) Gaussian Maximum Likelihood Classifier:** The maximum likelihood classifier is one of the most popular methods of classification in remote sensing, in which a pixel with the maximum likelihood is classified into the corresponding class. In this method it is assumed that the training sets of each class of each band have the normal distribution, which is Gaussian in nature and can be described by the mean vector and covariance matrix. From this information, the statistical probability is computed for a given pixel value being a member of a particular landuse landcover class. The probability is calculated for an unknown pixel for each class of the training sets. As an example, in the Fig. 12.3, the probability values for band 1 for the mean of the classes (A and B) and unknown pixel X is plotted in 2D graph. The bell-shaped curves are called probability density functions. In the example, it is found that the probability of the unknown pixel X for being in class B is greater than class A. So, the pixel X will be labeled as class B. The probability density functions are used to classify unidentified pixels by computing the probability of the pixels values belonging to each category. Thus it takes large number of computations to classify an image and is much slower in computation than the previous methods.



**Fig. 12.3. Concept of maximum likelihood classification method.**

(Source: [stlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp11/cp11-7.htm](http://stlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp11/cp11-7.htm))

## 12.2 Unsupervised Classification

In supervised classification, the image pixels are categorized as defined by the analyst specified landuse landcover classes and an algorithm thereafter. In contrast to supervised classification, in unsupervised classification an algorithm is first applied to the image to form some spectral classes (clusters); thereafter the image is classified based on these classes. Then the analyst tries to assign spectral classes to desirable information classes or landuse landcover classes. The spectral classes in an image may be numerous; it is difficult to obtain all of them as training areas in supervised classification, but in the unsupervised approach, they are found automatically. Unsupervised classified image must be compared with some reference data to identify the spectral classes which is automatically generated.

One of the commonly used algorithms of unsupervised classification is the “K-means” clustering method. It is an iterative method. The number of output cluster or spectral classes (assume  $k$  clusters) is specified by the analyst. The algorithm locates the centroids of  $k$  clusters in the multidimensional data. Each pixel in the image is assigned then to its nearest cluster. At this point the mean of each clusters re-calculated using the result obtained from the previous step. The image then reclassified using the revised mean of each clusters. This procedure continues until there is no significant change in the location of cluster mean vector between successive iterations.

The K-means algorithm is composed of the following steps:

1. Place  $K$  points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.



3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move in centroids.

One more useful algorithm in unsupervised classification is ISODATA algorithm. The ISODATA algorithm is similar to the k-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the k-means assumes that the number of clusters is known a priori. In ISODATA algorithm, clusters are merged if either the number of members (pixel) in a cluster is less than a certain threshold or if the centers of two clusters are closer than a certain threshold. Clusters are further split into two different clusters if the cluster standard deviation exceeds a predefined value and the number of members (pixels) is twice the threshold for the minimum number of members.

### 12.3 Fuzzy Classification

In remote sensing imagery, spectral reflectance/radiance of more than one features are recorded in one pixel. Thus one pixel contained information more than one feature. These pixels are known as mixed pixel. In case classifying an image contains mixed pixel, the mix pixels is a member in more than one class. Therefore, it requires a fuzzy logic based model to solve the unmixing problem in classification. Fuzzy classification and spectral mixture classification are two commonly used techniques.

There are many approaches of fuzzy classification, one most important technique is fuzzy c-means. Fuzzy c-means is similar to K-means clustering. In fuzzy c-means technique, fuzzy regions are formed between the boundaries of the clusters generated in K-means clustering. The fuzzy-ness (possibility of one pixel to be member of more than one class) of image pixels can be presented or a membership grade value assigned describes how close is a pixel to the means of the classes.

**Keywords:** Classification, Supervised, Maximum likelihood, Parallelepiped, Minimum distance to mean, Unsupervised, K-means, ISODATA, Fuzzy.



## Lesson 13 Classification Accuracy Assessment

### 13.1 Classification Accuracy and Classification Error Matrix

The classification made using classification algorithms discussed earlier does not give always perfect result. Therefore the classified image contains lot of errors due to: labeling clusters after unsupervised classification, preparing of training areas with wrong labeling, un-distinguishable classes and correlation between bands, imperfect classification algorithm, etc. Therefore a common question of digital satellite remote sensing is: "How accurate is the classification?". An accuracy assessment of a classified image gives the quality of information that can be obtained from remotely sensed data. Accuracy assessment is performed by comparing a map produced from remotely sensed data with another map obtained from some other source. Landscape often changes rapidly. Therefore, it is best to collect the ground reference as close to the date of remote sensing data acquisition as possible.

- **Classification Error Matrix**

One of the most common means of expressing classification accuracy is the preparation of a classification error matrix (confusion matrix or contingency table). To prepare a error matrix first task is to locate ground reference test pixels or sample collection, based on which an error matrix is formed. There are many mathematical approaches in this regard. Generally it is suggested that a minimum of 50 samples of each landuse landcover classes should be included. If the study area is large or the numbers of landuse landcover classes are more than 12, the sample should be 75 to 100. Data sampling can be done using various procedures such as: random, systematic, stratified random, stratified systematic unaligned, and cluster. An error matrix compares the relationship between known reference data (ground data) and the corresponding results obtained from classification.

Table 13.1 is an error matrix obtained from a data analysis (Randomly sample test pixels) and Table 13.2 presents results of various accuracy measurements.

**Table 13.1 Error Matrix resulting from a data analysis**

Classification Data	Water	Sand	Forest	Urban	Cultivated land	Barren land	Row Total
Water	150	12	0	0	0	0	162
Sand	0	56	0	10	0	0	66
Forest	0	0	130	0	17	0	147
Urban	0	0	0	126	0	15	141
Cultivated land	0	0	20	0	78	12	110

Barren land	0	0	5	24	15	115	159
Column Total	150	68	155	160	110	142	785

**Table 13.2 Different measurement from error matrix**

Omission error	Producer's Accuracy	Commission error	User's Accuracy
Water=0/150=0%	Water=150/150=100%	Water=12/162=7%	Water=150/162=93%
Sand=12/68=18%	Sand=56/68=82%	Sand=10/66=15%	Sand=56/66=85%
Forest=25/155=16%	Forest=130/155=84%	Forest=17/147=12%	Forest=130/147=88%
Urban=34/160=21%	Urban=126/160=79%	Urban=15/141=11%	Urban=126/141=89%
Cultivated land =32/110=29%	Cultivated land =78/110=71%	Cultivated land =32/110=29%	Cultivated land =78/110=71%
Barren land =27/142=19%	Barren land =115/142=81%	Barren land =44/159=28%	Barren land =115/159=72%
Overall accuracy=(150+56+130+126+78+115)/785=83%			

From the error matrix several measures of classification accuracy can be calculated using simple descriptive statistics (Table 13.2) as discussed below:

- (a) Omission error
- (b) Commission error
- (c) Overall accuracy
- (d) Producer's accuracy
- (e) User's accuracy
- (f) Kappa coefficient (k)

**(a) Omission error (exclusion):** represent pixels that belong to the actual class but fail to be classified into the actual class (e.g., 12 pixels which should be classified as sand but classified as water).

**(b) Commission error (inclusion):** represents the pixels that belong to another class but are classified to a class (e.g., 20 pixels of forest class and 12 pixels of barren land class included in cultivated land).

**(c) Overall accuracy:** represents the total classification accuracy. It is obtained by dividing the total numbers of correctly classified pixels by the total numbers of reference pixels. The

drawback of this measure is that it does not tell us about how well individual classes are classified. The producer and user accuracy are two widely used measures of class accuracy depends on the omission and commission accuracy respectively.

**(d) Producer's accuracy:** it refers to the probability that a certain feature of an area on the ground is classified as such. It results from dividing the numbers of pixels correctly classified in each category by the numbers of sample pixels taken for this category (column total).

**(e) User's accuracy:** it refers to the probability that a pixel labeled as a certain class in the map is really this class. It is obtained by dividing the accurately classified pixels by the total numbers of pixels classified in this category. The producer accuracy and user accuracy are not same typically. For example, the producer accuracy of water is 100% whereas the user accuracy is 93%.

**(e) Kappa coefficient (K):** it is a discrete multivariate method of use in accuracy assessment. In classification process, where pixels are randomly assigns to classes will produce a percentage correct value. Obviously, pixels are not assigned randomly during image classification, but there are statistical measures that attempt to account for the contribution of random chance when evaluating the accuracy of a classification. The resulting Kappa measure compensates for chance agreement in the classification and provides a measure of how much better the classification performed in comparison to the probability of random assigning of pixels to their correct categories.

$\hat{K}$  is theoretically expressed as:

$$\hat{K} = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}}$$

The can be calculated as:

$$\hat{K} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}$$

where,

r = the number of rows in the error matrix

$x_{ii}$  = the number of observations in the row i and column i

$x_{i+}$  = total of observations in row i

$x_{+i}$  = total of observations in column i

N = total number of observations included in matrix

$$\sum_{i=1}^r x_{ii} = 150 + 56 + 130 + 126 + 78 + 115 = 655$$

$$\begin{aligned} \sum_{i=1}^r (x_{i+} \cdot x_{+i}) &= (162 \times 150) + (66 \times 68) + (147 \times 155) + (141 \times 160) + (110 \times 110) + \\ &\quad (159 \times 142) \\ &= 108811 \end{aligned}$$

$$\hat{K} = \frac{785(655) - 108811}{(785)^2 - 108811} = 0.80$$

A value of 0.80 implies that the classification process was avoiding 80% of the errors that a completely random classification would generate.

### 13.2 Data Merging

Image acquisition by a remote sensing sensor of an area of study can be available in the following cases:

- Image acquired by different sensors;
- Image acquired by same sensor operating at different spectral bands;
- Image acquired by same sensor at different time;
- Image acquired by same sensor at different polarization.

In many applications it may be required to merge multiple data sets covering same geographical area. Data merging operation generally can be categorized as: multi-temporal data merging, multi-sensor image merging, and multi-polarization image merging.

**Multi-temporal data merging:** The term multi-temporal refers to the repeated imaging of an area over a time period. By analyzing an area through time, it is possible to develop interpretation techniques based on an object's temporal variations, and to discriminate different pattern classes accordingly. Multitemporal data merging is combining images of the same area taken on different time, depending on the nature of purpose. The principal advantage of multitemporal analysis is the increased amount of information for the study area. The information provided for a single image is, for certain applications, not sufficient to properly distinguish between the desired pattern classes. This limitation can sometimes be resolved by examining the pattern of 2 temporal changes in the spectral signature of an object. This is particularly important for vegetation applications (source: [heim.ifi.uio.no/inf5300/2008/datafusion\\_paper.pdf](http://heim.ifi.uio.no/inf5300/2008/datafusion_paper.pdf)).

For example, agricultural crop interpretation is often facilitated through merger of images taken early and late in the growing season. In early season images in the upper Midwest, bare soils often appear that late will probably be planted in such crops perennial alfalfa or winter wheat in an advanced state of maturity. In the late season images, substantial changes in the appearance of the crops present in the scene are typical. Merging various combinations

of bands from the two dates to create color composites can aid the interpreter in discriminating the various crop types present (Lillesand & Kiefer, 2000).

**Multi-sensor data merging:** Each sensor has its own characteristics. Depending on the nature of application, images obtained from different sensors can be merged, which requires different preprocessing. The image merging or fusion of two images obtained from multispectral image r (MSI) and hyperspectral imager (HSI) are shown in Fig. 13.1. Multispectral imager generally has high spatial resolution and low spectral resolution; whereas hyperspectral imager has low spatial resolution but high spectral resolution. Image fusion taken from these two sensors can be of both high spectral and spatial resolution.

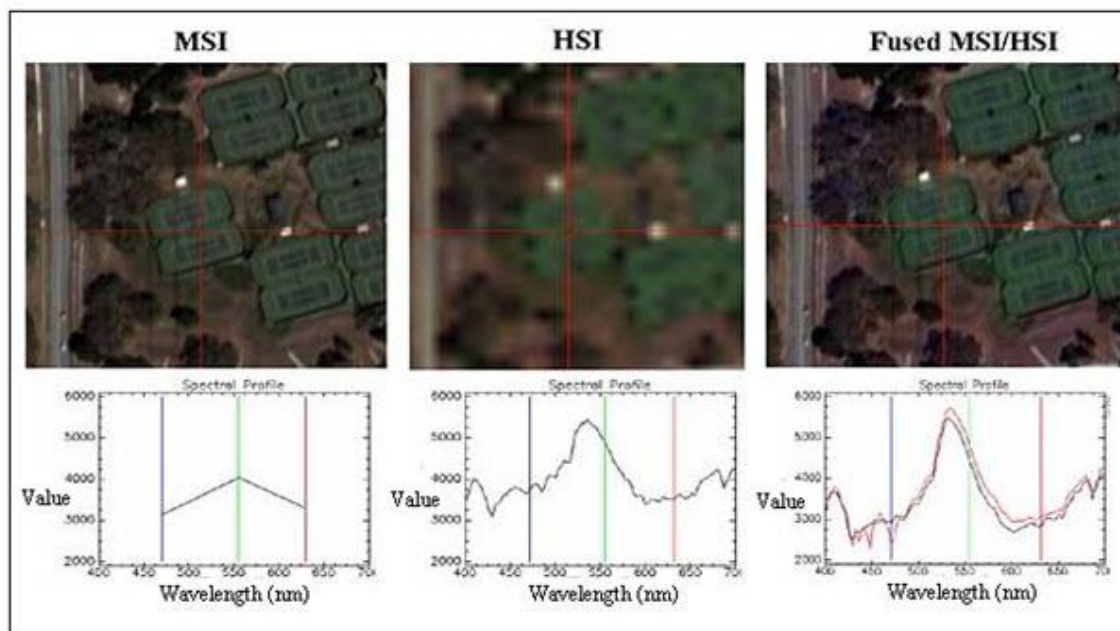


Fig. 13.1. Multisensor image fusion (merge).

(Source:

[www.weogeo.com/blog/Image\\_Fusion\\_and\\_Sharpening\\_With\\_Multi\\_and\\_Hyperspectral\\_Data.html](http://www.weogeo.com/blog/Image_Fusion_and_Sharpening_With_Multi_and_Hyperspectral_Data.html))

**Multipolarization image merging:** Multipolarization image merging is related to microwave image data, refers to image merging of different polarization.

### 13.3 GIS Integration

GIS integration refers to combining data of different types and from different sources. In a digital environment, where all the data sources are geometrically registered to a common geographic base, the potential for information extraction is extremely wide. This is the concept for analysis within a digital GIS database. The integration with GIS allows a synergistic processing of multisource spatial data.

Any data source which can be referenced spatially can be used in this type of environment. A DEM/DTM is just one example of this kind of data. Other examples could include digital maps of soil type, land cover classes, forest species, road networks, and many others,

depending on the application. The results form a classification of a remote sensing data set in map format could also be used in a GIS another data source to update existing map data.

As an example, GIS is now being widely used in crop management. Remote sensing and GIS have made huge impacts on how those in the agricultural planners are monitoring and managing croplands, and predicting biomass or yields. Map products, derived from remote sensing are usually critical components of a GIS. Remote sensing is an important technique to study both spatial and temporal phenomena (monitoring). Through the analysis of remotely sensed data, one can derive different types of information that can be combined with other spatial data within a GIS.

The integration of the two technologies creates a synergy in which the GIS improves the ability to extract information from remotely sensed data, and remote sensing in turn keeps the GIS up-to-date with actual environmental information.

As a result, large amounts of spatial data can now be integrated and analysed. This allows for better understanding of environmental processes and better insight into the effect of human activities. The GIS and remote sensing can thus help people to arrive at informed decisions about their environment. Like in all models, however, both maps and thematic data are abstractions or simplifications of the real world. Therefore, GIS and remote sensing can complement, but never completely replace field observations, (*Bhatta, 2008*).

**Keywords:** Accuracy assessment, Error matrix, Kappa statistics, Data merging, GIS integration.



## Module 5: Microwave and Radar System

### Lesson 14 Microwave Remote Sensing

#### 14.1 Microwave Sensor

The microwave region of interest to remote sensing covers the electromagnetic radiation of wavelength extending from a few millimeters to meters, in frequency interval from 40,000 to 300 MHz. The sensors are broadly classified as active and passive sensors. In passive mode, unlike in the visible and infrared regions the radiation from the sun does not play any direct part. Thus, the microwave sensors can operate during day and night. Due to its higher wavelength, the atmospheric haze, cloud, light rain (does not pass through heavy rain), dust/smoke is transparent to microwave, thus providing observation in all weather and environmental conditions. In addition, microwave remote sensing provides information about sea wind and wave direction derived from frequency characteristics, doppler effect, polarization, backscattering, etc which cannot be observed by visible and infrared sensors.

Microwave reflection or emission has no direct relation to the counterparts in visible or thermal portions of the spectrum for an object. Passive microwave sensing is similar in concept to thermal remote sensing. Due to higher wavelength, the field of view for a microwave sensor must be large to detect enough energy to record a signal. Most passive microwave sensors are therefore characterized by low spatial resolution.

#### 14.2 Antenna

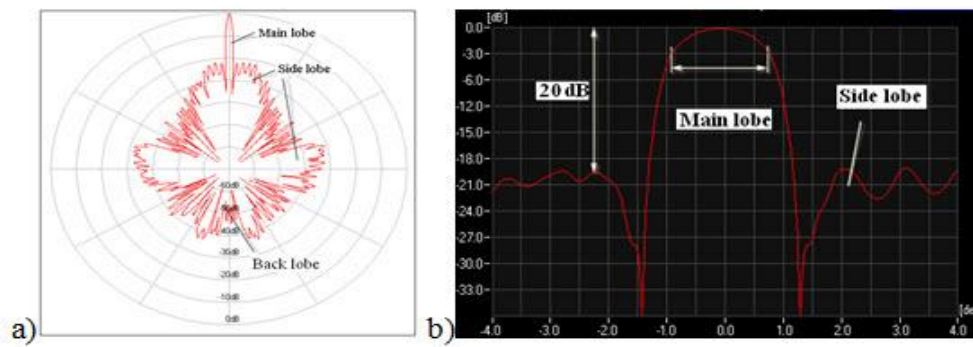
An antenna is a transducer to transform from a high frequency electric current to radio waves and vice-versa. An antenna is used to transmit and receive microwaves. In optical remote sensing the EM radiations used to focus with lenses as it cover wavelength 0.3 micrometer to 15 micrometer. Due to high wavelength, the microwaves are focused with antenna rather than lens.

Before we consider the realization of antennas, let us get familiarized with some of the terminologies used in antenna engineering.

#### Radiation pattern

A radiation pattern defines the variation of the power radiated (or received) by an antenna as a function of the direction away from the antenna. In an ideal isotropic antenna the radiation pattern is same in all directions; the plot of the antenna pattern would be formed as a sphere. Antennas with isotropic radiation patterns don't exist in practice, but are sometimes discussed as a means of comparison with real antennas. Radiation patterns are three-dimensional. However, they are usually measured in two orthogonal principal planes either in polar or rectangular coordinates. The patterns are normalized to the maximum value at 0 dB. In general, for any particular antenna for remote sensing, the pattern has a main lobe, where most of the energy is concentrated and sidelobes (Fig. 14.1), which are desirable.





**Fig. 14.1. Radiation pattern in a) polar coordinate, b) Cartesian system. (Source: [www.radartutorial.eu/18.explanations/ex42.en.html](http://www.radartutorial.eu/18.explanations/ex42.en.html))**

### Antenna gain

It is a measure of ability of the antenna to 'focus' the in a particular direction. Gain of an antenna is described as how much power is transmitted in a direction of peak radiation to that of an isotropic antenna. The total power of the antenna cannot be changed; but the energy can be redistributed in a particular direction of interest. Thus it is expressed relative to the performance of an isotropic antenna, which radiates equally in all direction and usually expressed in decibels (dB). An antenna with a gain of 3 dB means that the power received far from the antenna will be 3 dB higher than what would be received from an isotropic antenna with the same input power.

The gain  $G$  of an antenna with losses is given by,

$$G = \frac{4\pi\eta A}{\lambda^2} \quad (14.1)$$

Where,  $\eta$  is the efficiency, that is, the ratio of the total power radiated by the antenna to the net power fed to the antennas and  $\lambda$  is the wavelength.

$$A = \frac{\pi D^2}{4}$$

$A$  is the physical aperture area; for a particular aperture

### Beam width

Two points can be found on either side of the maximum power of the main lobe in the radiation pattern which represents half of the maximum power. These points are referred to as half power points. The angular range between the half power points is known as beam width. Since half power is 3 dB, the half power beam width is usually to as 3 dB beam width. The beam width ( $\beta$ ) is a function of the wavelength  $\lambda$  and the antenna aperture  $D$ .

$$\text{Antenna beam width } (\beta) \propto \frac{\lambda}{D} \quad (14.2)$$

That is beam width will be narrow if large antenna dimension increases.

Typical antennas in microwave remote sensing could be passive type of microwave radiometer, active types of microwave altimeter, scatterometer and imaging radar. There are three major types of antenna: horn antenna, parabolic antenna and array antenna.

### Horn antenna

It consists of a flaring metal waveguide shaped looks like a horn to direct radio waves in a beam. Due to its shape it is named as horn antenna. The flared portion can be square, rectangular, or conical. The maximum radiation and response corresponds with the axis of the horn. Horn antennas often have a directional radiation pattern with a high antenna gain, which can range up to 25 dB in some cases, with 10-20 dB being typical. Horn antennas have a wide impedance bandwidth, implying that the input impedance is slowly varying over a wide frequency range. Horn antennas have very little loss, so the directivity of a horn is approximately equal to its gain. This type of antenna is used in microwave radiometer.

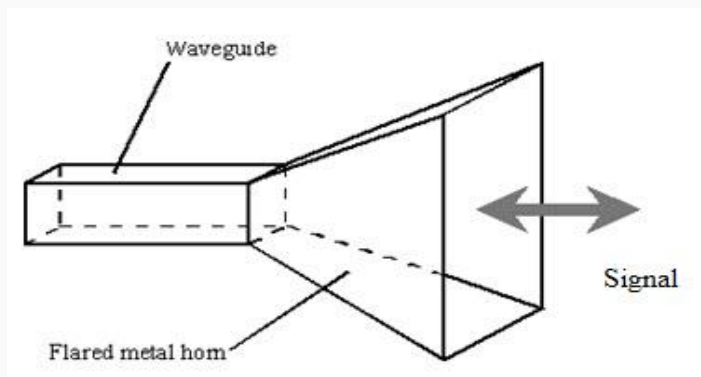


Fig. 14.2. Horn antenna.

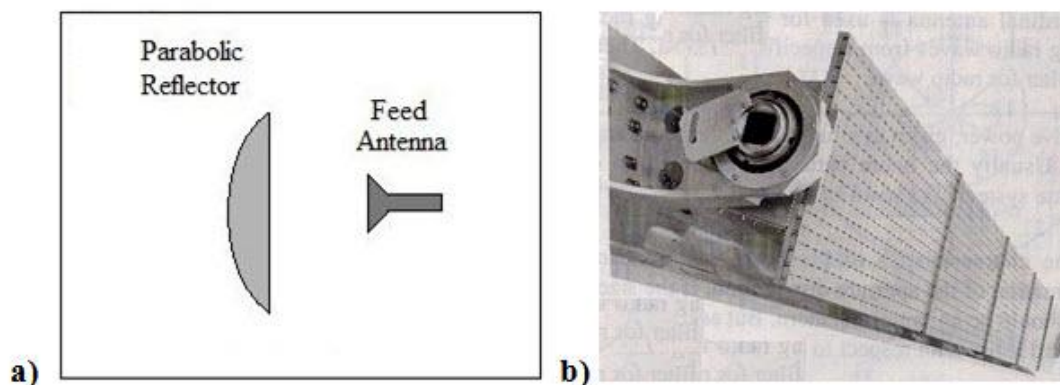
(Source: [searchmobilecomputing.techtarget.com/definition/horn-antenna](http://searchmobilecomputing.techtarget.com/definition/horn-antenna))

### Parabolic antenna

A parabolic antenna uses a parabolic reflector, a curved surface with the cross-sectional shape of a parabola, to direct the microwaves. The reflector is illuminated by a feed, usually a horn, which is located at the focal point of the reflector (Fig. 14.3. a). The main advantage of a parabolic antenna is that it has high directivity. Parabolic antenna has some of the highest gains, that is they can produce the narrowest beamwidth, of any antenna type. It is used in microwave radiometers, altimeters and scatterometers.

### Array antenna

An array antenna is composed of multiple element arrays for example, linear array, area array or nonformal array. The element antennas are half-wavelength dipoles, microstrip patches and wave guide slot (Fig. 14.3. b). The advantages of array antenna are to enable beam scanning without changing the looking angle of each array antenna and to generate an appropriate beam shaping by selective excitation of current distribution of each element. The array antenna is used in synthetic aperture radar (SAR) and real aperture radar.

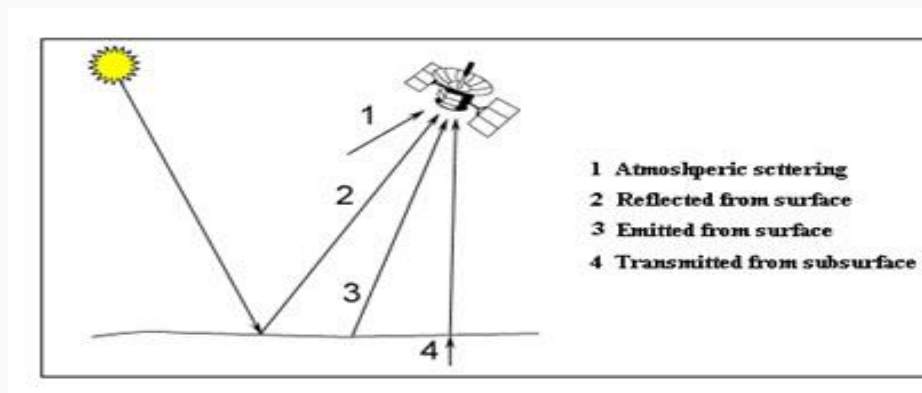


**Fig. 14.3. a) Parabolic antenna; b) array antenna.**

(Source: [www.antenna-theory.com/antennas/reflectors/dish.php](http://www.antenna-theory.com/antennas/reflectors/dish.php))

### 14.3 Passive Microwave Sensors

Passive microwave remote sensing is similar in concept to thermal remote sensing. All objects emit microwave energy of low magnitude. A passive microwave sensor detects the naturally emitted microwave energy within the field of view. This emitted energy is related to the temperature and moisture properties of the emitting object or surface. Passive microwave sensors are typically radiometers or scanners. The microwave energy recorded by a passive sensor can be emitted by the atmosphere (1), reflected from the surface (2), emitted from the surface (3), or transmitted from the subsurface (4), (Fig. 14.4.).

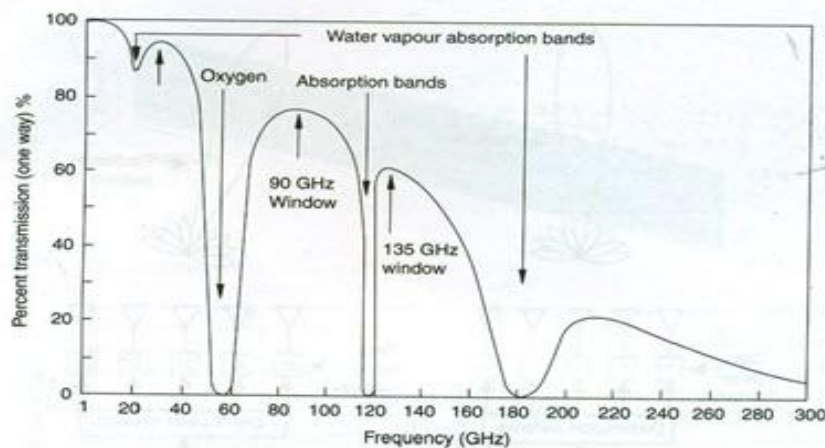


**Fig. 14.4. Components of passive microwave signal.**

The energy being too low in magnitude, in passive remote sensing the instantaneous field of view (IFOV) must be large to detect this energy, therefore most passive sensors are characterized by low spatial resolution. Microwave radiometer is a passive microwave sensor.

Passive microwave radiometers generally records energy in the region between 0.15 cm and 30 cm (between 1 GHz and 200 GHz), well beyond the thermal infrared region (3-14 micrometer). But the microwave is also radiated by thermal radiation from the objects on the earth. Thus microwave radiometers used in passive microwave remote sensing are also affected by atmosphere. In the Fig. 14.5 the atmospheric window for microwave remote

sensing is shown. Even measurements are at atmospheric window frequencies, the accurate estimation of surface radiance needs correction for these absorptions and emissions from the atmosphere. Most surface sensing radiometers include frequency channels also sensitive to water vapour and liquid water – mainly to correct for their effects. For observation of atmospheric parameters, frequencies are selected, which generally fall above 50 GHz.



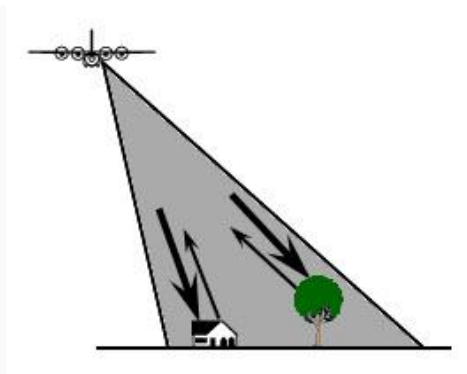
**Fig. 14.5. Atmospheric window in microwave remote sensing.**

(Source: Joseph, 2005).

Applications of passive microwave remote sensing include meteorology, hydrology, and oceanography. By looking "at", or "through" the atmosphere, depending on the wavelength, meteorologists can use passive microwaves to measure atmospheric profiles and to determine water and ozone content in the atmosphere. Hydrologists use passive microwaves to measure soil moisture since microwave emission is influenced by moisture content. Oceanographic applications include mapping sea ice, currents, and surface winds as well as detection of pollutants, such as oil slicks (Source: [www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/2021](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/2021)).

#### 14.4 Active Microwave Sensors

Active microwave sensors are generally divided into two distinct categories: imaging and non-imaging. The most common form of imaging active microwave sensors is RADAR (Radio Detection And Ranging), which is essentially characterizes the function and operation of a radar sensor. The sensor transmits a microwave signal towards the target and detects the backscattered portion of the signal. The time measurement gives the distance of the target from the sensor and hence gives the location of the target and the amount of energy backscattered depends on the target properties and hence gives a measure of some of the target characteristics (Fig. 14.6).



**Fig. 14.6. Active microwave remote sensing.**

(Source: [www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/2021](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/2021))

Non-imaging microwave sensors include microwave altimeters and scatterometers. In microwave scatterometer, the sensor sends short microwave pulse and measures the amount of energy backscattered from the surface of earth, used to make precise quantitative measurements of the amount of energy backscattered from targets. The amount of energy backscattered is dependent on the surface properties (roughness) and the angle at which the microwave energy strikes the target. As an example scatterometer measurements over ocean surfaces can be used to estimate wind speed based on the sea surface roughness. Ground-based scatterometers are used extensively to accurately measure the backscatter from various targets in order to characterize different materials and surface types. This is analogous to the concept of spectral reflectance curves in the optical remote sensing. Whereas microwave altimeters look straight down at nadir below the platform and thus measure height or elevation (if the altitude of the platform is accurately known).

**Keywords:** Microwave, Antenna, Antenna gain, Beam width, Radiometer, Scatterometer, Altimeter.

## Lesson 15 Radar Systems

Radar is an acronym for *radio detection and ranging*. Radar is an active sensor, which is very sensitive to terrain slope, and its characteristics. Radar can be imaging or non imaging. It transmits microwave signal and receives the reflection or echoes and determines the object distance. It also provides information about the surface features in case of remote sensing application. Radar can be mounted on aircraft or spacecraft. Depending on the working principle, a radar system in remote sensing can be categorized as Real Aperture Radar (RAR) or Side Looking Radar (SLR) and Synthetic Aperture Radar (SAR).

### 15.1 Real Aperture Radar

Imaging radar is classified into Real Aperture Radar (RAR) and Synthetic Aperture Radar (SAR). Side-Looking Airborne Radar (SLAR) was the first active sensor to produce imagery of the terrain from backscattered microwave radiation. In SLAR, an antenna is mounted underneath the platform to produce a fan beam (which is wide vertically and narrow horizontally) pointed to the side of the platform. The antenna produces a beam of narrow width in the along-track (azimuth) direction, and wide beam in the cross-track (range) direction. A short pulse (energy pulse over a time period on the order of microseconds) is transmitted radially by the antenna, which strikes the ground and backscattered to the antenna. Measuring the total time between transmitting and receiving the signal, the object distance from the platform is measured. The backscattered intensity which depends on the sensor and terrain characteristics is also measured by the radar.

There are few interrelated parameters of a radar system, defines the radar operation. These are

- (i) Frequency / wavelength
- (ii) Polarization
- (iii) Viewing geometry
- (iv) Spatial resolution

**(i) Frequency / wavelength:** The wavelengths used in microwave radar imaging are of in mm to m range. Different names are assigned to microwaves depending on its wavelength, primarily used for security purpose, which is continued for convenience use, given in Table 15.1.

**Table 15.1 Wavelength and frequency for radar bands**

Radar band name	Wavelength ( $\lambda$ ) in cm	Frequency ( $\nu$ ) in GHz
K <sub>a</sub>	0.75 – 1.18	40 – 26.5

K	1.19 – 1.67	26.5 – 18
K <sub>u</sub>	1.67 – 2.4	18 – 12.5
X	2.4 – 3.8	12.5 – 8
C	3.9 – 7.5	8 – 4
S	7.5 – 15	4 – 2
L	15 – 30	2 – 1
P	30 - 100	1 – 0.3

(Source: Bhatta, 2008)

Long wavelength microwaves can penetrate more than short wavelength microwaves. Atmospheric effect is observed on shorter wavelength microwaves. The L-band and P-band radar waves have the ability to penetrate the ground surface.

**(ii) Polarization:** Polarization refers to the orientation of the electric field which may be either vertical (V) or horizontal (H). Radar waves can be transmitted or received at different modes of polarization. There are four combinations of V, H polarized wave transmission and reception as follows:

1. horizontal (H) transmit & horizontal (H) receive: (HH)
2. vertical (V) transmit & vertical (V) receive: (VV)
3. horizontal (H) transmit & vertical (V) receive: (HV)
4. vertical (V) transmit & horizontal (H) receive: (VH)

HH and VV polarization are said like polarization; HV and VH polarization are said cross-polarization. The strength of backscattered signal for like-polarized wave is greater than cross-polarized wave.

**(iii) Viewing geometry:** In a radar microwave remote sensing the flight direction is said azimuth direction, and perpendicular to flight direction is said range direction. On the surface, range direction is said ground range direction. The point just below the platform is said nadir. Antenna transmit microwave in the range direction in a fan shape. The nearest point to the nadir of the fan shaped beam is known as near range and farthest point is known as far range. Distance between near range and far range is the swath. Ground range is the horizontal distance measured along the surface from the nadir track to the target.

Incidence angle is the angle between the radar beam and the perpendicular of the surface at a point. Depression angle is the angle of the radar beam to the target (i.e., line of sight from the antenna to the target) measured from a horizontal plane. Look angle is the angle of the radar beam to the target measured from vertical plane (Fig. 15.1).



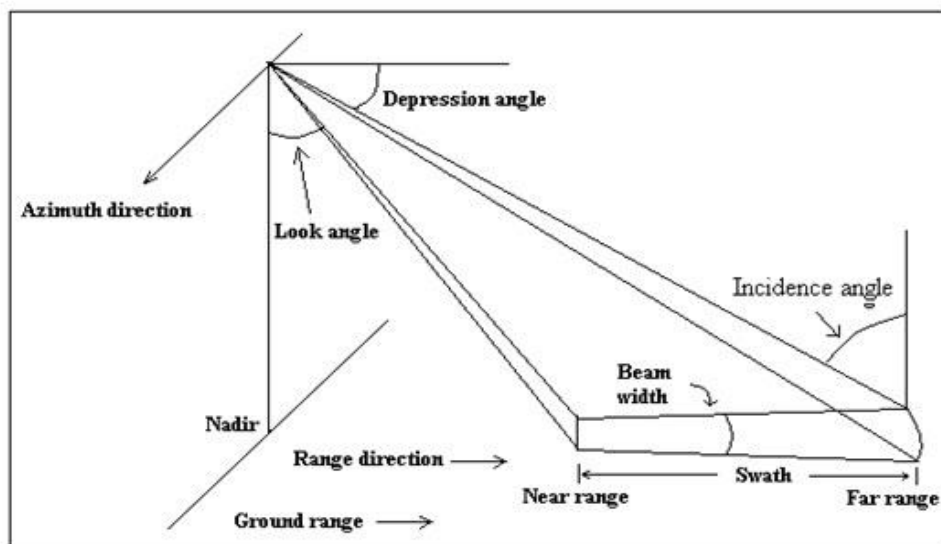


Fig. 15.1. Viewing geometry.

(iv) **Spatial resolution:** Spatial resolution is one of the most important parameter in radar imaging. There are two types of resolution associated with radar imaging: a. range resolution and b. azimuth resolution.

**a. Range resolution:** To distinguish two objects in a radar image in the range direction, it is necessary to reach the backscatter from the two objects separately. In the Fig. 15.2, a microwave of pulse length  $L$  is transmitted, which is backscattered from objects A and B, situated at a distance  $d$  in the slant range direction and  $R$  in the ground range direction.

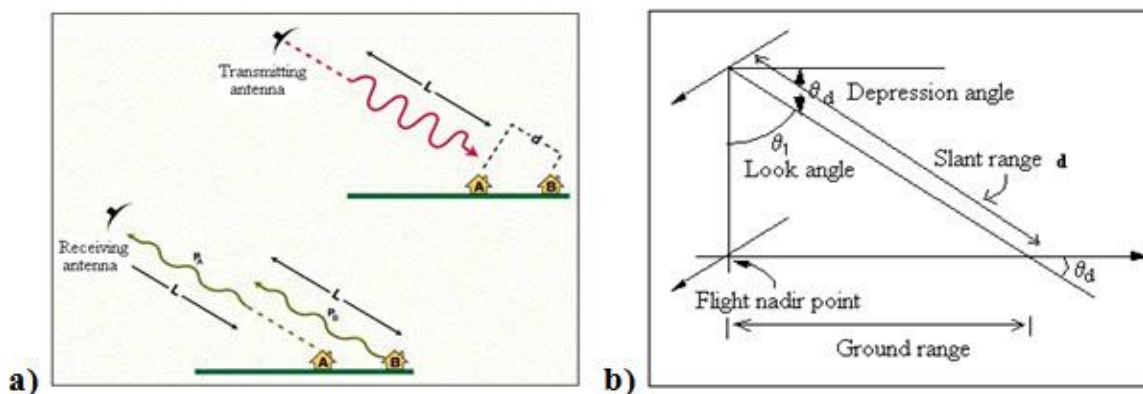


Fig. 15.2. Dependence of spatial resolution on pulse length.

(Source: [earth.esa.int/applications/data\\_util/SARDOCS/spaceborne/Radar\\_Courses/Radar\\_Course\\_II/real\\_aperture\\_radar\\_range\\_resolution.htm](http://earth.esa.int/applications/data_util/SARDOCS/spaceborne/Radar_Courses/Radar_Course_II/real_aperture_radar_range_resolution.htm))

So, the objects A and B can be discriminated in the imagery if the slant range distance between A and B is greater than half of the pulse length. If the front of the backscattered wave from B touches the rear of the backscattered from A (Fig. 15.2(a)). Therefore, they cannot be resolved in the imagery. If  $t$  be the pulse duration,  $c$  be the velocity of microwave, then the ground range resolution ( $R$ ) can be expressed as:



$$R = (ct)/(2\cos\theta_d) \quad (15.1)$$

where  $\theta_d$  is the depression angle and  $t$  is the time in  $\mu\text{sec}$ .

Further, this should be noted that the ground range resolution can be increased by increasing the depression angle and decreasing the pulse length. But depression angle increases,  $\cos\theta_d$  decreases, which will in turn increase the ground range.

**Example 15.1:** A SLAR system transmits pulse over duration of  $0.2 \mu\text{sec}$  at a depression angle of  $23^\circ$ . What would be the ground range resolution?

**Solution:**

From equation (1), ground range resolution  $R$

$$\begin{aligned} R &= (ct)/2\cos\theta_d \\ &= ((3 \times 10^8) (0.1 \times 10^{-6}))/ (2 \times \cos 23^\circ) \text{ m. sec}^{-1}.\text{sec} \\ &= 16.3 \text{ m} \end{aligned}$$

**b. Azimuth resolution:** The spatial resolution in the azimuth direction is defined by the signal beam width ( $\beta$ ). As the beam is fans on the ground, thus the spatial resolution in the azimuth direction varies in the ground range direction. In the Fig. 15.3, at the point A, three points are resolved spatially, where as the points in point B (spaced equally as at the point A) are not resolved. That means at near range the azimuth resolution is higher than far range. The azimuth resolution ( $r$ ) can be expressed as:

$$r = \beta P \quad (15.2)$$

where,  $P$  is the slant range distance.

$$P = \frac{h}{\cos \theta} \quad (15.2)$$

$h$  is the altitude of platform,  $\theta$  is look angle,

$$r = \frac{h\beta}{\cos \theta}$$

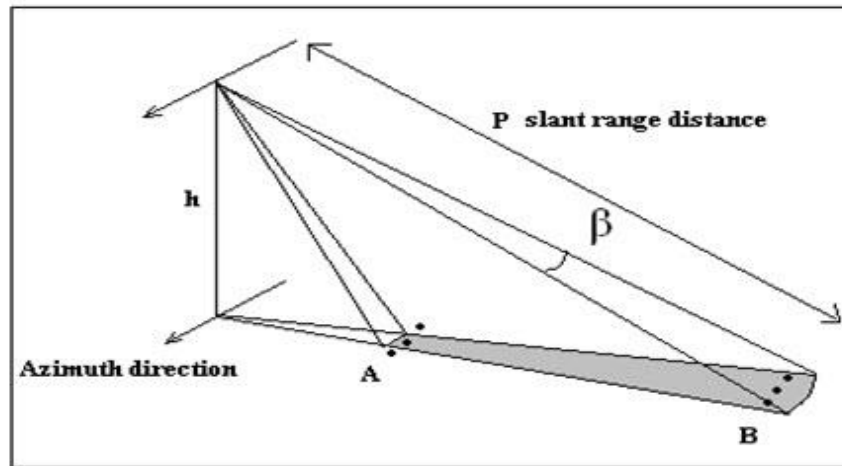
or,

If  $D$  be antenna length,  $\lambda$  be the operating wavelength, then antenna beam width ( $\beta$ )

$$\beta = \frac{\lambda}{D}$$

Thus we have azimuth resolution ( $r$ )

$$r = \frac{h\lambda}{D \cos \theta} \quad (15.3)$$



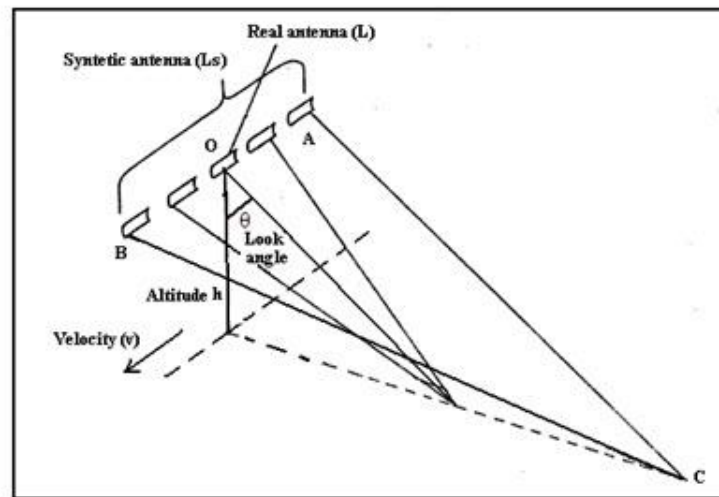
**Fig. 15.3. Azimuth resolution.**

To increase resolution by decreasing wavelength leads to cloud and atmospheric attenuation. For a particular wavelength, the azimuth resolution can be increased by decreasing the platform altitude, which is nearly constant for space borne satellites. Another way of increasing azimuth resolution is by increasing the antenna length. But increasing the antenna length physically at a big scale is impractical. For, an example, a SLAR system operating in 3 cm wavelength, having altitude 705 km. and 30° look angle. To obtain 30 m resolution, the antenna length should be 814 m. which is impractical. In case of airborne radar antenna length is usually 1-2 m, and for space borne 10-15m.

So, the resolution can be increased by increasing the antenna length virtually, which is known as synthetic aperture radar (SAR).

## 15.2 Synthetic Aperture Radar

The deficiency of real aperture radar (RAR) is overcome in synthetic aperture radar (SAR) systems. In SAR systems a short physical antenna is used as in RAR, which is synthesized so that it can act as a very long antenna. The result of this mode of operation is a very narrow effective antenna beamwidth, even at far ranges, without requiring physically long antenna or a short operating wavelength, which increases the azimuth resolution subsequently.



**Fig. 15.4. Concepts of real antenna positions forming a synthetic aperture.**

The concept of SAR operation is shown in Fig 15.4. The real antenna length ( $L$ ) is increased virtually at a synthetic aperture antenna length ( $L_s$ ). This results in essentially constant azimuth resolution irrespective of range. Through this process, the antenna length can be made several km. The antenna starts to transmit signal at position A and continues till B, thus backscatter from target C can be recorded during this entire time of signal transmission. Objects at far range, where the beam width is wider illuminated for longer time than the objects of near range where the beam width is narrower; thus it maintains constant range resolution. Due to relative motion of platform and object, a Doppler shift is observed in the signal. From the path A - O, the frequency of the signal will increase, at the points nearer to the point O, the frequency shift will be very less, and from point O - B the frequency will be decrease. This error can be processed accounting the Doppler shift.

**Example 15.2:** In the Fig. 15.4., a space borne radar operating at wavelength 3 cm, at an altitude 700 km. If the look angle is  $30^\circ$  and synthetic aperture length is 1 km, what would be the azimuth resolution of the data obtained?

**Solution:**

Using the equation (2), ground range resolution  $R$

$$\begin{aligned}
 r &= h\lambda / L\cos\theta \\
 &= ((700 \times 10^3) (3 \times 10^{-2})) / (1 \times 10^3 \times \cos 30^\circ) \text{ m.} \\
 &= 24 \text{ m.}
 \end{aligned}$$

### 15.3 Scatterometer

Scatterometer measures the power of the backscatter reflected from the surface of the earth. The amount of backscatter depends on the properties and characteristics of the surface of the earth. A microwave scatterometer may be a space borne sensor to measure the two dimensional velocity vectors of the sea wind, as well as ground based sensors to measure the

surface backscattering as well as volume scattering, such as rain radar. It is calibrated to measure accurately scattering the coefficient  $\sigma^0$ . The principle of a scatterometer may be pulse type or continuous wave type. A pulse type scatterometer transmits signal, the backscatter intensity and the time delay is measured, gives the information and location of the target.

The scattering coefficient  $\sigma^0$  is related to other instrument as follows

$$\sigma^0 = \frac{(4\pi r)^2 R^4}{G_o^2 \lambda^2 A_w} \quad (15.4)$$

where,

$G_o$ = the maximum gain of the antenna

$\lambda$ = wavelength

$R$ = Range

$P_t$ = transmit power

$P_r$ = received power

$A_w$ = weighted area

$g(\theta, \Phi)$ = normalized antenna gain (radiation pattern) such that the antenna gain

$G = G_o g(\theta, \Phi)$

$A_i$ = area of illumination

SESAT-A Scatterometer (SASS) was launched in 1978 is one of the typical scatterometer. SASS uses four beam antennas- two on both the sides of the sub-satellite track to receive the backscattering in subdivided cells through a Doppler filter. Two antennas on each side are aligned so that they are pointed  $45^\circ$  and  $135^\circ$  relative to the spacecraft flight direction. Thus, the footprint of the four antenna beam produce an 'X' shaped illumination pattern and on the earth (Fig. 15.5.). Thus, ant surface location is viewed by both the forward and aft antenna, near orthogonally (separated by approximately  $90^\circ$ ).

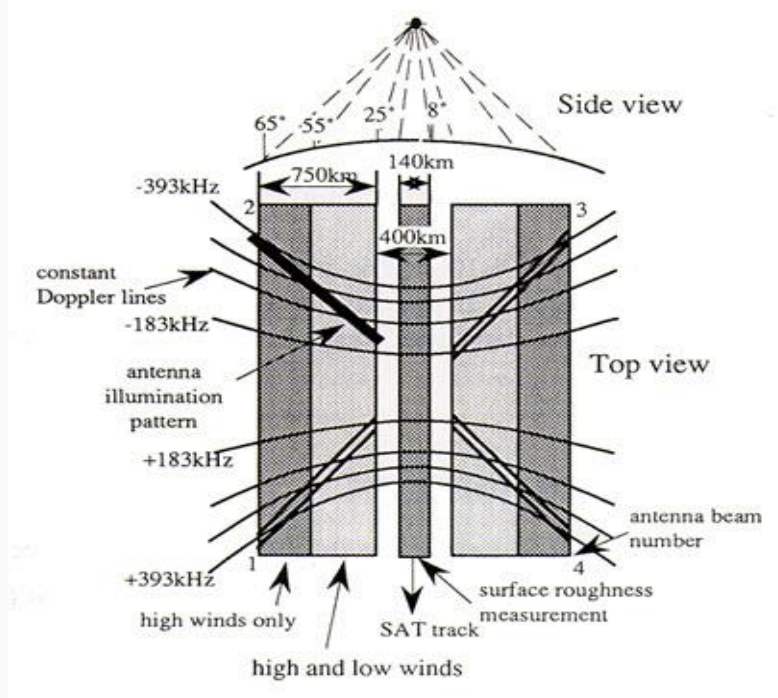


Fig. 15.5. Seasat SASS fan-beam coverage.

(Source: [stlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp4/4-9-1.gif](http://stlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp4/4-9-1.gif))

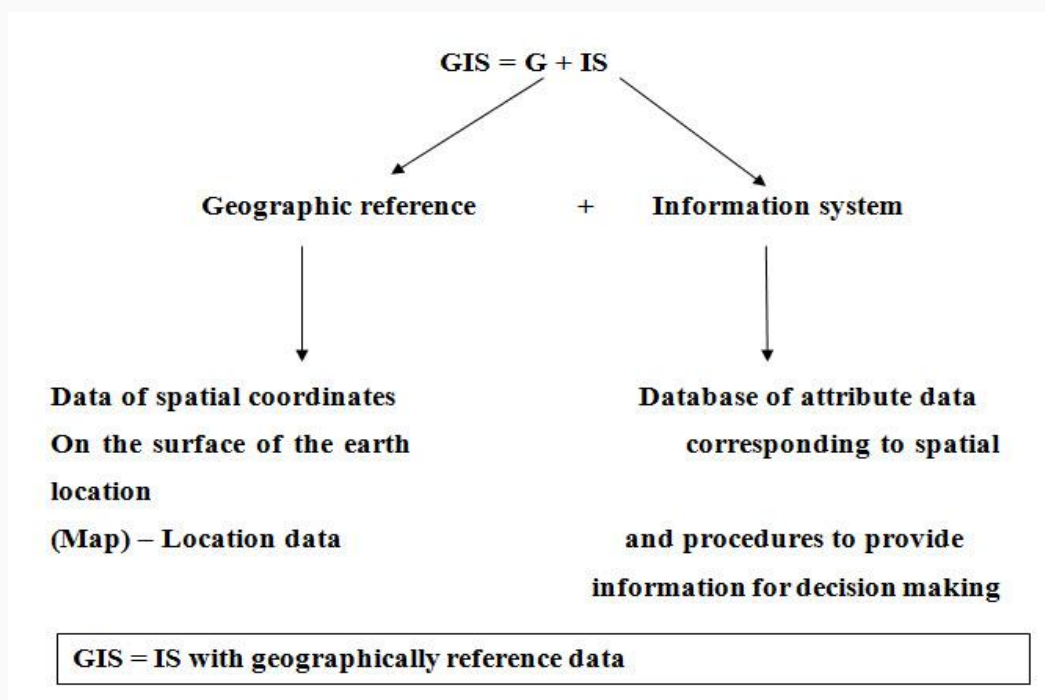
**Keywords:** Real Aperture radar (RAR), Synthetic Aperture Radar (SAR), Side-Looking Airborne Radar (SLAR), Range resolution, Azimuth resolution, Polarization, Scatterometer.

## Module 6: Geographic Information Systems (GIS)

### Lesson 16 Introduction to GIS

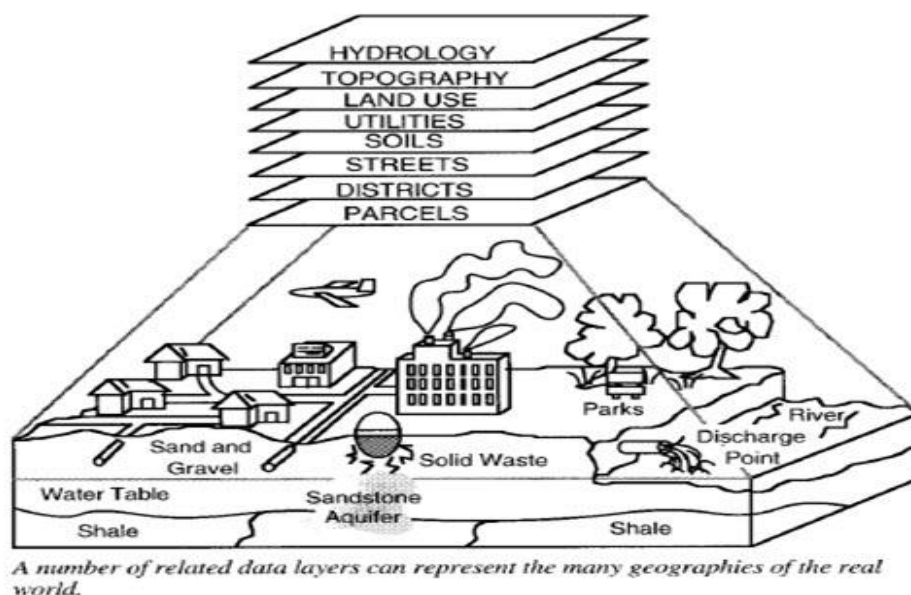
#### 16.1 Definition

A geographic information system (GIS) is basically a computerized information system like any other database, but with an important difference: all information in GIS must be linked to a geographic (spatial) reference (latitude/longitude, or other spatial coordinates).



There are many different definitions of GIS, as different users stress different aspects of its use. For example:

- (i) ESRI defined GIS as an organized collection of computer hardware, software, geographic data and personnel designed to efficiently capture, store, update, manipulate, analyze and display geographically referenced information.
- (ii) ESRI also provided a simpler definition of GIS as a computer system capable of holding and using data describing places on the earth's surface. (Shown in Fig. 16.1).



**Fig. 16.1. Real world data layers**

**(Source- IAI Summer Institute 19 July 2000).**

(iii) Duecker defined GIS as a special case of information systems where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines or areas. A GIS manipulates data about these points, lines or areas to retrieve data for ad hoc queries and analyses.

(<http://www.naarm.ernet.in/virtual/downloads/GIS%20concepts.pdf>)

### **16.1.1 Toolbox-based definitions.**

A powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world. (Burrough, 1986)

A system used for capturing, storing, checking, manipulating, analysing and displaying data which are spatially reference to the Earth. (Department of Environment, 1987)

An information technology stores, analyses, and displays both spatial and non-spatial data. (Parker, 1988)

### **16.1.2 Database definitions.**

A database system in which most of the data spatially indexed, and upon which a set of procedures operated in order to answer queries about spatial entities in the database. (Smith, 1987)

Any manual or computer based set of procedures used to store and manipulate geographically referenced data. (Aronoff, 1989)

### 16.1.3 Organization-based definitions

An automated set of functions that provides professionals with advanced capabilities for the storage, retrieval, manipulation and display of geographically locate data. (Ozemoy, Smith, and Sicherman, 1981)

An institutional entity, reflecting an organizational structure that integrates technology with a database, expertise and continuing financial support over time. (Crater, 1989)

A decision support system involving the integration of spatially referenced data in a problem solving environment. (Cowen, 1988)

([http://www.tiger.esa.int/TrainingCds/cd\\_01/content\\_2/sez\\_2\\_3/Unit-III-GIS.pdf](http://www.tiger.esa.int/TrainingCds/cd_01/content_2/sez_2_3/Unit-III-GIS.pdf))

### 16.1.4 System flow approach

From the very start, GIS practice emphasized the S word: systems perhaps the most pervasive metaphor for the twentieth century (Harvey, 1997b). Operations research, developed as a distinct practice during World War II, provided a technique of 'systems analysis' that helped bring the computer into nearly every part of modern life. GIS was not alone in being conceived as a series of procedures that lead from input to output; from data sources through processing to displays.

Marble (1990) elaborated the definition in more expansive form by detailing four 'subsystems':

1. A data input subsystem which collects and/or processes spatial data derived from existing maps, remote sensors, etc.
2. A data storage and retrieval subsystem which organizes the spatial data in a form which permits it to be quickly retrieved by the user for subsequent analysis, as well as permitting rapid and accurate updates and corrections to be made to the spatial database.
3. A data manipulation and analysis subsystem which performs a variety of tasks such as changing the form of the data through user-defined aggregation rules or producing estimates of parameters and constraints for various space-time optimization or simulation models.
4. A data reporting subsystem which is capable of displaying all or part of the original database as well as manipulated data and the output from spatial models in tabular or map form. The creation of these map displays involves what is called digital or computer cartography. This is an area which represents a considerable conceptual expansion of traditional cartographic approaches as well as a substantial change in the tools utilized in creating the cartographic displays.

This definition offers a sense of the stages of operation, but little about the internals. The adjectives 'quickly', 'rapid' and 'accurate' tell us little about the means of organization. The



definition of one system as a set of four 'subsystems', arranged in linear sequence, adopts a recursive strategy in which systems are explained by more systems.

While the basic system flow definition continues to be used, as the widespread adoption of GIS began in the 1980s, there was a sense that the definition was overly technical (Rhind 1996). One of the most commonly cited alternatives was developed by a Delphi panel of thirty specialists, Nicholas R Chrisman that Geographic Information System is a system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing and disseminating information about areas of the earth (Dueker and Kjerne, 1989).

While definitions pop up everywhere, there is remarkably little attention given to comparing definitions and to evaluating them. Cowen (1988) provided four approaches to a definition: process-oriented, application, toolbox, and database. Maguire (1991) described the approach to GIS from three viewpoints: the map view, the database view, and the spatial analysis view.

DeMers (1997) told that tools that allow for the processing of spatial data into information, generally information tied explicitly to, and used to make decisions about, some portion of the Earth.

Star and Estes (1990) described that an information system that is designed to work with data referenced by spatial or geographic coordinates; both a database system with specific capabilities for spatially-referenced data, as well as a set of operations for working with the data.

Clarke (1997), referring to Burroughs (1986): A powerful set of tools for storing and retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes. He also told that it is an automated system for the capture, storage, retrieval, analysis and display of spatial data.

Ron Abler (1988): GIS are simultaneously the telescope, the microscope, the computer, and the Xerox machine of regional analysis and synthesis of spatial data.

Although divisions of definitions occur in the literature, in practice it is rather arbitrary to assign a particular definition to just one approach. Some of the definitions reported come from the traditional literature sources, but some are drawn from the World-Wide Web, a forum of increasing importance for the description of the field and its promotion to newcomers.

(<http://www.naarm.ernet.in/virtual/downloads/GIS%20concepts.pdf>)

### **16.1.5 Changing the subject**

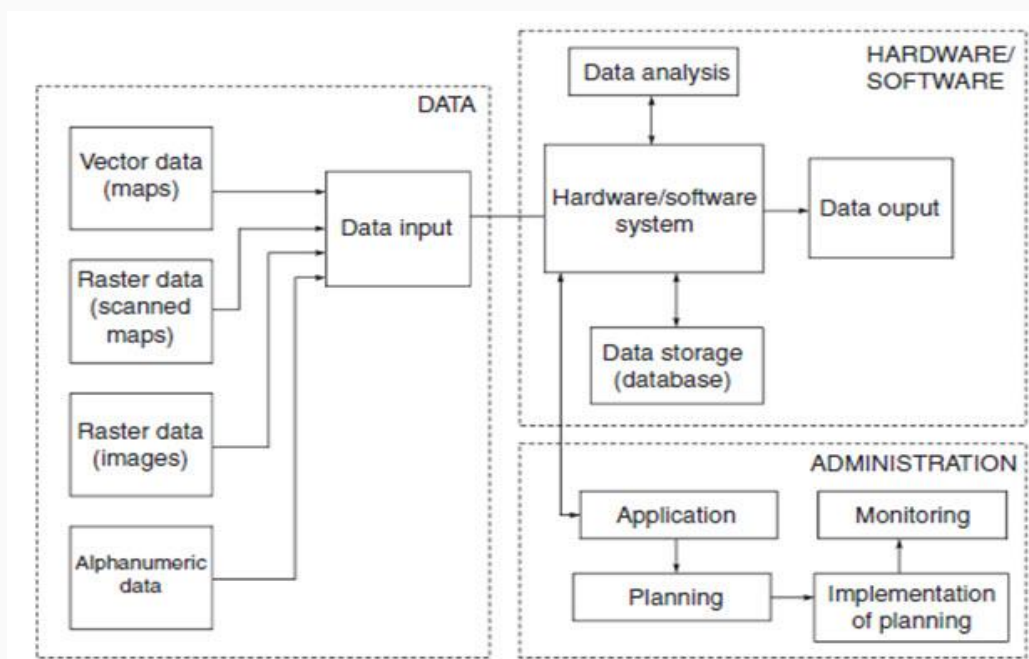
There has been a substantial effort to shift the emphasis by declaring that GIS should stand for 'Geographic Information Science' (Goodchild 1992).

Goodchild's original intent was neatly summarized in his unapologetic chapter in

Ground Truth:

GIS . . . has done much to remove the traditional isolation between photo- grammetry, remote sensing, geodesy, cartography, surveying, and geography (one could add to this list computer science, operations research, spatial statistics, cognitive science, behavioral psychology, and any other discipline with interests in the generic issues of spatial data). In an earlier paper, I argued that these were the disciplines of geographic information science, and that it made more sense for the research community to decode the GIS acronym in this way, focusing on the generic issues of spatial data, rather than on the limited solutions offered by today geographic information system products (Goodchild 1995, p 42 emphasis in original). (Nicholas, 1999)

A geographic information system (GIS), in a narrow definition, is a computer system for the input, manipulation, storage and output of digital spatial data. In a more broad definition it is a digital system for the acquisition, management, analysis and visualization of spatial data for the purposes of planning, administering and monitoring the natural and socioeconomic environment. It represents a digital model of geography in its widest sense (shown in Fig. 16.2). (Gottfried Konecny. Geoinformation-Remote Sensing Photogrammetry & Geographical Information Systems).



**Fig. 16.2. Concept of a geographic information system**

(Source-Geo-information, by Gottfried Conecny)

## 16.2. Component

There are four main components of a true GIS system (Marble 1990). These are:

1. **Data input system:** collects and/or processes spatial data from existing sources such as maps, remote sensing data, images, etc. Data can be "collected" through digitizing, scanning, interactive entry, etc.

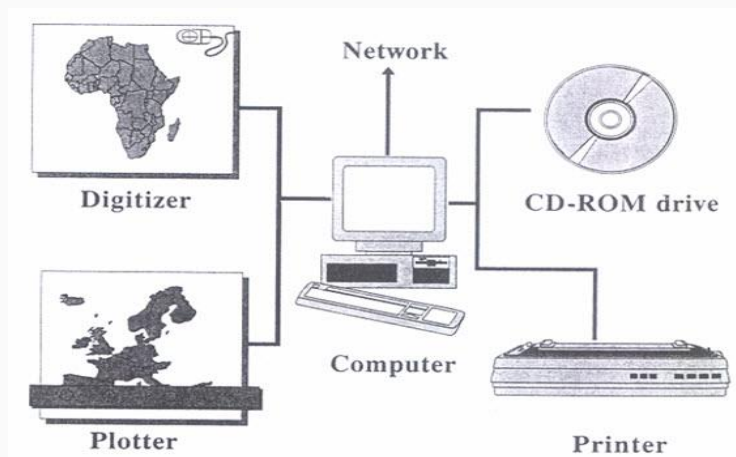
2. **Data storage and retrieval:** organizes spatial data and allows for quick retrieval and updates (i.e., editing).
3. **Data analysis and manipulation:** allows for changing form of data, simulation modeling, spatial-temporal comparison, etc.
4. **Output:** displays spatial database and analysis in graphic (i.e., map) or tabular form.

(Biology 483: Application of GIS (2002). Applications of geographic Information Systems (GIS): Introductory Lecture)

GIS component can also be divided by in this manner:

1. Computer hardware.
2. Sets of application software modules.
3. Skilled people to manage it.

### 1. Computer hardware



**Fig. 16.3. Major hardware components of a geographical information system.**

(Source- [http://www.tiger.esa.int/TrainingCds/cd\\_01/content\\_2/sez\\_2\\_3/Unit-III-GIS.pdf](http://www.tiger.esa.int/TrainingCds/cd_01/content_2/sez_2_3/Unit-III-GIS.pdf))

**Table: 16.1. The basic computer hardware components.**

Item	Usage
Hard disk drive	Storing data and programs
Digital tape cassettes, optical CD-ROM's etc.	Storage of data

Digitizer or Scanner	Converts maps and documents into digital form
Plotter, Printer or any other display device	Gives the output of data processing
Local & Global electronic network with either of the following:  1. Optical fibre data lines 2. Telephone lines with 'modem'	Provides Inter-Computer communication
Computer screen, Keyboard & mouse or other pointing device	To control the computer and the peripherals such as the digitizer, plotter, printer etc. which are linked to the computer

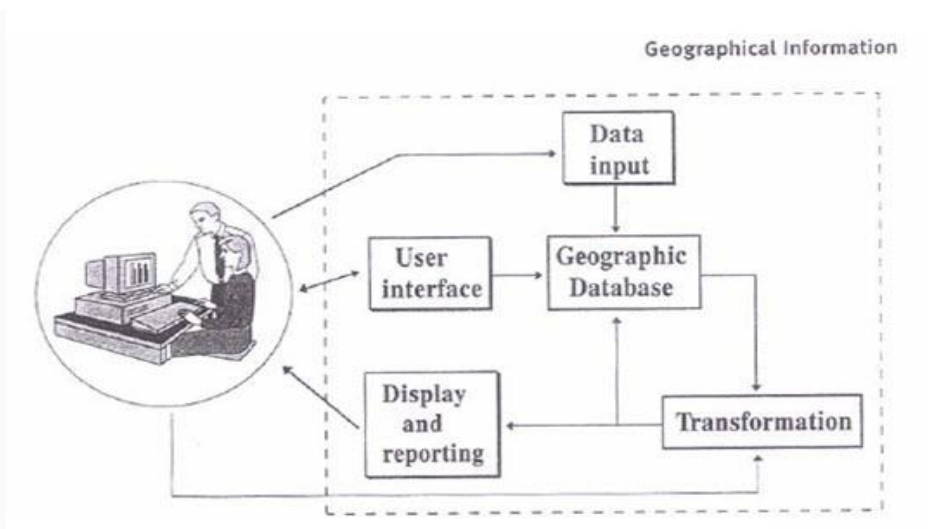
(Source: [http://www.tiger.esa.int/TrainingCds/cd\\_01/content\\_2/sez\\_2\\_3/Unit-III-GIS.pdf](http://www.tiger.esa.int/TrainingCds/cd_01/content_2/sez_2_3/Unit-III-GIS.pdf))

## 2. GIS Software

The software modules can be grouped as follow-

1. Data input and verification
2. Data storage and database management
3. Data output and presentation
4. Data transformation
5. interaction with the user

([http://www.tiger.esa.int/TrainingCds/cd\\_01/content\\_2/sez\\_2\\_3/Unit-III-GIS.pdf](http://www.tiger.esa.int/TrainingCds/cd_01/content_2/sez_2_3/Unit-III-GIS.pdf))



**Fig. 16.4. The main software components of a geographical information system.**

(Source-([http://www.tiger.esa.int/TrainingCds/cd\\_01/content\\_2/sez\\_2\\_3/Unit-III-GIS.pdf](http://www.tiger.esa.int/TrainingCds/cd_01/content_2/sez_2_3/Unit-III-GIS.pdf))

### 16.2.1 GIS Functional Components

Because of the different origins these systems have, it's important to understand each of the functional components of a GIS. Each of the various GIS software packages emphasizes certain aspects of spatial data handling and deemphasizes or omits others. The degree of emphasis placed on development of certain features depends on the marketplace a vendor is targeting. None of the current GIS software packages place their emphasis on the defense community; therefore, an understanding of the basic components of a GIS is needed. Understanding these underlying concepts will help users in their research of the technology and also in the development of applications for the defense community (NSA, 2008).

#### Data capture and input processing

The first GIS component is Data Capture and Input Processing. In order to successfully implement a GIS, one must have data available for the study area. This includes the acquisition of the data, its digitization, and the appropriate tagging of attributes. Sources may include hard copy maps, existing digital data, imaginary and tabular data. The format, coordinate system, and geographic projection of the data must be known before input into a GIS. One must also be able to generalize the data and select only the amount of data are necessary for a particular project. Digital data take up a huge amount of storage space on computer systems; therefore the thinning and proper selection of this data is important. Most of the GIS software packages have algorithms and methods to aid in. this process. One must also be aware of the amount of error that exists in the data being used and fully understand its limits and implications in project accuracy. Inherent data errors have been a traditional bottleneck in the development of GIS technology. Research is currently being done to help quantify this error (NSA, 2008).

## Digital data formats

"All data that can be mapped have both a locationl (x,y) and nonlocationl (i.e., attribute) characteristics.. These attributes can be both qualitative (e.g., the land use at a location) and quantitative (e.g. the elevation at the same location). In addition the attributes data location can be monitored through time. These three components location, attribute and time - represent the content of most GIS. This information has to be somehow represented inside the GIS. Map data in this particular format are called digital data, and the process of getting the data into this format is called digitization. Digital data are represented in the computer as large sets of numbers, not as analog images. Two different data structures are commonly used to represent map data inside the GIS: raster and vector representations. Four fundamental types of geographic data have to be stored within a GIS: point lines, polygons, and surfaces. T Raster and vector data representations use different techniques to store points, lines, polygons, and surfaces. These techniques will be discussed because they provide a basis for comparison of the two different types of digital data representations (NSA, 2008).

## Data storage and data management

The second functional component of a GIS is the role of data storage and the management of these data. Once the data encoded in their proper digital format, they must be stored in the GIS. Most of the GISs use a database mode to store these data. Today's Geographic Information Systems can use either a hierarchical network or a relational database model to achieve this component. The geographic information is arranged in files of related information, each file being called a layer. Each of these layers can be combined or overlaid upon each other to form new layers. These newly created layers form the basis for geographical analysis and can be queried to answer questions of interest to the user. Each of the attributes associated with this geographic data is stored in the database alongside its data structure and is queriable.

Since the amount of data needed is usually large, it's recommended that someone interested in setting up a GIS obtain the proper amount of disk storage. With the price of disk space decreasing and the storage capability on each hard disk drive increasing, the space requirements needed for these databases are becoming obtainable. One must always pay attention to the storage capacity demands of both the GIS package itself and the digital data being used for the project (NSA, 2008).

## Data manipulation

The third functional component of a GIS is the role of data manipulation. To extract meaningful information from a GIS .database, one must be able to query it and ask logical questions. The leading database model used in GIS technology is the relational database. Relational databases have the ability to join different attribute tables to create new relationships among the data. This concept is important to the GIS in that the geographic data are stored in the database along with its attribute tables, which enhance the geographic data. This relationship helps make all features within a GIS queriable. When layers of data are combined, the attribute information for this material is carried along and also becomes queriable. As new information is added to the database and geographic layers are combined

among themselves, the newly created geographic and spatial queries aid in performing actual analysis on the data. For example, a typical query could be to find all features of a defined type within a certain area. Another is to find all features that are adjacent to a specified feature. A third is to find all features that are a certain distance from another defined feature. A point and click type query could also be implemented (e.g., point to a road and tell me the attributes that are associated with that road) (NSA, 2008).

### **Data display and output**

The fourth and final functional component of a GIS is the role of data display and output. All GISs should include software for this capability and they should provide means for both soft- and hardcopy output. The ability to interface with output peripherals such as wax thermal printers and plotters to be able to produce a map depicting the results of analysis is important to say the least. Report generation and business graphic generation are necessary for some applications. Geographic Information Systems have been found to be lacking in this arena. Tabular data could be imported into desktop publishing packages or spreadsheet packages with little trouble to compensate for this deficiency. Spreadsheet packages could be used to produce graphical output displays such as histograms and time/frequency plots. One should also understand the types of maps he wants to produce when evaluating GIS software packages. Consider this list of maps when deciding what type of output capability a package has: polygon/choropleth maps, contour/isarithm maps, three-dimension/perspective map and grid-cell maps (NSA, 2008).

### **16.3 Architecture**

Geographic information systems (GIS) are becoming more usual due to the improved performance of computer systems. GIS applications are being developed using the three-tier software architecture traditionally used for general purpose information systems. Even though this architecture is suitable for GIS applications, the special nature and exclusive characteristics of geographic information pose special functional requirements on the architecture in terms of conceptual and logical models, data structures, access methods, analysis techniques, or visualization procedures. The architecture of general-purpose information systems must consist of three separate tiers, namely: the presentation tier, the application logic tier (or business logic tier), and the data tier. The main advantage of this architecture is that it enforces a strict separation of the functionality of the system into three different independent modules that interact only at well-defined interfaces. This enables a developer to modify each one of these modules of the application with little impact on the others. Therefore, this architecture provides increased performance, flexibility, maintainability, reusability and scalability. Even though the three-tier architecture for general-purpose information systems is suitable for GIS, the special nature and exclusive characteristics of geographic information pose special functional requirements on the architecture in terms of conceptual and logical models, data structures, access methods, analysis techniques, or visualization procedures. For instance:

- Special data types and operations are needed to represent and manipulate geographic information.
- Geographic information requires many different analysis and visualization procedures.
- Geographic information is typically voluminous with a naturally imposed hierarchical structure.
- Geographic information processing is characterized by transactions that are much longer than a typical standard relational database transaction.
- There are two different conceptual views of geographic space: an object-based view and a field-based view.

These and other features impact the overall architecture of a GIS. (Miguel et al., 2005)

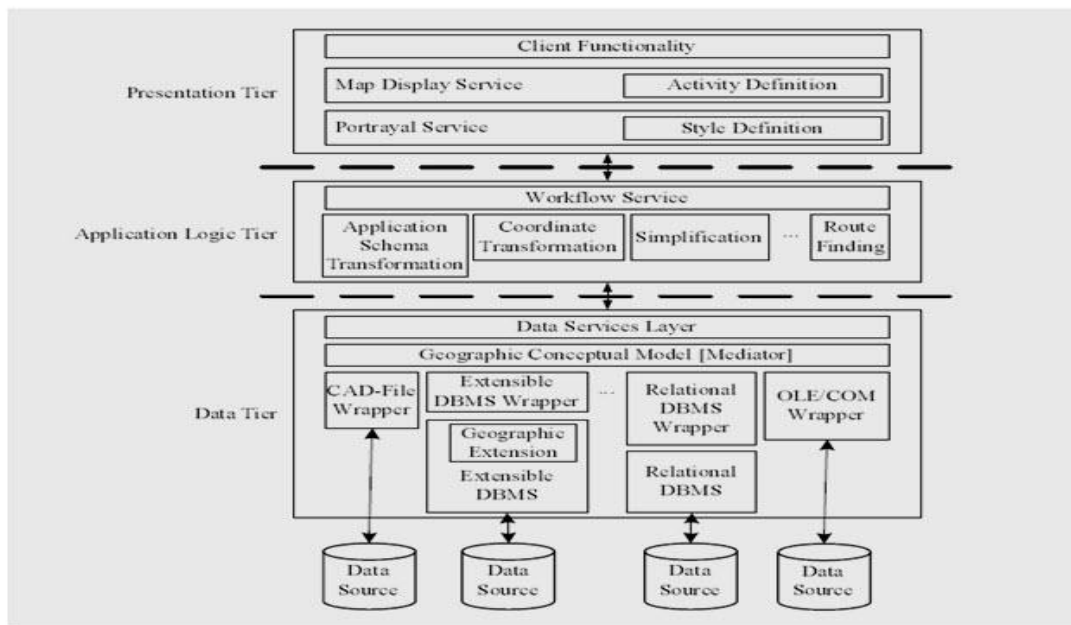
### **16.3.1 System Architecture**

The special nature of geographic information makes more important the fulfillment of some requirements of general-purpose information systems, such as flexibility, extensibility, reusability, scalability, reliability, and security. In order to provide these features, the architecture of the GIS must be based on an extensible DBMS providing geographic information management services, and a collection of modular, highly-distributed, geographic information processing and visualization services (Miguel et al., 2005)

### **16.3.2 A Generic Architecture**

This architecture is heavily influenced by the proposals of the ISO/TC 211 and OGC, and reuses the work of these organizations where their specifications are mature. Fig. 16.5 shows the proposal of a generic architecture for geographic information systems. The architecture separates the functionality of the system in three independent tiers, namely the Data Tier, the Application Logic Tier and the Presentation Tier. The Presentation Tier is responsible for implementing the user interface of the system, displaying the maps and providing some basic functionality over them. Finally, the Application Logic Tier implements the problem-solving functionality of the system. (Miguel et al., 2005)



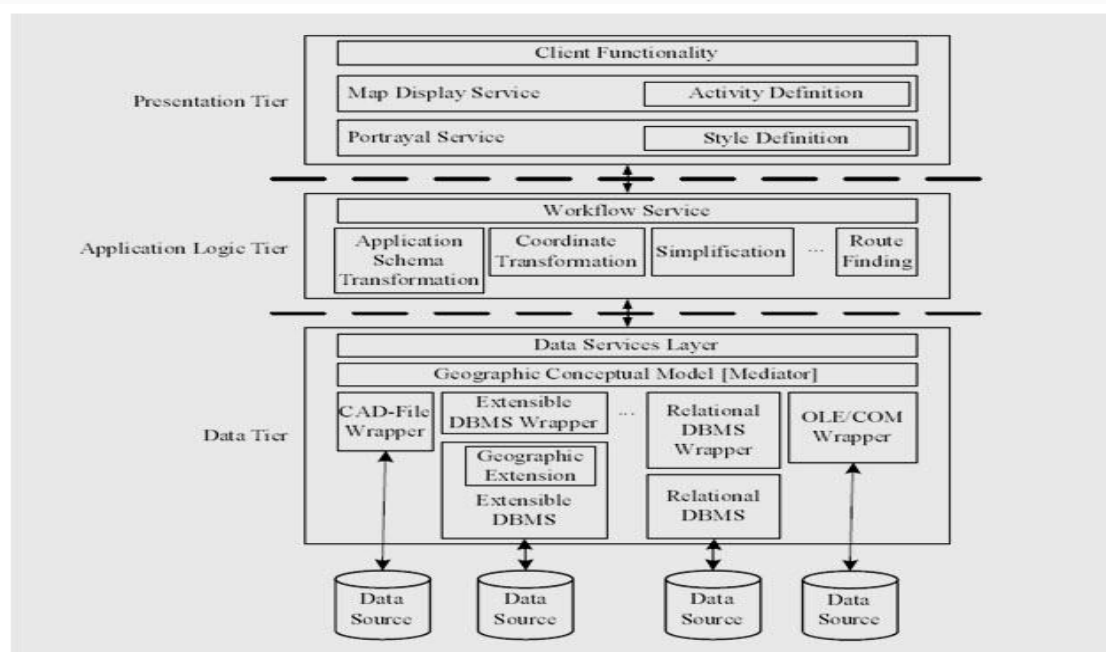


**Fig. 16.5. A Generic System Architecture for GIS.**

(Source: Miguel et al., 2005)

### 16.3.3 Web GIS Architecture:

In performing the GIS analysis tasks, WEB GIS is similar to the client/server typical three-tier architecture. The geo-processing is breaking down into server-side and client-side tasks. A client typically is a Web browser. The server-side consists of a Web server, Web GIS software and Database (Fig. 16.6) (Helali, 2001)



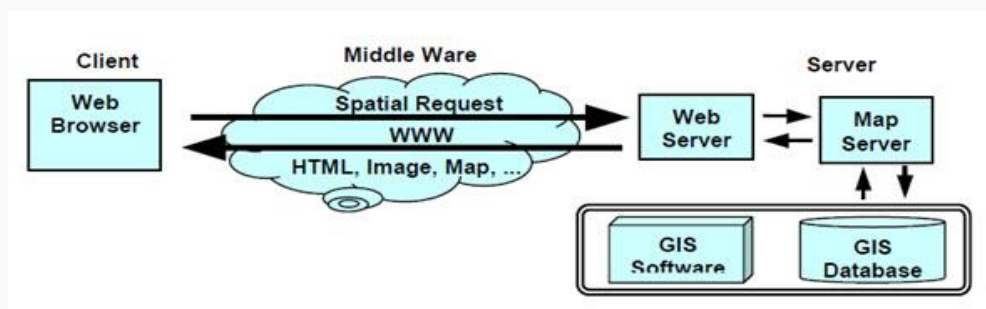
**Fig. 16.6. How a typical WEB GIS model works.**

(Source: Alesheikh et al, 2001)

This model of network widely exists within enterprises, in which some computers act as servers and others act as clients. Servers simply have the proprietary GIS running, and add a client interface at the client side and a middleware at the server side to communicate between the client and the proprietary GIS software. Recent development in object oriented programming make it possible to produce software components, and send them to the client before running it in the client machine, such as Java classes, ActiveX components and plug-ins. This comes out to the thick client GIS. The thick-client architecture let the client machine do the most processing works locally. Both thin and thick-client systems have some advantages and drawbacks, but they are not the best solution in terms of taking advantage of network resources. (Alesheikh et al, 2001)

### 16.3.3.1 Thin Client Architecture (Server Side Applications)

The thin client architecture is used in typical architecture. In a thin-client system, the clients only have user interfaces to communicate with the server and display the results. All the processing is done on the server actually as shown in Fig. 16.7. The server computers usually have more power than the client, and manage the centralized resources. Besides, the main functionality is on the Server side in thin architecture there is also the possibility for utility programs at the server side to be linked to the server software. Fig. 16.3 shows schematic communication between Web browser, Web Server and GIS server. On the Web Server side, there are some possibilities to realize the GIS connection to the World Wide Web; CGI, Web Server Application Programming Interface (API), Active Server Pages (ASP), Java Server Pages (JSP) and Java- Servlet. The descriptions of the five possibilities mentioned above are in Helali, (2001).



**Fig. 16.7. Server Side Application.**

**(Source: Alesheikh et al, 2001)**

The user on the client side does not need any knowledge about the linkage of the IMS at the server side, but the system administrator or application developers should be familiar with these techniques. This Architecture used in ESRI ArcView IMS, Map Objects IMS and MapInfo Map Xtreme systems.

Major advantages of this model driven form Data Base centralization and are:

- Central control
- Easy for data eminance/updating

- Keep the latest version
- Generally cheaper
- Integration possibilities
- Regarding some cartographic aspects such as font

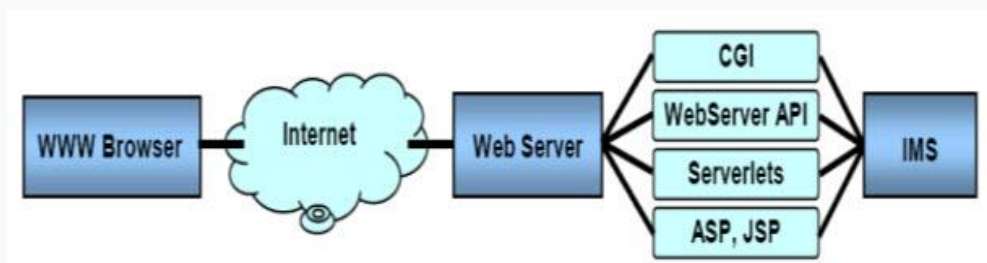
And disadvantages are:

- Not responsive to local needs: users have different invokes
- No local accountability: accountability need application in client side
- Large data volume (size of the database)
- Response time slow: users use a browser and it take long time to download new HTML frame
- Less interactive: in client side there is limited application and browsers abilities
- Vector data does not appear in client side: browsers without additional plug-in cannot read vector files.

(Alesheikh et al, 2001)

### 16.3.3.2 Thick Client Architecture (Client Side Applications)

In general, a Web browser can handle HTML documents, and embedded raster images in the standard formats. To deal with other data formats like vector data, video clips or music files, the browser's functionality has to be extended. Using exactly the same client sever communication in Thin Client architecture, vector files format could not be used. To overcome this problem most browser applications offer a mechanism that allows third tier programs to work together with the browser as a Plug-in. The user interface functionality has progressed from simple document fetching to more interactive applications. This progress is as follows: HTML, CGI, using HTML forms and CGI, Java script to increase user interface capabilities, Java applets to provide client-side functionality. Currently user interface capabilities combined with remote invocations (Fig.16.8) (Byong-Lyol, 1998).



**Fig. 16.8. Client Side Applications.**

(Source: Alesheikh et al, 2001)

Major advantages of this model are:

- Document/graphics standards are not required
- Vector data can be used
- Image quality not restricted to GIF and JPEG
- Modern interface is possible; it is not restricted to single-click Operations

And disadvantages to Client Side GIS

- Nonconformance cans limits
- User base
- Users require to obtain additional software
- Platform/browser are incompatible

(Alesheikh et al, 2001)

### 16.3.3.3 Medium Client Architecture

For avoiding vector data in client side and reducing problems of previous architectures, Medium Client is suggested. With using extensions in both client and server side, clients may have more functionally than Thin client architecture. In Fig. 16.9 these four components in interactive map are pictured as services, each with interfaces, which can be invoked by clients of that service. (Alesheikh et al, 2001)

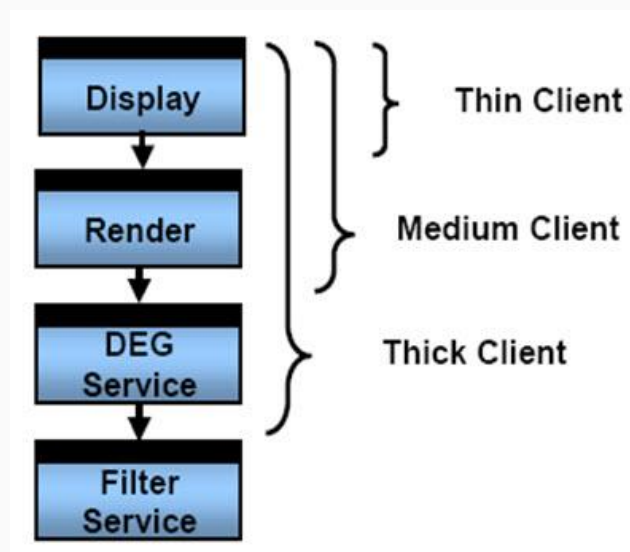


Fig. 16.9. Medium Client position in Open GIS point of view.

(Source: Alesheikh et al, 2001)

In other words, if a user's computer contains just the display service, then that user would be said to be using a thin client. If the user's computer additionally contained a render service, then that user would be said to be using a medium client. And finally, if the user's computer also contained the display element generator service that would indicate the user is using a thick client. After some consideration, it was decided that while this distinction may be somewhat helpful in describing web mapping, the terms "thick client" and "thin client" were already encumbered by very imprecise definitions used in marketing literature and were therefore not suitable for continued use in some cases (Doyle, 1999). (Alesheikh et al, 2001)

#### 16.3.3.4 Distributed Architecture

Recent developments in information technology have resulted in a number of distributed object architectures that provide the framework required for building distributed applications. The framework also supports a large number of servers and applications running concurrently. Many of such frameworks provide natural mechanism for interoperability (Kafatos, 1999). For example, Distributed Component Object Model architecture in windows platform and Java Remote Method Invocation (RMI) in Java Virtual Machine (JVM) are the most popular protocols that are used in different cases. These architectures may be applied to GIS to improve the traditional client/server GIS model and develop scalable distributed GIS model. Some attempts have been made in the academic area (Zhang, 1998).

The general idea of the distributed GIS service model is that a client program, in either an Internet browser or an independent application, should be able to access the resources distributed in the entire network. The resources here refer to both geo-data and geo-processing components available in the network. The client and the server in this context do not refer to a specific machine. Any machine, when it requests the remote resources during the processing, is a client, and any machine that provides such resources is a server. In a specific program, a client may connect to several servers if needed and a specific machine may be the client at one time and the server at another time. An ideal distributed GIS service model should be a "geo-data anywhere, geo-processing anywhere" model, which means the geo-data and geo-processing tool, could be distributed with the largest flexibility virtually anywhere in the network. The geo-data and geo-processing components do not have to be in the same site, but they should be able to cooperate or integrate whenever they are needed to finish a specific task (Yuan, 2000). (Alesheikh et al, 2001).

**Keywords:** GIS, GIS Functional Components, GIS Software, System Architecture, Web GIS.

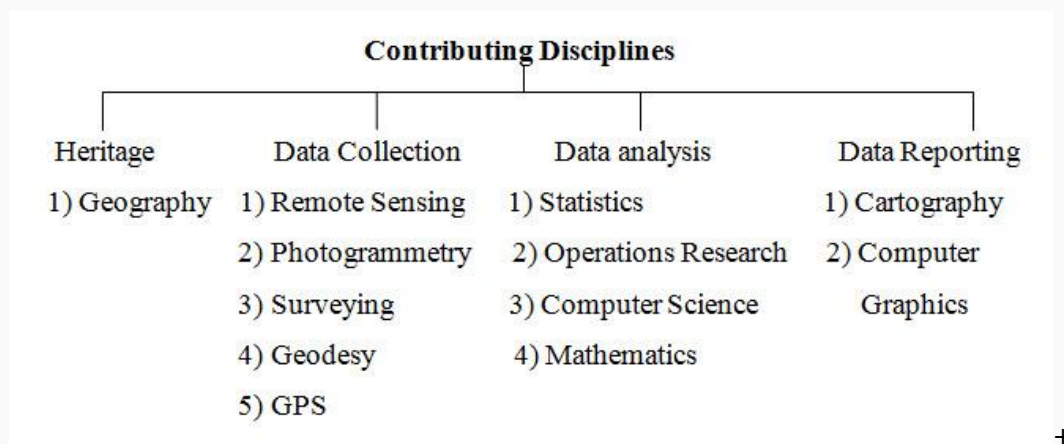


## Lesson 17 Applications of GIS

### 17.1 Contributing Disciplines

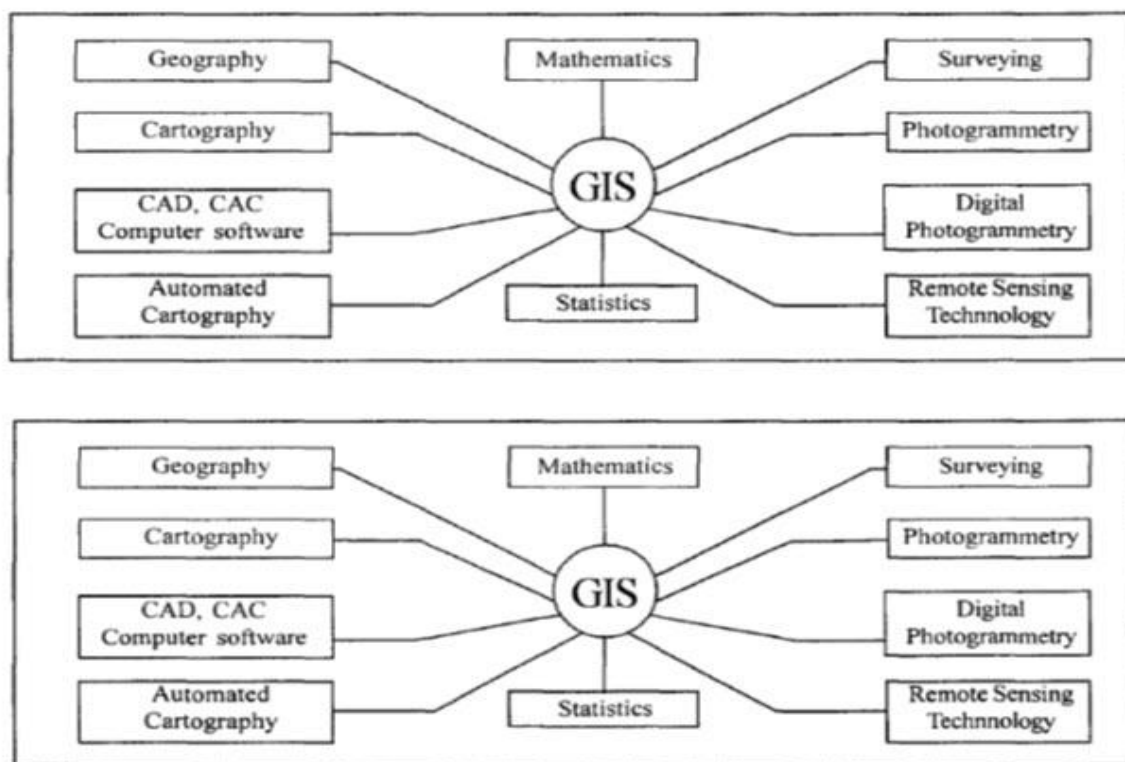
GIS is convergence of technological fields and traditional disciplines. GIS has been called an "enabling technology" because of the potential it offers for the wide variety of disciplines dealing with spatial data. Many related fields of study provide techniques which make up GIS. These related fields emphasise data collections while GIS brings them together by emphasising integration, modelling and analysis. Thus GIS often claims to be the science of spatial information. Fig. 17.1 shows the technical and conceptual development of GIS. (<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>). The list of contributing disciplines can be classified according to (1) Heritage (2) Data Collection (3) Data analysis (4) Data Reporting .

([http://gisserver.civil.iitd.ac.in/gislab/Lecture/Lecture%2001\\_Introduction%20to%20GIS.pdf](http://gisserver.civil.iitd.ac.in/gislab/Lecture/Lecture%2001_Introduction%20to%20GIS.pdf)).



The contributing disciplines for the evolution of a GIS (Burrough, 1998) are geography, cartography, remote sensing, surveying and photogrammetry, computer science technology, mathematics, and statistics.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>).



**Fig. 17.1 GIS: the result of linking parallel developments in many separate spatial data processing disciplines.**

(Source: <http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

**Geography:** is broadly concerned with understanding the world and man's place in it. Geography has a long tradition in spatial analysis, and provides techniques for conducting spatial analysis and a spatial perspective on research.

**Cartography:** is concerned with display of spatial information. It is now the main source of input data for GIS (maps) and has a long tradition in the design of maps which is an important form of output from GIS. It provides methods for digital representation and manipulation of cartographic features and methods of visualization.

**Remote Sensing:** is becoming an important source of geographical data by providing digital images derived from space and the air. Remote sensing provides techniques for data acquisition and processing anywhere on the globe at a low cost, and consistent update potential. While integrated with GIS, remotely sensed imagery can be merged with other data in a GIS providing real-time spatial information. The first part of this book enlightens the concepts and the potential utility of remote sensing.

**Surveying and Photogrammetry:** provide high quality data on positions of cadastral objects like land parcel and building, and topography. Aerial photogrammetry deals with the photographs taken by an aerial camera on board aircraft at different altitudes. Aerial photogrammetry is one of the most powerful data-capturing techniques for the creation of GIS spatial database. The relevant data can be extracted from the aerial photographs of various scales (Fig. 17.2), and may be used as input for GIS. Digital orthophotos provide the

source of digital data. These products are scanned airphotos that have been rectified to eliminate displacement caused by variable elevation of the ground surface and the tilt of the camera. Properly registered with other digital data sets, these images can be used directly as backdrops for vector data or to provide a basemap for onscreen digitising. The user may abstract information on land use, vegetation type and other aspects of the landscape from the photograph. Curran (1989) identifies six characteristics of aerial photographs of immense value as a data source for GIS. They are (i) wide availability, (ii) low cost, (iii) wide area views, (iv) time - freezing ability, (v) high spectral and spatial resolution, and (vi) three-dimensional perspective.

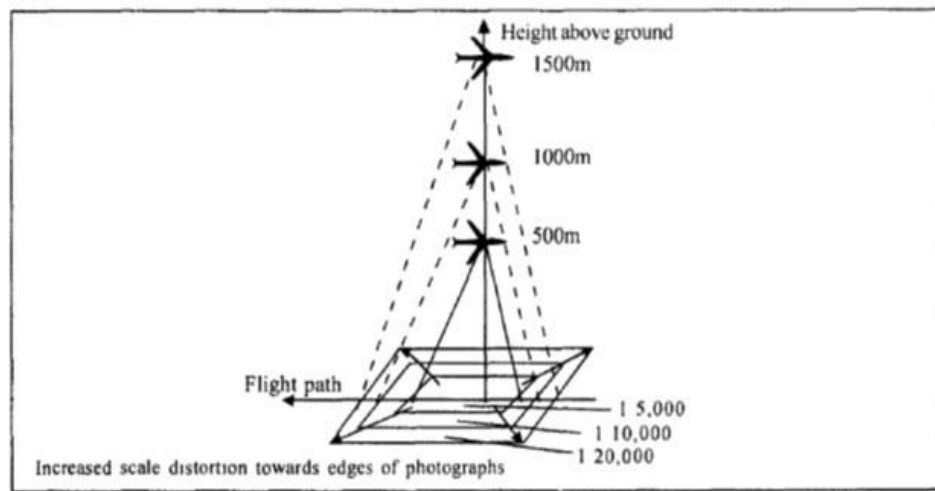


Fig. 17.2. Varying scale on Aerial photographs.

(Source: <http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

**Geodesy:** provides source of high accuracy positional control for GIS. It concentrates on placing objects accurately on global context. Many applications of GIS require use of optimizing techniques for decision-making which involves Operational research.

**Computer Sciences:** Computer Assisted Design (CAD) provides software, techniques for data input, display and visualisation, and representation, particularly in 3-dimensions. Advances in computer graphics provide hardware and software for handling and displaying graphic objects. Data Base Management System (DBMS) contributes methods for representing data in digital form and procedures for system design and update. Artificial intelligence (AI) uses the computer to make choices based on available data in a way that is seen to enhance the human intelligence and decision-making. Using AI, computer can act as an "expert" in such functions as designing maps, generalizing map features and classification.

**Mathematics:** Several branches of mathematics, especially geometry and graph theory, are used in GIS system design and analysis of spatial data.

**Statistics:** is used to build models and perform spatial data analysis in GIS. Statistics is also important in understanding issues of error, quality and uncertainty in GIS data.



Availability of large quantities of spatial data in the form of digital aerial photograph, digital remote sensing imagery, advancement of computer hardware, software and software development, increasing demand of spatial information for management, and infrastructure development parameters, lead to have a system to handle all these requirements. In order to handle such data to meet these demands, to store, retrieve, handle, analyse, manipulate, and display the results, it requires computer based system. Such a system is Geographical Information System (GIS).

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

## 17.2 GIS Workflow

There are five essential elements that a GIS must contain. They are data acquisition, preprocessing, data management, manipulation and analysis, and product generation. For any application of GIS, it is important to view these elements as a continuing process. Figs. 17.3 and 17.4 show the work flow process of GIS in procedural perspective. Data acquisition is the process of identifying and gathering the data required for any given application. Two main types of data acquisition are primary data sources are those collected in digital format specifically for use in a GIS project. Secondary sources are digital and analog datasets that were originally captured for another purpose and need to be converted into a suitable digital format for use in a GIS project. The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing. Data collection is a time consuming, tedious, and expensive process. Preparation involves obtaining data, redrafting poor-quality map sources, editing scanned map images, removing noise, setting up appropriate GIS hardware and software systems to accept data. Digitizing and transfer are the stages where the majority of the effort will be expended.

A GIS must be able to convert data from one structure to another. The Data acquisition stage collects raw imagery data that is redundant & needs to be stitched together into a more apt data stream. Data pre-processing involves digitization of the acquired data that includes:

- 3D data processing- Creating realistic models of buildings and structures by visualizing them in 3-Dimensional environment is termed as 3D data processing. Roof detail and elevation details of building or structure are truly represented in this process.
- Value addition to Image data for large scale mapping- In order to transfer the spatial components from images to maps, it is essential to have accurate base maps on large scales. It is necessary to generate large scale base maps for use in various thematic applications.

([http://www.albireotelematics.com/pre\\_processing.html](http://www.albireotelematics.com/pre_processing.html))

The other essential preprocessing procedures include: (a) format conversion, (b) data reduction and generalisation, (c) error detection and editing, (d) Merging of points into lines, and lines into polygons, (e) Edge matching and tiling, (f) Rectification/registration, (g) Interpolation, and (h) Interpretation.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>).

The functions of database management govern the creation of an access to the database itself. These functions provide consistent methods for data entry, update, deletion, and retrieval. Modern database management systems isolate the users from the details of data storage, such as, the particular data organisation on a mass storage medium. A modern Database Management System (DBMS) is used to create GIS database, that is, attribute database.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>).

Any DBMS makes assumptions about the data which it handles. Two ways to use DBMS within GIS

- Total DBMS solution: In this all data are accessed through the DBMS, so it must fit the DBMS designer.
- Mixed solution: In this some data (usually attribute tables and relationships) are accessed through the DBMS as they fit the model well, while some data(usually locational) are accessed directly because they do not fit the DBMS model.

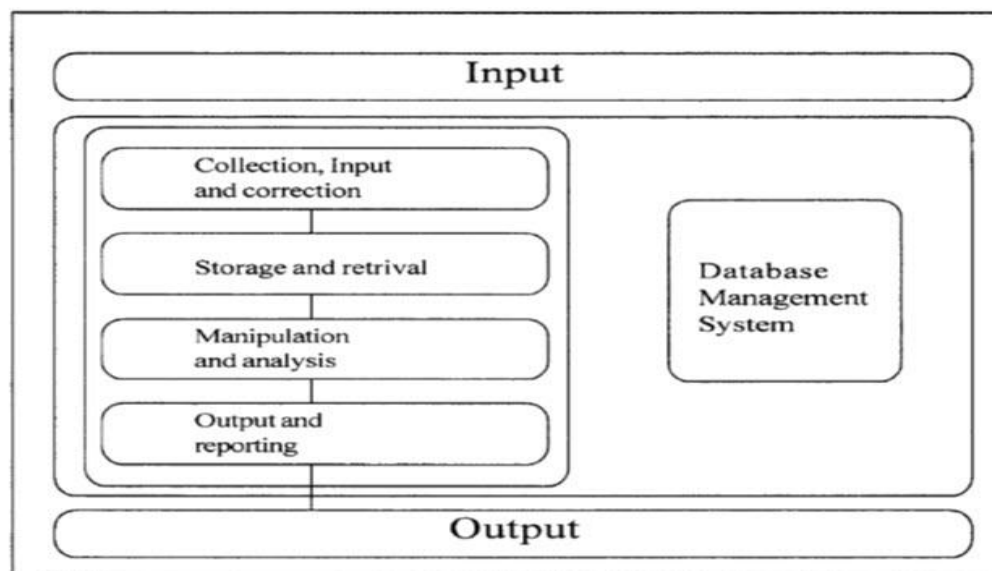
(<http://gis.nic.in/gisprimer/dbms.html>)

Storage and retrieval mechanisms include the control of physical storage of the data in memory, disk or tape, and mechanisms for its retrieval to serve the needs of the other three components. In a disaggregate GIS this data storage may be physically more from the rest of the system, and may meet the database requirements. This module includes the software structures used to organise spatial data into models of geographic reality.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

The development of new derived data layers, which may form the input to further analysis, is an important function of any GIS. The list of data manipulation and analysis operations are, (i) reclassification and aggregation, (ii) Geometric Operations: as rotation, translation and scaling, rectification, and registration, (iii) Controlled determination, (iv) Data structure conversion, (v) Spatial operations of connectivity and neighborhood operations, (vi) Measurement of Distance and Direction, (vii) Statistical analysis as descriptive statistics regression, correlation, and cross-tabulation, and (viii) Modelling.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

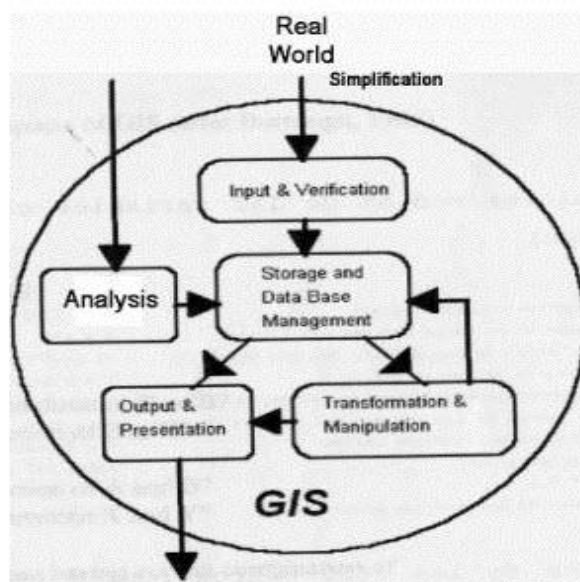


**Fig. 17.3. Workflow process of GIS (Procedural perspective)**

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)

This operation or subsystem represents the whole spectrum of techniques available for the transformation of the digital model by mathematical means. A library of data-processing algorithms is available for the transformation of spatial data, and incorporated in new visual maps. Using these techniques it is possible to deliberately change the characteristics of the data representation in order to meet theoretical requirements. It is equally possible to mishandle or unintentionally distort the digital map at this state.

(<http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u01.html#SEC1.2>)



**Fig. 17.4. Workflow process of GIS (Procedural perspective)**

(Source: <http://people.aub.edu.lb/~webeco/GIS%20course.htm>)

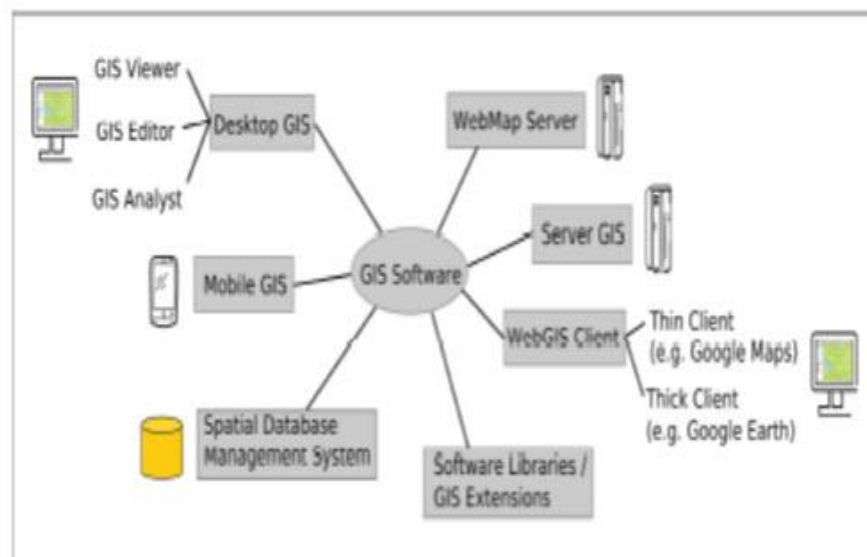
Output and Presentation is the phase where final outputs from the GIS are created. These output products might include statistical reports, maps, and graphics of various kinds. Some of these products are softcopy images and some are hardcopy.

### 17.3 Standard Software

Software that is used to create, manage, analyze and visualize geographic data is usually incorporated under the umbrella term 'GIS software' (Steiniger & Weibel, 2009). Typical applications for GIS software include the evaluation of places for the location of new stores, the management of power and gas lines, the creation of maps, the analysis of past crimes for crime prevention, route calculations for transport tasks, the management of forests, parks and infrastructure, such as roads and water ways, as well as applications in risk analysis of natural hazards, and emergency planning and response. For this multitude of applications different types of GIS functions are required and different categories of GIS software exist, which provide a particular set of functions needed to fulfill certain data management tasks. Different functional categories of GIS software can be identified with respect to the tool sets that GIS software offers, and with respect to the tasks that can be accomplished. Such tasks involve basic activities that are common to daily computer usage such as (1) data visualization and exploration, (2) data creation, (3) data editing and (4) data storage. Other common GIS tasks include (5) data conflation, i.e., integration of data from different sources (Blasby, Davis, & Kim, 2002), (6) data queries to select a subset of the data, (7) data analysis, which we consider to be the creation of new information (output) from existing data (input), (8) data transformation, as some analysis tasks require the user to transform, or manipulate, the data beforehand (e.g. transform the data into a different coordinate system, or convert them from raster to vector format), and lastly, (9) the creation of maps - the most common method used to visualize analysis and query results. (Hunter & Steiniger, 2012).

GIS software is the processing engine and a vital component of an operational GIS. It is made up of integrated collections of computer programs that implement geographic processing functions. The three key parts of any GIS software system are the user interface, the tools (functions), and the data manager. All three parts may be located on a single computer or they may be spread over multiple machines in a departmental or enterprise configuration. (Longley, 2005).

Steiniger and Weibel (2009) identified seven major types of GIS software: (i) Desktop GIS, (ii) Spatial Data Base Management Systems (SDBMS), (iii) Web Map Server, (iv) Server GIS, (v) Web GIS clients, (vi) Mobile GIS, and (vii) Libraries and Extensions. For the purpose of this survey we did split the last category "Libraries and Extensions" into the two categories: "Libraries" and "GIS Extensions, Plug-ins and APIs". Web Map Server and Web GIS clients have been subsumed under the category "Software for Internet Mapping Applications". Furthermore, we extend the set to include two additional categories (viii) Remote Sensing Software, which could be considered a special form of desktop GIS, and (ix) Exploratory Spatial Data Analysis (ESDA) software. Fig. 1 characterizes the different software types with respect to GIS functionality as defined above.



**Fig. 17.5. Different types of GIS software.**

(Source: Steiniger and Weibel, 2009)

### 17.3.1 Popular GIS packages

A large number of GIS software options are available as open-source or commercial products. Following is a brief summary of some of the more popular GIS packages. A large listing of GIS software can be found at [http://en.wikipedia.org/wiki/GIS\\_software](http://en.wikipedia.org/wiki/GIS_software).

#### 17.3.1.1 Proprietary GIS

- **ArcGIS:** ArcGIS is the name of a suite of GIS software product lines produced by ESRI (<http://www.esri.com/>). At the desktop GIS level, ArcGIS can include: ArcReader, which allows one to view and query maps created with the other Arc products; ArcView, which allows one to view spatial data, create maps, and perform basic spatial analysis; ArcEditor, which includes all the functionality of ArcView as well as more-advanced tools for manipulation of shape files and geodatabases; or ArcInfo, the most advanced version of ArcGIS, which includes added capabilities for data manipulation, editing, and analysis. There are also server-based ArcGIS products as well as ArcGIS products for personal digital assistants (PDAs). Extensions can be purchased separately to increase the functionality of ArcGIS.
- **AutoCAD:** AutoCAD is popular engineering CAD software produced by Autodesk (<http://usa.autodesk.com/>). AutoCAD Map 3D software is a leading engineering platform that bridges the gap between CAD and GIS. When combined with Autodesk Map Guide technology, AutoCAD Map 3D provides a way to publish data to the Web or an intranet.
- **Cadcorp:** Cadcorp (<http://www.cadcorp.com/>) is the developer of GIS software and Open GIS standard (e.g., Read/Write Open-Source Post GIS database). Products include a Spatial Information System (SIS), which runs on Microsoft Windows and

encompasses desktop GIS modules, ActiveX- and COM-based developer kits, Web-based GIS software (GeognoSIS), and a mobile data-capture solution (mSIS).

- **ERDAS IMAGINE:** ERDAS IMAGINE is a raster graphics editor and remote-sensing application designed by ERDAS, Inc. (<http://www.erdas.com/>). It is aimed primarily at geospatial raster data processing that allows the user to display and enhance digital images. It is a toolbox allowing the user to perform numerous operations on an image and generate an answer to specific geographical questions.
- **IDRISI:** GIS developed by Clark Labs (<http://www.clarklabs.org/products/>) at Clark University, Massachusetts. IDRISI Andes is an integrated GIS and image-processing software providing over 250 modules for spatial analysis and display. Originally developed under United Nations sponsorship, the IDRISI is widely used worldwide.
- **Intergraph:** Intergraph (<http://www.intergraph.com/>) provides software and services for infrastructure management for the electric, gas, water, pipeline, utility, and communications industries. Products include GeoMedia, GeoMedia Professional, GeoMedia WebMap, and add-on products for industry sectors, as well as photogrammetry.
- **MapInfo:** MapInfo (<http://www.mapinfo.com/>) GIS software products include the desktop GIS software, MapInfo Professional, MapXtreme 2005, and MapXtreme Java for Webbased and desktop client mapping, as well as developer tools such as MapBasic.
- **Micro Station:** Micro Station is a suite of CAD/GIS software products for 2-D and 3-D design and drafting, developed and sold by Bentley Systems (<http://www.bentley.com/>). It is used by engineering designers for transportation and for water and wastewater utilities. Bentley also offers GIS-based water resources modeling software for water, sewer, and storm water systems (Sewer CAD, Water CAD, and Storm CAD). (Johnson, 2009)

### 17.3.1.2 Open-Source GIS

- **GRASS:** GRASS (Geographic Resource Analysis Support System) is a public-domain open-source raster GIS developed as a general-purpose spatial modeling and analysis package (Neteler and Mitasova 2008). GRASS is a raster/vector GIS, image processing system, and graphics production system. GRASS contains over 350 programs and tools to render maps and images on monitor and paper; manipulate raster, vector, and sites data; process multispectral image data; and create, manage, and store spatial data. GRASS uses both an intuitive Windows interface as well as command-line syntax for ease of operation. (Johnson, 2009)

### 17.3.2 Desktop GIS Software

Desktop GIS software is probably the most common GIS software in use. The company ESRI (2012) defines desktop GIS as mapping software that is installed onto and runs on a personal computer and allows users to display, query, update, and analyze data about geographic locations and the information linked to those locations. All traditional GIS tasks, i.e., not tasks

related to web and remote processing applications, can be accomplished with a desktop GIS. Sometimes, proprietary GIS software vendors distinguish in software pricing between two or three categories of desktop GIS with respect to the functionality that the software offers. In the case of 3 desktop GIS categories so-called viewer applications offer functionality for viewing and exploring data, while editor applications provide, in addition to viewer capabilities, the ability to create and update spatial data. Analysis GIS software, also termed “Professional edition”, Typically offers the highest level of functionality by adding functions for data analysis, map creation, data conflation, etc., to editor applications. (DeMers, 2009).

### 17.3.3 Open Source GIS Software

Due to the lack of money for buying licenses one of the biggest considerations is the use of Open Source software. There are many analyzed software packages with a certified OSI Open Source License 10. The software is divided into different categories according to what it is used for. (DeMers, 2009).

#### 17.3.3.1 OGC Spatial servers

There is a set of available software that helps to create OGC services. This software is normally installed in a web server and is configured to publish your data through OGC standards. More specifically for our purpose we consider the following standards necessary: **WMS, WFS, CSW, and WCTS**.

**Degree 11:** free software initiative founded by the GIS and Remote Sensing unit of the Department of Geography, University of Bonn, and lat/long. It is the reference implementation for the WMS standard. With more than one single software package, degree is a set of building blocks to construct OGC services. The software is right now under a big rebuild and a new version 2.0 is expected to be released before the end of the year.

**.OGC Standards supported:** WMS, WFS, WCS, WCAS, WFS-G, WTS, WCTS, and CSW.

**.Read & Write interfaces:** ORACLE Spatial, PostGRES/PostGIS, MySQL, other JDBC-enabled databases, ESRI Shapefiles, several raster data formats (JPEG, GIF, PNG, (Geo) TIFF, PNM und BMP).

**.Architecture:** Java-servlet

**17.3.3.2 Geoserver 12:** again a free software initiative. It has a good community support behind and some projects funding its further development. It is the reference implementation for the WFS standard. The installation is easy and the documentation complete. Right now it only supports one feature per table and the mapping of complex schemas is not complete. There are people working on solving those limitations though.

- **OGC Standards supported:** WMS, WFS.
- **Read & Write interfaces:** PostGIS, ESRI Shapefile, ArcSDE and Oracle, VPF, MySQL, MapInfo, KML...
- **Architecture:** Java-servlet

(Javier de la Torre)

**17.3.3.3 Mapserver 13:** Originally developed at the University of Minnesota (UMN) through the NASA-sponsored ForNet project, a cooperative effort with the Minnesota Department of Natural Resources. The software has grown and is maintained by an increasing number of developers (nearing 20) from around the world and is supported by a diverse group of organisations funding enhancements.

The project started before the creation of these OGC standards and was later adapted to support them up to a certain degree. Right now the WFS service is read only and transactions are not supported. A bigger limitation is the lack of support of filters in feature attributes and POST operations.

- **OGC Standards supported:** WMS, non-transactional WFS, and WCS.
- **Read & Write interfaces:** ESRI shapefiles, PostGIS, ESRI ArcSDE, TIFF/GeoTIFF, EPPL7
- **Architecture:** CGI implemented in C, scripts in different scripting languages.

(Javier de la Torre)

#### 17.3.4 Software Manufacturers and Projects

GIS software is not only provided by companies but increasingly also by free and open source software projects. While commercial vendors usually offer products for all of software categories, open software projects often concentrate on a single category, e.g. desktop GIS or WebMap server. The key players in the GIS software market today are Autodesk, Bentley, ESRI Inc., GE (Smallworld), Pitney Bowes (MapInfo), and Intergraph. GIS software companies tend to target specific application domains. For instance, ESRI's ArcGIS product tends to be mainly used for business analysis, planning, and environmental applications, while Autodesk, GE and Bentley products are rather used in utility and facility management. Competitive GIS software that is developed by free software projects exists as well - especially with respect to server applications (MapServer, GeoServer) and spatial DBMS (PostGIS). Free desktop GIS projects, such as Quantum GIS and gvSIG, currently experience growing user communities. Such free GIS software rather complements the set of proprietary software instead of competing with it. (Steiniger and Weibel, 2009).

**Table 17.1. Sites of proprietary software suites and to those providing programs which are shareware, or low cost, or entirely free of license. (Source: Liu and Mason, 2009)**

Autodesk	<a href="http://www.usa.autodesk.com">www.usa.autodesk.com</a>
ERDAS	<a href="http://www.erdas.com">www.erdas.com</a>
ER Mapper	<a href="http://www.ermapper.com">www.ermapper.com</a>
ESRI	<a href="http://www.esri.com">www.esri.com</a>



FME Safe Software	<a href="http://www.safe.com">www.safe.com</a>
Geo tools	<a href="http://www.geotools.codehaus.org">www.geotools.codehaus.org</a>
Global Mapper	<a href="http://www.globalmapper.com">www.globalmapper.com</a>
GRASS	<a href="http://www.grass.itc.it">www.grass.itc.it</a>
Idrisi	<a href="http://www.clarklabs.org">www.clarklabs.org</a>
ILWISS	<a href="http://www.itc.nl/ilwis">www.itc.nl/ilwis</a>
ITTVis ENVI	<a href="http://www.ittvis.com/envi">www.ittvis.com/envi</a>
JUMP GIS	<a href="http://www.jump-project.org">www.jump-project.org</a>
Landserf	<a href="http://www.landserf.org">www.landserf.org</a>
Map Window	<a href="http://www.mapwindow.org">www.mapwindow.org</a>
MapInfo	<a href="http://www.mapinfo.com">www.mapinfo.com</a>
PCI Geomatics	<a href="http://www.pcigeomatics.com">www.pcigeomatics.com</a>
Quantum GIS opensource	<a href="http://www.qgis.org">www.qgis.org</a>
SAGA GIS	<a href="http://www.saga-gis.org">www.saga-gis.org</a>
Various independent	<a href="http://www.rockware.com">www.rockware.com</a>
Variowin	<a href="http://www.sst.unil.ch/research/variowin">www.sst.unil.ch/research/variowin</a>
Virtuozo	<a href="http://www.supresoft.com">www.supresoft.com</a>

**Table 17.2. Information and technical information on standards, best practice, formats, techniques and various publications. (Source: Liu and Mason, 2009)**

Association for Geographic Information (AGI)	<a href="http://www.agi.org.uk">www.agi.org.uk</a>
British Geological Survey (BGS)	<a href="http://www.bgs.ac.uk">www.bgs.ac.uk</a>
Committee on Earth Observation Satellites (CEOS)	<a href="http://www.coes.cnes.fr">www.coes.cnes.fr</a>
Digital Earth	<a href="http://www.dgeo.org">www.dgeo.org</a>
Digital National Framework	<a href="http://www.dnf.org">www.dnf.org</a>
ESRI ArcUser online	<a href="http://www.esri.com/news/arcuser">www.esri.com/news/arcuser</a>
ESRI online knowledge base	<a href="http://www.support.esri.com">www.support.esri.com</a>

Geospatial Analysis online	<a href="http://www.spatialanalysisonline.com">www.spatialanalysisonline.com</a>
Geospatial Information and Technology Association (GITA)	<a href="http://www.gita.org">www.gita.org</a>
GIS Day	<a href="http://www.gisday.com">www.gisday.com</a>
GIS Research UK (GISRUK)	<a href="http://www.geo.ed.ac.uk/gisruk">www.geo.ed.ac.uk/gisruk</a>
Grid Forum 2001	<a href="http://www.gridforum.org">www.gridforum.org</a>
International Association of Photogrammetry & Remote Sensing	<a href="http://www.isprs.org">www.isprs.org</a>
International DEM Service	<a href="http://www.cse.dmu.ac">www.cse.dmu.ac</a>
Open Geospatial Consortium	<a href="http://www.opengeospatial.org">www.opengeospatial.org</a>
Ordnance Survey (OS)	<a href="http://www.ordnancesurvey.co.uk">www.ordnancesurvey.co.uk</a>
Remote Sensing and Photogrammetry Society (RSPSoc)	<a href="http://www.rspsoc.org">www.rspsoc.org</a>
UKGeoForum (umbrella organization)	<a href="http://www.ukgeoforum.org.uk">www.ukgeoforum.org.uk</a>
Web 3D Consortium	<a href="http://www.geovrml.org">www.geovrml.org</a>
World Wide Web Consortium	<a href="http://www.w3.org">www.w3.org</a>

**Table 17.3. Data sources including online satellite imagery from major suppliers, Dem data plus GIS maps and data of all kinds. (Source: Liu and Mason, 2009)**

ALOS data search	<a href="http://www.cross.restec.or.jp">www.cross.restec.or.jp</a>
Asia Pacific Natural Hazards Network	<a href="http://www.pdc.org/mde/explorer.jsp">www.pdc.org/mde/explorer.jsp</a>
Digital Globe (Quickbird & WorldView)	<a href="http://www.browse.digitalglobe.com">www.browse.digitalglobe.com</a>
EarthExplorer	<a href="http://www.earthexplorer.usgs.gov">www.earthexplorer.usgs.gov</a>

**Table 17.4. Online resources for information, software and data. (Source: Liu and Mason, 2009)**

ALOS data search	<a href="http://www.cross.restec.or.jp">www.cross.restec.or.jp</a>
EOS Data Gateway edcims	<a href="http://www.cr.usgs.gov/pub/imswelcome/plain.html">www.cr.usgs.gov/pub/imswelcome/plain.html</a>
ESA EOLI catalogues	<a href="http://www.catalogues.eoportal.org">www.catalogues.eoportal.org</a>
Geo Community GIS free data depot	<a href="http://www.data.geocomm.com">www.data.geocomm.com</a>

GeoEye (GeoFUSE)	<a href="http://www.geofuse.geoeye.com/landing">www.geofuse.geoeye.com/landing</a>
GIS data depot	<a href="http://www.gisdepot.com">www.gisdepot.com</a>
GIS Lounge	<a href="http://www.gislounge.com">www.gislounge.com</a>
GLCF	<a href="http://wwwlcf.umiacs.umd.edu">wwwlcf.umiacs.umd.edu</a>
Glovis	<a href="http://www.glovis.usgs.gov">www.glovis.usgs.gov</a>
SPOT catalogue	<a href="http://www.sirius.spotimage.fr">www.sirius.spotimage.fr</a>
SRTM Public Data	<a href="http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html">www2.jpl.nasa.gov/srtm/cbanddataproducts.html</a>

GIS software is a fundamental and critical part of any operational GIS. The software employed in a GIS project has a controlling impact on the type of studies that can be undertaken and the results that can be obtained. There are also far reaching implications for user productivity and project costs. Today, there are many types of GIS software product to choose from and a number of ways to configure implementations. One of the exciting and at times unnerving characteristics of GIS software is its very rapid rate of development. This is a trend that seems set to continue as the software industry pushes ahead with significant research and development efforts.

#### 17.4 Applications of GIS

An important distinction between GIS applications is whether the geographic phenomena studied are man-made or natural. Clearly, setting up a cadastral information system, or using GIS for urban planning purposes involves a study of man-made things mostly: the parcels, roads, sidewalks, and at larger scale, suburbs and transportation routes are all man-made, those entities often have clear-cut boundaries. On the other hand, geomorphologists, ecologists and soil scientists often have natural phenomena as their study objects. They may be looking at rock formations, plate tectonics, distribution of natural vegetation or soil units. Often, these entities do not have clear-cut boundaries, and there exist transition zones where one vegetation type, for instance, is gradually replaced by another. (de et al, 2001).

The applications of GIS include mapping locations, quantities and densities, finding distances and mapping and monitoring change. Function of an Information system is to improve one's ability to make decisions. An Information system is a chain of operations starting from planning the observation and collection of data, to store and analysis of the data, to the use of the derived information in some decision making process. A GIS is an information system that is designed to work with data referenced to spatial or geographic coordinates. GIS is both a database system with specific capabilities for spatially referenced data, as well as a set of operation for working with data. There are three basic types of GIS applications which might also represent stages of development of a single GIS application. (ESRI White Paper, 2006).

### 17.4.1 Inventory Application

Many times the first step in developing a GIS application is making an inventory of the features for a given geographic area. These features are represented in GIS as layers or themes of data. The emphasis at this stage of application development consists of updating and simple data retrieval. (ESRI White Paper, 2006).

### 17.4.2 Analysis Application

Upon completion of the inventory stage, complex queries on multiple layers can be performed using spatial and aspatial analysis techniques. (ESRI White Paper, 2006).

### 17.4.3 Management Application

More advanced spatial and modeling techniques are required to support the decisions of managers and policy makers. This involves shifting of emphasis from basic geographic data handling to manipulation, analysis and modeling in order to solve real world problems.

### 17.4.4 Five Ms

There is, quite simply, a huge range of applications of GIS. They include topographic base mapping, socio-economic and environmental modeling, global modeling, and education. The five Ms of GIS application are Mapping, Measurement, Monitoring, Modeling, and Management.

In very general terms, GIS applications may be classified as traditional, developing, and new. Traditional GIS application fields include military, government, education and utilities. The mid-1990s saw the wide development of business uses, such as banking and financial services, transportation logistics, real estate, and market analysis. The early years of the 21st century are seeing new forward-looking application areas in small office/home office (SOHO) and personal or consumer applications, as well as applications concerned with security, intelligence, and counter-terrorism measures. This is a somewhat rough-and-ready classification, however, because the applications of some agencies (such as utilities) fall into more than one class. Many applications involve use of GIS by large numbers of people. Once GIS applications become established within an organization, usage often spreads widely. Integration of GIS with corporate information system (IS) policy and with forward planning policy is an essential prerequisite for success in many organizations.

- Government and public service
- Business and service planning
- Logistics and transportation
- Environment

(Longley, 2005)

### 17.4.5 Government and public service

Government users were among the first to discover the value of GIS. Today GIS is used at all levels of government from the national to the neighborhood, and government users still comprise the biggest single group of GIS professionals. As GIS has become cheaper, so it has come to be used in government decision making at all levels from the nation to the neighborhood. GIS is used to inventory resources and infrastructure, plan transportation routing, improve public service delivery, manage land development, and generate revenue by increasing economic activity. Typical GIS applications thus include monitoring public health risk, managing public housing stock, allocating welfare assistance funds, and tracking crime. It is convenient to group local government GIS applications on the basis of their contribution to asset inventory, policy analysis, and strategic modeling/planning.

(Longley, 2005)

**Table 17.5. GIS applications in Local Government**

	<b>Inventory Applications</b>	<b>Policy Analysis Applications</b>	<b>Management/ Policy Making Applications</b>
Economic development	Location of major business and their primary resource demands	Analysis of resource demand by potential local supplier	Informing the availability of local suppliers
Transportation and Services Routing	Identification of sanitation, truck routes, land fill and recycling area	Analysis of potential capacity development in certain area and accident patterns in certain sites	Identification of ideal high density development areas based on transportation
Housing	Inventory of housing condition, status, durability and demographics	Analysis of public support for housing by geographic area, drive time from low income areas to needed service facilities etc	Analysis of funding for housing rehabilitation, location of related public facilities
Infrastructure	Inventory of roads, sidewalks, bridges, utilities	Analysis of infrastructure by demographic variables such as income and population change	Analysis to schedule maintenance and expansion
Health	Location of person with particular health problems	Spatial, time series analysis of the spread of disease, effects of environmental condition on disease	Analysis of possible sources of diseases
Tax Maps	Identification of ownership data by land plot	Analysis of tax revenues by landuse within various distance from the city centre	Projecting land revenue change due to land use change

Law Enforcement	Inventory of location of police station, crimes, arrests, patrol car routing, alarm and security system locations	Analysis of police visibility and presence; officers in relation to density of criminal activity	Reallocation of police resources and facilities to areas where they are likely to be most efficient and effective.
Land-use planning	Parcel inventory of zoning areas, floodplains, industrial parks, land uses, trees, green space etc	Analysis of percentage of land used in each category, density levels by neighbourhoods, threats to residential amenities	Evaluation of land-use plan based on demographic characteristics of nearby population
Parks and Recreation	Inventory of park holdings/playscapes, trails by type etc	Analysis of neighbourhood access to parks and recreation opportunities, age-related proximity to relevant playscapes	Modeling population growth projections and potential future recreational needs
Environmental Monitoring	Inventory of environmental hazards in relation to vital resources such as groundwater; layering of nonpoint pollution sources	Analysis of spread rates and cumulative pollution level; analysis of potential years of life lost in a particular area due to hazards	Modeling potential environmental harm to specific local areas; analysis of place- specific multilayered pollution abatement plans
Emergency Management	Location of key emergency exit routes, their traffic flow capacity and critical danger points	Analysis of potential effects of emergencies of various magnitudes on exit routes, traffic flows etc	Modeling effect of placing emergency facilities and response capacities in particular locations
Geodemographics	Location of persons with specific demographic characteristics such as voting patterns, service usage and preferences, commuting routes occupations	Analysis of voting characteristics of particular areas	Modeling effect of placing information kiosks at particular locations

Source:<http://www.scribd.com/doc/55973482/Geographic-Information-Systems-and-Science>

#### 17.4.6 Business and service planning

Business and service planning (sometimes called retailing) applications focus upon the use of geographic data to provide operational, tactical, and strategic context to decisions that involve the fundamental question, where? Geodemographics is a shorthand term for composite indicators of consumer behavior that are available at the small area level (e.g., census output area, or postal zone). Geodemographic data are the basis for much market area analysis.

The tools of business applications typically range from simple desktop mapping to sophisticated decision support systems. Tools are used to analyze and inform the range of operational, tactical, and strategic functions of an organization. These tools may be part of

standard GIS software, or they may be developed in-house by the organization, or they may be purchased as a 'business solution' product. Operational functions concern the day-to-day processing of routine transactions and inventory analysis in an organization, such as stock management. Tactical functions require the allocation of resources to address specific (usually short term) problems, such as store sales promotions. Strategic functions contribute to the organization's longer-term goals and mission, and entail problems such as opening new stores or rationalizing existing store networks. Early business applications were simply concerned with mapping spatially referenced data, as a general descriptive indicator of the retail environment. This remains the first stage in most business applications, and in itself adds an important dimension to analysis of organizational function. (Longley, 2005)

#### **17.4.7 Logistics and transportation**

Knowing where things are can be of enormous importance for the fields of logistics and transportation, which deal with the movement of goods and people from one place to another and the infrastructure (highways, railroads, canals) that moves them. Logistics companies (e.g., parcel delivery companies, shipping companies) need to organize their operations, deciding where to place their central sorting warehouses and the facilities that transfer goods from one mode to another (e.g., from truck to ship), how to route parcels from origins to destinations, and how to route delivery trucks. All of these fields employ GIS, in a mixture of operational, tactical, and strategic applications.

The field of logistics addresses the shipping and transportation of goods. Each of these applications has two parts: the static part that deals with the fixed infrastructure, and the dynamic part that deals with the vehicles, goods, and people that move on the static part. Of course, not even a highway network is truly static, since highways are often rebuilt, new highways are added, and highways are even some-times moved. But the minute-to-minute timescale of vehicle movement is sharply different from the year-to-year changes in the infrastructure. Historically, GIS has been easier to apply to the static part, but recent developments in the technology are making it much more powerful as a tool to address the dynamic part as well. Today, it is possible to use GPS to track vehicles as they move around, and transit authorities increasingly use such systems to inform their users of the locations of buses and trains. (Longley, 2005)

#### **17.4.8 Environment**

GIS allows us to compare the environmental conditions prevailing in different nations. Generally, it is understood that the 21st century will see increasing proportions of the world's population resident in cities and towns, and so understanding of the environmental impacts of urban settlements is an increasingly important focus of attention in science and policy. Researchers have used GIS to investigate and understand how urban sprawl occurs, in order to understand the environmental consequences of sprawl and to predict its future consequences. Such predictions can be based on historic patterns of growth, together with information on the locations of roads, steeply sloping land unsuitable for development, land that is otherwise protected from urban use, and other factors that encourage or restrict urban development. Each of these factors may be represented in map form, as a layer in the GIS, while specialist software can be designed to simulate the processes that drive growth. These urban growth models are examples of dynamic simulation model, or computer programs

designed to simulate the operation of some part of the human or environmental system. (Longley, 2005)

#### 17.4.9 Other Applications

GIS applications involve both natural and man-made entities. Examples are common in areas where we study the effect of human activity on the environment. Railroad construction is such an area: it may involve parcels to be reclaimed by Government; it deals with environmental impact assessment and will usually be influenced by many restrictions, such as not crossing seasonally flooded lands, and staying within inclination extremes in hilly terrain.

A second distinction in applications of GIS stems from the overall purposes of use of the system. A prototypical use of GIS is that of a research project with an explicitly defined project objective. Such projects usually have an a priori defined duration. Feasibility studies like site suitability, but also simulation studies, for instance in erosion modeling, are examples. We call all of these project-based GIS applications.

In contrast to these are institutional GIS applications. They can be characterized in various ways. The life time (duration) of these applications is either indefinite or at least not a priori defined. Their goal is usually to provide base data to others, not to address a single research issue. Good examples of this category are monitoring systems like early warning systems for food/water scarcity, or systems that keep track of weather patterns. Indeed, our El Niño example is best qualified under this heading, because the SST and WS measurements continue. Another class of examples is found in governmental agencies like national topographic surveys, cadastral organizations and national census bureaus. They see it as their task to administer (geographic) changes, and their main business is to stay up-to-date, and provide data to others, in the form of printed material such as maps or in the form of digital data. (DeMers, 2009)

##### 17.4.9.1 Managing business activities

GIS is about business. In fact, a whole new industry of business geographic is developing because GIS can affect the bottom line, improve product quality, and provide new opportunities. Here are a few examples of how different sectors of the business world are using GIS:

**17.4.9.2 Banking:** GIS activities in banking and financial institutions include regulatory compliance customer prospecting, and locating new branches and ATMs.

**17.4.9.3 Business locations and customer behavior:** You can more effectively Fig. out where to place a business if you know your customer base's location. The closer or more accessible you are to the people whose buying patterns match your products, the more apt you are to be successful. Businesses can also compare market share with other surrounding businesses and adjust to changing demographic conditions.

**17.4.9.4 Insurance claims adjustment:** GIS can perform all the typical business and marketing tasks, including identifying potential clients, and determining risk factors. And if



you're an insurance broker, imagine being able to match applications for storm and flood insurance with a map of the 100-year flood zone.

**17.4.9.5 Journalism:** The media, especially the television news media, use flythrough, zoom-ins, and visual overlays of maps on imagery, providing a media-rich and enticing product.

**17.4.9.6 Real estate:** Real estate agents, both residential and commercial, use a GIS to search for and select properties that match client needs. Some appraisers use GIS to perform mass appraisals of whole regions at the same time.

**17.4.9.7 Trucking and delivery:** Moving products and material is getting more expensive all the time. Minimizing route lengths reduces cost, speeds delivery times, and increases customer satisfaction. GIS has tools specifically designed to work with road and rail networks.

**17.4.9.8 Planning city operations and expansion:** City, county, and regional planning has long used GIS to track development, zone land parcels, assess available resources, and plan for future growth. GIS allows planners to evaluate master plans, monitor expansion and traffic patterns, predict change, monitor population, and even decide the best place to put the new government planning office. Many GIS operations across the world are called planning departments, not GIS departments. But they often use GIS as their primary tool.

**17.4.9.9 Providing protection and emergency services:** Police, human services, and emergency services are beginning to use GIS. Crime mappers can identify crime hot spots and move officers where needed, corrections officers can track their parolees, hospitals can be placed where they meet the most need, and dispatchers can route emergency services (such as ambulance and fire) to their destinations — all with the power of GIS.

**17.4.9.10 Land management and conservation:** The first major implementation of GIS managed the enormous expanse of Canada's natural and mineral resources. In this and similar applications, the software can help monitor fires and dispatch firefighters, monitor and manage disease and insect outbreaks, control land use and land inventory, select set asides and easements, track and manage wildlife, plan for ecotourism, and much more. Combined with today's expanding complement of Earth-sensing satellites, the role of GIS in land management is sure to continue to expand.

**17.4.9.11 Military and defense-related tasks:** The military and intelligence communities are taking advantage of the GIS toolkit, which includes tools specifically targeted to those users. By combining top-secret satellite data and visual evaluation from unmanned aerial vehicles (UAVs) with the power of GIS and existing datasets, defense departments can evaluate troop movements, target artillery fire, test scenarios, perform supply and logistics operations, and monitor borders. The military and intelligence communities often rely on the same geospatial tools available to the general public — but they have exclusive access to certain data and data sources, as well as some additional and quite sophisticated software. (DeMers, 2009)

#### **17.4.9.12 Uses Of Coastal GIS Packages**

1. Retrieval of information of any specific site.
2. Provision of summary data for planning purpose.
3. Graphical display for educational and public relation exercise.
4. Classification of coasts and management zone.
5. Predicting modeling to determine.
  - A. Projections of coastal changes.
  - B. Impacts of individual schemes.
  - C. Impacts from changes of use.
  - D. Impacts from natural calamities.

(ESRI White Paper, 2006)

#### **17.4.9.13 GIS for Fire Prevention**

Wildfire prevention efforts can be focused where wildfires pose the greatest risk of resource loss. When intense fire areas (highly flammable landscapes) exist near high-risk areas (ignition sources) and high values, fire prevention becomes critical. Historical fire information can be viewed with all the other landscape information. Fire prevention officers can begin to determine an appropriate program strategy. This fire prevention strategy may be one of education, enforcement, or engineering, depending on the type of land use and historical fire causes. As housing development continues to expand and encroach into wooded and brush-covered areas, it becomes fire prone. The "urban interface" requires extensive fire prevention and fire protection measures. GIS can model and display potential fire prevention/protection strategies. (Ramachandran, 2005)

#### **17.4.9.14 GIS for Vegetation Management**

Landscapes that require vegetation management treatments (e.g., prescribed fire, mechanical treatments) will stand out when using GIS. Landscapes with high flammability characteristics (high hazard, high risk, and high value) become obvious candidates for vegetation or fuel treatment programs. GIS technology allows fire managers to identify prescribed fire and vegetation management projects with the highest benefit (meeting multiple goals for resource and fire management). Vegetation management tactics can include mechanical, chemical, or prescribed burning techniques. GIS can assist in modeling how a fire will behave and spread under a variety of conditions to assist in developing fire prescriptions. (Ramachandran, 2005)

#### 17.4.9.15 Education and Training

GIS is beneficial for education and training. Wildfire management personnel require several years of experience and training to become proficient. GIS provides access to detailed landscape information during wildfire events. Fire personnel begin to understand the complexities, fire effects, and fire behavior characteristics of various wildfires much sooner when using GIS. Modeling provides a better understanding of what a fire might do and what elements influence the wildfire most. Wildfire knowledge has traditionally been gained through years of experience, formal training, and discussions with experienced fire personnel. GIS is now another resource fire personnel can use to expand their understanding of the variables and complexities that affect wildfires. As modeling becomes more precise, dispatchers will be able to determine the potential of new starts and possible dispatch requirements.

GIS is a vital tool for wildfire information management. GIS is a primary repository of information that can be quickly accessed and viewed when required. GIS is becoming more suitable for emergency field operation use and is integrating tools that allow real-time display of information. Rapid access to information, safety, efficiency, and better resource management decisions are being made with the use of GIS for wildfire management. Information is critical for wild land fire management. GIS is information, all in one place, easy to visualize and understand.

(Ramachandran, 2005)

**Keywords:** GIS Workflow, Database Management System (DBMS), GIS packages, Open-Source GIS, Applications of GIS.



## ***Module 7: Data Models and Structures***

### **Lesson 18 Data Models**

A data model describes in an abstract way how data is represented in an information system or in a database management system. It is also a description of structural properties that define all entities represented in the database, and all the relationships that exist among them. The manner in which data is generally organized in the database management system is sometimes also called as database model. A database model is a theory or specification describing how database is structured and used (also called data structure). Several such models have been suggested. Common model include hierarchical model, network model, relational model, entity-relationship model and object oriented model.

A data model aims to identify and organize the required data logically and physically, used as part of a database design phase. A model provides attributes of a data element and also the relationship between them. Study of data model is necessary to understand the data in a better way.

Maps may be represented by several GIS data models such as raster, vector, etc. knowledge about the data models of GIS is important because it provides us information on how data are stored on the computer and their representation. Depending on the GIS data model and representation, the way in which our analysis is done and result obtained may be different (Bhatia, 2010).

#### **18.1 Spatial Data**

Data that describe a part of the Earth's surface or the features found on it could be described as geographical 'spatial' data. It includes cartographic and scientific data with photographs, videos, land records, travel information, customer databases, property records, legal documents and so on. Another term used, Features in reference to objects, located at the surface of the Earth, whose positions have been measured and described. Features may be naturally occurring objects or anthropogenic constructions and classifications. Features appearing on the map are stored as separate entities which have other intelligence stored with them as 'attributes' (Jian et al, 2009).

##### **18.1.1 Spatial Data Model**

The real world too complex for our immediate and direct understanding, so we create models or abstractions of reality that are intended to have some similarity with selected aspects of real world. A spatial database is a collection of spatially referenced data that act as model of reality.

There are two types of data models used for spatial data:

**(A) Conceptual data models:** It organizes principles that translate the real world into functional descriptions of how phenomena are represented and related to one another.

- Objects: Objects with discrete boundaries represented by geometric features.
- Fields: continuous phenomena such as elevation, temperature and soil chemistry; they exist everywhere (every point has an elevation or temperature); they are not discrete entities.

Spatial feature may be discrete (objects) or continuous (fields). Discrete features are those that do not exist between observations, those that form separate entities, and are individually distinguishable. Roads, buildings, water bodies, etc. are the examples of continuous features.

**(B) Logical data models:** It provides the explicit forms, which the conceptual model can take.

- Raster: A grid of cells or pixels.
- Vector: Objects are represented as geometric entities, including points, spaghetti lines, arc/node lines, networked lines with directionality, whole polygons, and topological polygons. Continuous fields may be also represented by vector such as triangular irregular network (TINs).
- Object oriented: Real-world objects are used as basis for abstraction, instead of geometric objects such as points lines and polygons.

Understanding the type of data model by which particular dataset is created and stored is important, since the type of functions or transformations that can be performed on the data is dependent on how it is represented and organized. The logical data model should not only describe how data is represented, but it should also describe entities of features are related to one another (Bhatia, 2010).

### 18.1.2 Spatial Data File:

1) Spatial data files are somewhat like other files you work with on a computer. They can be:

- a) stored on a hard drive, memory stick, CD, DVD
- b) Assigned either a user-defined file name, or are given default file name by a software application
- c) Organized into folders
- d) Have the ability to be opened, viewed and edited by one or more GIS software applications that understand their format.

2) However, that is where the similarities end. Spatial data files are unique in that they store “georeferenced” information – information that defines location or place. In addition, descriptive information about the georeferenced information is stored in each spatial data file.

3) Thus rather than just text (like a word processing document) or numbers (like a spreadsheet), an individual spatial data file is a digital representation of a similar group of geographic features on the surface of the earth (or any other planetary body!).

4) The geographic features can be actual physical entities or events, or they can represent conceptual features.

5) Examples of individual spatial data files representing real geographic features or events are lakes, rivers, wetlands, elevation contours, roads, forested areas, rare species habitats, soils, earthquakes, vehicle thefts, electricity distribution lines, and groundwater reservoirs.

6) Examples of individual spatial data files representing conceptual geographic features are census tract boundaries, zoning boundaries, or parcel boundaries (i.e. conceptual features do not physically exist on the landscape, but are imposed by us for various reasons and can be represented in a geographic context).

7) Each spatial data file is uniquely constructed to work within GIS software applications. Each one consists of unique characteristics:

- a) “Shapes” that attempt to reflect / convey the appearance and position of individual geographic features as accurately as possible.
- b) Records within a related tabular database that contain numeric and/or textual descriptions of each feature.
- c) A coordinate system that defines the true location of all the features on the earth’s surface (i.e. the latitude/longitude).

### 18.1.3 Formats of Spatial Data Files

Spatial data files come in several different formats. You may need to use only one, or you may need to use a combination of them, depending on your particular application and/or type of analysis. Each format falls under one of two different categories: vector or raster.

- **Vector** spatial data files are ones in which the geographic features being represented are built by a collection of vertices and lines.
- **Raster** spatial data files are ones in which the geographic features across an entire area are represented by a continuous set of “pixels” or “cells”.

#### 1. The Shapefile spatial data file format

- a) This is a very common format for spatial data files in the vector category.
- b) In this format, geographic features can be represented in one of three ways:
  - i) Points
  - ii) Lines (aka arcs)

iii) Polygons (aka areas, polylines)

## **2. The Coverage spatial data file format**

a) This was the original spatial data file format used in GIS software. While this format has taken a “backseat” to the shapefile format due to the simplicity of shapefiles, coverages are still very viable and have many advantages.

b) Along with shapefiles, the coverage format is being superseded by the new geodatabase format (see below).

c) However, many web sites still offer spatial data for download in the coverage format, so you should at least know that they exist, and know a little bit about their structure in the event you need to use one.

d) Just like shapefiles, geographic features are represented as points, lines, or polygons and many factors come into play when deciding which representation is best (see B-1-c above). Coverages also fall within the vector category.

e) Unlike shapefiles, a single coverage is actually comprised of two folders. Each folder contains a multitude of other files that the GIS software “puts together” in order to represent geographic features and associated tabular information when it is opened.

f) If either folder is missing, or if files from within either folder are missing, the coverage will be “corrupt” and not useable.

g) Coverages and shapefiles are often used almost interchangeably in GIS. They each can represent the same geographic features. It is only the internal file structure that is different. An analogy would be a Microsoft Word document vs. a Corel Word Perfect document. Both files are used to contain text (primarily) and you can import and export them at will, but they have different underlying structures which are, most of the time, invisible to you. Similarly, coverages and shapefiles are both used to contain geographic data of the vector type. It is possible to convert a shapefile to a coverage and vice-versa.

## **3. The Grid spatial data file format**

a) In most respects, grids are very different from either shapefiles or coverages. Grids fall into the raster category; they are constructed of rows and columns of pixels instead of vertices and arcs.

b) Like coverages however, grids are comprised of two folders, each containing files that the software “puts together” for display and manipulation

c) Grids can be either:

i) Integer Grids – in this case, the pixel values are integers and each integer may also be associated with one or more textual descriptions.

ii) Floating Point Grids – in this case, the pixel values will be expressed as decimals. Floating point grids cannot have textual descriptions.

#### 4. Images as Spatial Data Files

- a) Many different image formats can be used in GIS. Some of these may be more familiar than others: .jpg, .tif, .bil, .png, .img, .sid
- b) All image formats fall within the raster category of spatial data.
- c) In some cases, images are not used specifically as “spatial data”, but are used to enhance spatial data by providing a digital photograph of a place or object. For example, a shapefile representing all land parcels within a city may have links to digital photographs of each house on each parcel.
- d) In other cases, the images themselves are spatial data. Data provided from the Landsat satellite is an example of imagery that is spatial. If you have ever used Google Earth, the images that appear when you zoom in are spatial data.
- e) When an image is “georeferenced” - meaning that information is embedded within the image that describes its position on the surface of the earth in real world coordinates (Latitude/longitude) – It becomes spatial data.
- e) In addition to being “georeferenced”, many images may also be “orthorectified”. This term refers to a complex process wherein distortions caused by differences in terrain, camera tilt, and edge effects are removed from the image. Images that are both georeferenced and orthorectified are frequently called “orthophotographs” or just “Orthos” for short (<http://gisatbrown.typepad.com/gis/files/spatialdatafiles.pdf>).

#### 18.2 Attributes and Measurement Scales

Descriptive attributes can also be described as being spatial or non-spatial, though the difference between them may be subtle and ambiguous. The nature of the information stored, or rather the scale of measurement to which they belong, dictates what kind of processing or analysis can be performed with them. Measurement scales describe how values are assigned to features and objects represented in GIS. The type of scale chosen is dictated by the intended use of the recorded information. There are five scales commonly used in GIS, namely nominal, ordinal, interval, and ratio and cyclic, and these are summarized, along with the numerical operators appropriate to each case, in Table 18.1 (Jian et al, 2009).

Nominal data are discrete (i.e., mutually exclusive) and are classed according to type or quality. For example, a line could represent either a road or river, and a land use polygon could be residential, commercial, or a recreational area. Nominal data are often labeled with numbers or letters, but these labels do not imply ranking. A nominal datum can only be examined for its physical similarity to, or its difference from, other occurrences, or for the frequency of its occurrence (<http://www.rain.org/gis/map-classroom/maps-symbols-for-maps.html>). Nominal or categorical scales include numerical values used to represent real-world objects or qualitative descriptions. They can be used as ‘pointers’ to other descriptive (textual) information held in attribute tables (Jian et al, 2009).

Ordinal data provide information about rank or hierarchy, in other words, relative values. Therefore, it is possible to describe one item as larger or smaller than another, or as low,



medium or high. However, it is not possible to measure the differences between ordinal data, because there are no specific numerical values attached to them. An example of ordinal data is roads ranked as expressway, main thoroughfare, and secondary road (<http://www.rain.org/gis/map-classroom/maps-symbols-for-maps.html>). Ordinal measures involve values ranked or ordered according to a relative scale and which generally have unequal intervals. Greater than or less than operators are therefore useful but addition, subtraction, multiplication and division are not appropriate. One example is multi-element geochemical data where element concentrations are given on a percentile scale and the intervals between classes are not constant but arbitrary (Jian et al, 2009).

Scale	Operators	Examples
Nominal	=, ≠ and mode	Categorical (class) identifiers (e.g. 5=forest, 4=pasture, 9= urban)
Ordinal	<, ≤, ≥, > and median	Sequences of natural order, for example 1, 2, 3, 4
Interval	+, -, ×, ÷ and mean	Ranges between, and sections along, distance measures, for example temperature scales
Ratio	All the above	Distance measures, and subdivisions thereof, along lines and routes
Cyclic	All the above	Special measures, for example 360° bearings (azimuth), flow directions

**Table 18.1 Measurement of scale: methods for describing and operating on thematic information**

**(Source: Jian et al, 2009)**

Interval data, in addition to being ranked, include numerical values. The information can be arranged along a scale using a standard unit. Therefore, it is possible to calculate the distance or difference between ranks, which must be expressed in terms of a standard unit (<http://www.rain.org/gis/map-classroom/maps-symbols-for-maps.html>). Interval measures are used to denote quantities like distances or ranges but in this case the intervals between the values are based on equal or regular units. There is, however, no true zero on an interval scale because the position of zero depends on the units of the quantity being described. Temperature scales are a good example because the position of zero temperature depends on the units of measurement, Fahrenheit or Celsius (Jian et al, 2009).

Ratio data are the same as interval data, except there is a natural zero; therefore, it is possible to express data as ratios. Physical measurements of height, weight, and length are examples of ratio variables. With this type of data it is meaningful to state that a measurement is twice that of another. This ratio remains true no matter what the unit of measurement (e.g., meters or feet) because this type of data has a natural zero. A natural zero is a non-arbitrary starting point for data. For example, a measurement of distance at zero units has no length; furthermore, it makes sense to state that two metres are twice as long as one metre. Whereas, with the measurement of time, the year zero is arbitrary, so it is not sensible to state that the

year 2000 is twice as old as the year 1000 (<http://www.rain.org/gis/map-classroom/maps-symbols-for-maps.html>).

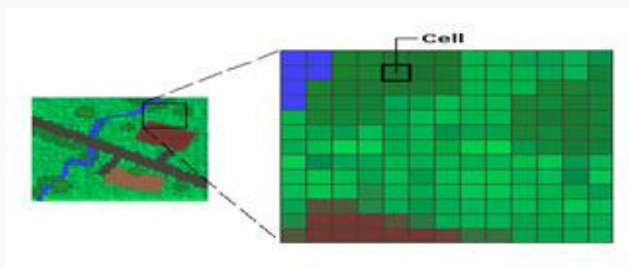
Ratio measures are similar to interval scales and are often used for distances or quantities but the zero value represents absolute zero, regardless of the units.

Cyclic measures are a special case describing quantities which are measured on regular scales but which are circular or cyclic in their framework, such as aspect or azimuth directions of slopes, or flow directions, both of which are angular measures made with respect to north. Appropriate operators are then any or all of the previously mentioned arithmetic and average operators (Jian et al, 2009).

### 18.2.1 Data structures

There are two basic types of structures used to represent the features or objects, namely raster and vector data, and as a consequence of this split, there are different types of GIS software, and different types of analysis, which have been designed in such a way as to be effective with one or the other type (Jian et al, 2009).

In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps.



**Fig. 18.1. Raster data structures.**

(Source: [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What\\_is\\_raster\\_data%3F](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What_is_raster_data%3F))

Data stored in a raster format represents real-world phenomena, such as:

- Thematic data (also known as discrete), representing features such as land-use or soils data
- Continuous data, representing phenomena such as temperature, elevation or spectral data such as satellite images and aerial photographs
- Pictures, such as scanned maps or drawings and building photographs

Thematic and continuous rasters may be displayed as data layers along with other geographic data on your map but are often used as the source data for spatial analysis with the ArcGIS Spatial Analyst extension. Picture rasters are often used as attributes in tables –

they can be displayed with your geographic data and are used to convey additional information about map features.

While the structure of raster data is simple, it is exceptionally useful for a wide range of applications. Within a GIS, the uses of raster data fall under four main categories.

### **(1) Raster as a base map**

A common use of raster data in a GIS is as a background display for other feature layers. For example, orthophotographs displayed underneath other layers provide the map user with confidence that map layers are spatially aligned and represent real objects, as well as representing additional information.

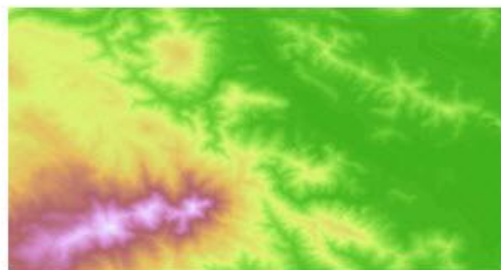


**Fig. 18.2. Rasters as base maps.**

([http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What\\_is\\_raster\\_data%3F](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What_is_raster_data%3F))

### **(2) Rasters as surface maps**

Rasters are well suited for representing data that changes continuously across a landscape (surface). They provide an effective method of storing the continuity as a surface. They also provide a regularly spaced representation of surfaces. Elevation values measured from the earth's surface are the most common application of surface maps, but other values, such as rainfall, temperature, concentration, and population density, can also define surfaces that can be spatially analyzed. The raster below displays elevation using green to show lower elevation and red, pink and white cells to show higher elevation.



**Fig. 18.3. Raster as a surface map.**

(Source: [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What\\_is\\_raster\\_data%3F](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What_is_raster_data%3F))

### (3) Rasters as thematic maps

Rasters representing thematic data can be derived from analyzing other data. A common analysis application is classifying a satellite image by land-cover categories. Basically, this activity groups the values of multispectral data into classes (such as vegetation type) and assigns a categorical value. Thematic maps can also result from geoprocessing operations that combine data from various sources, such as vector, raster, and terrain data. For example, you can process data through a geoprocessing model to create a raster dataset that maps suitability for a specific activity. Below is an example of a classified raster dataset showing land use.

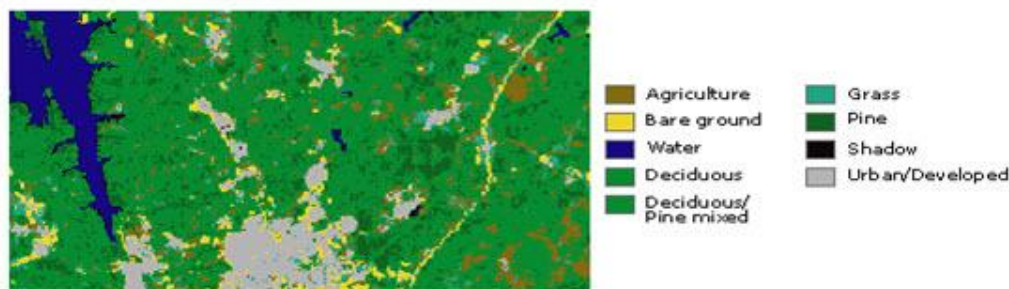


Fig. 18.4. Raster as a thematic map.

(Source: [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What\\_is\\_raster\\_data%3F](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What_is_raster_data%3F))

### (4) Rasters as attributes of a feature

Rasters used as attributes of a feature may be digital photographs, scanned documents, or scanned drawings related to a geographic object or location. A parcel layer may have scanned legal documents identifying the latest transaction for that parcel, or a layer representing cave openings may have pictures of the actual cave openings associated with the point features. Below is a digital picture of a very large, old tree that could be used as an attribute to a landscape layer that a city may maintain.

Vector data provide a way to represent real world features within the GIS environment. A feature is anything you can see on the landscape. Imagine you are standing on the top of a hill. Looking down you can see houses, roads, trees, rivers, and so on. Each one of these things would be a feature when we represent them in a GIS Application. Vector features have attributes, which consist of text or numerical information that describe the features.

A point symbol represents a geographic feature or event characterized by location and attributes. Its location can be represented by a single "x, y" coordinate pair, while attributes can be many. For example, a populated place or a mine site can be represented by a point symbol.

A line symbol represents a geographic feature characterized by linear dimension, but not area. In digital technology, a line is a sequence or stream of point coordinates with a node at each end (vector data) that symbolizes a linear feature such as a road, river, or boundary.

An area symbol represents a closed geographic surface feature, two-dimensional geometric region, or a polygon. A census division, a lake, or a province (any geographically defined surface) is examples of area features. Rasters, images or grids consist of a regular array of digital numbers or DN's, representing picture elements or which are usually square. The basic unit of such data is the pixel, or grid cell, such that points, lines and areas are represented in raster form as individual or groups of pixels. Vector or discrete data store the geometric form and location of a particular feature, along with its attribute information describing what the feature represents. Vector data typically resemble cartographic data. Points and pixels represent discrete geographic features of no or limited area, or which are too small to be depicted in any other way, such as well locations, geochemical sample points, towns or topographic spot heights. Lines are linear features consisting of connected positions which do not in themselves represent area, such as roads, rivers, railways or elevation contours. Areas are closed features that represent the shape, area and location of homogeneous features such as countries, land parcels, buildings, rock types or land-use categories.

A surface describes a variable which has a value for every position within the extent of the dataset, such as elevation or rainfall intensity, and implies data of a 'continuous' nature. Surfaces are typically represented on conventional cartographic maps as a series of isolines or contours; within GIS there are other possibilities. Deciding how these features should be stored in the database, and represented on the map, depends on the nature of that information and the work it will be required to do.

The two most basic components of GIS are therefore the pixel and the point. Every other, more complex, structure in GIS stems from, and depends on, one or other of these two basic structures. GIS operations and spatial analysis can be performed on either type of data, but that analysis will be performed slightly differently as a result of this difference. We will now describe these structures in turn  
[http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What\\_is\\_raster\\_data%3F](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=What_is_raster_data%3F).

**Keywords:** Spatial Data Model, Data structures, Formats of Spatial Data Files, Raster, Vector.



## **Lesson 19 Data Model Classification**

### **19.1 Raster Data**

Raster data represent a regular grid or array of digital numbers, or pixels, where each has a value depending on how the image was captured and what it represents. One important aspect of the raster data structure is that no additional attribute information is stored about the features it shows.

#### **19.1.1 Data quantization and storage**

The range of values that can be stored by image pixels depends on the quantization level of the data, i.e. the number of binary bits used to store the data. The more the number of bits, the greater the range of possible values. The most common image data quantization formats are 8 bit and 16 bit. The binary quantization level selected depends partly on the type of data being represented and what it is used for. In many cases, the image data file contains a header record that stores information about the image, such as the number of rows and columns in the image, the number of bits per pixel and the geo referencing information. The main issue in connection with raster data storage is the disk space potentially required. The goal of raster compression is then to reduce the amount of disk space consumed by the data file while retaining the maximum data quality.

#### **19.1.2 Spatial variability**

The raster data model can represent discrete point, line and area features but is limited by the size of the pixel and by its regular grid-form nature. A point's value would be assigned to and represented by the nearest pixel; similarly, a linear feature would be represented by a series of connected pixels; and an area would be shown as a group of connected pixels that most closely resembles the shape of that area.

#### **19.1.3 Representing spatial relationships**

Because the raster data model is a regular grid, spatial relationships between pixels are implicit in the data structure since there can be no gaps or holes in the grid. Each raster is referenced at the top-left corner; its location is denoted by its row and column position and is usually given as 0, 0. All other pixels are then identified by their position in the grid relative to the topleft. The upper left pixel being used as the reference point for 'raster space' is in contrast to 'map space' where the lower left corner is the geographical coordinate origin; this difference has an effect on the way raster images are geo referenced. Another benefit of implicit spatial relationships is that spatial operations are readily facilitated.

#### **19.1.4 The effect of resolution**

The accuracy of a map depends on the scale of that map. In the raster model the resolution, scale and hence accuracy depends on the real-world area represented by each pixel or grid

cell. The pixel can be thought of as the limit beyond which the raster becomes discrete, and with computer power becoming ever greater we may have fewer concerns over the manageability of large, high-resolution raster files. Providing we maintain sufficient spatial resolution to describe adequately the phenomenon of interest, we should be able to minimize problems related to accuracy.

### 19.1.5 Representing surfaces

Raster's are ideal for representing surfaces since a value, such as elevation, is recorded in each pixel and the representation is therefore 'continuously' sampled across the area covered by the raster. The input dataset representing the surface potentially contributes two pieces of information to this kind of perspective viewing. The first is the magnitude of the DN which gives the height and the second is the way the surface appears or is encoded visually, i.e. the DN value is also mapped to colour in the display. [(Liu, and Mason, 2009) Essential Image Processing and GIS for Remote Sensing, UK.]

#### (A) Advantage of Raster Data

- The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.
- Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.
- The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.
- Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.
- Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.

([http://bgis.sanbi.org/gis-primer/page\\_19.htm](http://bgis.sanbi.org/gis-primer/page_19.htm) Accessed on 03.12.2012.)

#### (B) Disadvantage of Raster Data

- The cell size determines the resolution at which the data is represented.
- It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
- Processing of associated attribute data may be cumbersome if large amounts of data exist. Raster maps inherently reflect only one attribute or characteristic for an area.
- Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization and choice of inappropriate cell size.



- Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

## 19.2 Vector Data

Vector data is a series of discrete features described by their coordinate positions rather than graphically or in any regularly structured way. Tabular data represent a special form of vectordata which can include almost any kind of data, whether or not they contain a geographic component. Tabular data are not necessarily spatial in nature. The vector model could be thought of as the opposite of raster data in this respect, since it does not fill the space it occupies; not every conceivable location is represented, only those where some feature of interest exists.

There have been a number of vector data models developed over the past few decades, which support topological relationships to varying degrees, or not at all. The representation, or not, of topology dictates the level of functionality that is achievable using those data. These models include spaghetti (unstructured), vertex dictionary, dual independent map encoding (DIME) and arc-node (also known as POLYVRT). To understand the significance of topology it is useful to consider these models, from the simplest to the more complex.

### 19.2.1 Unstructured or spaghetti data

Spaghetti form of vector data is stored without relational information. There is no mechanism to describe how their features relate to one another i.e. there is no topology. The advantages of unstructured data are that their generation demands little effort and that the plotting of large unstructured vector files is potentially faster than the structured data. Disadvantages are that storage is insufficient.

### 19.2.2 Vertex dictionary

Vertex dictionary is a minor modification of the 'spaghetti' model. It involves the use of two files to represent the map instead of one. This prevents duplication, since each coordinate pair is stored only once, but it does not allow any facility to store the relationships between the features, i.e. topology is still not supported.

### 19.2.3 Dual Independent Map Encoding (DIME)

The DIME structure was developed by the US Bureau of the Census for managing its population Databases. Both street addresses and UTM coordinates were assigned to each entity in the database. Here again, additional files (tables) are used to describe how the coordinate pairs are accessed and used.

### 19.2.4 Arc-node Structure or "POLYVRT" (POLYgonconVeRTer)

A more efficient model for the storage of vector data, and one which supports topological relationships, is the 'arc-node' data structure. Here vector entities are stored separately but are linked using pointers. An arc is a line which, when linked with other arcs, forms a polygon. Arcs may be referred to as edges and sometimes as chains. A point where arcs terminate or connect is described as a node. Polygons are formed from an ordered sequence



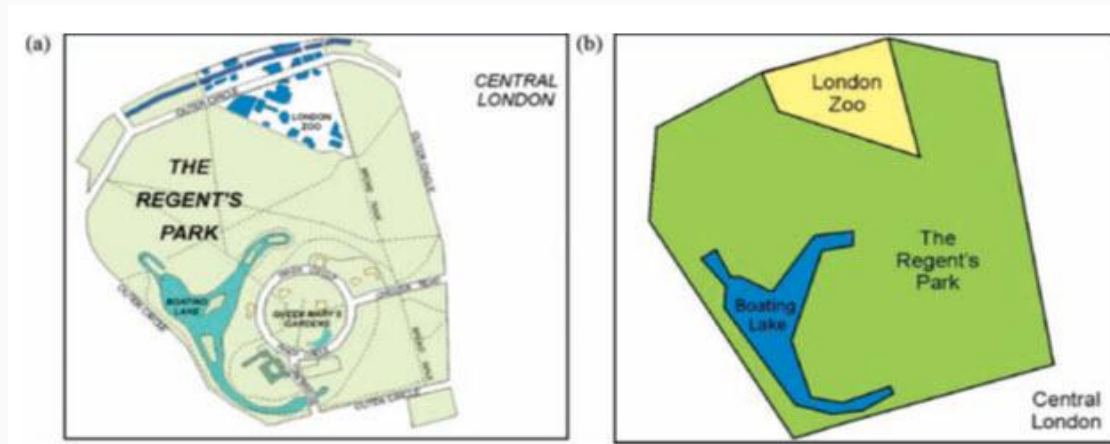
of arcs and may be termed 'simple' or 'complex' depending on their relationship to other polygons. When the features are digitized each arc is digitized between nodes, in a consistent direction; it has a start node, and is a given identifying number.

### 19.2.5 Connectivity

Connectivity allows the identification of a pathway between two locations, between your home and the airport, along a bus, rail and/or underground route, for instance. Using the arc-node data structure, a route along an arc will be defined by two end points, the start or from-node and the finish or to-node. Network connectivity is then provided by an arc node list that identifies which nodes will be used as the from and to positions along an arc

### 19.2.6 Area Definition

This is the concept by which it is determined that the Boating Lake lies completely within Regent's Park, i.e. it represents an island polygon inside it, as shown in Fig. below.



**Fig. 19.1. (a) A topographic map showing the Regent's Park in London; and (b) a topological map showing the location of the London zoo and the Boating Lake which lie inside the Regent's Park in London .**

(Source: Liu, and Mason, 2009)

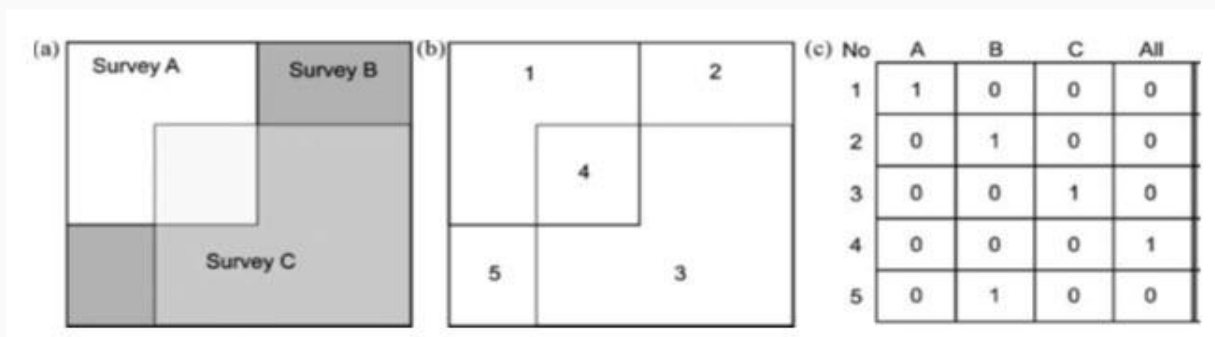
### 19.2.7 Contiguity or Adjacency

Contiguity, a related concept, allows the determination of adjacency between features. Two features can be considered adjacent if they share a boundary. Hence, the polygon representing London Zoo can be considered adjacent to Regent's Park. The from-node and to-node define an arc's direction, so that the polygons' left and right sides must also be known, left-right topology describes this relationship and therefore adjacency.

### 19.2.8 Extending the Vector Data Model

Topology allows us to define areas and to model three types of association, namely connectivity, area definition and adjacency (or contiguity), but we may still need to add further complexity to the features we wish to describe. For instance, a feature may represent a composite of other features, so that a country could be modelled as the set of its counties,

where the individual counties are also discrete and possibly geographically disparate features. Alternatively, a feature may change with time, and the historical tracking of the changes may be significant. For instance, a parcel of land might be subdivided and managed separately but the original shape, size and attribute information may also need to be retained. Other examples include naturally overlapping features of the same type, such as the territories or habitats of several species, or the marketing catchments of competing supermarkets, or surveys conducted in different years as part of an exploration program (as illustrated in Fig. 19.2). The 'spaghetti' model permits such area subdivision and/or overlap but cannot describe the relationships between the features. Arc-node topology can allow overlaps only by creating a new feature representing the area of overlap, and can only describe a feature's relationship with its subdivisions by recording that information in the attribute table. Several new vector structures have been developed by ESRI and incorporated into its ArcGIS technology. These support and enable complex relationships and are referred to as regions, sections, routes and events.

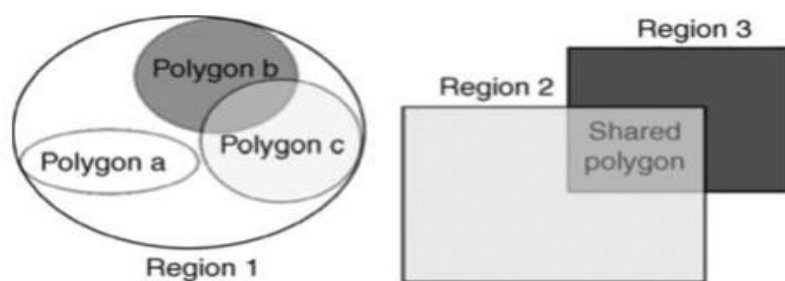


**Fig. 19.2. (a) Map of the boundaries of three survey areas, carried out at different times. Notice that the areas overlap in some areas; this is permitted in 'spaghetti' data but not in arc-node structures; (b) the same survey maps after topological enforcement to create mutually exclusive polygonal areas; (c) the attribute table necessary to link the newly created polygons (1 to 5) to the original survey extents (A, B and C).**

(Source: Modified after Bonham-Carter, 2002)

### 19.2.8.1 Regions

A region consists of a loose association of related polygons and allows the description of the relationships between them. A series of arcs and vertices construct a polygon and a series of polygon forms a region. The polygon comprising the region may be listed in any order. Point's lines and polygon has a unique identifier. The polygons representing the features within the region are independent, they may overlap and they do not necessarily cover the entire area represented by the region. So overlapping survey areas in fig 19.2 could simply be associated within a survey region. Constructing overlapping regions is rather similar to constructing polygons; where regions overlap, they share a polygon in the same way that polygons share an arc where they meet, as shown in Fig. 19.3

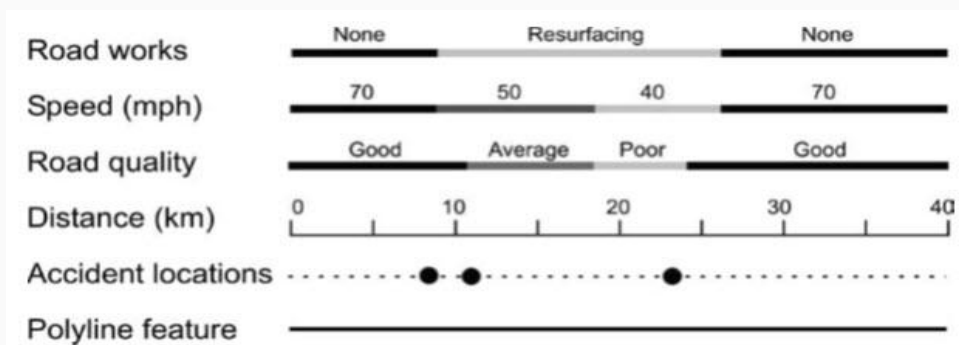


**Fig. 19.3. Illustration of different types of region: associations of polygons and overlapping polygons which share a polygon.**

(Source: Liu, and Mason, 2009)

### 19.2.8.2 Linear referencing

Routes, sections and events can be considered together since they tend not to exist on their own, and together they constitute a system of linear referencing as it is termed in ESRI's ArcGIS. The constructed route defines a new path along an existing linear feature or series of features, as illustrated in Fig. 19.4.



**Fig. 19.4. Several routes representing different measures (linear and point events), created from and related to a pre-existing polyline feature representing, in this case, a road network. (Source: Liu, and Mason, 2009)**

- ROUTE**

A route may be circular, beginning and ending in the same place. Routes may be disconnected, such as one that passes through a tunnel and so is not visible at the surface. A further piece of information necessary for the description of a route is the unit of measurement along the route. This could be almost any quantity and for the example of a journey the measure could be time or distance.

- SECTION**

A section describes particular portions of a route, such as a where road works are in progress on a motorway, where speed limits are in place on a road, or where a portion of a pipeline is currently undergoing maintenance. Again, starting and ending nodes of the section must be defined according to the particular measure along the route.

- **EVENT**

An event describes specific occurrences along a route, and events can be further subdivided into point and linear events. A point event describes the position of a point feature along a route, such as an accident on a section of motorway or a leak along a pipeline. The point event's position is described by a measure of, for instance, distance along the route. A linear event describes the extent of a linear feature along a route, such as speed restrictions along a motorway, and is rather similar in function to a section. A linear event is identified by measures denoting the positions where the event begins and ends along the route. Route and event structures are of use in the description of application-specific entities such as seismic lines and shot-point positions. Since conventional vector structures cannot inherently describe the significance of discrete measurements along such structures. Along seismic lines the shot points are the significant units of measurement but they are not necessarily regularly spaced or numbered along that line, so they do not necessarily denote distance along it or any predictable quantity.

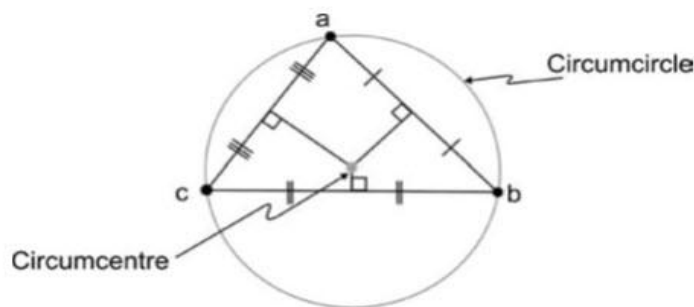
### **19.2.9 Representing surfaces**

The vector data model provides several options for surface representation. Iso-lines (or contours), the triangulated irregular network, or TIN, and, Thiessen polygons (although less commonly used). Contours can only describe the surfaces from which they were generated and so do not readily facilitate the calculation of further surface parameters, such as slope angle, or aspect (the facing direction of that slope); both of these are important for any kind of 'terrain' or surface analysis. The techniques surrounding the calculation of contours are comprehensively covered in many other texts and so we will skirt around this issue here.

#### **19.2.9.1 TIN Surface Model (Triangulated Irregular Network)**

The TIN data model describes a 3D surface composed of a series of irregularly shaped and linked but non-overlapping triangles. The TIN is also sometimes referred to as the 'irregular triangular mesh' or 'irregular triangular surface model'. The points which define the triangles can occur at any location, hence the irregular shapes. This method of surface description differs from the raster model in three ways. Firstly, it is irregular in contrast with the regular spacing of the raster grid; secondly, the TIN allows the density of point spacing (and hence triangles) to be higher in areas of greater surface complexity (and requires fewer points in areas of low surface complexity); and lastly it also incorporates the topological relationships between the triangles.

The process of Delaunay triangulation is used to connect the input points to construct the triangular network. The triangles are constructed and arranged so that no point lies inside the circumcircle of any triangle (Fig. 19.5). Delaunay triangulation maximizes the smallest of the internal angles and so tends to produce 'fat' rather than 'thin' triangles.



**Fig. 19.5. The Delaunay triangle constructed from three points by derivation of the circumcircle and circumcentre; the position of the latter is given by the intersection of the perpendicular bisectors from the three edges of the triangle. (Source: Liu, and Mason, 2009)**

As with all other vector structures, the basic components of the TIN model are the points or nodes, and these can be any set of mass points with which are stored values other than x, y and a unique identifying number, i.e. a z value in their attribute table. Nodes are connected to their nearest neighbours by edges, according to the Delaunay triangulation process.

The input mass points may be located anywhere. Of course the more carefully positioned they are, the more closely the model will represent the actual surface. TINs are sometimes generated from raster elevation models, in which case the points are located according to an algorithm that determines the sampling ratio necessary to describe the surface adequately. Well-placed mass points occur at the main changes in the shape of the surface, such as ridges, valley floors, or at the tops and bottoms of cliffs

TINs allow rapid display and manipulation but have some limitations. The detail with which the surface morphology is represented depends on the number and density of the mass points and so the number of triangles. So to represent a surface as well and as continuously as a raster grid, the point density would have to match or exceed the spatial resolution of the raster. Further, while TIN generation involves the automatic calculation of slope angle and aspect for each triangle, in the process of its generation, the calculation and representation of other surface morphological parameters, such as curvature, are rather more complex and generally best left in the realm of the raster. (Liu & Mason, 2009)

#### **A) Advantage of Vector Data**

- Data can be represented at its original resolution and form without generalization.
- Graphic output is usually more aesthetically pleasing (traditional cartographic representation).
- Accurate geographic location of data is maintained.
- Hard copy maps, in vector form no data conversion is required.

## (B) Disadvantage of Vector Data:

- The cell size determines the resolution at which the data is represented.
- It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
- Processing of associated attribute data may be cumbersome if large amounts of data exists. Raster maps inherently reflect only one attribute or characteristic for an area.
- Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introducedata integrity concerns due to generalization and choice of inappropriate cell size.
- Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

([http://bgis.sanbi.org/gis-primer/page\\_19.htm](http://bgis.sanbi.org/gis-primer/page_19.htm) Accessed on 03.12.2012.)

**Keywords:** Data quantization and storage, Spatial variability, Spaghetti data, Unstructured or Vertex dictionary, Dual Independent Map Encoding (DIME), Arc-node Structure or "POLYVRT" (POLYgonconVeRTer) Contiguity or Adjacency, Triangulated Irregular Network.



## Lesson 20 Data Structures

### 20.1 Introduction

The raster and vector methods of spatial data structure are two different approaches for modelling graphical information. Earlier the notion was that the raster and vector data structures were irreconcilable alternatives. However, now-a-days it is accepted that these two approaches have to be used in a synergetic manner for optimal results. Raster method requires huge computer memories to store and process image at the level of spatial resolution obtained by vector structures. Certain operations such as polygon intersection or spatial averaging presented enormous technical problems with the choice of raster methods that allowed easy spatial analysis but resulted in poor maps or vector methods that could provide database of manageable size and elegant graphics but in which spatial analysis was extremely difficult. The problem of raster or vector disappears once it is realised that both are valid methods for representing spatial data, and that both structures are inter-convertible. Conversion from vector to raster (rasterization) is the simplest. The reverse operation i.e. raster to vector (vectorization) is also well understood but is much more complex operation that is complicated by the need to reduce the number of coordinates in the resulting line by a process known as weeding. ([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm)).

In this unit we will focus on the conversion of data structures from vector-to-raster (rasterization) as well as from raster-to-vector (vectorization). A little description of the various errors produced due to the conversion of data structure is also included in the end of the chapter.

#### 20.1.1 Geo-data

As Geo-data we can define every dataset that has a spatial aspect or component. It can also be called as "spatial data", "geographic data", "geographic data sets" or "GIS data". The syllable "Geo" implies that the dataset has a spatial component that allows to geo-reference the described phenomena to a location or region on the earth.

#### 20.1.2 Geo-data structure

As geo-data structure we can define the logical, internal data organization of our geographic information, the means of representing a real-life entity inside a geo-data model. Data structures should enable data storage and data management, as well as quick retrieval of the data. Unique identifier, links, relationships and dependencies help to build consistent and normalized data structures and enable links within the dataset or to external data sources.

#### 20.1.3 Geo-data model

A geo-data model is an abstract, artificially created mapping of a part of the real world relevant to a geo-informatics project. The goal of geo-data modelling is to map the relevant

conditions and processes in the real world to geo-data structure. A data model not only describes the content, properties and data structures, but also rules and relations between the entities of a data model. ([https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/L2GeodataStructuresAndDataModels\\_glossary.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/L2GeodataStructuresAndDataModels_glossary.html)).

## 20.2 Conversion between Data Models and Structures

Spatial data can be represented in two formats, raster (grid cell) or vector (polygon). Raster data such as satellite images and scanned maps are comprised of numerically coded grid cells. Vector data are comprised of coded points, lines, and polygons. There are sometimes circumstances when conversion from raster to vector formats is necessary for display and/or analysis. Data may have been captured in raster form through scanning, for instance, but may be needed for analysis in vector form (e.g. elevation contours needed to generate a surface, from a scanned paper topographic map). Data may have been digitized in vector form but subsequently needed in raster form for input to some multi-criteria analysis. In such cases it is necessary to convert between models and some consideration is required as to the optimum method, according to the stored attributes or the final intended use of the product. Moreover, there are different types of vector and raster formats for data structures, which make it necessary to have intra format data conversion procedures. Thus there are four sets of conversion methods (Adam and Gangopadhyay, 2000):

- (i) Raster to raster
- (ii) Raster to vector
- (iii) Vector to raster
- (iv) Vector to vector

Most geographic information systems (GIS) now provide software for such a conversion. There are a number of processes which fall under this description and these are summarized in Table 20.1

**Table 20.1 Summary of general conversions between feature types (points, lines and areas), in vector/raster form**

Conversion type	To point/pixel	To line	To polygon/area
From point/pixel	Grid or lattice creation	Contouring, line scan conversion/filling	Building topology, TIN, Thiessen polygons/interpolation, dilation
From line	Vector intersection, line splitting	Generalizing, smoothing, thinning	Buffer generation, dilation



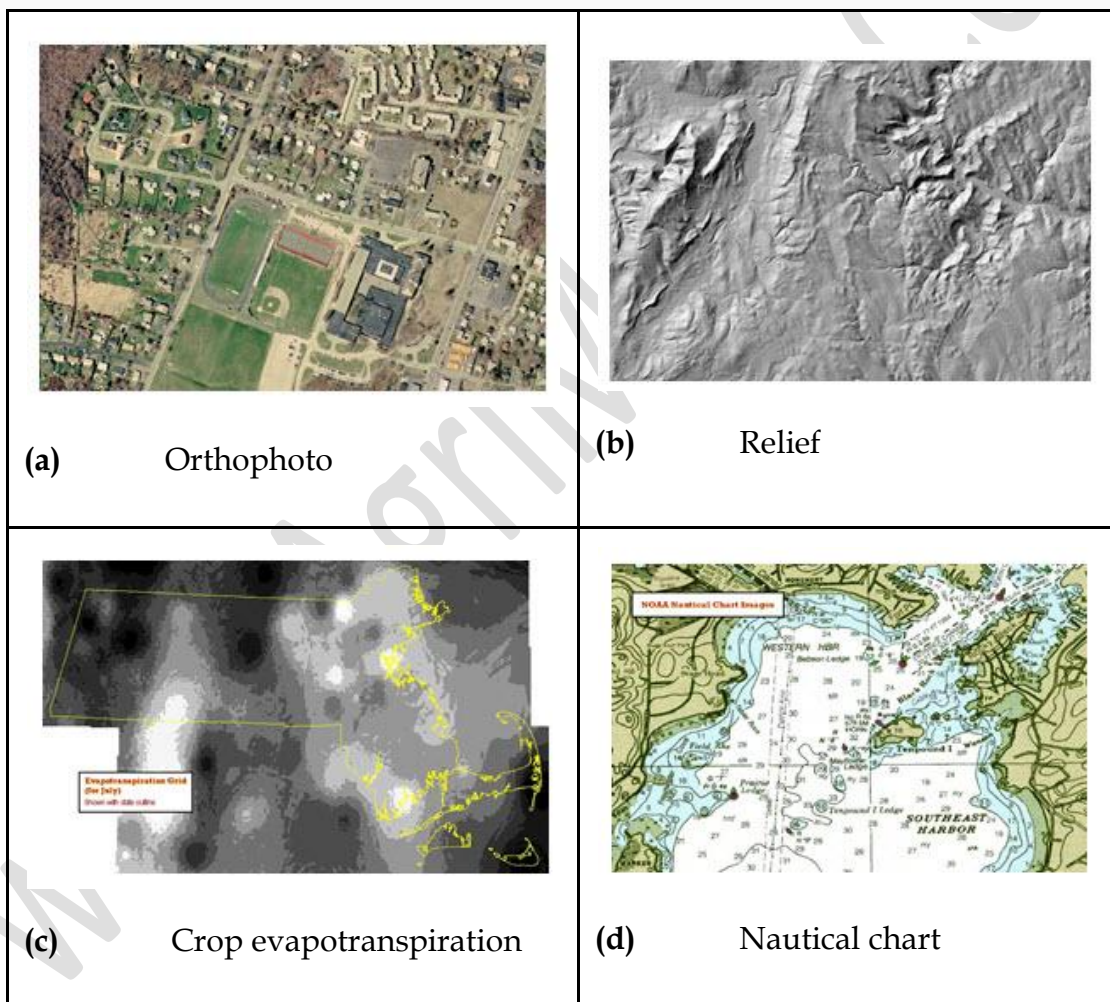
From polygon/area	Centre point derivation, vector intersection	Area collapse, skeletonization, erosion, thinning	Clipping, subdivision, merging
-------------------	---	---	-----------------------------------

(Source: Liu and Mason, 2009)

### 20.3 Different uses of the Data in GIS

Since raster data refers directly to spatial extensions instead of lines or points, as it is in vector data, it is difficult to overlay with other raster data information, that's why it is often used as background information. The difference between typical GIS raster data sets and vector data sets are illustrated in following section:

#### *Typical GIS raster data sets*



**Fig. 20.1. Examples of typical GIS raster dataset.**

(Source: [https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit\\_u4VecVsRas.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit_u4VecVsRas.html); accessed on Dec 02, 2012)

#### *Advantages of Raster Data Structures*

- Simple data structure
- Easy to generate (e.g. from remote sensing or scan-digitizing)

- Easy workflows and analysis
- Technology is cheap and is being energetically developed
- Overlay and limitation of mapped data with remotely sensed data is easy
- Various kinds of analytical (spatial) operations are easy
- Simulation is very easy as each spatial unit has the same shape and size
- Same set of grid cells are used for several variables
- Simple when doing your own programming

([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm))

### **Disadvantages of Raster Data Structures**

- Non-adaptive data structure
- Tends to generate huge files, depending on resolution
- Cell arrangement is usually random and does not respect natural borders
- Limited interactivity and more primitive analysis algorithms
- Errors in evaluating perimeter of shape
- Topology or network linkages are difficult to establish
- Geometric transformations are difficult to handle
- Use of large cells to reduce data volumes result into loss of information

([https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit\\_u2Raster.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit_u2Raster.html))

([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm))

### **Usage Scenarios of Raster Data Structures**

- Photos
- Photogrammetry and remote sensing
- Scanned images of maps
- Terrain modelling
- Landcover analysis

- Hydrologic modelling and analysis
- General GIS surface modelling and analysis for continuous surfaces

([https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit\\_u2Raster.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit_u2Raster.html))

### Typical GIS vector data sets

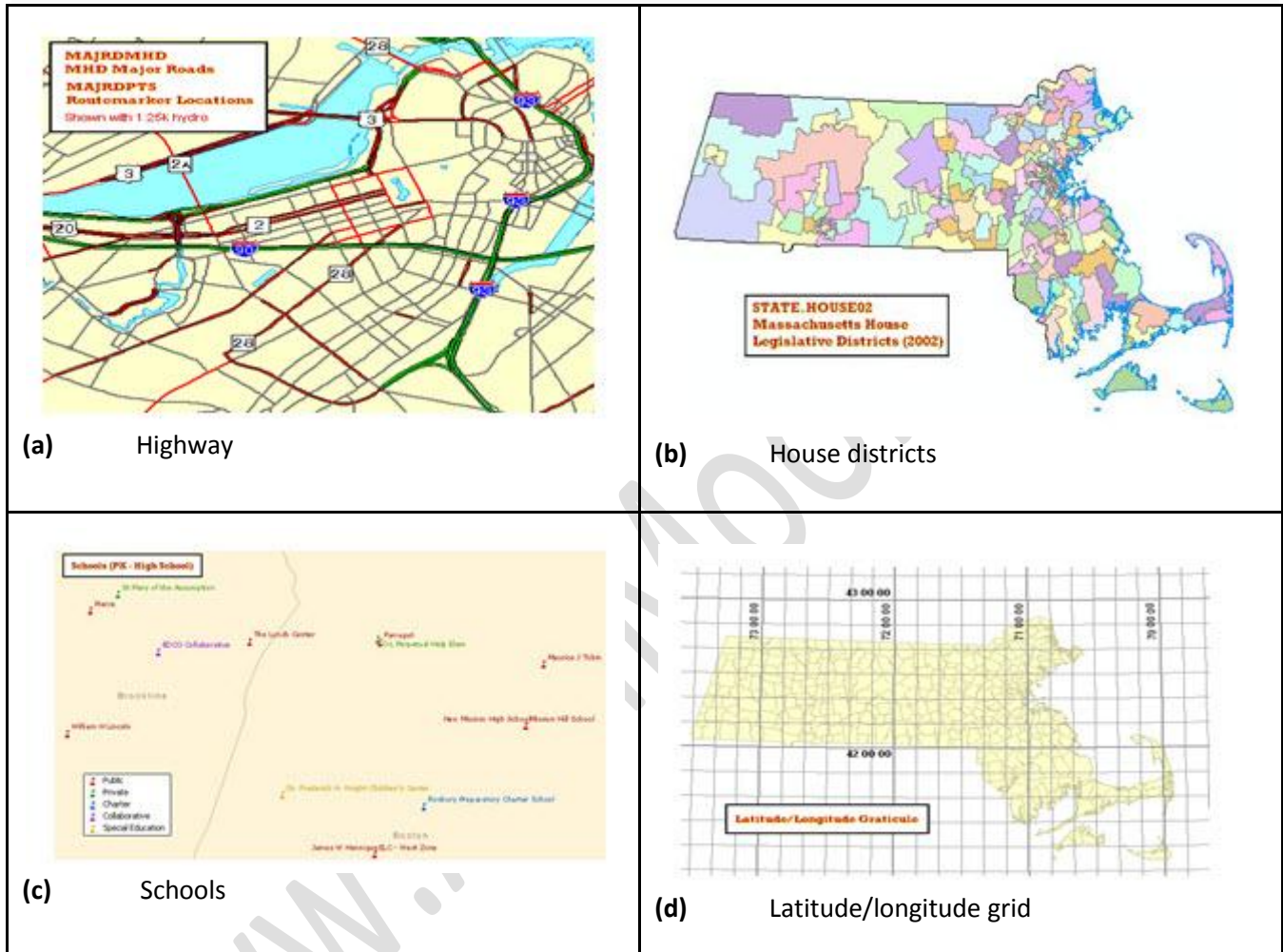


Fig. 20.2. Examples of typical GIS vector dataset.

(Source: Liu and Mason, 2009)

### Advantages of Vector Data Structures

- Small amount of data
- Logical data structure
- Attributes are combined with objects
- Preserves quality after interactivity (e.g. scaling)
- More sophisticated in spatial analysis

- Topology described with network linkages
- Retrieval, updating and generalization of graphics and attributes are possible
- Widely used to described administrative zones

([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm))

([https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit\\_u3Vector.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit_u3Vector.html))

### **Disadvantages of Vector Data Structures**

- Complex data structure
- Continuous data is not represented effectively
- Spatial variability is not implicitly represented
- Spatial analysis and filtering within polygons is impossible
- Needs a lot of manual editing to get good quality
- It always introduces hard boundaries
- Simulation is difficult as each unit has different topological form
- Overlaying of several polygon maps or polygon and raster maps is difficult
- Display and plotting can be expensive

([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm))

([https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit\\_u3Vector.html](https://geodata.ethz.ch/geovite/tutorials/L2GeodataStructuresAndDataModels/en/html/unit_u3Vector.html))

### **Usage Scenarios of Vector Data Structures**

- CAD, technical drawings
- Street or river networks, cadastral maps
- Network analysis
- Cartography

([http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_37/E\\_37\\_4.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_37/E_37_4.htm))

## 20.4 Intra-Format Conversion

Raster to raster and vector to vector data conversions are the types of intra-format conversion methods. Raster formats differ in the way they are stored. Three commonly used storage strategies for data in raster formats are band sequential (BSQ), band interleaved by pixels (BIP) and band interleaved by lines (BIL). The purpose of all these strategies is to efficiently stored raster structure data for different thematic layers. For example, a given geographic study area may have three thematic data layers: land elevation, rainfall and slope. Each of these layers will be stored in a separate raster structure with the same resolution. These individual raster structures can be stored in different files, which is called the BSQ strategy; or the pixels (raster cells) can be stored sequentially with all data value stored after each pixels, which is the BIL strategy. Conversion between raster formats requires reorganization of the raster cells and their corresponding data values, which is a relatively simple operation as compare to inter format conversion.

Vector to vector conversion is necessitated due to the usage of different types of vector formats. Data structures in vector formats can be classified into two primary categories. In the first category is the whole polygon method in which the polygons are stored in terms of the coordinates of the vertices. In this method the nodes and the arc are implicitly represented. In contrast other methods, including DIME (Dual Independent Map Encoding), arc-node and relational structures are variation of the arc node structure, where nodes arcs and polygons are stored in the tabular formats.

## 20.5 Inter-Format Conversion

Inter format conversion consists of two types (Adam and Gangopadhyay, 2000):

- (a) Vector to raster
- (b) Raster to vector

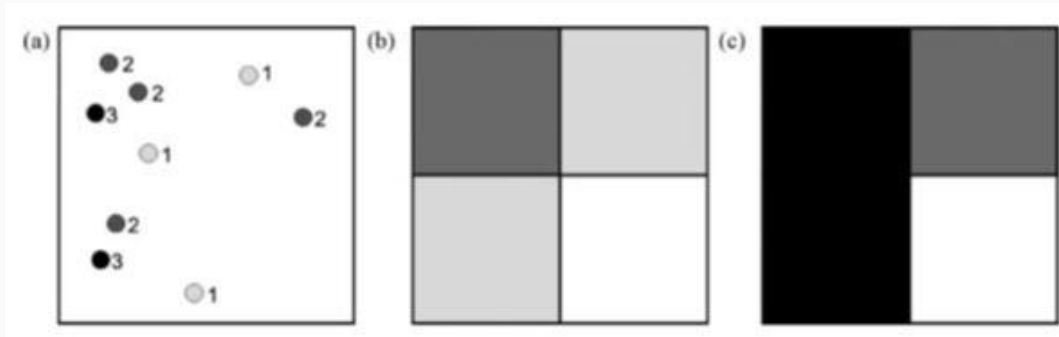
### 20.5.1 Vector to raster conversion (rasterization)

These processes begin with the identification of pixels that approximate significant points, and then pixels representing lines are found to connect those points. The locations of features are precisely defined within the vector coordinate space but the raster version can only approximate the original locations, so the level of approximation depends on the spatial resolution of the raster. The finer the resolution, the more closely the raster will represent the vector feature. Many GIS programs require a blank raster grid as a starting point for these vector-raster conversions where, for instance, every pixel value is '0' or has a null or no data value. During the conversion, any pixels that correspond to vector features are then 'turned on': their values are assigned a numerical value to represent the vector feature.

#### 20.5.1.1 Point to raster

For conversions between vector points and a discrete raster representation of the point data, there are several ways to assign a point's value to each pixel (as shown in Fig 20.3). The first is to record the value of the unique identifier from each vector point. In this case, when more than one vector feature lies within the area of a single pixel, there is a further option to accept

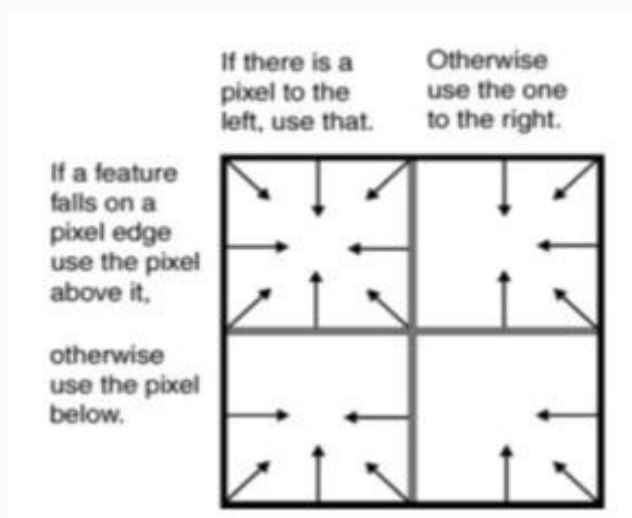
the value of either the first or the last point encountered since there may be more than one within the area of the pixel. Another option is to record a value representing merely the presence of a point or points. The third choice is to record the frequency of points found within a pixel. The fourth is to record the sum of the unique identifying numbers of all vector points that fall within the area of the output pixel. The last is to record the highest priority value according to the range of values encountered.



**Fig. 20.3. (a) Input vector point map (showing attribute values). (b) and (c) Two different resulting raster versions based on a most frequently occurring value rule (if there is no dominantly occurring value, then the lowest value is used) (b), and a highest priority class rule (where the attribute values 1–3 are used to denote increasing priority) (c).**

(Source: Liu and Mason, 2009)

Point, line and polygon features can be converted to a raster using either textual or numerical attribute values. Only numbers are stored in the raster file– numbers in a value range which dictates how the raster data are quantized, as byte or integer data for instance. So if text fields are needed to describe the information in the output raster, an attribute table must be used to relate each unique raster DN to its text descriptor. When pixels do not encounter a point, they are usually assigned a null (No Data) or zero value. The last is to record the highest priority value according to the range of values encountered.



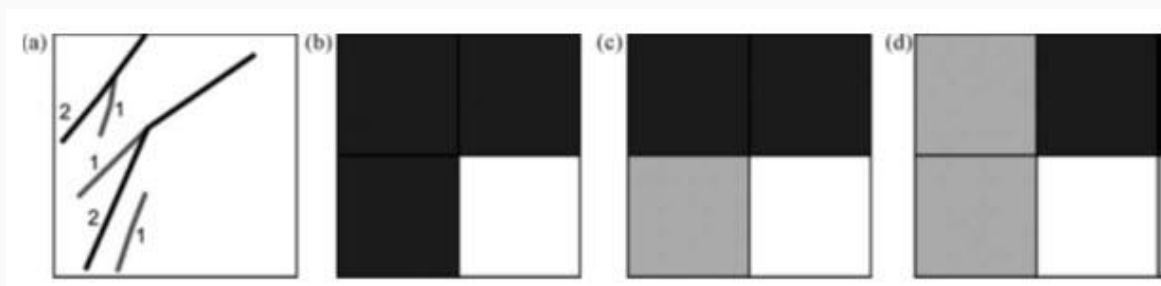
**Fig. 20.4. Boundary inclusion rules applied when a feature falls exactly on a pixel boundary. Arrows indicate the directional assignment of attribute values. (Source: Modified after ESRI's ArcGIS online knowledge base)**

The above rules decide the value assigned to the pixel but further rules are required when points fall exactly on the boundary between pixels. These are used to determine which pixel will be assigned the appropriate point value. The scheme used within ESRI's ArcGIS is illustrated in Fig 20.4, in which a kind of kernel and associated logical rules provide consistency by selecting the edge and direction to which the value will be assigned.

Point to raster area conversions also includes conversion to the continuous raster model. This category generally implies interpolation or gridding, of which there are many different types.

### 20.5.1.2 Polyline to raster

A typical line rasterizing algorithm first finds a set of pixels that approximate the locations of nodes. Then lines joining these nodes are approximated by adding new pixels from one node to the next one and so on until the line is complete. As with points, the value assigned to each pixel when a line intersects it is determined by a series of rules. If intersected by more than one feature, the cell can be assigned the value of the first line it encounters, or merely the presence of a line (as with point conversions above), or of the line feature with the maximum length, or of the feature with the maximum combined length (if more than one feature with the same feature ID cross it), or of the feature that is given a higher priority feature ID (as shown in Fig 20.5).



**Fig. 20.5. (a) Input vector line map. (b) to (d) Three different resulting raster versions based on a maximum length rule (or presence/absence rule) (b), maximum combined length rule (c) and a highest priority class rule (d) where the numbers indicate the priority attribute values.**

**(Source: Liu and Mason, 2009)**

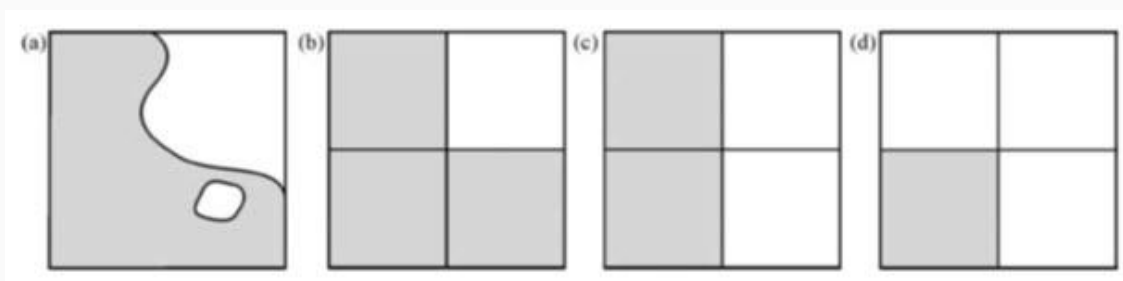
Again, pixels which are not intersected by a line are assigned a null or No Data value. Should the feature fall exactly on a pixel boundary, the same rules are applied to determine which pixel is assigned the line feature value, as illustrated in Fig 20.4.

The rasterizing process of a linear object initially produces a jagged line, of differing thickness along its length, and this effect is referred to as aliasing. This is visually unappealing and therefore undesirable but it can be corrected by anti-aliasing techniques such as smoothing. When rasterizing a line or arc the objective is to approximate its shape as closely as possible, but, of course, the spatial resolution of the output raster has a significant effect on this.

### 20.5.1.3 Polygon to raster

The procedures used in rasterizing polygons are sometimes referred to as 'polygon scan conversion' algorithms. These processes begin with the establishment of pixel representations of points and lines that define the outline of the polygon. Once the outline is found, interior pixels are identified according to inclusion criteria; these determine which pixels that are close to the polygon's edge should be included and which ones should be rejected. Then the pixels inside the polygon are assigned the polygon's identifying or attribute value. This value will be found from the pixel that intersects the polygon centre. The inclusion criteria in this process may be one of the following, whose effects are illustrated in Fig 20.6:

- a) Central point rasterizing, where the pixel is assigned the value of the feature which lies at its centre.
- b) Dominant unit or largest share rasterizing, where a pixel is assigned the value of the feature (or features) that occupies the largest proportion of that pixel.
- c) Most significant class rasterizing, where priority can be given to a certain value or type of feature, such that if it is encountered anywhere within the area of a pixel, the pixel is assigned its value.



**Fig. 20.6. (a) Input vector polygons map. (b) to (d) Three different resulting raster versions based on a dominant share rule (b), a central point rule (c) and a most significant class rule (d). (Source: Liu and Mason, 2009)**

When viewed in detail (as in Figures 20.3, 20.5 and 20.6) it can be seen that the inclusion criteria have quite different effects on the form of the raster version of the input vector feature. Again, if the polygon feature's edge falls exactly on a pixel edge, special boundary rules are applied to determine which pixel is assigned the line feature value, as illustrated in Fig 20.4.

## 20.5.2 Raster to vector conversion (vectorization)

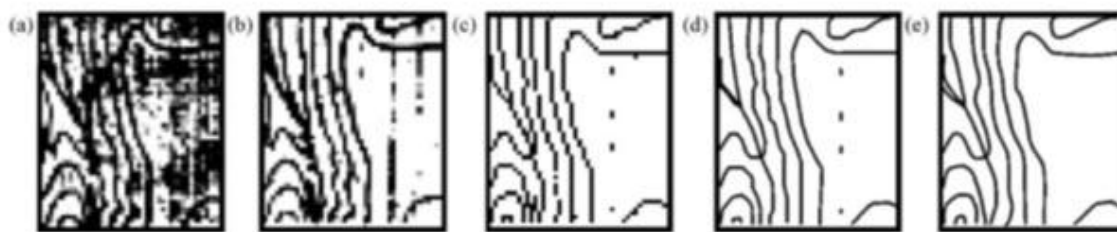
### 20.5.2.1 Raster to point

All non-zero cells are considered points and will become vector points with their identifiers equal to the DN value of the pixel. The input image should contain zeros except for the cells that are converted to be points. The x, y position of the point is determined by the output point coordinates of the pixel centroid.



### 20.5.2.2 Raster to polyline

This process essentially traces the positions of any non-zero or non-null raster pixels to produce a vector polyline feature, summarized in Fig 20.7.



**Fig. 20.7. Illustration of polygon scan vectorization procedures: (a) image improvement by conversion to binary (bi-level) image; (b) thinning or skeletonization process to reduce thickness of features to a single line of pixels; (c) vectorized lines representing complex areas; (d) collapsed but still including spurious line segments; and (e) the lines are smoothed or 'generalized' to correct the pixelated appearance and line segments removed. (Source: Liu and Mason, 2009)**

One general requirement is that all the other pixel values should be zero or a constant value. Not surprisingly, the input cell size dictates the precision with which the output vertices are located. The higher the spatial resolution of the input raster, the more precisely located the vertices will be. The procedure is not simple and generally involves several steps and is summarized as follows:

1. **Filling:** The image is first converted from a greyscale to a binary raster (through reclassification or thresholding), then any gaps in the features are filled by dilation.
2. **Thinning:** The line features are then thinned, skeletonized or eroded, i.e. the edge pixels are removed in order to reduce the line features to an array or line of single but connected pixels.
3. **Vectorizing:** The vertices are then created and defined at the centroids of pixels representing nodes, that is where there is a change in orientation of the feature. The lines are produced from any connected chains of pixels that have identical DN value. The resultant lines pass through the pixel centres. During this process, many small and superfluous vertices are often created and these must be removed. The vectors produced may also be complex and represent area instead of a true linear feature.
4. **Collapsing:** Complex features are then simplified by reducing the initial number of nodes, lines and polygons and, ideally, collapsing them to their centre lines. One commonly adopted method is that proposed by Douglas and Peucker (1974), which has subsequently been used and modified by many other authors.
5. **Smoothing:** The previous steps tend to create a jagged, pixelated line, producing an appearance which is rather unattractive to the eye; the vector features are then smoothed or generalized, to smooth this appearance and to remove unnecessary

vertices. This smoothing may be achieved by reducing the number of vertices, or using an averaging process (e.g. a three- or five-point moving average).

### 20.5.2.3 Raster to polygon

This is the process of vectorizing areas or regions from a raster. Unless the raster areas are all entirely discrete and have no shared boundaries, it is likely that the result will be quite complex. Commonly, therefore, this process leads to the generation of both a line and a polygon file, in addition to a point file representing the centres of the output polygons. The polygon features are constructed from groups of connected pixels whose values are the same. The process begins by determining the intersection points of the area boundaries and then follows this by generating lines at either external pixel centroids or the boundaries. A background polygon is also generated; otherwise any isolated polygons produced will float in space. Again, such vectorization procedures from raster images are usually followed by a smoothing or generalization procedure, to correct the 'pixelated' appearance of the output vectors. There are now a great many software suites available which provide a wealth of tools to perform these raster-vector conversions, some of which are proprietary and some 'shareware', such as MATLAB (MathWorks), AutoCAD (Autodesk), R2V (developed by Able Software Corp.), Illustrator, Freehand, etc.

## 20.6 Errors in Data Conversion

During vector to raster conversion both the size of the raster and the method of rasterization used have important implications for positional error and, in some cases, attribute uncertainty. The smaller the cell size the greater the precision of the resulting data. Finer raster sizes can trace the path of a line more precisely and therefore help to reduce classification error- a form of attribute error. Positional and attribute errors as a result of generalization are seen as classification error in cells along the vector polygon boundary. This is seen visually as the 'stepped' appearance of the raster version when compared with the vector original (Fig 20.8).

The conversion of data from raster to vector format is largely a question of geometric conversion; however certain topological ambiguities can occur- such as where different coded raster cells join at corner as in Fig 20.9. In this it is impossible to say, without returning to the original source data, whether the vector polygon should join. Where vector maps have been derived from raster data, conversion may result in a stepped appearance in the output map. This can be reduced, to some extent by line smoothing algorithms but these makes certain assumptions about topological relationships and detail that may not be present in the raster source data. (Heywood et al., 2010)

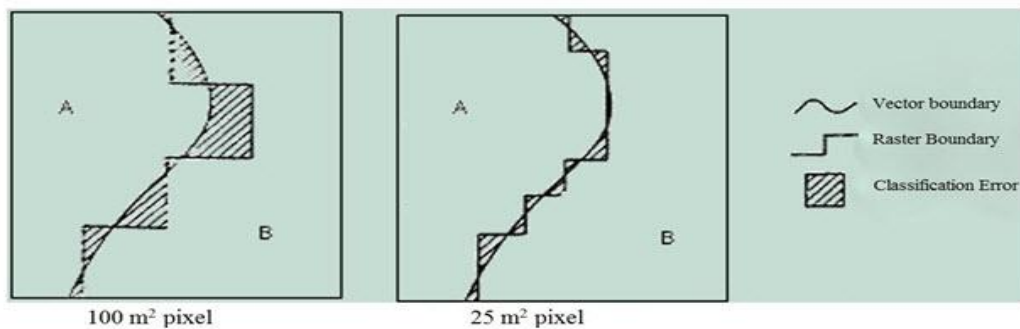


Fig. 20.8. Effect of size of raster cell (resolution) on representation of a feature. (Source: [http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_40/E\\_40\\_3.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_40/E_40_3.htm))

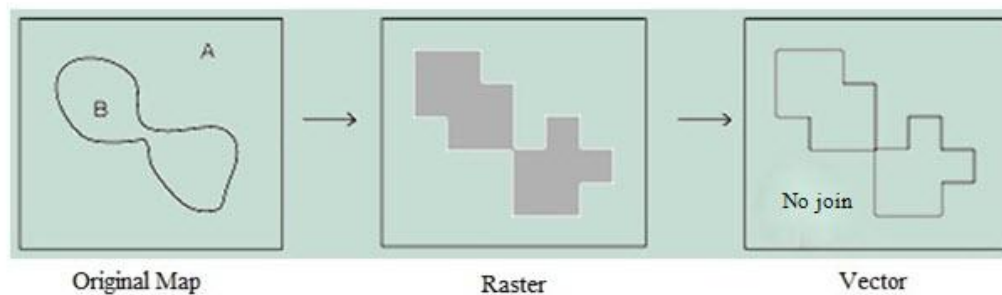


Fig. 20.9. Topological ambiguities. (Source: [http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_40/E\\_40\\_3.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_40/E_40_3.htm))

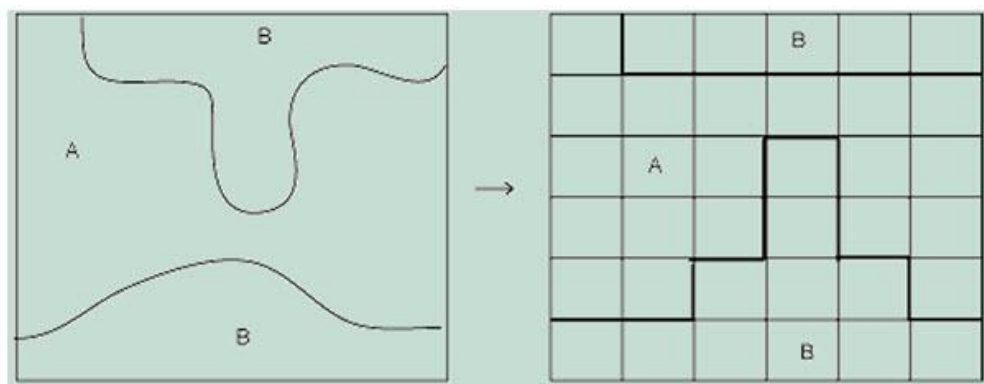
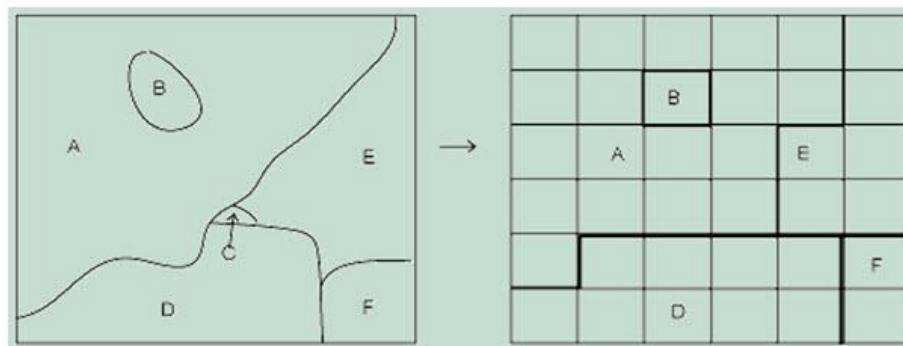


Fig. 20.10. Loss of connectivity and creation of false connectivity. (Source: [http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_40/E\\_40\\_3.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_40/E_40_3.htm))



**Fig. 20.11. Loss of information (What happened to 'C'?).** (Source: [http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE\\_40/E\\_40\\_3.htm](http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/ModernSurveyingTech/lectureE_40/E_40_3.htm))

**Keywords:** Geo-data structure, Geo-data model, Intra-format conversion, Inter-format conversion, rasterization, vectorization, Thinning, Collapsing.



## *Module 8: Map Projections and Datum*

### **Lesson 21 Map Projections**

#### **21.1 Introduction**

Coordinate systems, also known as map projections, are arbitrary designations for spatial data. Their purpose is to provide a common basis for communication about a particular place or area on the earth's surface. The most critical issue in dealing with coordinate systems is knowing what the projection is and having the correct coordinate system information associated with a dataset. There are two types of coordinate systems - geographic and projected.

A geographic coordinate system uses a three-dimensional spherical surface to define locations on the earth. It includes an angular unit of measure, a prime meridian and a datum (based on a spheroid). In a geographic coordinate system, a point is referenced by its longitude and latitude values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. The angles are often measured in degrees (or in radians). (<http://webhelp.esri.com>)

Map projections and datums have been described very comprehensively by many other authors and this topic is an extremely important one in the understanding of GIS construction and functionality, and as such cannot be ignored here. We therefore prefer to overview the main principles, concentrating on the practical applications, and refer the reader to other more detailed texts.

To make GIS function, we must be able to assign 'coordinates' of time and location in a way that is generally understood. Several terms are commonly used to denote the positioning of objects: georeference, geolocation, georegistration and geocoding. The main requirement for a georeference is that it is unique, to avoid any confusion. Hence the address, or georeference, of the Royal School of Mines, Prince Consort Road, London SW72AZ, United Kingdom, refers only to one building; no other in the world has this specific address. Georeferencing must also be persistent through time, again to avoid both confusion and expense. (Liu and Mason, 2009)

##### **21.1.1 Datums and projections**

The Earth is a 3D object, roughly oblately spherical in shape, and we need to represent that 3D shape in a 2D environment, on paper or on a computer screen. This is the reason for the existence of a multitude of map projections - since this cannot be done without distorting information, accurate measurements become potentially ambiguous. To achieve this 2D representation, two things need to be approximated: the shape of the Earth and the transformations necessary to plot a location's position on the map. (Liu and Mason, 2009)

A projected coordinate system is defined on a flat, two-dimensional surface on a flat, two-dimensional surface. Unlike a geographic coordinate system; a projected coordinate system has constant lengths, angles, and areas across the two dimensions. A projected coordinate system is always based on a geographic coordinate system that is in turn based on a sphere or spheroid.

In a projected coordinate system, locations are identified by x and y coordinates on a grid, with the origin at the center of the grid. Each position has two values that reference it to that central location. One specifies its horizontal position and the other its vertical position.

When the first map projections were devised, it was assumed, incorrectly, that the earth was flat. Later the assumption was revised, and the earth was assumed to be a perfect sphere. In the 18th century, people began to realize that the earth was not perfectly round. This was the beginning of the concept of the cartographic spheroid.

To more accurately represent locations on the earth's surface, mapmakers studied the shape of the earth (geodesy) and created the concept of the spheroid. A datum links a spheroid to a particular portion of the earth's surface. Recent datums are designed to fit the entire earth's surface well.

The most commonly used datums in North America are:

- NAD 1927 (North American Datum 1927) using the Clarke 1866 spheroid
- NAD 1983 (North American Datum 1983) using the GRS 1980 spheroid
- WGS 1984 (World Geodetic System 1984) using the WGS 1984 spheroid

Newer spheroids are developed from satellite measurements and are more accurate than those developed in the 19th and early 20th centuries. (<http://webhelp.esri.com>)

You will find that the terms "geographic coordinate system" and "datum" are used interchangeably (<http://webhelp.esri.com>). A geographic coordinate system (GCS) defines locations on the earth using a three-dimensional spherical surface. A GCS is often incorrectly called a datum, but a datum is only one part of a GCS. A GCS includes an angular unit of measure, a prime meridian, and a datum (based on a spheroid). A feature is referenced by its longitude and latitude values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. (<http://php.auburn.edu/academic/classes/fory/7470/lab08/projections.pdf>)

The coordinates for a location will change depending on the datum and spheroid on which those coordinates are based, even if using the same map projection and projection parameters. (<http://webhelp.esri.com>).

For example, the geographic coordinates below are for the city of Bellingham, Washington, using 3 different datum:

**Table 21.1 Geographic coordinates of Bellingham city**

Datum	Longitude	Latitude
NAD 1927	-122.466903686523	48.7440490722656
NAD 1983	-122.46818353793	48.7438798543649
WGS 1984	-122.46818353793	48.7438798534299

(Source: webhelp.esri)

## 21.2 Describing and Measuring the Earth

Geodesy (shape of the earth) developed for practical needs: It was always necessary to establish property lines, mainly for tax purposes. Roads and buildings need advance planning. The location of a nation's resources must be determined and recorded. To get from one place to another, we want to know which way and how far to go. We can outline these needs as follows:

CADASTRAL - property lines, taxes

CIVIL ENGINEERING - roads, buildings

RESOURCES - What? Where? How much?

NAVIGATION - Which way? How far?

For such needs it was assumed that the earth is flat – as long as we stay within our immediate vicinity. But for long distances and large areas, this simple assumption does not work.

**It contradicts the following observations like:**

The ancient seafarers had noticed that a boat coming in from the horizon is not in full view all at once; its superstructure is visible long before the hull is seen.

Traveling north at night, the north pole star appears to get higher in the sky, as measured from the horizon up and many more such deviations are there. So from all these observations scientists jumped into a conclusion that the earth is not flat rather curved or sphere.

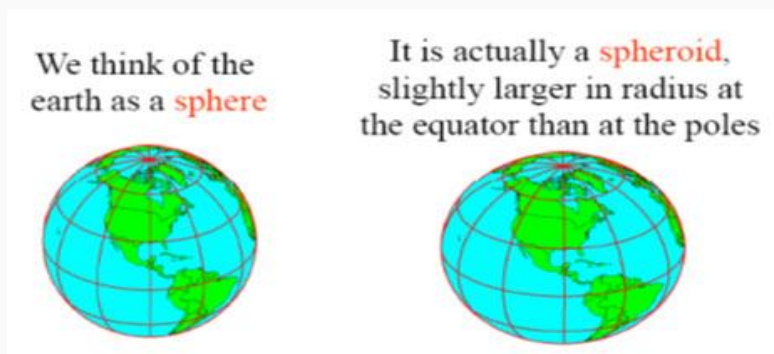
(NOAA (National Oceanic and Atmospheric Administration), September 1997, NOAA Reprint of Basic Geodesy, Rockville Md. US Department of Commerce, United States of America.)

They describe the earth with many types of coordinate systems like:

- (1) Global Cartesian coordinates (x, y, z) for the whole earth
- (2) Geographic coordinates ( $\phi$ ,  $\lambda$ , z)
- (3) Projected coordinates (x, y, z) on a local area of the earth's surface

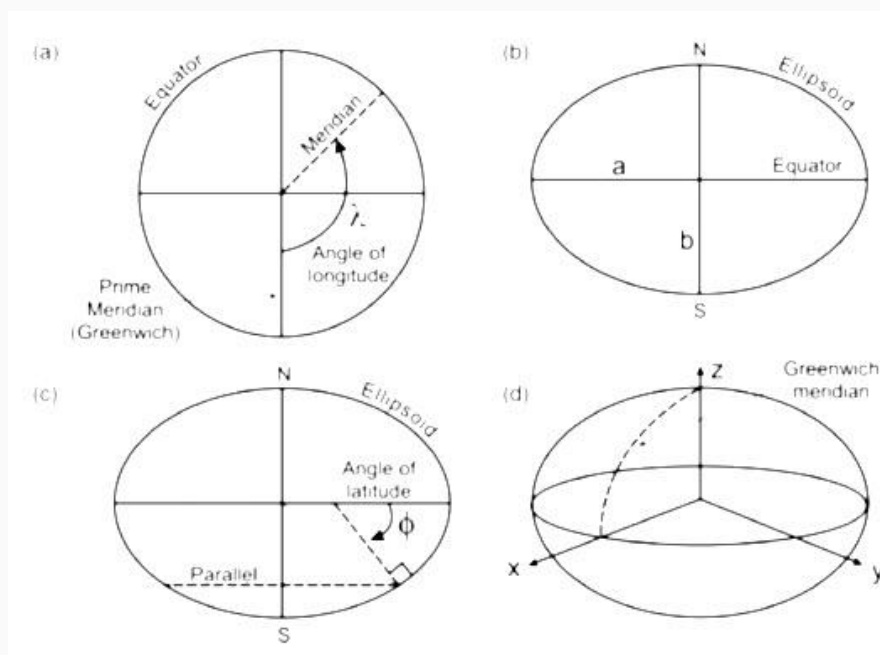
The z-coordinate in (1) and (3) is defined geometrically; in (2) the z-coordinate is defined gravitationally. (<http://webhelp.esri.com>).

The shape of the earth apart from the assumed sphere, shown below



**Fig. 21.1.** Shows the shape of the earth. (Source: <http://webhelp.esri.com>).

The system of latitude and longitude is considered the most comprehensive and globally constant method of description and is often referred to as the geographic system of coordinates, or geodetic system, and it is the root for all other systems. It is based on the Earth's rotation about its centre of mass.



**Fig. 21.2.** Schematic representation of the Earth, looking (a) down the pole, perpendicular to the equator, (b) and (c) perpendicular to the pole from the equator, and (d) obliquely at the Earth. This illustrates the relationship between longitude and the meridians (a), between the equator and major and minor semi-axes (b), and between latitude and parallels (c), and the locations of the x, y and z axes forming Cartesian coordinates (d). (Source: Liu and Mason, 2009)

To define the centre of mass and so latitude and longitude (see Fig. 21.2), we must first define the Earth's axis of rotation and the plane through the centre of mass perpendicular to the axis



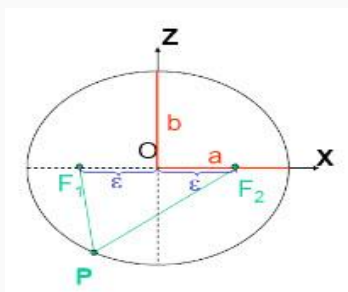
(the equator). Slices parallel to the axis but perpendicular to the plane of the equator are lines of constant longitude; these pass through the centre of mass and are sometimes also referred to as great circles. The slice through Greenwich defines zero degrees longitude and the angle between it and any other slice defines the angle of longitude, so that longitude then goes from 180° west to 180° east of Greenwich. A line of constant longitude is also called a meridian. Perpendicular to a meridian, a slice perpendicular to the axis and passing through the Earth but not through its centre is called a parallel, also referred to as a small circle, except for the equator (which is a great circle). We also need to describe the shape of the Earth, and the best approximation of this is the ellipsoid of rotation or spheroid. An ellipsoid is a type of quadric surface and is the 3D equivalent of an ellipse. It is defined, using x, y, z Cartesian coordinates, by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad \dots \quad (21.1)$$

The Earth is not spherical but oblate and so the difference between the ellipsoid or spheroid and a perfect sphere is defined by its flattening (f), or its reduction in the shorter minor axis relative to the major axis. Eccentricity (e) is a further phenomenon which describes how the shape of an ellipsoid deviates from a sphere (the eccentricity of a circle being zero) (Liu and Mason, 2009). Flattening and eccentricity then have the following relationships:

**An ellipse is defined by**

Z Focal length =  $\epsilon$



**Fig. 21.3. Ellipse.** (Source: <http://webhelp.esri.com>)

Distance (F1, P, F2) is constant for all points on ellipse

When  $\epsilon = 0$ , ellipse = circle

**For the earth: at point P**

Major axis,  $a = 6378$  km

Minor axis,  $b = 6357$  km

$$\text{Flattening ratio, } f = \frac{(a-b)}{a} \quad (21.2)$$

~ 1/300

(Source: <http://webhelp.esri.com>)

$$f = \frac{(a - b)}{a} \text{ and } e^2 = \frac{a^2 - b^2}{a^2} \text{ or } e^2 = 2f - f^2 \quad \dots \dots \dots (21.3)$$

Where a and b are the lengths of the major and minor axes respectively (usually referred to as semi-axes or half lengths of the axes). The actual flattening for the Earth's case is about 1 part in 300. (Liu and Mason, 2009)

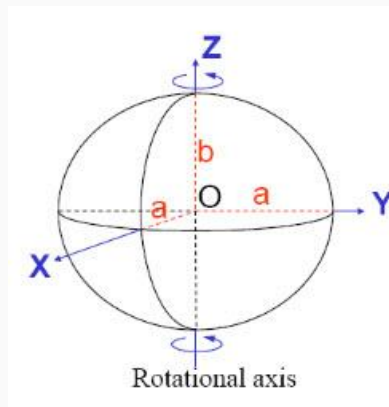
**Table 21.2 Standard ellipse**

Ellipsoid	Major axis	Minor axis	Flattening ratio
Clark (1886)	6,378,206	6,356,584	1/294.98
GRS (80)	6,378,137	6,356,752	1/298.57

(Source: [webhelp.esri.com](http://webhelp.esri.com))

### 21.2.1 Ellipsoid or Spheroid

It is formed by rotating ellipse along an axis, as shown in Fig. 21.4



**Fig .21.4. Ellipsoid or Spheroid formed by rotating ellipse around an axis. (Source: <http://webhelp.esri.com>)**

Some of the first ellipsoids to be established were not particularly accurate and were not actually centered on the Earth's centre of mass. Fortunately, and rather ironically, the Cold War, the nuclear arms race and the need to target intercontinental missiles helped to drive the development of an international standard ellipsoid. Out of all the datums defined in former pages, the World Geodetic System of 1984 (or WGS84) is now accepted as this standard although many others are in use.

### 21.2.2 Defining Latitude of earth

Latitude can now be defined as the angle between the equator and a line perpendicular to the ellipsoid, which ranges from  $90^0$  north or south of the equator (see Fig. 21.5). Latitude is commonly given the Greek symbol phi ( $\phi$ ). ) (Liu and Mason, 2009).

Or

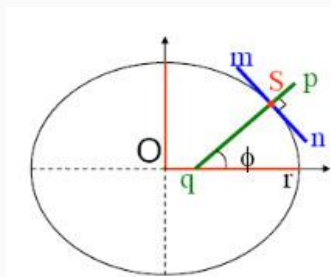


Fig. 21.5. Showing latitude of earth. (Source: <http://webhelp.esri.com>)

It can also be defined in this way also:

- Take a point S on the surface of the ellipsoid and define there the tangent plane, mn.
- Define the line pq through S and normal to the tangent plane
- Angle pqr which this line makes with the equatorial plane is the latitude  $\lambda$ , of point S.

A line of constant latitude is known as a parallel. Parallels never meet since they are parallel to one another, whereas meridians (lines of longitude) converge at the poles. (Source: <http://webhelp.esri.com>)

### 21.2.3 Defining latitude of earth, $\lambda$

$\lambda$  = the angle between a cutting plane on the prime meridian and the cutting plane on the meridian through the point

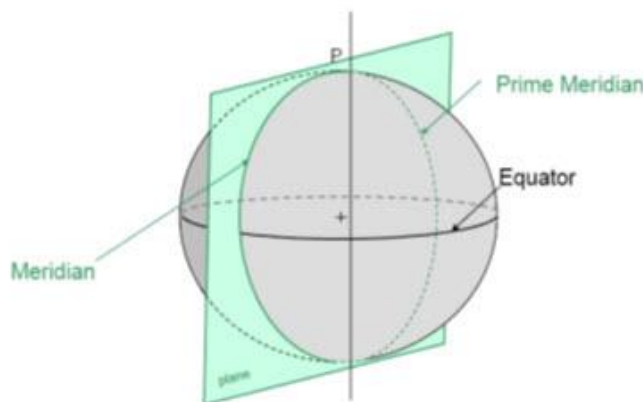
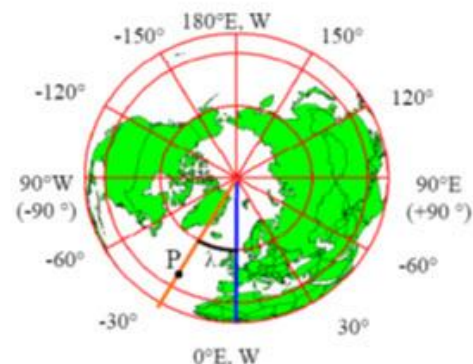


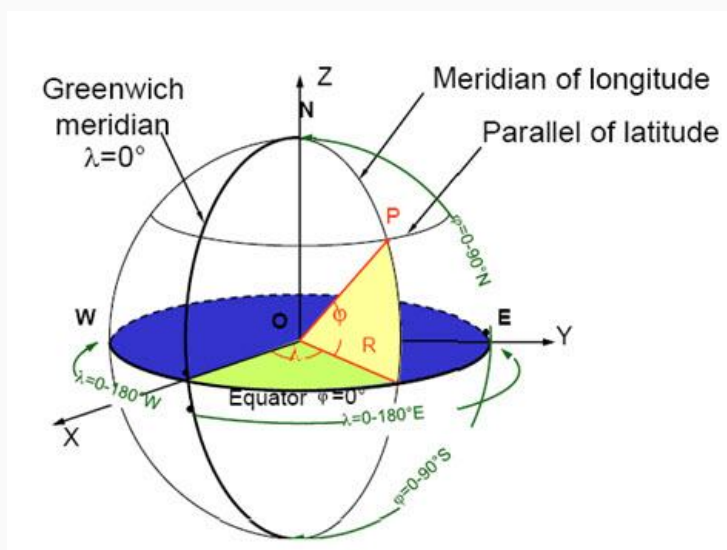
Fig. 21.6. Cutting plane of prime meridian. (Source: <http://webhelp.esri.com>)



**Fig .21.7. Showing the latitudes of earth. (Source: <http://webhelp.esri.com>)**

Longitude is more complex and only east–west measurements made at the equator are true. Away from the equator, where the lines of latitude decrease in length, measures are increasingly shortened, by approximately the cosine of the latitude. This means that at 30° north (or south), shortening is about 0.866, 0.707 at 45° and 0.5 at 60°. At 60° north or south, 1° of longitude will represent 55 km ground distance. (Liu and Mason, 2009).

#### 21.2.4 Combined latitude and longitude on a sphere



**Fig.21.8. Showing latitude and longitude on a sphere.**

(Source: <http://webhelp.esri.com>)

#### 21.2.5 Length of meridians and parallels.

(Lat, Long) = ( $\phi$ ,  $\lambda$ ), here considering the shape of the earth as spheroid and using geographic coordinates length of meridians and parallels are defined as:

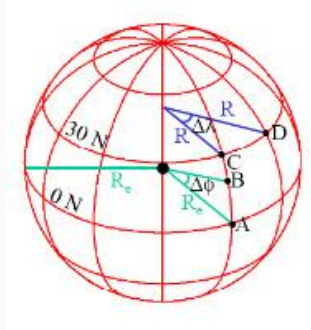


Fig.21.9. Showing the parallels and meridian length.

(Source: <http://webhelp.esri.com>)

Length on a Meridian:

$$AB = R_e \Delta\phi \quad (\text{same for all latitudes}) \quad (21.4)$$

Length on a Parallel:

$$CD = R \Delta\lambda = R_e \Delta\lambda \cos\phi \quad (\text{varies with latitude}) \quad (21.5)$$

**Example:** What is the length of 2° increment on a meridian and on a parallel at 30°N, 90°W? Radius of Earth = 6370km.

**Solution:**

First 2° angle has to be converted into radians,  $\pi$  radians = 180°

So 1° = = = 0.0175 radians

Therefore, 2° = 0.035 radians

For the meridian,  $\Delta L = R_e \Delta\phi = 6370 * (0.035) = 222.96$

For the parallel,  $\Delta L = R_e \Delta\lambda \cos\phi = 6370 * 0.035 * \cos 30^\circ$

$$= 193.08 \text{ km.}$$

#### 21.2.5 Curved Earth Distance (From A to B)

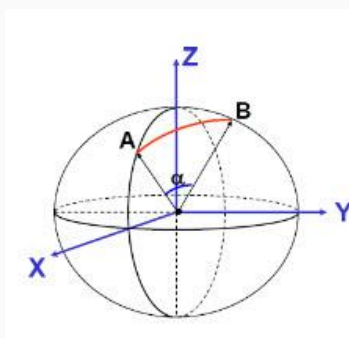


Fig.21.10 Curved surface of earth. (Source: <http://webhelp.esri.com>)

Shortest distance is along a “Great Circle” A “Great Circle” is the intersection of a sphere with a plane going through its AB center.

1. Spherical coordinates converted to Cartesian coordinates.
2. Vector dot product used to calculate angle  $\alpha$  from latitude and longitude
3. Great circle distance is  $R\alpha$ , where  $R=6370 \text{ km}^2$

$$\text{Therefore, } R\alpha = R \cos^{-1} (\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2)) \quad (21.6)$$

Where latitude & longitude are  $\lambda$  and  $\phi$ .

### 21.2.6 Measuring height: the geoid

The true shape of the Earth forms a surface which is perpendicular to the direction of gravity, and is described as an equipotential surface, in which there are fluctuations and irregularities according to variations in the density of the crust and mantle beneath. The spheroid or ellipsoid can therefore be thought of as a reasonable representation of the shape of the Earth but not the true shape; this we refer to as the geoid and it is defined as an ‘equipotential surface which most closely resembles mean sea level.

Or

The geoid is a surface of constant potential energy that coincides with mean sea level over the oceans. This definition is not very rigorous. First, mean sea level is not quite a surface of constant potential due to dynamic processes within the ocean. Second, the actual equipotential surface under continents is warped by the gravitational attraction of the overlying mass. But geodesists define the geoid as though that mass were always underneath the geoid instead of above it. The main function of the geoid in geodesy is to serve as a reference surface for leveling. The elevation measured by leveling is relative to the geoid.

([http://www.ceri.memphis.edu/people/smalley/ESCI7355/Li\\_G\\_Tut.pdf](http://www.ceri.memphis.edu/people/smalley/ESCI7355/Li_G_Tut.pdf))

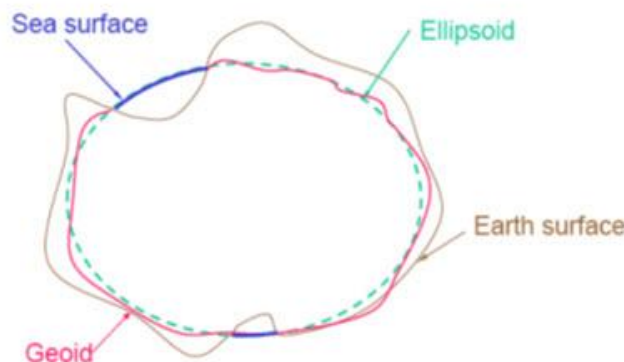
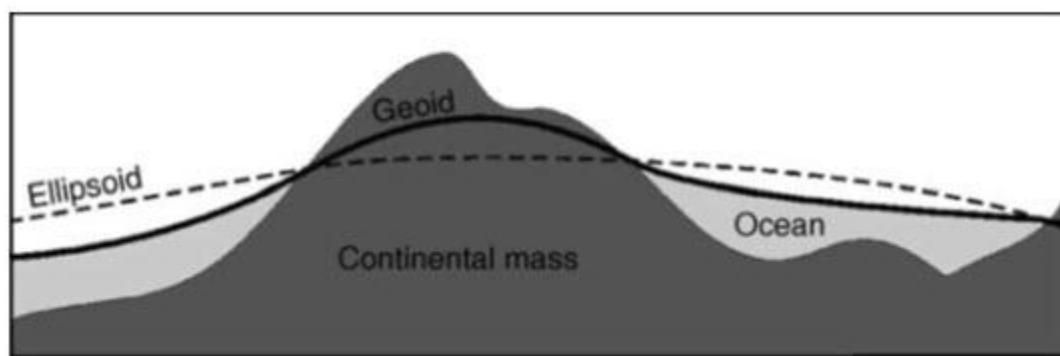


Fig. 21.9 Representation of earth. (Source: <http://webhelp.esri.com>)

Mean sea level is used in this context since it refers to everywhere and is a surface perpendicular to gravity. In general, differences between mean sea level and the geoid (referred to as separation) are greatest where undulations in the terrain surface are of the greatest magnitude but are generally less than 1 m.

The significance of the geoid's variability is that it leads to different definitions of height from one place to another, since mean sea level also varies. Different countries may define slightly different equipotential surfaces as their reference. We should therefore take some care to distinguish between heights above geoid or spheroid (Fig. 21.8).



**Fig. 21.10. Heights above the spheroid and geoid. (Source: Liu and Mason, 2009)**

Fortunately the differences are small so that only highly precise engineering applications should be affected by them. For reference, orthometric heights and spheroidal heights are those defined with respect to the geoid and spheroid respectively. The variation in height from the geoid gives us topography. (Liu and Mason, 2009)

## Lesson 22 Coordinate and Datum System

### 22.1 Introduction

The oldest global coordinate system is the Latitude/ Longitude system (also referred to as Geographic coordinates). It is the primary system used for determining positions in surveying and navigation. A grid of east-west latitude lines (parallels) and north-south longitude lines (meridians) represent angles relative to standard reference planes. Latitude is measured from 0 to 90 degrees north and south of the equator. Longitude values range from 0 to 180 degrees east or west of the Prime Meridian, which by international convention passes through the Royal Observatory at Greenwich, England. Because the Latitude/Longitude system references locations to a spheroid rather than to a plane, it is not associated with a map projection.

Use of latitude/longitude coordinates can complicate data display and spatial analysis. One degree of latitude represents the same horizontal distance anywhere on the Earth's surface. However, because lines of longitude are farthest apart at the equator and converge to single points at the poles, the horizontal distance equivalent to one degree of longitude varies with latitude. If you have data of regional or smaller extent with Geographic coordinates, you will achieve better results by warping or resampling the data to a planar coordinate system. A large-scale map represents a small portion of the Earth's surface as a plane using a rectilinear grid coordinate system to designate location coordinates.

A planar Cartesian coordinate system has an origin set by the intersection of two perpendicular coordinate axes and tied to a known location. The coordinate axes are normally oriented so that the x-axis is east-west and the y-axis is north-south. Grid coordinates are customarily referred to as easting (distance from the north-south axis) and northing (distance from the east-west axis). The definition of a Cartesian coordinate system also includes the units used to measure distances relative to these axes (such as meters or feet). For very small areas (such as a building construction site) the curvature of Earth's surface is so slight that ground locations can be referenced directly to an arbitrary planar Cartesian coordinate system without introducing significant positional errors. For areas larger than a few square kilometers, however, the difference between a planar and curving surface becomes important when relating ground and map locations. A map projection must then be selected to relate surface and map coordinates to reduce undesirable distortions in the map. Cartesian coordinate system is related to the Earth via a map projection is termed a projected coordinate system. A number of projected coordinate systems have been set up to represent large areas (states, countries, or larger areas) by subdividing the coverage area into geographic zones, each of which has its own origin. To minimize variations in scale associated with the map projection, one of the coordinate axes (sometimes both) typically bisects the zone. In that case, in order to force all locations within the zone to have positive coordinate values, a large number is added to one or both coordinates of the origin, termed false easting and false northing. This procedure moves the (0, 0) position of the coordinate



system outside the zone to create a false origin, as illustrated in Fig. 22.1. (<http://www.microimages.com/documentation/Tutorials/project.pdf>)

Many different generic types of coordinate systems can be defined and the calculations necessary to move between them may sometimes be rather complex. In order of increasing complexity they can be thought of as follows. Spherical coordinates are formed using the simplest approximation of a spherical Earth, where latitude is the angle north or south of the equatorial plane, longitude is the angle east or west of the prime meridian (Greenwich) and height is measured above or below the surface of the sphere.

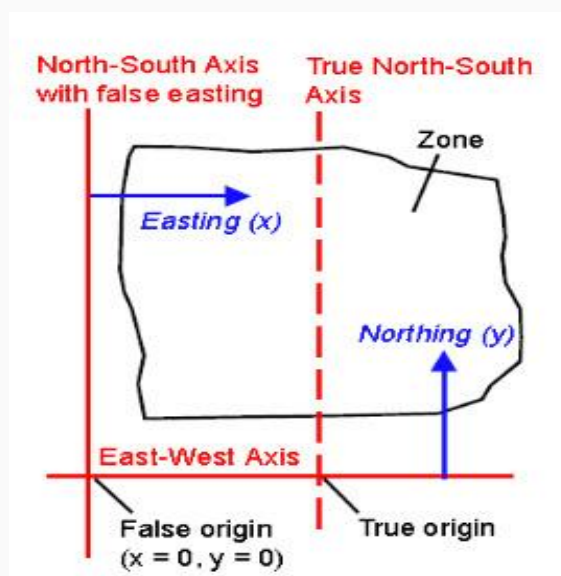


Fig. 22.1. Planar coordinate system with false easting.

(Source: Smith, 2011)

If high accuracy is not important, then this simple model may be sufficient for your purposes. Spheroidal coordinates are formed using a better approximation based on an ellipsoid or spheroid, with coordinates of true latitude, longitude and height; this gives us the system of geodetic coordinates. Cartesian coordinates involve values of  $x$ ,  $y$  and  $z$ , and are defined with their origin at the centre of a spheroid. The  $x$  and  $y$  axes lie in the equatorial plane, with  $x$  aligned with the Greenwich meridian, and  $z$  aligned with the polar axis. Projection coordinates are then defined using a simple set of  $x$  and  $y$  axes, where the curved surface of the Earth is transformed onto a plane, the process of which causes distortions. Polar coordinates, generically, are those which are defined by distance and angle, with distance usually denoted  $r$  and angle  $u$ . Planar coordinates refer to the representation of positions, as identified from polar coordinate positions, on a plane within which a set of orthogonal  $x$ ,  $y$  axes is defined. The conversion between these polar and planar coordinates, for any particular datum, is relatively straightforward and the relationship between them is illustrated in Fig. 22.2. The following expression can be used to derive the distance ( $d$ ) between two points  $a$  and  $b$  on an assumed spherical Earth:

$$d(a,b) = R \arccos[\sin\theta_A \sin\theta_B + \cos\theta_A \cos\theta_B \cos(\lambda_A - \lambda_B)] \quad (22.1)$$

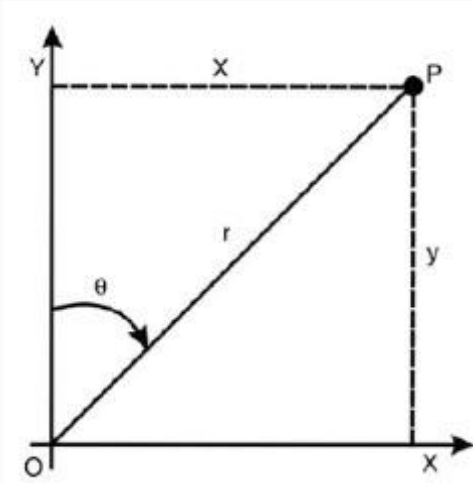
where  $R$  is the radius of the Earth,  $A$  and  $B$  denote the positions of points  $a$  and  $b$  on the sphere,  $\lambda$  is the longitude and  $\phi$  the latitude. Generically, the  $x, y$  (planar) positions of the two points can be derived from the polar coordinates as:

$$x = r \sin\theta, y = r \cos\theta \quad (22.2)$$

$$r = \sqrt{x^2 + y^2}, \theta = \arctan\left(\frac{y}{x}\right) \quad (22.3)$$

where  $u$  is measured clockwise from north. The Pythagorean distance between two points ( $a$  and  $b$ ) can then be found by the following, where the two points are located at  $(x_a, y_a)$  and  $(x_b, y_b)$ :

$$d(a, b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (22.4)$$



**Fig. 22.2. The relationship between polar and planar co-ordinates in a single datum.**  
(Source: Jian and Philippa, 2009)

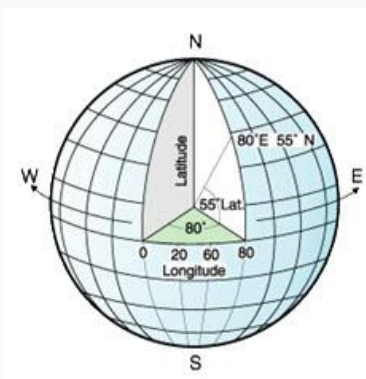
## 22.1 Geographic Coordinate System

A geographic coordinate system is a reference system that uses a three-dimensional spherical surface to determine locations on the earth. Any location on earth can be referenced by a point with longitude and latitude coordinates. The values for the points can have the following units of measurement:

- Linear units when the geographic coordinate system has a spatial reference system identifier (SRID) that DB2<sup>(R)</sup> Geodetic Extender recognizes.
- Any of the following units when the geographic coordinate system has an SRID that DB2 Geodetic Extender does not recognize.
  - Decimal degrees

- Decimal minutes
- Decimal seconds
- Gradians
- Radians

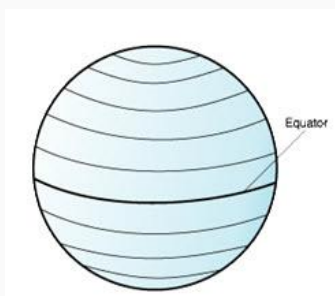
For the range of values for these units, refer to Supported coordinate systems. For example, Fig. 22.3 shows a geographic coordinate system where a location is represented by the coordinates longitude 80 degree East and latitude 55 degree North.



**Fig. 22.3. A geographic coordinate system.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

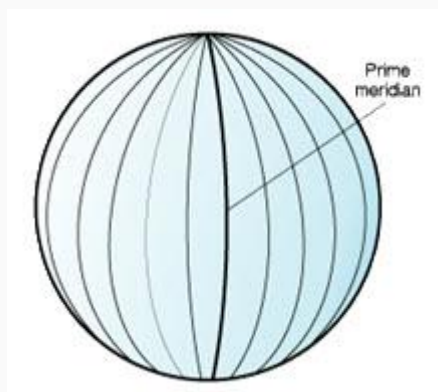
The lines that run east and west each have a constant latitude value and are called parallels. They are equidistant and parallel to one another, and form concentric circles around the earth. The equator is the largest circle and divides the earth in half. It is equal in distance from each of the poles, and the value of this latitude line is zero. Locations north of the equator have positive latitudes that range from 0 to +90 degrees, while locations south of the equator have negative latitudes that range from 0 to -90 degrees. Fig. 22.4 illustrates latitude lines. The lines that run north and south each have a constant longitude value and are called meridians. They form circles of the same size around the earth, and intersect at the poles. The prime meridian is the line of longitude that defines the origin (zero degrees) for longitude coordinates. One of the most commonly used prime meridian locations is the line that passes through Greenwich, England.



**Fig. 22.4. Latitude lines.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

However, other longitude lines, such as those that pass through Bern, Bogota, and Paris, have also been used as the prime meridian. Locations east of the prime meridian up to its antipodal meridian (the continuation of the prime meridian on the other side of the globe) have positive longitudes ranging from 0 to +180 degrees. Locations west of the prime meridian have negative longitudes ranging from 0 to -180 degrees. Fig. 22.5 illustrates longitude lines.

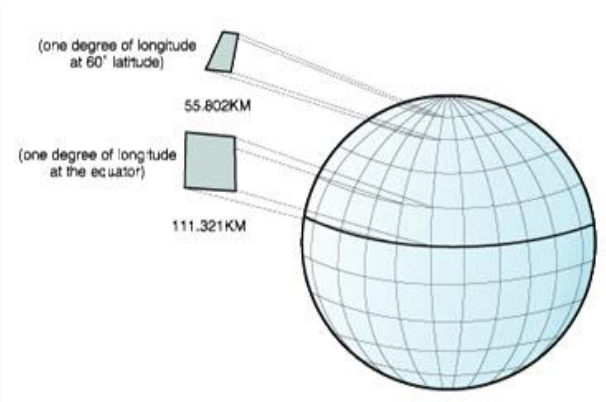


**Fig. 22.5. Longitude lines.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

The latitude and longitude lines can cover the globe to form a grid, called a graticule. The point of origin of the graticule is (0, 0), where the equator and the prime meridian intersect. The equator is the only place on the graticule where the linear distance corresponding to one degree latitude is approximately equal the distance corresponding to one degree longitude. Because the longitude lines converge at the poles, the distance between two meridians is different at every parallel. Therefore, as you move closer to the poles, the distance corresponding to one degree latitude will be much greater than that corresponding to one degree longitude.

It is also difficult to determine the lengths of the latitude lines using the graticule. The latitude lines are concentric circles that become smaller near the poles. They form a single point at the poles where the meridians begin. At the equator, one degree of longitude is approximately 111.321 kilometers, while at 60 degrees of latitude, one degree of longitude is only 55.802 km (this approximation is based on the Clarke 1866 spheroid). Therefore, because there is no uniform length of degrees of latitude and longitude, the distance between points cannot be measured accurately by using angular units of measure. Fig. 22.6 shows the different dimensions between locations on the graticule.



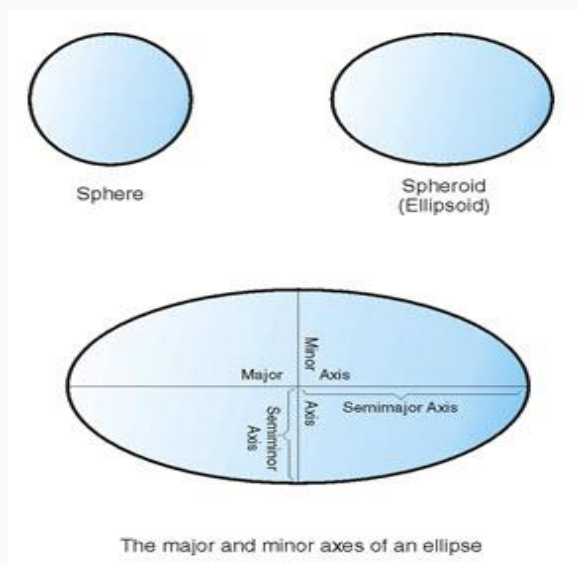
**Fig. 22.6. Different dimensions between locations on the graticule.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

A coordinate system can be defined by either a sphere or a spheroid approximation of the earth's shape. Because the earth is not perfectly round, a spheroid can help maintain accuracy for a map, depending on the location on the earth. A spheroid is an ellipsoid that is based on an ellipse, whereas a sphere is based on a circle.

The shape of the ellipse is determined by two radii. The longer radius is called the semimajor axis, and the shorter radius is called the semiminor axis. An ellipsoid is a three-dimensional shape formed by rotating an ellipse around one of its axes. Fig. 22.7 shows the sphere and spheroid approximations of the earth and the major and minor axes of an ellipse.

(<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>).

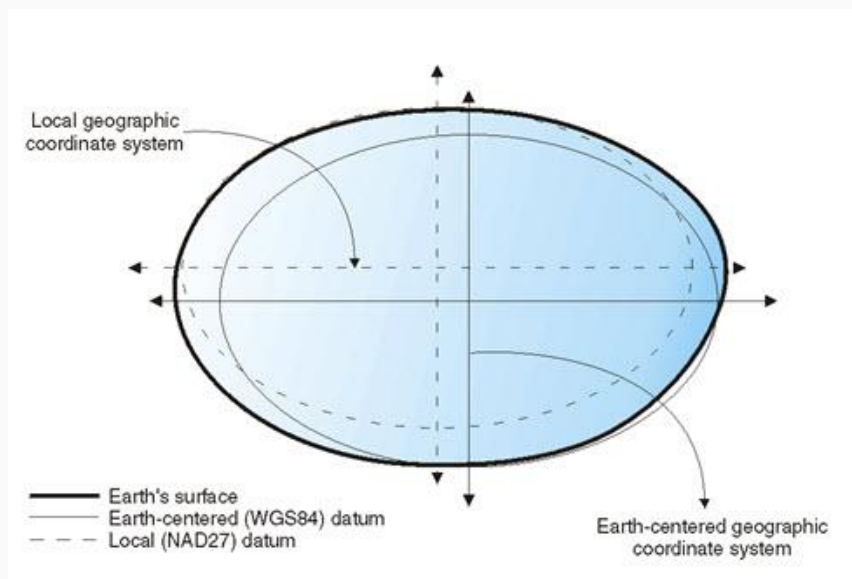


**Fig. 22.7. Sphere and spheroid approximations.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

## 22.2 Datum

A geodetic datum is a mathematical approximation of the Earth's 3D surface and a reference from which other measurements are made. A datum is a set of values that defines the position of the spheroid relative to the center of the earth. The datum provides a frame of reference for measuring locations and defines the origin and orientation of latitude and longitude lines. Some datums are global and intend to provide good average accuracy around the world. A local datum aligns its spheroid to closely fit the earth's surface in a particular area. Therefore, the coordinate system's measurements are not be accurate if they are used with an area other than the one that they were designed. Fig. 22.8 shows how different datums align with the earth's surface. The local datum, NAD27, more closely aligns with Earth's surface than the Earth-centered datum, WGS84, at this particular location.



**Fig. 22.8. Datum alignments.**

(Source:<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=%2Fcom.ibm.db2.udb.spatial.doc%2Fcsb3022a.html>; 3 December 2012)

Every spheroid has a major axis and a minor axis, with the major axis being the longer of the two but is not in itself a datum. The missing information is a description of how and where the shape deviates from the Earth's actual surface. This is provided by the definition of a tie point, which is a known position on the Earth's surface (or its interior, since the Earth's centre of mass could be used), and its corresponding location on or within the ellipsoid.

Complications arise because datums may be global, regional or local, so that each is only accurate for a limited set of conditions. For a global datum, the tie point may well be the centre of mass of the Earth, meaning that the ellipsoid forms the best general approximation of the Earth's shape, and that at any specific positions and accuracies may be quite poor. Such generalizations would be acceptable for datasets which are of very large or global extent. In contrast, a local datum, which uses a specific tie point somewhere on the surface, near the area of interest, would be used for a 'local' projector data set. Within this area, the deviation of the ellipsoid from the actual surface will be minimal but at some distance from it may be considerable. This is the reason behind the development of the great number of datums and



projections worldwide. In practice, we choose a datum which is appropriate to our needs, according to the size and location of the area we are working with, to provide us with optimum measurement accurac. It is worth noting that in many cases there may be several different versions of datum and ellipsoid under the same name, depending on when, where, by whom and for which purpose they were developed. The differences between them may seem insignificant at first glance but, in terms of calculated ground distances, could produce very significant differences between measurements.

### 22.3 Geometric Distortions and Projection Models

Since paper maps and geospatial databases are flat representations of data located on a curved surface, the map projection is an accepted means for fitting all or part of that curved surface to the flat surface or plane. This projection cannot be made without distortion of shape, area, distance, direction or scale. We would ideally like to preserve all these characteristics but we cannot, so we must choose which of them should be represented accurately at the expense of others, or whether to compromise on several characteristics. Measuring exact distances from any map features in any direction would require constant scale throughout the map, but no map projection can achieve this. In most projections scale remains constant along one or more standard lines, and careful positioning of these lines can minimize scale variations elsewhere in the map. Specialized equidistant map projections maintain constant scale in all directions from one or two standard points. In many types of spatial analysis it is important to compare the areas of different features. Such comparisons require that surface features with equal areas are represented by the same map area regardless of where they occur. An equal-area map projection conserves area but distorts the shapes of features. A map projection is conformal if the shapes of small surface features are shown without distortion. This property is the result of correctly representing local angles around each point, and maintaining constant local scale in all directions. Conformality is a local property; while small features are shown correctly, large shapes must be distorted. A map projection cannot be both conformal and equal-area. No map projection can represent all great circle directions as straight lines. Azimuthal projections show all great circles passing through the projection center as straight lines. There are probably 20 or 30 different types of map projections in common usage. These have been constructed to preserve one or other characteristics of geometry, as follows:

- 1. Area:** Many map projections try to preserve area, so that the projected region covers exactly the same area of the Earth's surface no matter where it is placed on the map. To achieve this the map must distort scale, angles and shape.
- 2. Shape:** There are two groups of projections which have either: (a) a conformal property where the angles and the shapes of small features are preserved, and the scales in x and y are always equal (although large shapes will be distorted); or (b) an equal area property where the areas measured on the map are always in the same proportion to the areas measured on the Earth's surface but their shapes may be distorted.
- 3. Scale:** No map projection shows scale correctly everywhere on the map, but for many projections there are one or more lines on the map where scale is correct.

**4. Distance:** Some projections preserve neither angular nor area relationships but distances in certain directions are preserved.

**5. Angle:** Although conformal projections preserve local angles, one class of projections (called azimuthal projections) preserve the easting and northing pair, so that angle and direction are preserved. Scale factor ( $k$ ) is useful to quantify the amount of distortion caused by projection and is defined by the following:

$$k = \frac{\text{Projected distance}}{\text{Distance on the sphere}} \quad (22.5)$$

This relationship will be different at every point on the map and in many cases will be different in each direction. It only applies to short distances. The ideal scale factor is 1, which represents no distortion at all. Most scale factors approach but are less than 1. Any projection can achieve one or other of these properties but none can preserve all, and the distortions that occur in each case are illustrated schematically in Fig. 22.9. Once the property to be preserved has been decided, the next step is to transform the information using a 'projectable' or 'flattenable' surface. The transformation or projection is achieved using planar, cylindrical or cone-shaped surfaces that touch the Earth in one of a few ways; these form the basis for the three main groups of projection. Where the surface touches the Earth at a point (for a plane), along a great circle (for a cylinder) or at a parallel (for a cone), projections of the tangent type are formed. Where the surface cuts the earth, rather than just touching it, between two parallels, a secant type of projection is formed. The conic is actually a general form, with azimuthal and cylindrical forms being special cases of the conic type. In the planar type (such as a stereographic projection) where the surface is taken to be the tangent to one of the poles, the following relationship can be used to derive polar positions (all the equations given here assume a spherical Earth for simplicity):

$$\theta = \lambda, r = 2 \tan\left(\frac{x}{2}\right) \quad (22.6)$$

where  $x$  represents the colatitude ( $x = 90^\circ - \theta$ ); the resultant polar coordinates can then be converted to planar coordinates using Equation (22.3). Of course, the plane could be a tangent to the Earth at any point, not just at one of the poles.

For cylindrical projections, the axis of the cylinder may pass through the poles, for example, so that it touches the sphere at the equator (as in the Mercator). In this case, positions may be derived as:

$$x = \lambda, y = \log \tan\left(45^\circ + \frac{\phi}{2}\right) \quad (22.7)$$

For conic projections of the tangent type, the following can be used to derive positions, assuming one standard parallel at a colatitude of  $X_0$ :



$$r = \theta = \lambda \cos(\chi_0) \text{ and } \tan(\chi_0) + \tan(\chi - \chi_0) \quad (22.8)$$

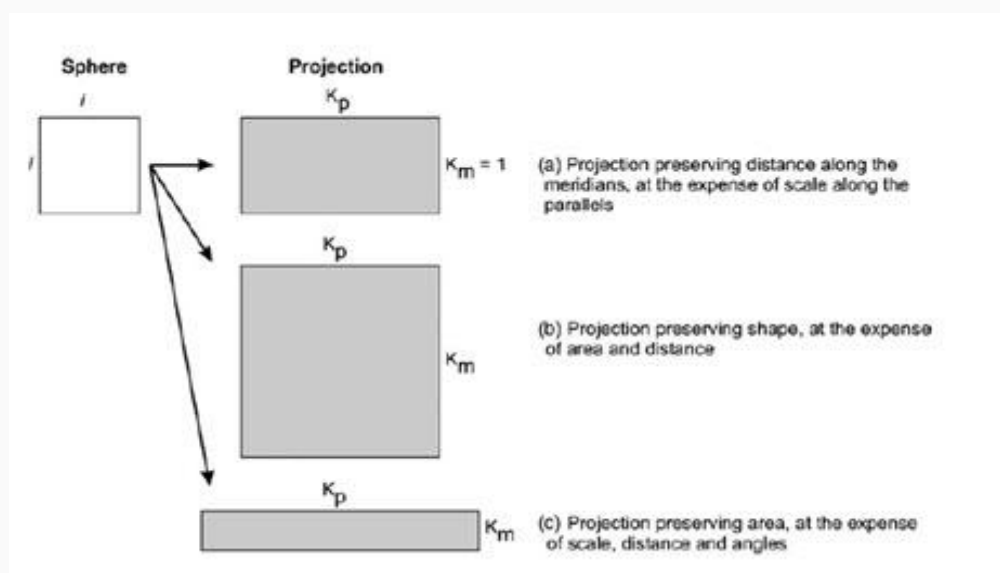


Fig. 22.9. Schematic illustration of the projection effects on a unit square of side length  $i$ , where  $K$  represents the scale along each projected side, and subscripts  $m$  and  $p$  represent meridian and parallel.

(Source: Jian & Philippa, 2009).

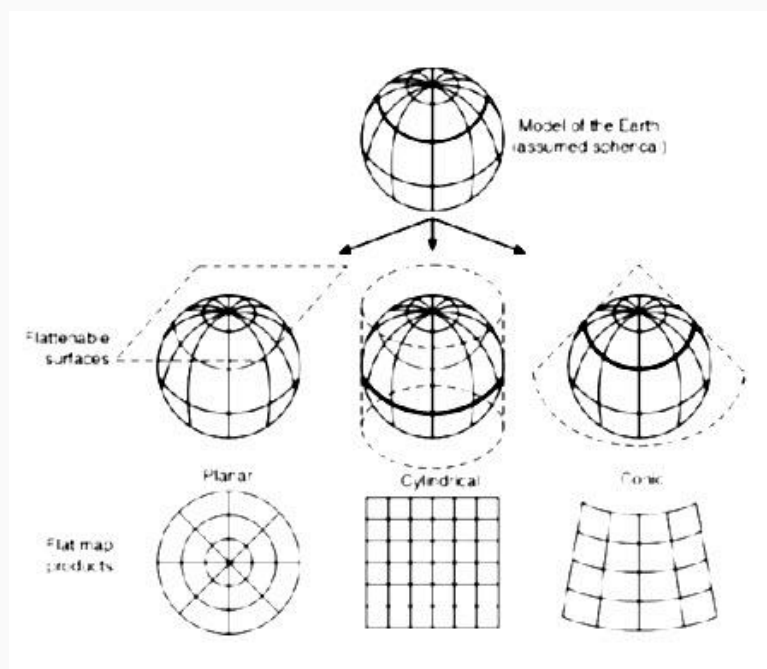


Fig. 22.10. Three main types of projections which are based on tangent case: planar (left), cylindrical (centre) and conical (right).

(Source: Jian & Philippa, 2009).

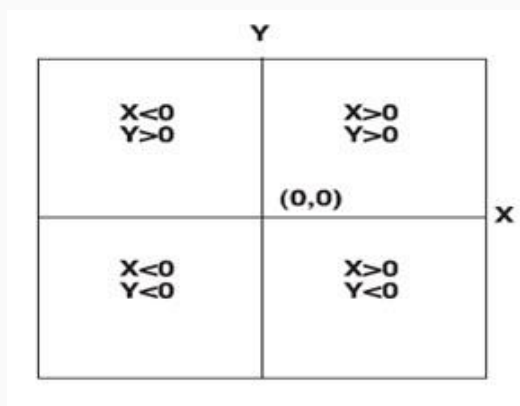
## Lesson 23 Projection System

### 23.1 Projection System

A projected coordinate system is defined on a flat, two-dimensional surface. Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. A projected coordinate system is always based on a geographic coordinate system that is based on a sphere or spheroid.

In a projected coordinate system, locations are identified by  $x$ ,  $y$  coordinates on a grid, with the origin at the center of the grid. Each position has two values that reference it to that central location. One specifies its horizontal position and the other its vertical position. The two values are called the  $x$ -coordinate and  $y$ -coordinate. Using this notation, the coordinates at the origin are  $x = 0$  and  $y = 0$ .

On a gridded network of equally spaced horizontal and vertical lines, the horizontal line in the center is called the  $x$ -axis and the central vertical line is called the  $y$ -axis. Units are consistent and equally spaced across the full range of  $x$  and  $y$ . Horizontal lines above the origin and vertical lines to the right of the origin have positive values; those below or to the left have negative values. The four quadrants represent the four possible combinations of positive and negative  $X$  and  $Y$  coordinates. (ESRI Map projections.pdf)



**Fig. 23.1. The signs of  $x$ - and  $y$ -coordinates in a projected coordinate system. (Source: ESRI Map projections.pdf)**

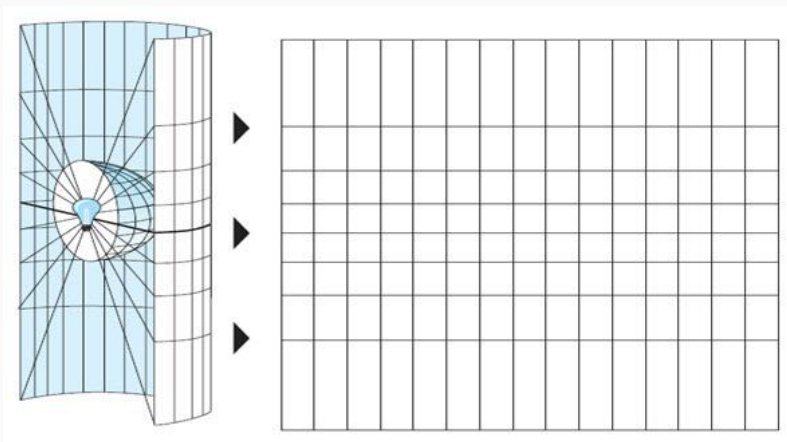
#### 23.1.1 Map projection

A map projection is used to portray all or part of the round Earth on a flat surface. This cannot be done without some distortion.

(<http://rnlx635.er.usgs.gov/mac/isb/pubs/MapProjections/projections.html>). Whether you treat the earth as a sphere or a spheroid, you must transform its three-dimensional surface to create a flat map sheet.

One easy way to understand how map projections alter spatial properties is to visualize shining a light through the earth onto a surface, called the projection surface. Imagine the earth's surface is clear with the graticule drawn on it. Wrap a piece of paper around the earth. A light at the center of the earth will cast the shadows of the graticule onto the piece of paper. You can now unwrap the paper and lay it flat. The shape of the graticule on the flat paper is different from that on the earth. The map projection has distorted the graticule.

A map projection uses mathematical formulas to relate spherical coordinates on the globe to flat, planar coordinates. Different projections cause different types of distortions. Some projections are designed to minimize the distortion of one or two of the data's characteristics. A projection could maintain the area of a feature but alter its shape. In the graphic below, data near the poles is stretched. (ESRI Map projections.pdf)



**Fig. 23.2. The graticule of a geographic coordinate system is projected onto a cylindrical projection surface. (Source: Kennedy, 1994)**

### 23.1.2 Major map projections

The three major types of projections developed from this method are the:

- (1) Conic
- (2) Planar
- (3) Cylindrical

#### (1) Conic projections

Conic projections, points from the globe graticule are transferred to a cone which has been enveloped around the sphere. The cone is then unrolled into a flat plane. The normal aspect is the north or south pole where the axis of the cone (the point) coincides with the pole. Conic projections can only represent one hemisphere, or a portion of one hemisphere, for the cone does not extend far beyond the center of the sphere. Conic projections are often used to project areas that have a greater east-west extent than north-south, e.g., the United States. When projected from the center of the globe, the typical grid appearance for Conic projections shows parallels forming arcs of circles facing up in the Northern Hemisphere and down in the Southern Hemisphere; and meridians are either straight or curved and radiate

outwards from the direction of the point of the cone. (Laurie A. B. Garo, 1997, Introduction to Map Projections).

The most simple conic projection is tangent to the globe along a line of latitude. This line is called the standard parallel. The meridians are projected onto the conical surface, meeting at the apex, or point, of the cone. Parallel lines of latitude are projected onto the cone as rings. The cone is then 'cut' along any meridian to produce the final conic projection, which has straight converging lines for meridians and concentric circular arcs for parallels. The meridian opposite the cut line becomes the central meridian.

Conic projections are used for mid latitude zones that have an east-to-west orientation. Somewhat more complex conic projections contact the global surface at two locations. These projections are called secant conic projections and are defined by two standard parallels. It is also possible to define a secant projection by one standard parallel and a scale factor. The distortion pattern for secant projections is different between the standard parallels than beyond them. Generally, a secant projection has less overall distortion than a tangent case. On still more complex conic projections, the axis of the cone does not line up with the polar axis of the globe. These are called oblique. (ESRI Map projections.pdf)

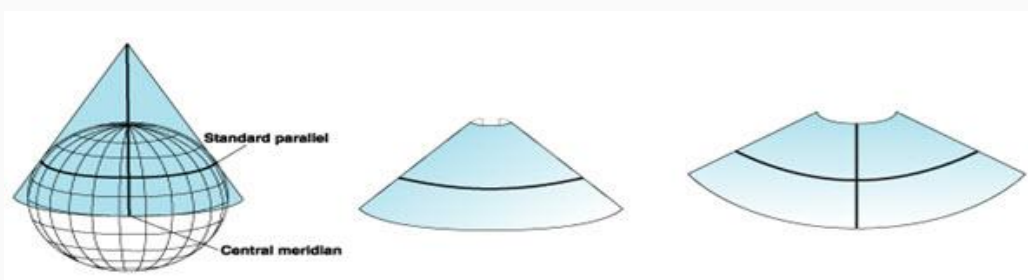


Fig. 23.3. (a) Tangent conic projection. (Source: [www.arcgis.com](http://www.arcgis.com))

With the **tangent** case Fig. 3(a), A cone is placed over a globe. The cone and globe meet along a latitude line. This is the standard parallel. The cone is cut along the line of longitude that is opposite the central meridian and flattened into a plane.

In **secant** case Fig. 3(b), A cone is placed over a globe but cuts through the surface. The cone and globe meet along two latitude lines. These are the standard parallels. The cone is cut along the line of longitude that is opposite the central meridian and flattened into a plane.

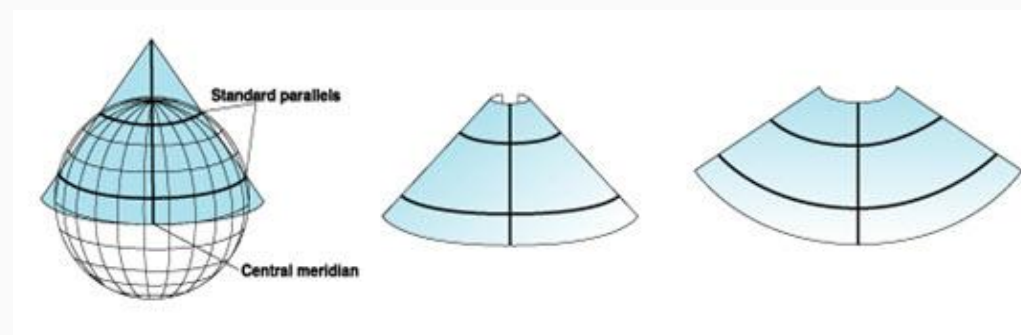


Fig. 23.3. (b) Secant conic projection. (Source: [www.arcgis.com](http://www.arcgis.com))

The representation of geographic features depends on the spacing of the parallels. When equally spaced, the projection is equidistant in the north-south direction but neither conformal nor equal area such as the **Equidistant Conic projection**. For small areas, the overall distortion is minimal. On the **Lambert Conic Conformal projection**, the central parallels are spaced more closely than the parallels near the border, and small geographic shapes are maintained for both small-scale and large-scale maps. Finally, on the **Albers Equal Area Conic projection**, the parallels near the northern and southern edges are closer together than the central parallels, and the projection displays equivalent areas (Kennedy, 1994).

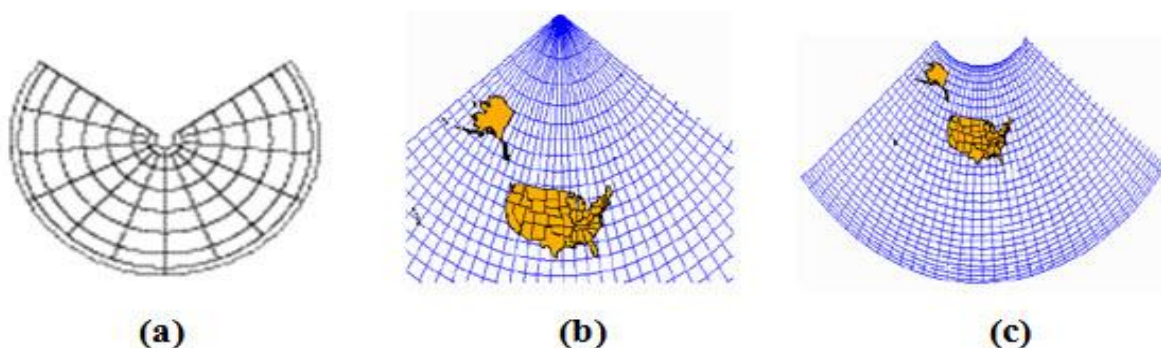


Fig. 23.4. (a) Equidistant conic, (b) Lambert Conformal Conic Projection, and (c) Albers Equal Area Conic projection.

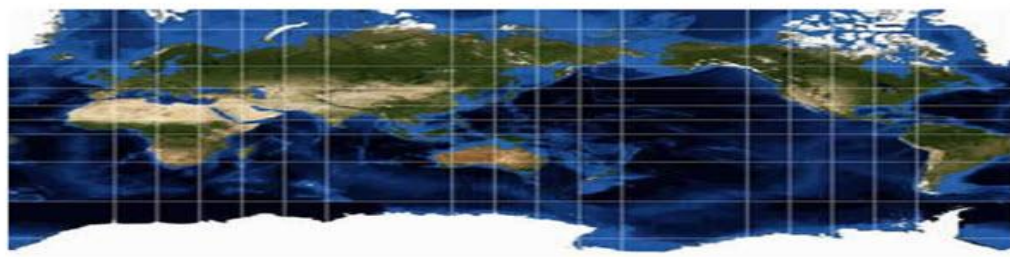
(Source: <http://www.progonos.com/furuti/MapProj/Normal/ProjTbl/projTbl.htm> and Class notes on Map projection)

## (2) Cylindrical projections

Cylindrical projections are formed by wrapping a large, flat plane (e.g., a large sheet of paper) around the globe to form a cylinder. The points on the spherical grid are transferred to the cylinder which is then unfolded into a flat plane. The equator is the "normal aspect" or viewpoint for these projections.

This family of projections are typically used to represent the entire world. When projected from the center of the globe with the normal aspect, the typical grid appearance for cylindrical projections shows parallels and meridians forming straight perpendicular lines. The spacing varies depending on the type of cylindrical projection. (Laurie A. B. Garo, 1997, Introduction to Map Projections).

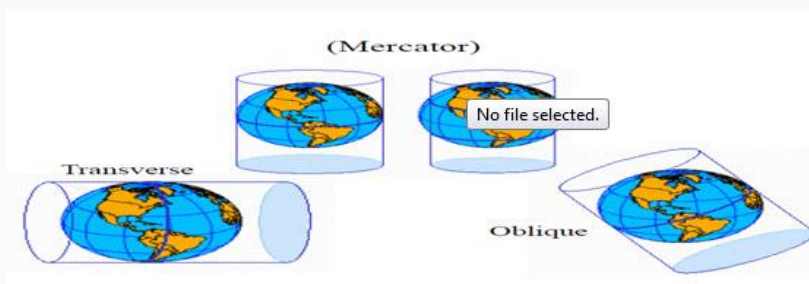




**Fig. 23.5.** This is an example of a cylindrical map projection and it is one of the most famous projections ever developed. It was created by a Flemish cartographer and geographer – Gerardus Mercator in 1569. It is famous because it was used for centuries for marine navigation. The sole reason for this is that any line drawn on the map was a true direction. However, shapes and distances were distorted. Notice the huge distortions in the Arctic and Antarctic regions, but the reasonable representation of landmasses out to about 50° north and south. (Source: [www.icsm.gov.au](http://www.icsm.gov.au))

Cylindrical projections can also have tangent or secant cases. The **Mercator projection** is one of the most common cylindrical projections, and the equator is usually its line of tangency. Meridians are geometrically projected onto the cylindrical surface, and parallels are mathematically projected, producing graticular angles of 90 degrees. The cylinder is 'cut' along any meridian to produce the final cylindrical projection. The meridians are equally spaced, while the spacing between parallel lines of latitude increases toward the poles. This projection is conformal and displays true direction along straight lines. Rhumb lines, lines of constant bearing, but not most great circles, are straight lines on a Mercator projection. (Kennedy, 1994).

For more complex cylindrical projections the cylinder is rotated, thus changing the tangent or secant lines. Transverse cylindrical projections such as the **Transverse Mercator** use a meridian as the tangential contact or lines parallel to meridians as lines of secancy. The standard lines then run north and south, along which the scale is true. **Oblique cylinders** are rotated around a great circle line located anywhere between the equator and the meridians. In these more complex projections, most meridians and lines of latitude are no longer straight. of equidistance. Other geographical properties vary according to the specific projection.

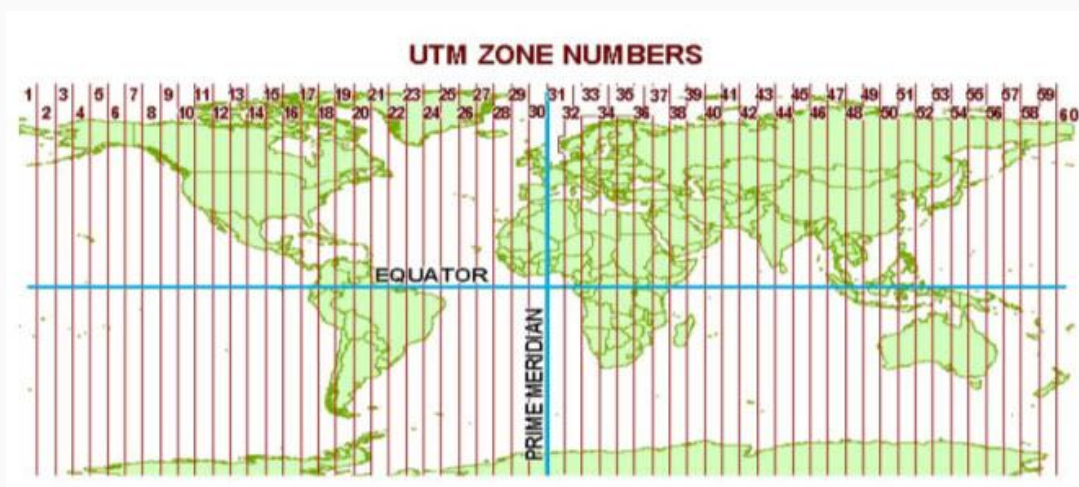


**Fig. 23.6.** A cylinder is placed over a globe. The cylinder can touch the globe along a line of latitude (normal case or Mercator), a line of longitude (transverse case), or another line (oblique case).

(Source: Class notes on Map Projection)

The National Imagery and Mapping Agency (NIMA) (formerly the Defense Mapping Agency) adopted a special grid for military use throughout the world called the **Universal Transverse Mercator (UTM)** grid.

([egsc.usgs.gov/isb/pubs/factsheets/fs07701.html](http://egsc.usgs.gov/isb/pubs/factsheets/fs07701.html)).



**Fig. 23.7. Universal Transverse Mercator system.**

(Source: [www.luomus.fi/english/botany/afe/map/utm\\_ups.pdf](http://www.luomus.fi/english/botany/afe/map/utm_ups.pdf))

The **UNIVERSAL TRANSVERSE MERCATOR** is also very widely used. In this grid, the world is divided into 60 north-south zones, each covering a strip 6° wide in longitude. These zones are numbered consecutively beginning with Zone 1, between 180° and 174° west longitude, and progressing eastward to Zone 60, between 174° and 180° east longitude. Thus, the conterminous 48 States are covered by 10 zones, from Zone 10 on the west coast through Zone 19 in New England (fig.7). In each zone, coordinates are measured north and east in meters. (One meter equals 39.37 inches, or slightly more than 1 yard.) The northing values are measured continuously from zero at the Equator, in a northerly direction. To avoid negative numbers for locations south of the Equator, NIMA's cartographers assigned the Equator an arbitrary false northing value of 10,000,000 meters. A central meridian through the middle of each 6° zone is assigned an easting value of 500,000 meters. Grid values to the west of this central meridian are less than 500,000; to the east, more than 500,000.

([egsc.usgs.gov/isb/pubs/factsheets/fs07701.html](http://egsc.usgs.gov/isb/pubs/factsheets/fs07701.html)).

### (3) Planar or Azimuthal Projection

Azimuthal projections, the spherical (globe) grid is projected onto a flat plane, thus it is also called a plane projection. The poles are the "normal aspect" (the viewpoint or perspective) which results in the simplest projected grid for this family of projections. That is, the plane is normally placed above the north or south pole. Normally only one hemisphere, or a portion of it, is represented on Azimuthal projections. When projected from the center of the globe with the normal aspect, the typical grid appearance for Azimuthal projections shows parallels forming concentric circles, while meridians radiate out from the center. (Laurie A. B. Garo, 1997, Introduction to Map Projections).

Planar projections project map data onto a flat surface touching the globe. This type of projection is usually tangent to the globe at one point but may be secant. The point of contact may be the North Pole, the South Pole, a point on the equator, or any point in between. This point specifies the aspect and is the focus of the projection. The focus is identified by a central longitude and a central latitude. Possible aspects are polar, equatorial, and oblique.



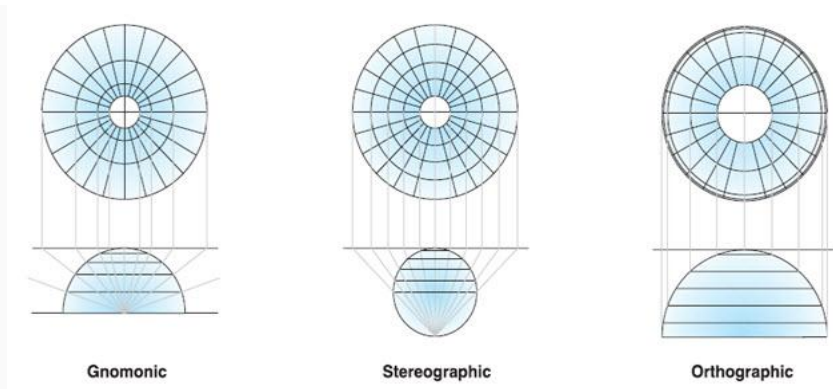
**Fig. 23.8. Polar, Equatorial and oblique projections. (Source: Kennedy, 1994)**

Polar aspects are the simplest form. Parallels of latitude are concentric circles centered on the pole, and meridians are straight lines that intersect at the pole with their true angles of orientation. In other aspects, planar projections will have graticular angles of 90 degrees at the focus. Directions from the focus are accurate. Great circles passing through the focus are represented by straight lines; thus the shortest distance from the center to any other point on the map is a straight line. Patterns of area and shape distortion are circular about the focus. For this reason, azimuthal projections accommodate circular regions better than rectangular regions. Planar projections are used most often to map polar regions. Some planar projections view surface data from a specific point in space. The point of view determines how the spherical data is projected onto the flat surface. The perspective from which all locations are viewed varies between the different azimuthal projections. Perspective points may be the center of the earth, a surface point directly opposite from the focus, or a point external to the globe, as if seen from a satellite or another planet.

Azimuthal projections are classified in part by the focus and, if applicable, by the perspective point.

The graphic below compares three planar projections with polar aspects but different perspectives. The Gnomonic projection views the surface data from the center of the earth, whereas the Stereographic projection views it from pole to pole. The Orthographic projection views the earth from an infinite point, as if viewed from deep space. Note how the differences in perspective determine the amount of distortion toward the equator. (ESRI Map projections.pdf)





**Fig. 23.9. Types of Azimuthal projection based on focus and perspective point.**

(Source: Kennedy, 1994)

### 23.2 Projection Specification

A map projection by itself isn't enough to define a projected coordinate system. One can state that a dataset is in Transverse Mercator, but that's not enough information. Where is the center of the projection? Was a scale factor used? Without knowing the exact values for the projection parameters, the dataset can't be reprojected.

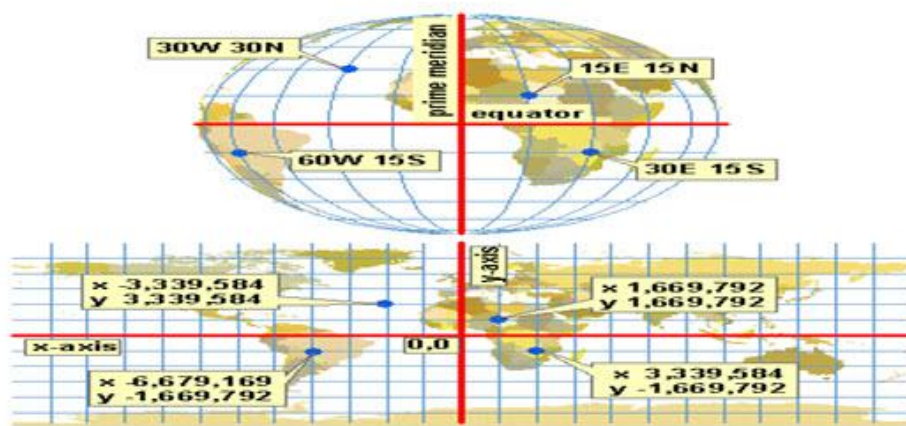
You can also get some idea of the amount of distortion the projection has added to the data. If you're interested in Australia but you know that a dataset's projection is centered at 0,0, the intersection of the equator and the Greenwich prime meridian, you might want to think about changing the center of the projection.

Each map projection has a set of parameters that you must define. The parameters specify the origin and customize a projection for your area of interest.

([gistutorial.blogspot.com/2011/04/projection-parameters.html](http://gistutorial.blogspot.com/2011/04/projection-parameters.html))

On the round surface of the earth, locations are described in terms of latitude and longitude. Some projection parameters, called angular parameters, are set with these latitude-longitude values. Once the earth's back has been broken with a projection, locations are described in terms of constant units like meters or feet. Some projection parameters, called linear parameters, use these constant units (or they use ratios, such as 0.5 or 0.9996).

([www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm)).



**Fig. 23.10. Round data is described with meridians, parallels, and latitude-longitude values. Bottom: Flat data is described with x, y units. Projection parameters use both kinds of descriptions. The projection at bottom is Plate Carrée.** (Source: [www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm))

### 23.2.1 Angular parameters

#### (i) Central meridian

Central meridian defines the origin of the x-coordinates. Every projection has a central meridian, which is the middle longitude of the projection. In most projections, it runs down the middle of the map and the map is symmetrical on either side of it. It may or may not be a line of true scale. (True scale means no distance distortion.)

In ArcGIS, you can change the central meridian of any projection. (Occasionally, it's the only angular parameter you can change.) The central meridian is also called the 'longitude of origin' or the 'longitude of center'. Its intersection with the latitude of origin (see below) defines the starting point of the projected (x, y) map coordinates. ([www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm)).

#### (ii) Latitude of origin

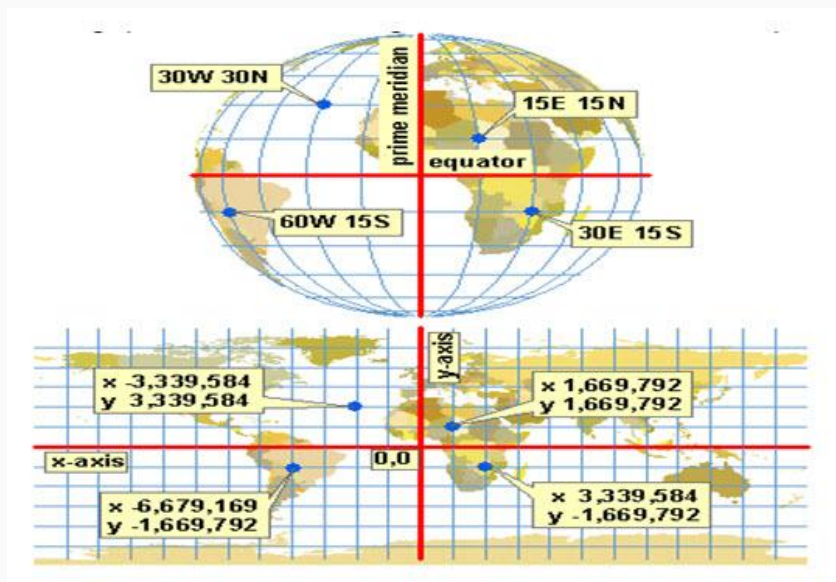
Latitude of origin defines the origin of the y-coordinates. This parameter may not be located at the center of the projection. In particular, Conic projections use this parameter to set the origin of the y-coordinates below the area of interest. In that instance, you don't need to set a false northing parameter to ensure that all y-coordinates are positive.

([gistutorial.blogspot.com/2011/04/projection-parameters.html](http://gistutorial.blogspot.com/2011/04/projection-parameters.html)).

Every projection also has a latitude of origin. The intersection of this line with the central meridian is the starting point of the projected coordinates. In ArcGIS, you can put the latitude of origin wherever you want for most conic and transverse cylindrical projections. (In many world projections, on the other hand, it is defined to be the equator and can't be changed.) The latitude of origin may or may not be the middle latitude of the projection and may or may not be a line of true scale. The important thing to remember about the latitude and

longitude of origin is that they don't affect the distortion pattern of the map. All they do is define where the map's x, y units will originate.

When data is unprojected, it doesn't have x, y units. Locations are measured in latitude and longitude, as you know from the previous module. But when you set a projection and flatten everything out, you also start using a new way to measure location. This new way is in terms of constant distance units (like meters or feet) measured along a horizontal x-axis and a vertical y-axis. A location like  $x = 500,000$ ,  $y = 100,000$  would refer to a point 500,000 meters (or whatever units of measure you are using) along the x-axis and 100,000 meters along the y-axis. The place where the axes cross is the coordinate origin, or 0,0 point. Commonly, this is in the middle of the map but it doesn't have to be.

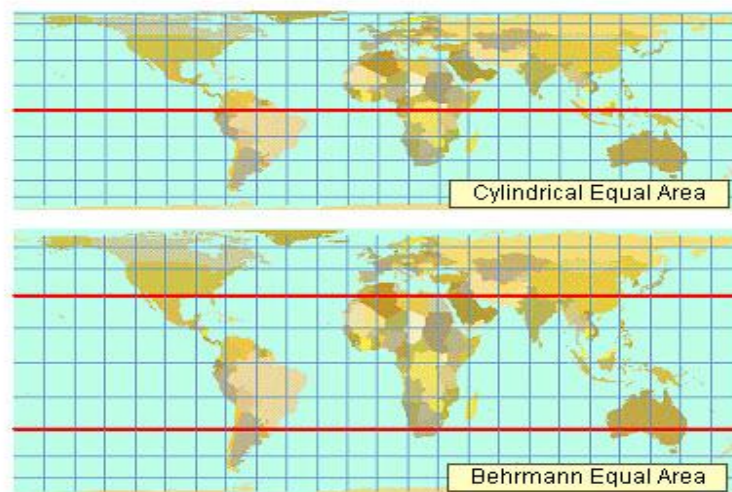


**Fig. 23.11.** In the top graphic below, the intersection of the central meridian (longitude of origin) and the latitude of origin is marked with a cross. This point becomes the origin of the x, y coordinates. The bottom graphic shows the grid (normally invisible) on which the x, y coordinates are located. The heavy lines are the x- and y-axes, which divide the grid into four quadrants. Coordinates are positive in one direction and negative in the other for each axis. (Source: [www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm))

### (iii) Standard parallel

A standard parallel is a line of latitude that has true scale. Not all projections have standard parallels, but many common ones do. Conic projections often have two. In a few projections, like the Sinusoidal and the Polyconic, every line of latitude has true scale and is therefore a standard parallel.

In ArcGIS, you can change the standard parallel for some projections and not for others. Many world projections, for instance, have fixed standard parallels. A standard parallel may or may not coincide with the latitude of origin.



**Fig. 23.12. Top: The Cylindrical Equal Area projection has a single standard parallel. By default, it is the equator, but you can change it. Bottom: The Behrmann projection is the same projection, but with two standard parallels at 30° N and 30° S. These standard parallels define the projection and cannot be changed.** (Source: [www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm))

#### (iv) Latitude of center

Latitude of center is used with the Hotine Oblique Mercator Center (both Two-Point and Azimuth) cases to define the origin of the y-coordinates. It is almost always the center of the projection ([gistutorial.blogspot.com/2011/04/projection-parameters.html](http://gistutorial.blogspot.com/2011/04/projection-parameters.html)).

#### (v) Central parallel

Central parallel defines the origin of the y-coordinates.

**NOTE:** In some projections, you will also see parameters called the latitude of center and the central parallel. These two terms seem to have the same meaning. Like the latitude of origin, they define the starting point of the y-coordinates; unlike it, they are nearly always the middle parallel of the projection. These parameters are used mainly with projections that have single points (rather than lines) of zero distortion, such as the Gnomonic and Orthographic. The intersection of the latitude of center (or central parallel) with the central meridian defines both the origin of the x, y coordinates and the point of zero distortion for the projection.

([www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm)).

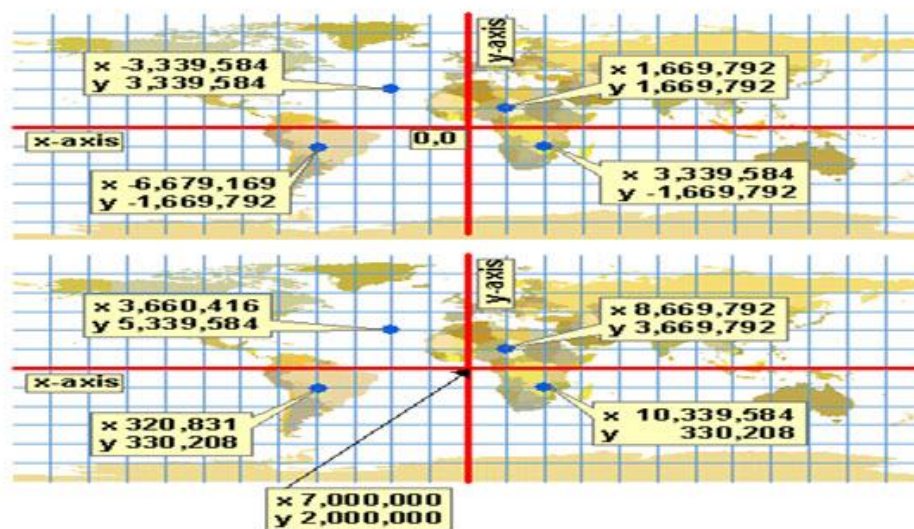
### 23.2.2 Linear parameters

#### (i) False easting and false northing

Linear parameters false easting is a linear value applied to the origin of the x coordinates. False northing is a linear value applied to the origin of the y coordinates. These are nothing but two big numbers that are added to each x- and y-coordinate, respectively. The numbers are big enough to ensure that all coordinate values or at least all those in your area of interest,



come out positive. You can also use the false easting and northing parameters to reduce the range of the x or y coordinate values. For example, if you know all y values are greater than 5,000,000 meters, you could apply a false northing of -5,000,000. ([gistutorial.blogspot.com/2011/04/projection-parameters.html](http://gistutorial.blogspot.com/2011/04/projection-parameters.html)).



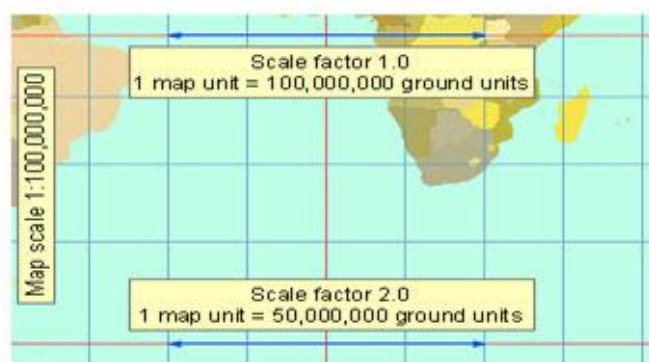
**Fig. 23.13. Top:** Projected coordinates are positive or negative, depending on their location. **Bottom:** A false easting value of 7,000,000 and a false northing value of 2,000,000 have been set. Every x-coordinate is now its original value plus 7,000,000. Every y-coordinate is its original value plus 2,000,000. The projection is Plate Carrée.

(Source:[www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm))

## (ii) Scale factor

A scale factor is the ratio of the true map scale to the stated map scale for a particular location. Scale factor is a unitless value applied to the centerpoint or line of a map projection. The scale factor is usually slightly less than one. The UTM coordinate system, which uses the Transverse Mercator projection, has a scale factor of 0.9996. Rather than 1.0, the scale along the central meridian of the projection is 0.9996. This creates two almost parallel lines approximately 180 kilometers, or about  $1^\circ$ , away where the scale is 1.0. The scale factor reduces the overall distortion of the projection in the area of interest. Learn more about the Transverse Mercator projection ([gistutorial.blogspot.com/2011/04/projection-parameters.html](http://gistutorial.blogspot.com/2011/04/projection-parameters.html)). Remember that no map has true scale everywhere.

A line of true scale is defined as having a scale factor of 1.0. Along this line, the actual map scale is equal to the stated scale (there is no distortion of distance). A scale factor of 2.0 means that distance measurements on the map are twice too long – if your scale bar tells you it's a hundred kilometers from A to B, it's really only fifty kilometers. A scale factor of 0.5 means that distance measurements are twice too short. ([www.geo.hunter.cuny.edu/.../Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/.../Projection%20parameters.htm))



**Fig. 23.14.** A Mercator projection with a stated map scale of 1:100,000,000. Along a line of true scale, such as the equator in this projection, the scale factor is 1.0. One map unit equals the number of ground units that the map says it does. At 60° north or south, the scale factor increases to 2.0 along the parallels. The blue double-headed arrow at the bottom of the map measures only half as much ground as the one at the top.

(Source: [www.geo.hunter.cuny.edu/~/Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/~/Projection%20parameters.htm))

**Table 23.1.** Some examples of projected co-ordinate system (Source: [www.geo.hunter.cuny.edu/~/Projection%20parameters.htm](http://www.geo.hunter.cuny.edu/~/Projection%20parameters.htm))

Projection	Central meridian (longitude)	Central parallel (latitude)	False easting	False northing	Scale factor
OS GB	2° W	49° N	+400 000	-100 000	0.999 601 2
UTM $\varphi > 0^\circ$	Zonal	0°	+500 000	0	0.999 6
UTM $\varphi < 0^\circ$	Zonal	0°	+500 000	-100 000	0.999 6
GK TM zone 11	63° E	0°	+500 000	0	1

## Module 9: Operations on Spatial Data

### Lesson 24 Map Algebra

#### 24.1 Concepts of Map Algebra

Map algebra is an informal and commonly used scheme for manipulating continuously sampled (i.e. raster) variables defined over a common area. It is also a term used to describe calculations within and between GIS data layers, according to some mathematical expression, to produce a new layer; it was first described and developed by Tomlin (1990). Map algebra can also be used to manipulate vector map layers, sometimes resulting in the production of a raster output. Although no new capabilities are brought to GIS, map algebra provides an elegant way to describe operations on GIS datasets. It can be thought of simply as algebra applied to spatial data which, in the case of raster data, are facilitated by the fact that a raster is a georeferenced numerical array.

Map Algebra models the surface of the earth as a multitude of independent, coincident layers or themes. The layers interact according to mathematical models and are typically based on real world observations. Planners develop layers on development and population (Steinitz et al. 1976). Social scientists develop layers on demographics, ethnicity, and economic factors (McHarg 1969). Applying Map Algebra model to input layers produces a new layer, which may be a physical map sheet, a vision perceived through a stack of mylars on a light table, or an electronic dataset displayed on a computer screen. Regardless of mechanism, the result allows its users to explain complex phenomena, predict trends, or make adjustments to the model.

However it is the mechanism which bounds usability of Map Algebra. How easy it is for scientists to perform simple tasks? Can complex models be developed and tested? Historically layers were plotted on individual transparent maps which, when superimposed and registered provide a visually integrated view of the data. The manual process of map overlay is slow and tedious.

##### 24.1.1 Data types

This section focuses on map algebra operations available for gridded datasets, such as those implemented in ESRI's Spatial Analyst extension for ArcView. Similar operations are available in other grid-based GISes, such as GRASS (Geographic Resources Analysis Support System).

The data associated with any grid cell can be of any type whatsoever. It is conceptually useful to divide data types into several classes, however. These include:

- **Categorical data:** These are non-numerical data. Grids that classify land use or land cover exemplify this category. Other examples are **proximity grids** (values identify the nearest object) and **feature grids** (only two values are possible: one value for cells

where features occur, another value--typically zero or NoData--where features do not occur).

- **Integral data:** These data may be relative **ranks** or preferences or they may be counts of occurrences or observations, for example. Thus, what they measure is inherently integral.
- **Floating-point ("real" data).** These typically represent a real surface, such as elevation, or the values of a scalar function (a "conceptual surface," if you will). Examples of such functions would be temperature, slope, amount of sunlight received per year, distance to the nearest feature, population density.
- **Vector data:** These are ordered tuples of real values that represent fields of directions. For example, hydraulic gradients (for two-dimensional groundwater models), wind velocities (again for two-dimensional models), and ocean currents are two-dimensional vector fields. Vector data may have more than two dimensions, even though they are defined over a strictly two-dimensional domain. For example, models using astronomical data, such as climate models, may make use of information about the three-dimensional location (on the earth's surface) of each grid point.

(Scientific visualization systems usually have built-in support for vector data, whereas most GISes require the modeler to represent vector data as an ordered collection of floating-point grids).

([http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))

### 24.1.2 Working with null data

An essential part of map algebra or spatial analysis is the coding of data in such a way as to eliminate certain areas from further contribution to the analysis. For instance, if the existence of low-grade land is a prerequisite for a site selection procedure, we then need to produce a layer in which areas of low-grade land are coded distinctively so that all other areas can be removed. One possibility is to set the areas of low-grade land to a value of 1 and the remaining areas to 0. Any processes involving multiplication, division or geometric mean that encounter the zero value will then also return a zero value and that location (pixel) will be removed from the analysis. The opposite is true if processing involves addition, subtraction or arithmetic mean calculations, since the zero value will survive through to the end of the process. The second possibility is to use a null or No Data value instead of a zero. The null is a special value which indicates that there is no digital numerical value. In general, unlike zero, any expression will produce a null value if any of the corresponding input pixels have null values. Many functions and expressions simply ignore null values, however, and in some circumstances this may be useful, but it also means that a special kind of function must be used if we need to test for the presence of (or to assign) null values in a dataset. For instance, within ESRI's ArcGIS, the function ISNULL is used to test for the existence of null values and will produce a value of 1 if null, or 0 if not. Using ER Mapper's formula editor, null values can easily be assigned, set to other values, made visible or hidden. Situations where the presence of nulls is disadvantageous include instances where there are unknown gaps in the dataset, perhaps produced by measurement error or failure. Within map algebra,



however, the null value can be used to great advantage since it enables the selective removal or retention of values and locations during analysis.

**Table 24.1 Operations categorized according to their spatial or non-spatial nature.**

Output	Spatial attributes involved?	
	Yes	No (not necessarily)
Map or image	Neighbourhood processing(filtering), zonal and focal operations, mathematical morphology	Reclassification, rescaling (unary operations), overlay (binary operations), thresholding and density slicing.
Tabular	Spatial autocorrelation and variograms	Various tabular statistics (aggregation, variety) and tabular modeling (calculation of new fields from existing ones), scattergraphs

(Source: After Bonham-Carter, 2002)

### 24.1.3 Logical and conditional processing

These two processes are quite similar and they provide a means of controlling what happens during some function. They allow us to evaluate some criterion and to specify what happens next if the criterion is satisfied or not. Logical processing describes the tracking of true and false values through a procedure. Normally, in map algebra, a non-zero value is always considered to be a logical true, and zero, a logical false. Some operators and functions may return either logical true values (1) or logical false values (0), for example relational and Boolean operators. The return of a true or false value acts as a switch for one or other consequence within the procedure. Conditional processing allows that a particular action can be specified, according to the satisfaction of various conditions; if the conditions are evaluated as true then one action is taken, and an alternative action is taken when the conditions are evaluated as false. The conventional if-then-else statement is a simple example of a conditional statement:

if  $i < 16$  then 1 else null where  $i$  = input pixel dn

Conditional processing is especially useful for creating analysis 'masks'. In Fig. 24.1, each input pixel value is tested for the condition of having a slope equal to or less than  $15^\circ$ . If the value tests true (slope angle is  $15^\circ$  or less), a value of 1 is assigned to the output pixel. If it tests false (exceeds  $15^\circ$ ), a null value is assigned to the output pixel. The output could then be used as a mask to exclude areas of steeper slopes and allow through all areas of gentle slopes, such as might be required in fulfilling the prescriptive criteria for a site selection exercise.

### 24.1.4 Other types of operator

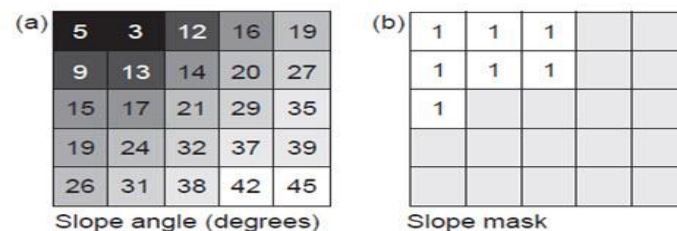
Expressions can be evaluated using arithmetic operators (addition, subtraction, logarithmic, trigonometric) and performed on spatially coincident pixel DN values within two or more input layers (Table 24.2). Generally speaking, the order in which the input layers are listed denotes the precedence with which they are processed; the input or operator listed first is given top priority and is performed first, with decreasing priority from left to right.

A relational operator enables the construction of logical functions and tests by comparing two numbers and returning a true value (1) if the values are equal or false (0) if not. For example, this operator can be used to find locations within a single input layer with DN values representing a particular class of interest. These are particularly useful with discrete or categorical data.

A Boolean operator, for example AND, OR or NOT, also enables sequential logical functions and tests to be performed. Like relational operators, Boolean operators also return true (1) and false (0) values. They are performed on two or more input layers to select or remove values and locations from the analysis. For example, to satisfy criteria within a slope stability model, Boolean operators could be used to identify all locations where values in one input representing slope are greater than 40° AND where values in an elevation model layer are greater than 2000m (as in Fig. 24.2a).

Logical operators involve the logical comparison of the two inputs and assign a value according to the type of operator. For instance, for two inputs (A and B) A DIFF B assigns the value from A to the output pixel if the values are different or a zero if they are the same. An expression A OVER B assigns the value from A if a non-zero value exists; if not then the value from B is assigned to the output pixel. A combinatorial operator finds all the unique combinations of values among the attributes of multiple input rasters and assigns a unique value to each combination in the output layer. The output attribute will contain fields and attributes from all the input layers.

All these operators can be used, with care, alone or sequentially, to remove, test, process, retain or remove values (and locations) selectively from datasets alone or from within a spatial analysis procedure.

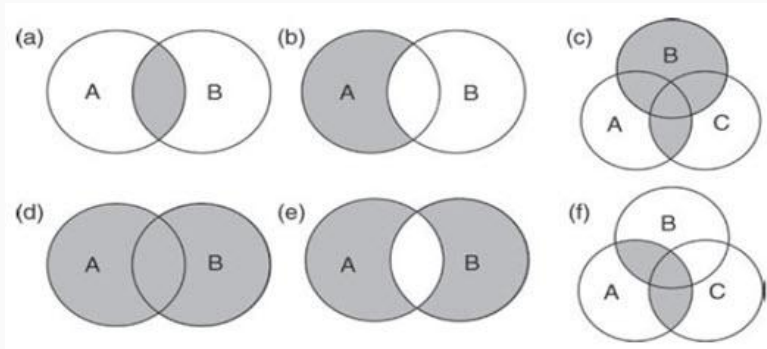


**Fig. 24.1. Logical test of slope angle data, for the condition of being no greater in value than 15°: (a) slope angle raster and (b) slope mask (pale grey blank cells indicate null values). (Source: Liu, and Mason, 2009)**

**Table 24.2. Summary of common arithmetic, relational, Boolean, power, logical and combinatorial operator.**

Arithmetic	Relational (return true/false)	Boolean (return true/false)
+, Addition -, Subtraction *, Multiplication /, Division MOD, Modulus	=, EQ Equal ^=, <>, NE Not equal <=, LT Less than/equal to <=LE Less than/equal to >, GT Greater than	^, Not Logical complement Logical AND OR !, XOR Logical XOR & AND  , OR Logical

	$\geq$ , GE greater than/ equal to	
Power	Logical	Combianational
Sqrt, Square root Sqr, Square Pow, Raised to a power	DIFF, Logical difference IN{list}, Contained in list OVER, Replace	CAND, Combinational AND COR, Combinational OR CXOR, Combinational XOR



**Fig. 24.2.** Use of Boolean rules and set theory within map algebra; here the circles represent the feature classes A, B and C, illustrating how simple Boolean rules can be applied to geographic datasets, and especially rasters to extract or retain values, to satisfy a series of criteria: (a) A AND B (intersection or minimum ); (b) A NOT B; (c) (A AND C) OR B; (d) A OR B (union or maximum); (e) A XOR B; and (f) A AND (B OR C).

(Source: Liu, and Mason, 2009)

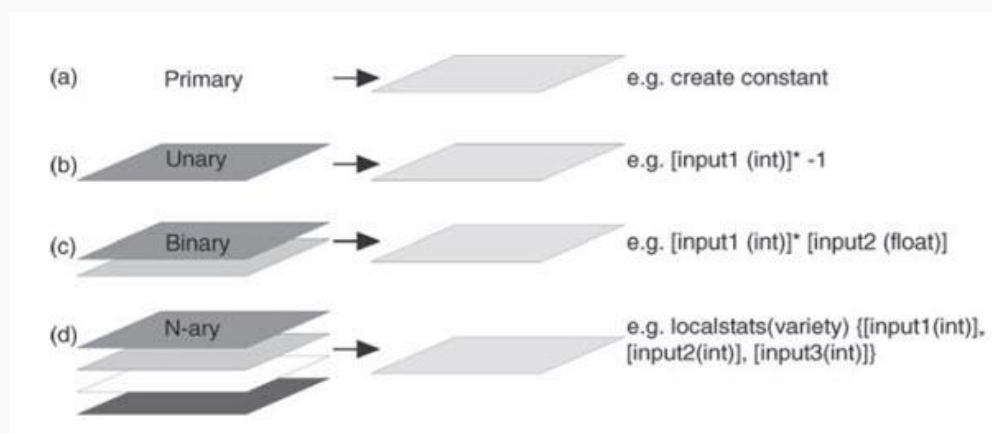
**Table 24.3.** Summary of local operations.

Type	Includes	Example
Primary	Creation of a layer from nothing	Rasters of constant value or containing randomly generated values
Unary	Conversion of units of measurement and as intermediary steps of spatial analysis	Rescaling, negation, comparing or applying mathematical functions reclassification
Binary	Operations on ordered pairs of numbers in matching pixels between layers	Arithmetic and logical combinations of rasters
N-ary	Comparison of local statistics between several rasters (many to one or many to many)	Change or variety detection

(Source: Liu and Mason, 2009)

## 24.2 Local Operations

A local operation involves the production of an output value as a function of the value(s) at the corresponding locations in the input layer(s). These operations can be considered point operations when performed on raster data, i.e. they operate on a pixel and its matching pixel position in other layers, as opposed to groups of neighbouring pixels. They can be grouped into those which derive statistics from multiple input layers (e.g. mean, median, minority), those which combine multiple input layers, those which identify values that satisfy specified criteria or the number of occurrences that satisfy specified criteria (e.g. greater than or less than), or those which identify the position in an input list that satisfies a specified criterion. All types of operator previously mentioned can be used in this context. Commonly they are subdivided according to the number of input layers involved at the start of the process. They include primary operations where nothing exists at the start, to n-ary operations where n layers may be involved; they are summarized in Table 24.3 and illustrated in Fig. 24.3.



**Fig. 24.3. Classifying map algebra operations in terms of the number of input layers and some examples.**

(Source: Liu and Mason, 2009)

### 24.2.1 Primary operations

This description refers primarily to operations used to generate a layer, conceptually from nothing, for example the creation of a raster of constant value, or containing randomly generated numbers, such as could be used to test for error propagation through some analysis. An output pixel size, extent, data type and output DN value (either constant or random between set limits) must be specified for the creation of such a new layer.

### 24.2.2 Unary operations

These operations act on one layer to produce a new output layer and they include tasks such as rescaling, negation, comparison with other numbers, application of functions and reclassification. Rescaling is especially useful in preparation for multi-criteria analysis where all the input layers should have consistent units and value range: for instance, in converting from byte data, with 0 to 255 value range, to a percentage scale (0-100) or a range of between 0 and 1, and vice versa. Negation is used in a similar context, in modifying the value range of a dataset from being entirely positive to entirely negative and vice versa. Comparisons create

feature grids: the places where the comparison is true can be considered features on the earth's surface. They map the regions where a logical condition (the comparison) holds. These could be regions where, say, ozone concentrations exceed a threshold, ocean depths are below a certain target, or land use equals a given code. Mathematical functions are useful for changing the visualization of a grid. An equal interval classification using the square roots of the values will differ from an equal interval classification of the values themselves, for instance. Functions are also important as intermediate steps in many models. Reclassification is especially significant in data preparation for spatial analysis, and so deserves rather more in-depth description, but all these activities can be and are commonly carried out in image processing systems.

To illustrate different applications succinctly, suppose that three grids appear in the current view: "Integer" is an integer grid, "Float" is a floating-point grid, and "Indicator" is an integer grid containing only 0, 1, and NoData values. A value of 0 can be interpreted as a logical "false" and a value of 1 as a logical "true". In practice, of course, we will replace these names by the names of our themes.

- **Rescale** a grid: that is, Multiply all its values by a constant value.

[Float] \* 3.1415927 Multiply all values by Pi

[Integer].Float \* (39.37/12) Convert meters to feet

[Integer] \* (-1) Negate all values

Not [Indicator] Negate all logical values: 0 becomes 1, 1 becomes 0

- **Compare** a grid to a constant value. The result of a comparison is true, 0 where the comparison is false, and NoData where the original value is NoData

[Float] < 1 Returns 1 where values are less than 1, otherwise returns 0

[Integer] = 0 Converts all zeros to ones and all other values to zeros

- Apply a mathematical (or logical) **function** to a grid, cell by cell.

[Float].Cos Computes the cosine of each value (interpreted as radians).

[Float].Int Rounds all values and converts the result to an integer grid.

[Float].Sqrt Computes the square root of each value. Negative values return NoData (because the square root is not defined for negative values).

[Float].IsNull Returns 1 at all cells with NoData values, otherwise returns 0.

#### 24.2.2.1 Reclassification

This involves the process of reassigning a value, a range of values, or a list of values in a raster to new output values, in a new output raster. If one class (or group or range of classes) is more interesting to us than the other classes, its original values can be assigned a specific

value and all the others can be changed into a different (background) value. This involves the creation of a discrete raster from either a continuous one or another discrete raster. Reclassification can be applied to both vector and raster objects.

In the case of discrete raster data, a reclassification may be required to produce consistent units among a set of input raster images, in which case a one-to-one value change may be applied. The output raster would look no different, spatially, from the input, having the same number of classes, but the values would have changed.

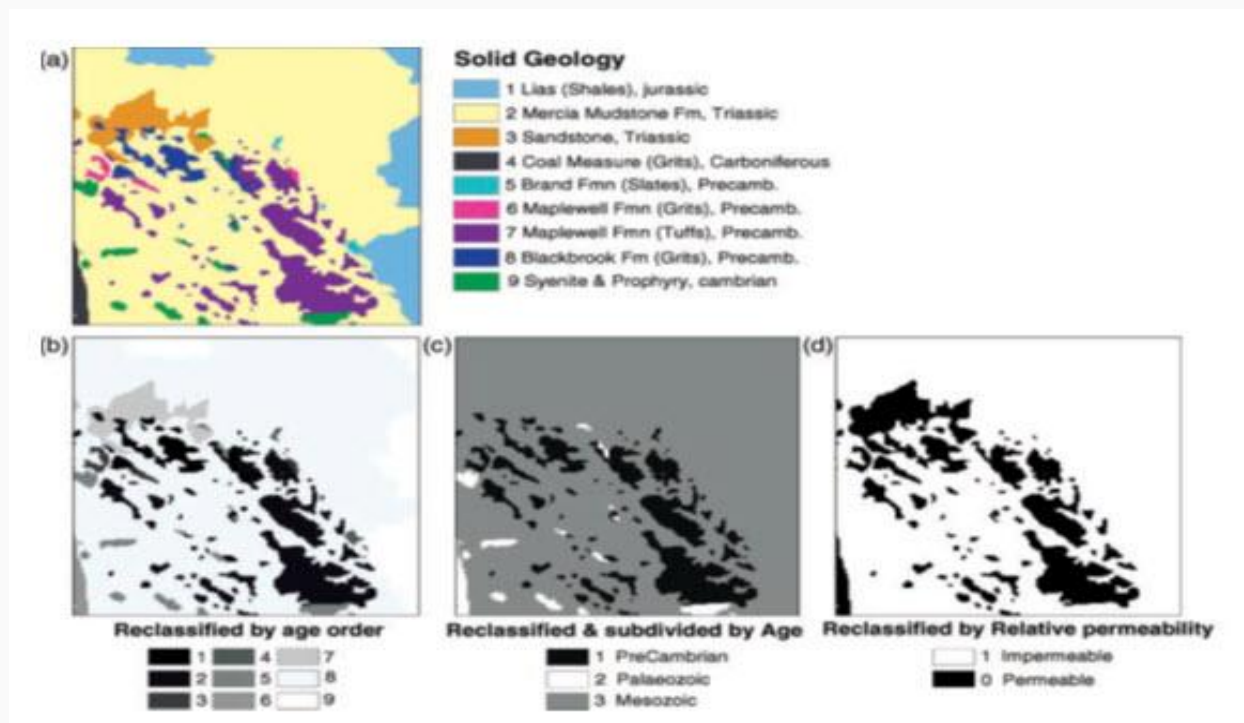
Different classes or types of feature may be reclassified according to some criteria that are important to the overall analysis. During the reclassification process, weighting can be applied to the output values to give additional emphasis to the significant classes, and at the same time reducing the significance of other classes.

The example in Fig. 24.4a shows a discrete raster representation of a geological map in which nine lithological units are coded with values 1 to 9 and labelled for the purposes of presentation, according to their name, rock type and ages. For the purposes of some analysis it may be necessary to simplify this lithological information, for example according to the broad ages of the units, PreCambrian, Palaeozoic and Mesozoic, for instance. The result of such a simplification is shown in Fig. 24.4c; now the map has only three classes and it can be seen that the older rocks (Precambrian and Palaeozoic) are clustered in the south-western part of the area, with the younger rocks (Mesozoic) forming the majority of the area as an envelope around the older rocks. So the simplification of the seemingly quite complex lithological information shown in Fig. 24.4a has revealed spatial patterns in that information which are of significance and which were not immediately apparent beforehand. Fig. 24.4d shows a second reclassification of the original lithological map, this time on the basis of relative permeability. The information is again simplified by reducing the number of classes to two, impermeable and permeable. Such a map might form a useful intermediary layer in an exercise to select land suitable for waste disposal but also illustrates that subjective judgements are involved at the early stage of data preparation. In the very act of simplifying information, we introduce bias and, strictly speaking, error into the analysis. We also have to accept the assumptions that the original classes are homogeneous and true representations everywhere on the map, which they may not be. In reality there is almost certainly heterogeneity within classes and the boundaries between the classes may not actually be as rigid as our classified map suggests.

Continuous raster data can also be reclassified in the same way. The image in Fig. 24.5a shows a DEM of the same area with values ranging between 37 and 277, representing elevation in meters above sea level. Reclassification of this dataset into three classes of equal interval to show areas of low, medium and high altitude produces the simplified image in Fig. 24.5b. Comparison with Fig. 24.5b shows that the areas of high elevation coincide with the areas where older rocks exist at the surface in the south-west of the area, again revealing spatial patterns not immediately evident in the original image. Reclassification of the DEM into three classes, this time with the classes defined according to the natural breaks in the image histogram (shown in Fig. 24.6), produces a slightly different result, Fig. 24.5c. The high-elevation areas are again in the south-west but the shape and distribution of those areas are different. This demonstrates several things. Firstly, that very different results can be produced when we simplify data so that (and secondly) we should be careful in doing so,

and, thirdly, that the use of the image histogram is fundamental to the understanding of and sensible use of reclassification of continuous raster data.

Reclassification forms a very basic but important part of spatial analysis, in the preparation of data layers for combination, in the simplification of layer information and especially when the layers have dissimilar value ranges. Reclassification is one of several methods of producing a common range among input data layers that hold values on different measurement scales.



**Fig. 24.4** (a) Discrete raster representation of a geological map, with nine classes representing different lithologies; (b) one-to-one reclassification by age order (1 representing the oldest, 9 the youngest); (c) a reclassified and simplified version where the lithological classes have been grouped and recoded into three broad age categories (Pre-Cambrian, Palaeozoic and Mesozoic); (d) a second reclassified version where the lithologies have been grouped according to their relative permeability, with 1 representing impermeable rocks and 0 permeable; such an image could be used as a mask. (Source: Liu and Mason, 2009)



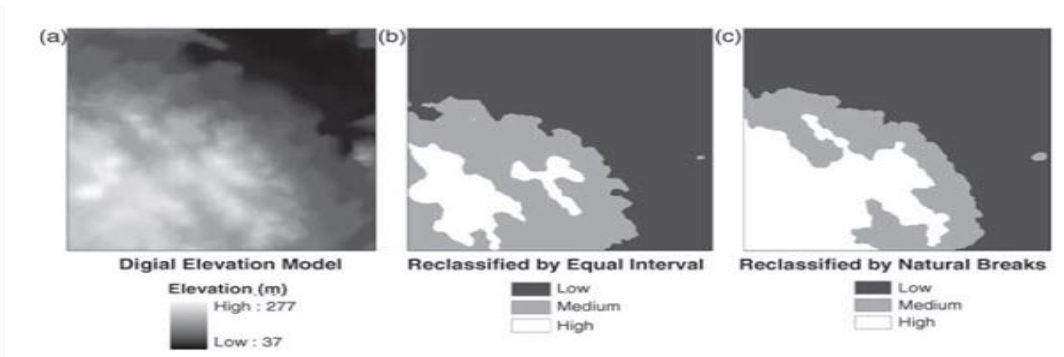


Fig. 24.5. (a) A DEM; (b) a DEM reclassified into three equal interval classes; and (c) a DEM reclassified into three classes by natural breaks in the histogram (shown in Fig. 24.6).

(Source: Liu and Mason, 2009)

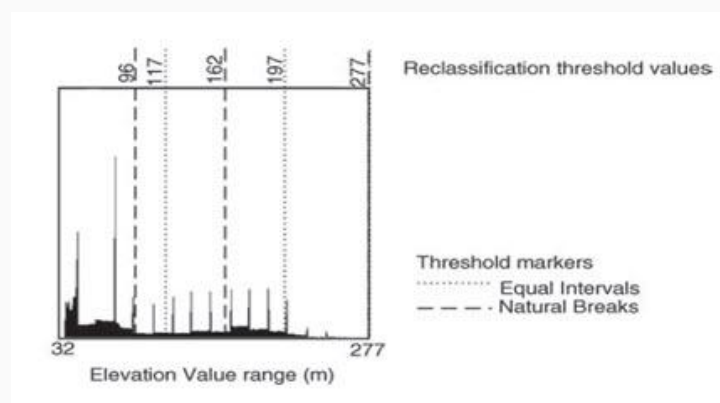


Fig. 24.6. Image histogram of the DEM shown in Fig. 24.5a and the positions of the reclassification thresholds set by equal interval and natural break methods (shown in Fig. 24.5b and c, respectively).

(Source: Liu and Mason, 2009)

### 24.2.3 Binary operations

Binary numeric operations act on ordered pairs of numbers. Likewise, binary grid operations act on the pairs of numbers obtained in each set of matching cells. The resulting grid is defined only where the two input grids overlap.

Suppose there are several floating-point grids represented by themes named "Float", "Float1", "Float2", and so on; with a similar supposition for integer and logical grids.

- **Mathematical operators**

[Float] + [Integer] Converts the values in [Integer] to floats, then performs the additions.

- **Logical operators**

[Float1] < [Float2] Returns 1 in each cell where [Float1]'s value is less than [Float2]'s value; otherwise, returns 0.



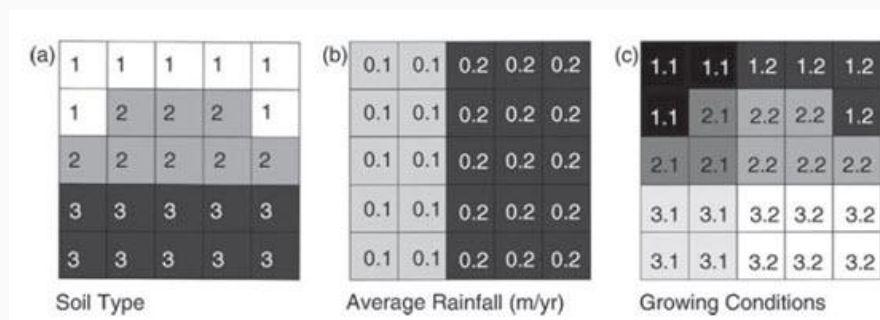
([http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))

[Indicator1] And [Indicator2] Returns 1 where both values are nonzero otherwise returns 0.

This description refers to operations in which there are two input layers, leading to the production of a single output layer. Overlay refers to the combination of more than one layer of data, to create one new layer. The example shown in Fig. 24.7 illustrates how a layer representing average rainfall, and another representing soil type, can be combined to produce a simple, qualitative map showing optimum growing conditions for a particular crop. Such operations are equivalent to the application of formulae to multiband images, to generate ratios, differences and other inter-band indices and as mentioned in relation to point operations on multi-spectral images, it is important to consider the value ranges of the input bands or layers, when combining their values arithmetically in some way. Just as image differencing requires some form of stretch applied to each input layer, to ensure that the real meaning of the differencing process is revealed in the output, here we should do the same. Either the inputs must be scaled to the same value range, or if the inputs represent values on an absolute measurement scale then those scales should have the same units.

The example shown in Fig. 24.7 represents two inputs with relative values on arbitrary nominal or ordinal (Fig. 24.7a) and interval (Fig. 24.7b) scales. The resultant values are also given on an interval scale and this is acceptable providing the range of potential output values is understood, having first understood the value ranges of the inputs, since they may mean nothing outside the scope of this simple exercise.

Another example could be the combination of two rasters as part of a cost-weighted analysis and possibly as part of a wider least cost pathway exercise. The two input rasters may represent measures of cost, as produced through reclassification of, for instance, slope angle and land value, cost here being a measure of friction or the real cost of moving or operating across the area in question. These two cost rasters are then aggregated or summed to produce an output representing total cost for a particular area (Fig.24.8).



**Fig. 24.7.** An example of a simple overlay operation involving two input rasters: (a) an integer raster representing soil classes (class 2, representing sandy loam, is considered optimum); (b) a floating-point raster representing average rainfall, in metres per year (0.2 is considered optimum); and (c) the output raster derived by addition of a and b to produce a result representing conditions for a crop; a value of 2.2 (2 + 0.2), on this rather arbitrary scale, represents optimum growing conditions and it can be seen that there are five pixel positions which satisfy this condition.

(Source: Liu and Mason, 2009)



**Fig. 24.8 (a) Slope gradient in degrees; (b) ranked (reclassified) slope gradient constituting the first cost or friction input; (c) ranked land value (produced from a separate input land-use raster) representing the second cost or friction input; and (d) total cost raster produced by aggregation of the input friction rasters (f1 and f2). This total cost raster could then be used within a cost-weighted distance analysis exercise.**

(Source: Liu and Mason, 2009)

#### 24.2.4 N-ary operations

Here we deal with a potentially unlimited number of input layers to derive any of a series of standard statistical parameters, such as the mean, standard deviation, majority and variety. Ideally there should be a minimum of three layers involved but in many instances it is possible for the processes to be performed on single layers; the result may, however, be rather meaningless in that case. The more commonly used statistical operations and their functionalities are summarized in Table 24.4. As with the other local operations, these statistical parameters are point operations derived for each individual pixel position, from the values at corresponding pixel positions in all the layers, rather than from the values within each layer.

**Table 24.4. Summary of local pixel statistical operations, their functionality and input/output data format.**

Statistic	Input format	Functionality	Data type
Variety	Only rasters. If a number is input, it will be converted to a raster constant for that value	Reports the number of different of different DN values occurring in the input rasters	Output in integer
Mean		Reports the average DN value among the input rasters	Output is floating point
Standard deviation	Rasters, numbers and constants	Reports the standard deviation of the DN values among the input rasters	Output is floating point
Medium		Reports the middle DN value among the input raster pixel values. With an even number of inputs, the values are ranked and the middle two values are averaged. If inputs are all integer, output will be truncated to integer Reports the total DN value among the input	

Sum		rasters Reports the difference between maximum and minimum DN	
Range		Reports the highest DN value among the input rasters	
Maximum		Reports the lowest DN value among the input rasters	
Minimum	Only rasters. If a number is input, it will be converted to a raster constant for that value	Reports the DN value which occurs most frequently among the input rasters. If no clear majority, output = null, for example if there are three inputs all with different values. If all inputs have equal value, output=input	If inputs are all integer, output will be integer, unless one is a float, there the input will be a float
Minority		Reports the DN value which occurs least frequently among the input rasters. If no clear minority, as majority If only two inputs, where different, output=null. If all inputs equal, output = input. If only one input, output= input	

(Source: Liu and Mason, 2009)

#### 24.2.4.1 Local statistics

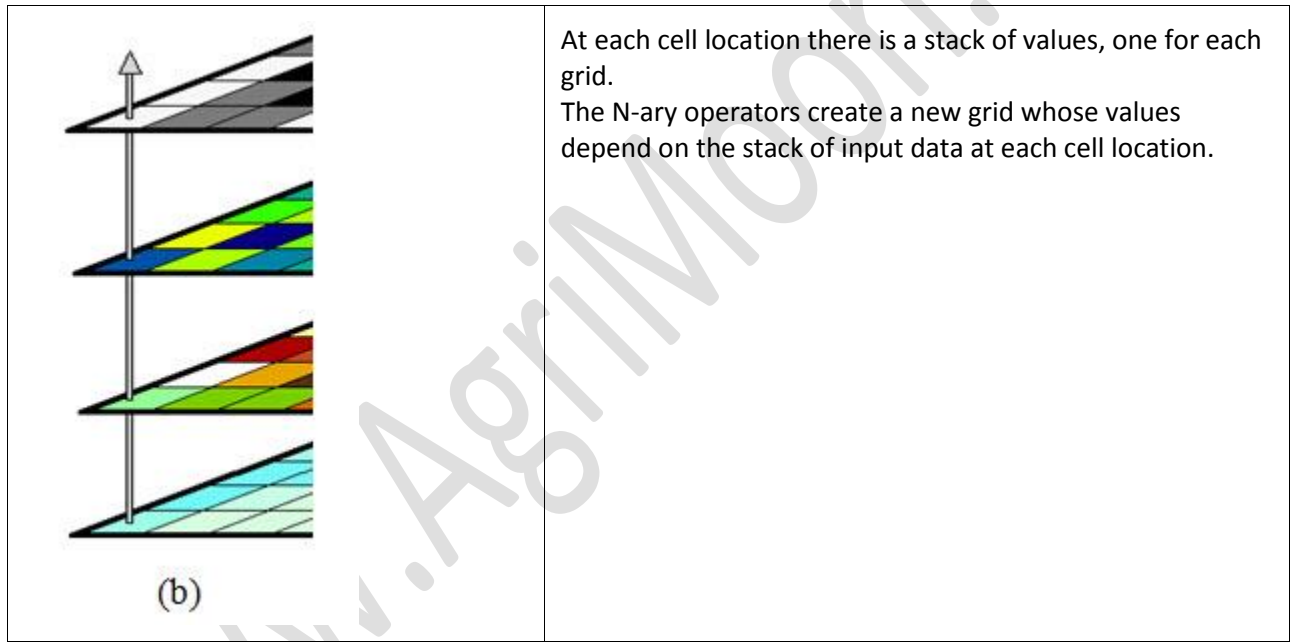
When we have many related grids defined in the same region, we often want to assess change: at each cell, how varied are the grid results? How large do they get? How small? What is the average? These questions make sense for numerical data.

For grids with ordinal data--that is, values that can be ordered, but which may not have any absolute meaning--you can still ask about order statistics. These are the relative rankings of values within the ordered collections of values observed at each cell.

For grids with categorical data, you might want to know at each cell whether one category predominates throughout the collection of grids and how many different categories actually appear at the cell's location. (Liu, and Mason, 2009)



**Fig. 24.9. (a) stack of grids with common mesh.** (Source: [http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))



**Fig. 24.9. (b) New grid by N-ary operator.**  
(Source: [http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))

The Spatial Analyst syntax for some of these requests is strange, because it wants to force expressions into the form "aGrid. Request (list of other grids)". This is inherently asymmetric because it singles out one grid in the collection to play the role of the object ("aGrid") to which the calculation is applied and leaves the other grids in the role of a list of arguments ("list of other grids"). Despite this syntax, for some requests, such as the local statistics, there is no asymmetry in the calculation itself: all the grids are equivalent. For some other requests, there is an asymmetry in the calculation: one grid plays a special role.

Spatial Analyst constructs lists with curly braces {} and separates the elements by commas.

- **Compute local statistics**

[Float]. LocalStats (#GRID\_STATYPE\_MAX, {[Float1], [Float2], [Float3]}) Computes the largest value among four grids.

[Float]. LocalStats (#GRID\_STATYPE\_MEDIAN, {[Float1], [Float2]}) Computes the median of three values.

[Integer]. LocalStats (#GRID\_STATYPE\_MAJORITY, {[Integer1], [Integer2]}) Computes the value occurring the most times (out of the three input values at each cell). If two or more values occur an equal number of times, Spatial Analyst returns NoData.

The Majority statistic evidently is not very useful when many ties occur: that is, when there are many cells where two or more values occur equally often.

([http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))

- **Compare one grid** (a “base” grid) to many others simultaneously

[Float]. Grids Greater than ([Float1], [Float2], [Float3], [Float4]) For each base cell in [Float], computes the number of times corresponding cells from [Float1], ..., [Float4] exceed (and do not equal) the base cell's value. There is a corresponding Grids Less Than operator.

- **Combine** the values of two grids based on values at a third grid

[Indicator]. Con ([Float1], [Float2]) Creates a grid with the values of [Float1] where [Indicator] is nonzero and with the values of [Float2] where [Indicator] is zero.

The **Con** request is especially useful. The result of **Con**, by default, is the second grid ([Float2] or [Mosaic] in the examples). However, at cells where [Indicator] is true, the values of the first grid ([Float1] or [Average]) are "painted" over the default values. Thus the **Con** request is a natural vehicle for selectively editing grids.

([http://www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm))

**Keywords:** Map algebra, proximity grids, null data, Local operations, Primary operations, Unary operations, Reclassification, Binary operations Mathematical operators, Logical operators, N-ary operations, Local statistics



## Lesson 25 Spatial Data

### 25.1 Neighbourhood Operations

#### 25.1.1 Local Neighbourhood

The general objective of a neighbourhood operation is to analyze the characteristics and/or spatial relationships of locations surrounding some specific (control) locations. Note that the control locations are actually part of the neighbourhood to be analyzed. Neighborhood operations are those that combine a small area or neighborhood of pixels to generate an output pixel. Thus, in fact, spatial interpolation techniques are a type of neighbourhood operations, because they aim to estimate the values at unsampled locations based on values at sampled locations (Carranza, 2009). These are incremental in their behaviour or operation. They work within a small neighbourhood of pixels to change the value of the pixel at the Centre of that neighbourhood, based on the local neighbourhood statistics. Then this process is incremented to the next pixel in the same row and it continued until the whole raster has been processed. Here the dealing data are of implemented 3-D character. Neighbourhood processes can be used to simplify or generalize discrete rasters (Liu and Mason, 2009). Neighbourhood operations applied to raster maps are basically filtering operations. Filtering can be performed in the time domain, frequency domain or spatial domain. Filtering in the spatial domain is a basic function in GIS. Filtering of a raster map involves an equal-sided filter window, also called a “kernel” or “template”, which moves across a raster map one pixel at a time. A filter has an odd number of pixels on each of its sides so that it defines a symmetrical neighbourhood about central pixel. The simplest filter is a square of pixels (Carranza, 2009). The applications of neighborhood operators are many, ranging from digital filters to techniques for sharpening, transforming, and warping images.

**Table 25.1. Summary of local pixel statistical operations, their functionality and input/output data format**

Statistic	Input format	Functionality	Data type
Variety	Only rasters. If a number is input, it will be converted to a raster constant for that value	Reports the number of different DN values occurring in the input rasters	Output is integer
Mean	values occurring in the input rasters	Reports the average DN value among Output is floating point the input rasters	Output is floating point
Standard deviation	Rasters, numbers and constants	Reports the standard deviation of the DN values among the input rasters	Output is floating point
Median		Reports the middle DN value among the input raster pixel values. With an even number of inputs, the values are ranked	

Sum	Only rasters. If a number is input, it will be converted to a raster constant for that value	and the middle two values are averaged. If inputs are all integer, output will be truncated to integer	If inputs are all integer, output will be integer, unless one is a float, then the output will be a float
Range		Reports the total DN value among the input rasters	
Maximum		Reports the difference between maximum and minimum DN value among the input raster	
Minimum		Reports the highest DN value among the input rasters	
Majority		Reports the lowest DN value among the input rasters	
Minority		Reports the DN value which occurs most frequently among the input rasters. If no clear majority, output ¼ null, for example if there are three inputs all with different values. If all inputs have equal value, output ¼ input Reports the DN value which occurs least frequently among the input rasters. If no clear minority, as majority If only two inputs, where different, output ¼ null. If all inputs equal, output ¼ input. If only one input, output ¼ input	

(Source: Liu and Mason, 2009)

#### 25.1.1.1 Distance

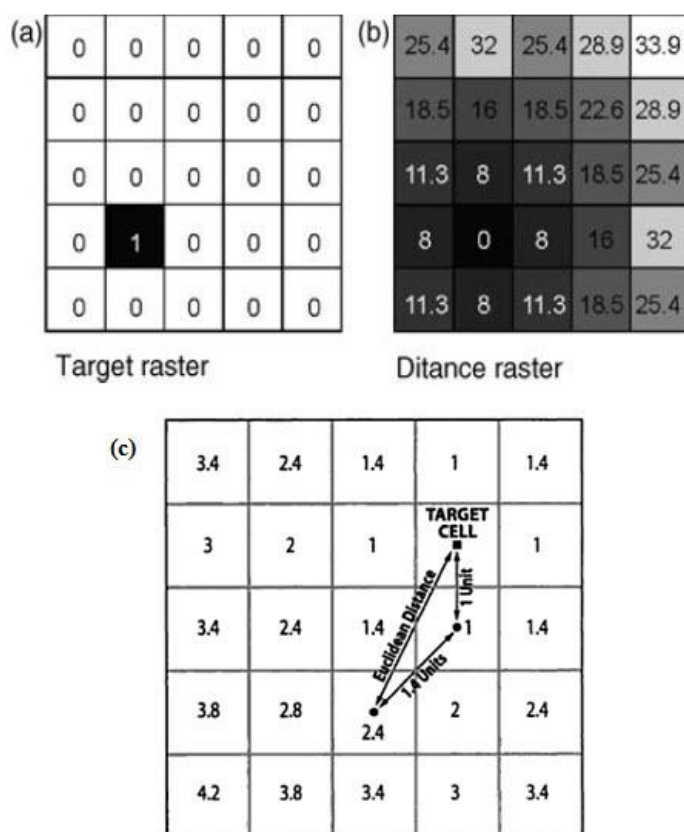
Mapping distance allows the calculation of the proximity of any raster pixel to/from a set of target pixels, to determine the nearest or to gain a measure of cost in terms of distance. Here the value assigned to the output pixel is a function of its position in relation to another pixel. The input is a discrete raster image, in which the target pixels are coded, with a value of 1 against a background of 0 (Fig. 25.1a). This operation involves the use of a straight line distance function, which calculates the Euclidean distance from every pixel to the target pixels (Fig. 25.1b).

The distance transformation transforms a feature or area of raster cells into an area based on given distances. One single distance can be used-for example, 10 m from the well or multiple distances for example 50 m, 100 m, 150 m, 200 m and 250 m from the road for the transformation. The distance transformation geographic information analysis is often used to show the geographic extent of events (e.g., noise from traffic, leaking of oil tanks into the ground) as a thing. In these uses the distances correspond to model or assumed values regarding the process underlying the events. This ability to transform from process to pattern is perhaps the single most important reason for the significance of this geographic information analysis type (Harvey, 2008).

The output pixel values represent the Euclidean distance from the target pixel centers to every other pixel Centre and are coded in the value units of the input raster, usually meters,



so that the input raster will usually contain integers and the output normally floating-point numbers. Then the calculated distance raster may then be further reclassified for used as input to more complex multi-criteria analysis or used within a cost-weighted distance analysis.



### Spread operation: calculating the distance

**Fig. 25.1. (a) An input discrete (binary) raster, (b) the straight line or Euclidean distance calculated from a single target or several targets are coded to every other pixel in an input and (c) Spread operation: calculating the distance. (Source: Malczewski, 1999)**

#### 25.1.1.2 Cost pathways

This moving window or kernel procedure is used to derive a cost-weighted distance and cost-weighted direction as part of a least cost pathway exercise. The cost-weighted distance function operates by evaluating each input pixel value of a total cost raster and comparing it with its neighbouring pixels. The average cost between each is multiplied by the distance between them. Cost-weighted direction is generated also from the total cost raster, where each pixel is given a value using a direction-encoded 3×3 kernel, which indicates the direction to the lowest cost pixel value from among its local neighbours. Then these two rasters or surfaces are combined to derive the least cost pathway or route across the raster, to the target.

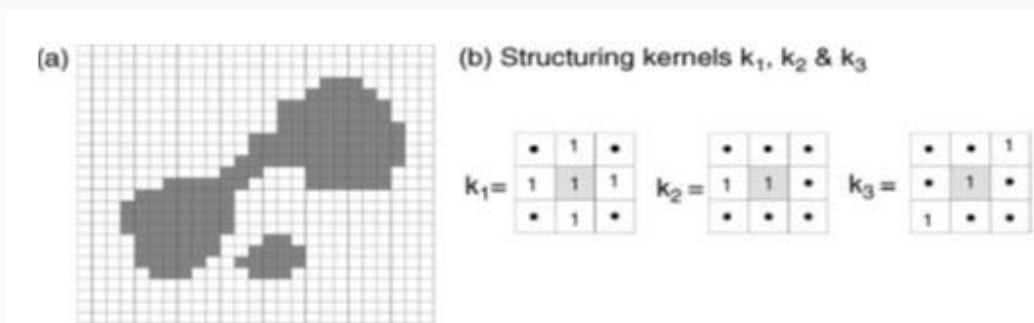


### 25.1.1.3 Mathematical morphology

This concept was first developed by Matheron (1975). Mathematical morphology is the combination of map algebra and set theory or of conditional processing and convolution filtering. This concept describes the spatial expansion and shrinking of objects through neighbourhood processing and extends the concept of filtering. These changes include erosion or shrinking, dilation or expansion, opening and closing of raster images. The size and shape of the neighbourhoods used are controlled by structuring elements or kernels which may be of varying size and form. The processing may not be reversible; for instance, after eroding such an image, using an erosion kernel, it is generally not possible to return the binary image to its original shape through the dilation kernel.

Mathematical morphology can be applied to vector point, line and area features but more often involves raster data, commonly discrete rasters and sometimes continuous raster surfaces, such as DEMs. It has also been used in mineral prospection mapping, to generate evidence maps, and in the processing of rock thin section images, to find and extract mineral grain boundaries. This method has applications in raster topology and networks, in addition to pattern recognition, image texture analysis, terrain analysis and also can be used for edge feature extraction and image segmentation.

To illustrate the effects, consider a simple binary raster image showing two classes, shown in Fig. 25.2, the values in the raster of the two classes are 1 (inner, dark grey class) and 0 (surrounding, white class). This input raster is processed using a series of 3×3 structuring elements or kernels (k), which consist of the values 1 and null (instead of 0).



**Fig. 25.2. (a) Simple binary raster image (i); and (b) the three structuring kernels (k<sub>1</sub>, k<sub>2</sub> and k<sub>3</sub>) the effects of which are illustrated in Figs. 25.3–25.5. The black dots in the kernels represent null values.**

(Source: Liu and Mason, 2009)

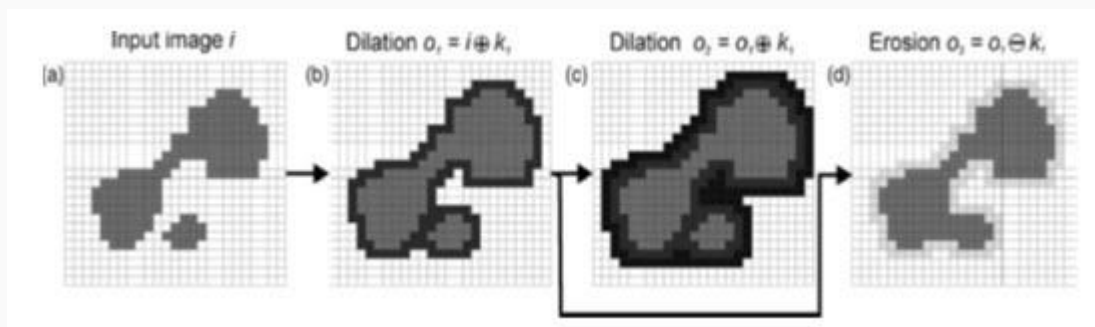
According to the pattern of its neighbouring values, the kernels are passed incrementally over the raster image, changing the central pixel each time. Therefore the incremental neighbourhood operation is similar to spatial filtering but with conditional rather than arithmetic rules controlling the modification of the central value. A simple dilation operation involves the growth or expansion of an object and can be described as:

$$o = i \oplus k \text{ or } \delta k(i) \quad (25.1)$$

Where  $o$  is the output binary raster,  $i$  is the input binary raster and  $k$  is the kernel which is centred on a pixel at  $i$ , and  $d$  indicates a dilation. The Minkowski summation of sets (a b) refers to all the pixels in  $a$  and  $b$ , in which  $i$  is the vector sum, and  $a$  belongs to set  $b$ , and  $b$  belongs to set  $a$  (Minkowski, 1911). The values of  $i$  are compared with the corresponding values in the kernel  $k$ , and are modified as follows: the value in  $o$  is assigned a value of 1 if the central value of  $i$  equals 1, or if any of the other values in  $k$  match their corresponding values in  $i$ ; if they differ, the resultant value in  $o$  will be 0. The result of this is only to modify the surrounding outer values by the morphology of the kernel. The effect of a dilation, using kernel  $k_1$ , is to add a rim of black pixels around the inner shapes and in doing so the two shapes in the binary image are joined into one, both having been dilated, as in Fig. 25.3b.

If the output  $o_1$  is then dilated again using  $k_1$ , then a second rim of pixels is added and so on.

By this process, the features are merged into one. Dilation is commonly used to create a map or image that reflects proximity to or distance from a feature or object, such as distance from road networks or proximity to major faults. These distance or 'buffer' maps often form an important part of multi-layer spatial analysis. (Source: Liu and Mason, 2009)



**Fig. 25.3. Dilation, erosion and closing: (a) the original image ( $i$ ); (b) dilation of  $i$  using  $k_1$  to produce  $o_1$ ; (c) dilation of  $o_1$  also using  $k_1$  to produce  $o_2$ ; and (d) erosion of  $o_1$  using  $k_1$  to produce  $o_3$ . Notice that  $o_3$  cannot be derived from  $i$  by a simple dilation using  $k_1$ ; the two objects are joined and this effect is referred to as closing. The pixels added by dilation are shown black and those pixels lost through erosion are shown with pale grey tones.**

(Source: Liu and Mason, 2009)

A simple erosion operation (a b) has the opposite effect, where  $i$  is a vector subtraction, so that it involves the shrinking of an object using the Minkowski subtraction, and is described by

$$o = i \ominus k \text{ or } \varepsilon k(i) \quad (10.2)$$

where  $\varepsilon$  indicates an erosion. The values in  $o$  are compared with those in  $k$  and if they are the same then the pixel is 'turned off' i.e. the value in  $o$  will be set to 0. The effect of using kernel  $k_1$ , is the removal of a rim of value 1 (grey) pixels from the edges of the feature shown in Fig. 25.3b to produce that shown in Fig. 25.3d. It can be noticed that the output  $o_3$ , which is the product of the sequential dilation of  $i$ , then erosion of  $o_1$ , results in the amalgamation of the two original objects and that the subsequent shrinking produces a generalized object which

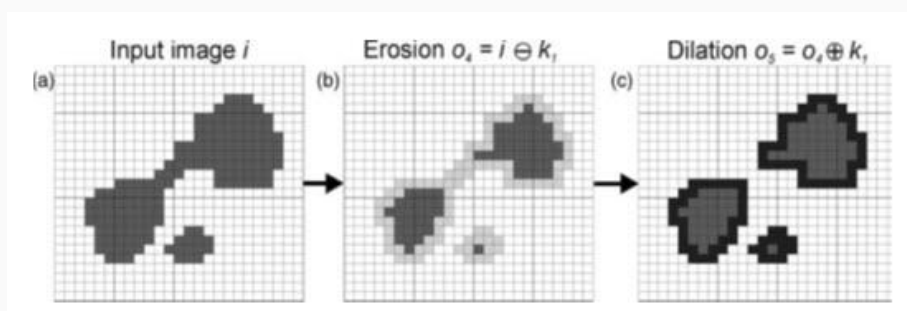
covers approximately the area of the original, an effect known as closing (dilation followed by erosion).

In Fig. 25.4b, an erosion operation is performed on the original  $i$ , removing one rim of pixels and causes the feature to be subdivided into two. When this is followed by a dilation, the result is to restore the two features to more or less their original size and shape except that the main feature has been split into two. This splitting is known as an opening and is shown in Fig. 25.4c.

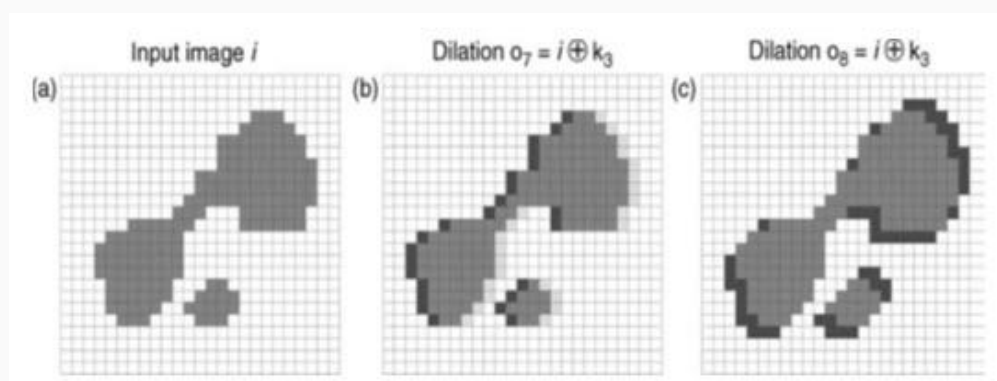
$$\text{Opening, } \gamma_k(i) = \delta_k[\varepsilon_k(i)] \quad (10.3)$$

$$\text{Opening, } \emptyset_k(i) = \varepsilon_k[\delta_k(i)]$$

Again, repeated dilations of the features after opening will not restore the features to their appearance in  $i$ .



**Fig. 25.4. Erosion, dilation and opening: (a) the original image ( $i$ ); (b) erosion of  $i$  using  $k$  to produce  $o_4$ ; (c) subsequent dilation of  $o_4$ , using  $k_1$ , to produce  $o_5$ . Note that the initial erosion splits the main object into two smaller ones and that the subsequent dilation does not restore the object to its original shape, an effect referred to as opening. (Source: Liu and Mason, 2009)**

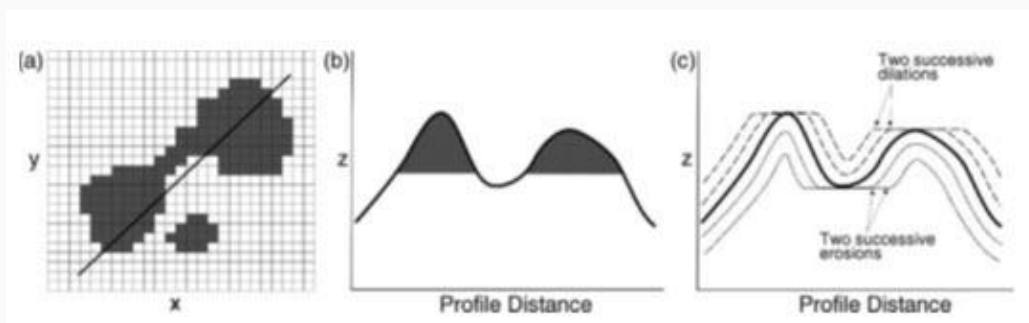


**Fig. 25.5. Anisotropic effects: (a) the original image ( $i$ ); (b) dilation of  $i$  using  $k$  to produce  $o_6$ , causing a westward shift of the object; and (c) dilation of  $i$  using  $k_3$ , producing an elongation in the NE-SW directions to produce  $o_7$ .**

(Source: Liu and Mason, 2009)

Closing can be used to generalize objects and to reduce the complexity of features in a raster. Opening can be used to perform a kind of sharpening or to add detail or complexity to the image. Dilation and erosion operations can also be carried out anisotropically in specific directions. Such directional operations are often relevant in geological applications where there is some kind of structural or directional control on the phenomenon of interest.

For example, the effect of kernel  $k_2$  on  $i$  is shown in Fig. 25.5a, where the effect is a westward shift of the features by 1 pixel. The effect of kernel  $k_3$  is to cause dilation in the NW-SE directions, resulting in an elongation of the feature (Fig. 25.5b). To consider the effect of mathematical morphology on continuous raster data, we can take the binary image (i) shown in Figs. 25.2–25.5 to represent a density slice through a raster surface, such as an elevation model. In this case, the darker class represent the geographical extent of areas exceeding a certain elevation value. Fig. 25.6a shows the binary image and a line of profile (Fig. 25.6b) across a theoretical surface which could be represented by image (i). The effect of simple dilation and erosion of the surface is shown in Fig. 25.6c.



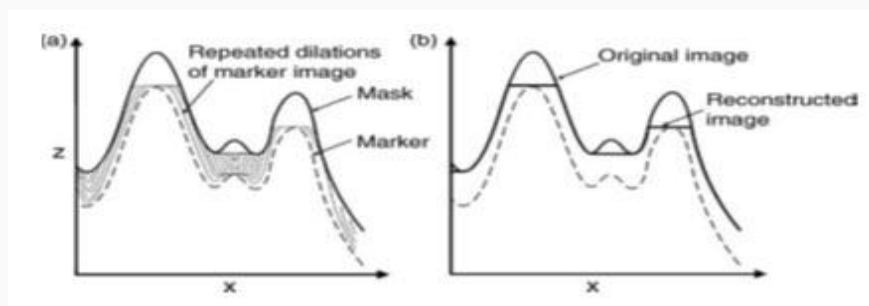
**Fig. 25.6 (a) The original input image with the position of a profile line marked; (b) the theoretical cross-sectional profile with the shaded area representing the geographical extent of the darker class along the line shown in (a); and (c) the effect on the profile of dilations and erosions of that surface.**

**(Source: Liu and Mason, 2009)**

Here it can be seen that dilations would have the effect of filling pits or holes, and broaden peaks in the surface, while erosions reduce the peaks or spikes, and widen depressions. Such techniques could therefore be used to correct errors in generated surfaces such as DEMs. But the errors in DEMs cannot be properly corrected by merely smoothing. So for correction of DEM errors a modification of the mathematical morphology technique, known as morphological reconstruction has been proposed. In this case, the original image is used as a mask and the dilations and erosions are performed iteratively on a second version of the same image until stability between the mask and marker images is reached and the image is fully reconstructed and no longer changes, when the holes are corrected (Fig. 25.7).

Since morphological reconstruction is based on repeated dilations, rather than directly modifying the surface morphology, it works by controlling connectivity between areas. The marker could simply be created by making a copy of the mask and either subtracting or adding a constant value. The error-affected raster image is then used as the mask, and the marker is dilated or eroded repeatedly until it is constrained by the mask, i.e. until there is no change between the two, and the process then stops. By subtracting a constant from the

marker and repeatedly dilating it, extreme peaks can be removed, whereas by adding a constant and repeatedly eroding the marker, extreme pits would be removed. The extreme values are effectively reduced in magnitude, relative to the entire image value range, in the reconstructed marker image. This technique can be used selectively to remove undesirable extreme values from DEMs.



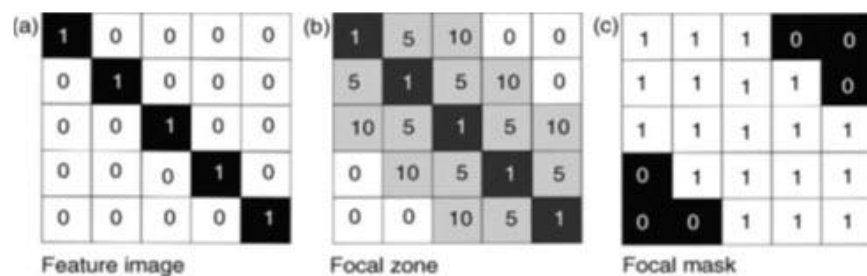
**Fig. 25.7. Mechanism of morphological reconstruction of an image, as illustrated by a profile across the image: (a) in this case, by repeated dilations of the marker until it is constrained by the mask image; (b) the extreme peaks are reduced in magnitude in the reconstructed image.**

(Source: Liu and Mason, 2009)

### 25.1.2 Extended neighbourhood

The term 'extended neighbourhood' is used to describe operations whose effects are constrained by the geometry of a feature in a layer and performed on the attributes of another layer. These extended neighbourhood operations can be further described as focal and zonal.

If for instance slope angles must be extracted from within a corridor along a road or river, the corridor is defined from one layer and then used to constrain the extent of the DEM from which the slope angle is then calculated (Fig. 25.8).



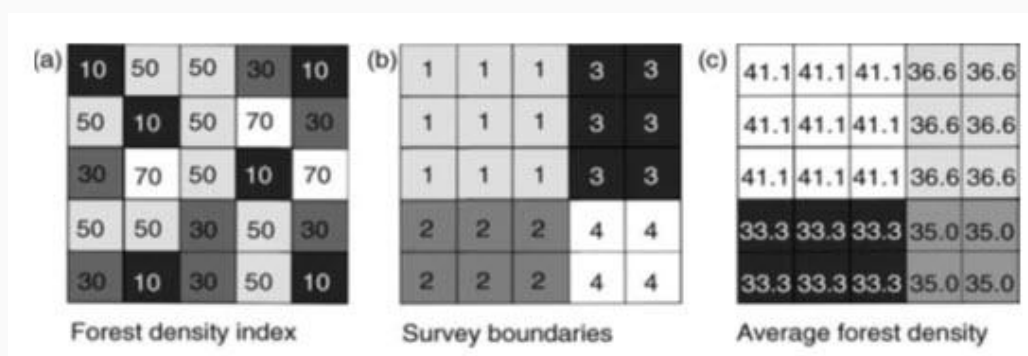
**Fig. 25.8. Focal statistics: (a) a binary image representing a linear target feature (coded with a value of 1 for the feature and 0 for the background); (b) a 10m focal image created around the linear feature, where each pixel is coded with a value representing its distance from the feature (assuming that the pixel size is 5m×5 m), areas beyond 10m from the feature remain coded as 0; and (c) binary focal zone mask with values of 1 within the mask and zero outside it. This has a similar effect to a dilation followed by a reclassification, to produce a distance buffer. (Source: Liu and Mason, 2009)**

### 25.1.2.1 Focal operations

Working within “neighborhoods” this class of operations implements a moving window algorithm to modify cell values in the input layer (Lein). A focal operation is used for generating corridors and buffers around features. Focal operations are those that derive each new value as a function of the existing values, distances and/or directions of neighbouring locations. The relationships may be defined by Euclidean distance, travel cost, engineering cost or inter-visibility. Such operations could involve measurement of the distance between each pixel (or point) position and a target feature(s). A buffer can then be created by reclassification of the output ‘distance’ layer. This allows specific values to be set for the original target features, with the buffer zones and for the areas beyond the buffers.

### 25.1.2.2 Zonal operations

These operations are applied to define regions or zones in the input surface. Zones explain collection of cells that exhibit similar attributes which can be created using (1) reclassification calculations based on area, shape or perimeter or (2) categorical overlay developed from binary “cookie cutters” to extract cell values from a raster layer (Lein). A zonal operation also involves the use of the spatial characteristics of a zone or region defined on one layer, to operate on the attribute(s) of a second layer or layers. Since zonal operations most commonly involve two layers, this process falls into the binary operations category. An example is given in Fig. 25.9 where zonal statistics are calculated from an input layer representing the density of forest growth, within the spatial limits defined by a second survey boundary layer, to provide an output representing, in this case, the average forest density within each survey unit. Here it can be seen that the two raster inputs contain integer values but that the output values are floating-point numbers, as is always the case with mean calculations.



**Fig. 25.9. Zonal statistics: (a) forest density integer image; (b) survey boundaries (integer) image; and (c) the result of zonal statistics (in this case a zonal mean) for the same area. Note that this statistical operation returns a non-integer result. (Source: Liu and Mason, 2009)**

## Lesson 26 Vector and Raster Map Algebra

A Geographic Information System stores two types of data that are found on a map the geographic definitions of earth surface features and the attributes or qualities that those features possess. Not all systems use the same logic for achieving this. Most, however, use one of two fundamental map representation techniques: vector and raster. (Ronald, 1993)

### 26.1 Vector

With vector representation, the boundaries or the course of the features are defined by a series of points that, when joined with straight lines, form the graphic representation of that feature. The points themselves are encoded with a pair of numbers giving the X and Y coordinates in systems such as latitude/longitude or Universal Transverse Mercator grid coordinates. The attributes of features are then stored with a traditional database management (DBMS) software program. For example, a vector map of property parcels might be tied to an attribute database of information containing the address, owner's name, property valuation and land use. The link between these two data files can be a simple identifier number that is given to each feature in the map.

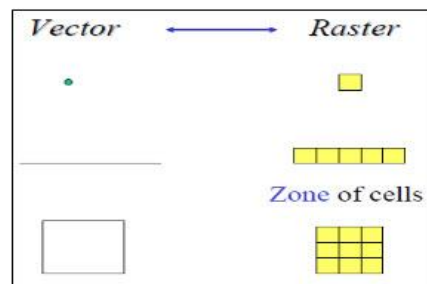
(Ronald, 1993)

### 26.2 Raster

The second major form of representation is known as raster. With raster systems, the graphic representation of features and the attributes they possess are merged into unified data files. In fact, they typically do not define features at all. Rather, the study area is subdivided into a fine mesh of grid cells in which the condition or attribute of the earth's surface at each cell point is recorded. Each cell is given a numeric value which may then represent a feature identifier, a qualitative attribute code or a quantitative attribute value.

For example, a cell could have the value "6" to indicate that it belongs to District 6 (a feature identifier), or that it is covered by soil type 6 (a qualitative attribute) or that it is 6 meters above sea level (a quantitative attribute value). Although the data we store in these grid cells do not necessarily refer to phenomena that can be seen in the environment, the data grids themselves can be thought of as images -- images of some aspect of the environment that can be made visible through the use of a raster display. In a raster display, such as the screen on your computer, there is also a grid of small cells called pixels. The term pixel is a contraction of picture element. Pixels can be made to vary in their colour, shape or grey tone. To make an image, the cell values in the data grid are used to regulate directly the graphic appearance of their corresponding pixels. Thus in a raster system, the data directly control the visible form we see. Vector and raster data can be represented as shown in Fig. 26.1. (Ronald, 1993).





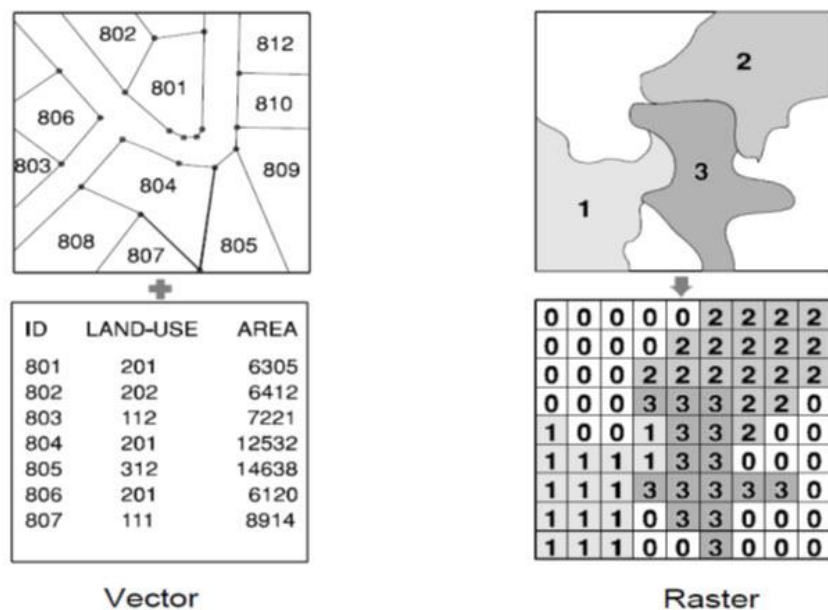
**Fig. 26.1. Vector and Raster Data Representation. (Source: Ronald, 1993).**

### 26.3 Raster versus Vector

Raster systems are typically data intensive (although good data compaction techniques exist) since they must record data at every cell location regardless of whether that cell holds information that is of interest or not. However, the advantage of the raster data structure is that geographical space is uniformly defined in a simple and predictable fashion. As a result, raster systems have substantially more analytical power than their vector counterparts in the analysis of continuous space and are thus ideally suited to the study of data that are continuously changing over space such as terrain, vegetation biomass, rainfall and the like. The second advantage of raster is that its structure closely matches the architecture of digital computers. As a result, raster systems tend to be very rapid in the evaluation of problems that involve various The basic data structure of vector systems can best be described as a network. As a result, it is not surprising to find that vector systems have excellent capabilities for the analysis of network space. Thus the difference between raster and vector is less one of inherent ability as it is of the difference in the types of space they describe. Mathematical combinations of the data in multiple grids. Hence they are excellent for evaluating environmental models such as those for soil erosion potential and forest management suitability. In addition, since satellite imagery employs a raster structure, most raster systems can easily incorporate these data and some provide full image processing capabilities. While raster systems are predominantly analysis oriented, vector systems tend to be more database management oriented.

Vector systems are quite efficient in their storage of map data because they only store the boundaries of features and not what is inside those boundaries. Because the graphic representation of features is directly linked to the attribute database, vector systems usually allow one to roam around the graphic display with a mouse and inquire about the attributes of any displayed feature: the distance between points or along lines, the areas of regions defined on the screen, and so on. In addition, they can produce simple thematic maps of database queries such as, "show all sewer line sections over one meter in diameter installed before 1940." (Ronald, 1993)





**Fig. 26.2. Comparison between Vector and Raster data representation. (Source: Ronald, 1993)**

Compared to their raster counterparts, vector systems do not have as extensive a range of capabilities for analysis over continuous space. However, they do excel at problems concerning movements over a network and can undertake the most fundamental of GIS operations. For many, it is the simple database management functions and excellent mapping capabilities that make vector systems attractive. Because of the close affinity between the logic of vector representation and traditional map production, a pen plotter can be driven to produce a map that is indistinguishable from that produced by traditional means. As a result, vector systems are very popular in municipal applications where issues of engineering map production and database management predominate. Raster and vector systems each have their special strengths. Some GIS incorporate elements from both representational techniques. Many systems provide most functions for one technique and provide limited display and data transfer functions using the other. Which technique is most appropriate depends upon the application. A complete GIS setup may include a vector system, a raster system, or both, depending upon the types of tasks that must be done. While some applications are suitable to either vector or raster, usually one is more appropriate. Using a system that is not well suited to a particular task can be very frustrating and lead to unsatisfactory results. (Ronald, 1993)

## 26.4 Raster Calculator

The Raster Calculator provides you a powerful tool for performing multiple tasks. You can perform mathematical calculations using operators and functions, set up selection queries, or type in Map Algebra syntax. Up to 4 rasters can be used in a single expression. Inputs can be raster datasets or raster layers, coverage's, shape files, tables, constants and numbers. The expressions are evaluated by the Raster Calculator on the fly and the user is provided with as status of the formula as he/she builds it. (<http://www.ian-ko.com>)

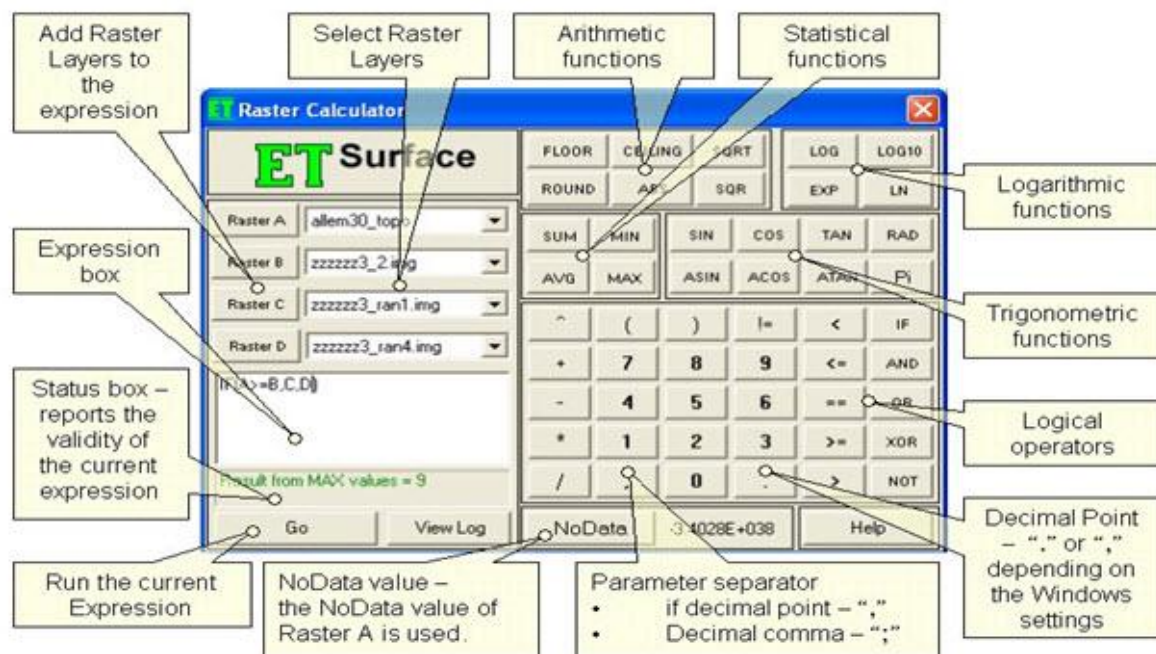


Fig. 26.3. Raster calculator showing expressions, functions and layers.  
(Source: <http://www.ian-ko.com>)

### Inputs:

- **Rasters** - Up to 4 rasters can be used. If the Raster Calculator is used from the GUI, the rasters are selected from the raster layers loaded in Arc Map. In the Toolbox implementation the input can be a raster layer or raster dataset. The 4 rasters are called Raster A, Raster B, Raster C and Raster D. Raster A is required, the other rasters are optional.
- **Expression** - the formula to be used for the calculation to be performed. For shortness the rasters should be entered with their letters in the expression - A for Raster A, B for Raster B, etc. All the functions available can be typed in the expression box or selected from the calculator buttons provided. The functions are not case sensitive - SIN, Sin and sin will be accepted as correct entries. Note that the operator for EQUAL is "==" and NOT "=" (which is operator for assignment). The syntax of all functions is discussed below.

### Output:

If the Raster Calculator is used from the GUI, the raster dataset created when an expression is executed is a temp raster and is stored in the temp folder of ET Surface. If you want to save it as a permanent raster, use the Export Data tool.

If the Toolbox implementation is used, the user is asked for an output name and location and the raster dataset created is permanent.

(Source: <http://www.ian-ko.com>)

The output raster dataset will

- Be FLOAT type
- Have the cell size of the Raster A (if any of the other rasters used have a different cell size, it will be resampled).
- The extent will be calculated as the intersection of the extents of the input rasters.

(Source: <http://www.ian-ko.com>)

### Functions performed by raster Calculator:

- The Raster Calculator tool allows for creating and executing a Map Algebra expression that will output a raster.
- Use the Layers and variables list to select the datasets and variables to use in the expression.
- Numerical values and mathematical operators can be added to the expression by clicking the respective buttons in the tool dialog box. A list of commonly used conditional and mathematical tools is provided, allowing you to easily add them to the expression.
- Full paths to data or data existing in the specified current workspace environment setting can be entered in quotes (""). Numbers and scalars can be directly entered into an expression.

(<http://help.arcgis.com>)

Most conventional vector data models maintain data as multiple attribute maps, e.g. forest inventory polygons linked to a database table containing all attributes as columns. This basic distinction of raster data storage provides the foundation for quantitative analysis techniques. This is often referred to as raster or map algebra. (<http://planet.botany.uwc.ac.za>).

#### 26.4.1 Map Algebra

This is in contrast to most conventional vector data models that maintain data as multiple attribute maps, e.g. forest inventory polygons linked to a database table containing all attributes as columns. This basic distinction of raster data storage provides the foundation for quantitative analysis techniques. This is often referred to as raster or map algebra. (<http://planet.botany.uwc.ac.za>).

Map algebra is a simple and an elegant set-based algebra for manipulating geographic data, proposed by Dr. Dana Tomlin in the early 1980s. It is a set of primitive operations in a geographic information system (GIS) which allows two or more raster layers ("maps") of similar dimensions to produce a new raster layer (map) using algebraic operations such as addition, subtraction etc. A set of tool that a GIS will typically provide is that for combining map layers mathematically. Modelling, in particular, requires that we be able to combine maps according to various mathematical combinations.

For example, we might have an equation that predicts mean annual temperature as a result of altitude. Or, as another example, consider the possibility of creating a soil erosion potential map based on factors of soil erosion, slope gradient and rainfall intensity. Clearly we need the ability to modify data values in our maps by various mathematical operations and transformations and to combine factors mathematically to produce the final result.

The Map Algebra tools will typically provide three different kinds of operations:

1. The ability to arithmetically modify the attribute data values over space by a constant (i.e., scalar arithmetic).
2. The ability to mathematically transform attribute data values by a standard operation (such as the trigonometric functions, log transformations and so on).
3. The ability to mathematically combine (such as add, subtract, multiply, divide) different data layers to produce a composite result.

This third operation is simply another form of overlay mathematical overlay, as opposed to the logical overlay of database query. To illustrate this, consider a model for snow melt in densely forested areas:

$$M = (0.19T + 0.17D) \quad (26.1)$$

where M is the melt rate in cm/day, T is the air temperature and D is the dew point temperature. Given maps of the air temperatures and dew points for a region of this type, we could clearly produce a snow melt rate map. To do so would require multiplying the temperature map by 0.19 (a scalar operation), the dew point map by 0.17 (another scalar operation) and then using overlay to add the two results. This ability to treat maps as variables in algebraic formulas is an enormously powerful capability.

Map algebra provides one method to run spatial analyst tool. The Raster Calculator provides a powerful tool for performing multiple tasks. One can perform mathematical calculations using operators and functions, set up selection queries, or type in Map Algebra syntax. Inputs can be raster datasets or raster layers, coverages, shape files, tables, constants, and numbers. The set of operators is composed of arithmetical, relational, Boolean, bitwise, and logical operators that support both integer and floating-point values and combinatorial operators. (Ronald, 1993)

#### 26.4.2 Raster data formats

Raster data models incorporate the use of a grid-cell data structure where the geographic area is divided into cells identified by row and column. This data structure is commonly called raster. While the term raster implies a regularly spaced grid other tessellated data structures do exist in grid based GIS systems. In particular, the quadtree data structure has found some acceptance as an alternative raster data model. (<http://planet.botany.uwc.ac.za>).

A raster data structure is in fact a matrix where any coordinate can be quickly calculated if the origin point is known, and the size of the grid cells is known. Since grid-cells can be handled as two-dimensional arrays in computer encoding many analytical operations are easy to program. This makes tessellated data structures a popular choice for many GIS software. Several tessellated data structures exist, however only two are commonly used in GIS's. The most popular cell structure is the regularly spaced matrix or raster structure. The use of raster data structures allow for sophisticated mathematical modelling processes while vector based systems are often constrained by the capabilities and language of a relational DBMS. (<http://planet.botany.uwc.ac.za>).

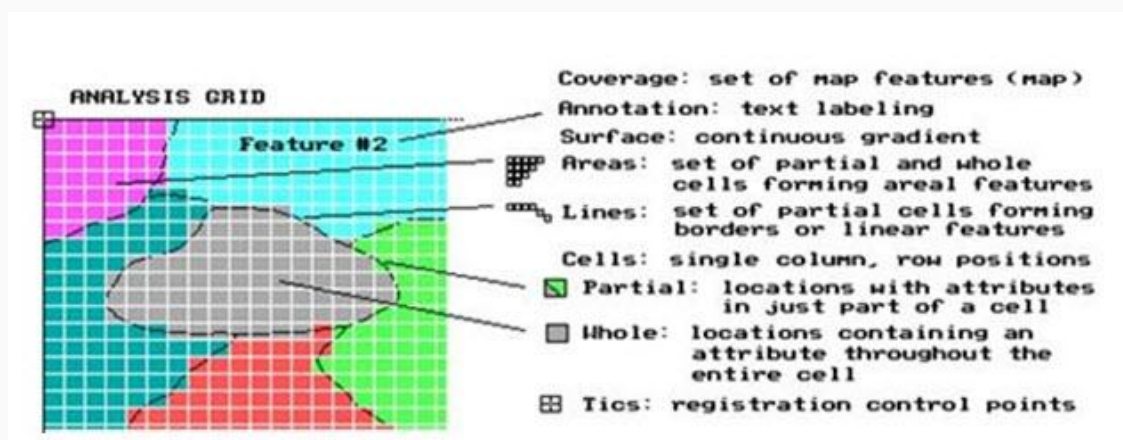


Fig. 26.4. Raster data formats. (Source: <http://planet.botany.uwc.ac.za>)

Since geographic data is rarely distinguished by regularly spaced shapes, cells must be classified as to the most common attribute for the cell. The problem of determining the proper resolution for a particular data layer can be a concern. If one selects too coarse a cell size then data may be overly generalized. If one selects too fine a cell size then too many cells may be created resulting in a large data volume, slower processing times, and a more cumbersome data set. As well, one can imply accuracy greater than that of the original data capture process and this may result in some erroneous results during analysis. So analysis mask. (<http://planet.botany.uwc.ac.za>).

Resampling or interpolation (and reprojection) of inputs to target extent, cell size, and projection within region defined by analysis mask.

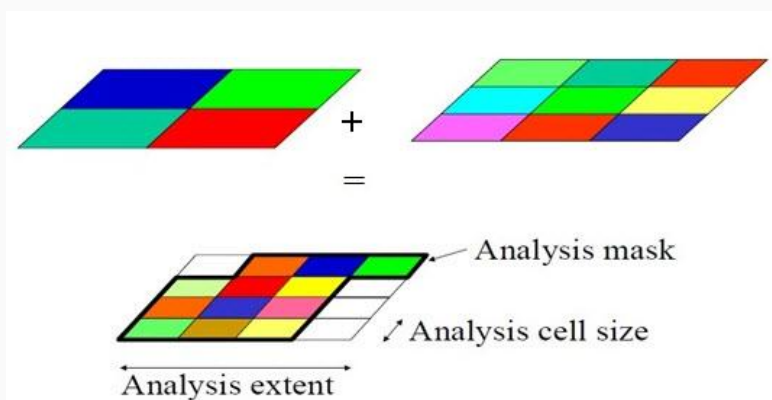


Fig. 26.5. Mask Analysis. (Source: Slides)



### 26.4.3 Vector-raster conversion

As well, since most data is captured in a vector format, e.g. digitizing, data must be converted to the raster data structure. This is called vector-raster conversion. Most GIS software allows the user to define the raster grid (cell) size for vector-raster conversion. It is imperative that the original scale, e.g. accuracy, of the data be known prior to conversion. The accuracy of the data, often referred to as the resolution, should determine the cell size of the output raster map during conversion. (<http://planet.botany.uwc.ac.za>).

#### 26.4.3.1 Extracting information from surface

Some tools extract vector features from surfaces, or produce tabular summaries or smaller raster samples of surfaces. (<http://planet.botany.uwc.ac.za>).

#### Sampling rasters

The Sample tool creates a table that shows the values of a raster, or several rasters, at a set of sample point locations. The points can be in a point feature class or the cells in a raster that have values other than No Data. You might use this tool to get information about what occurs at a set of points, such as bird nesting sites, from terrain, distance to water, and forest type rasters.

(<http://resources.arcgis.com>).

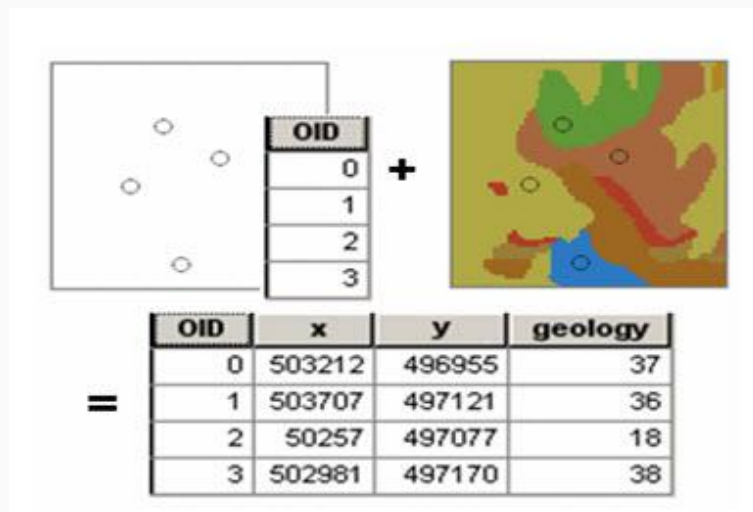
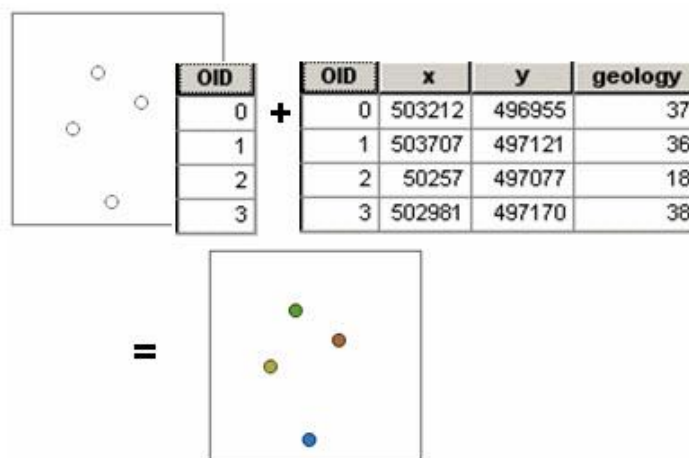


Fig. 26.6. Geology Raster Being Sampled at a set of points.

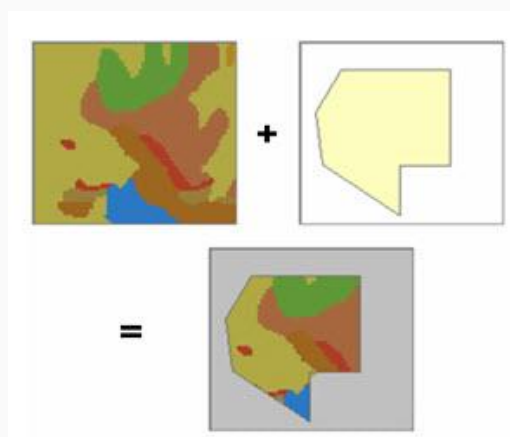
(Source: <http://resources.arcgis.com>)

The output table can be analyzed on its own or joined to the sample point features. Below is an example of the sample results table joined back to the original sample points.



**Fig. 26.7. Sample results Table joined back to the original sample points.**  
(Source: <http://resources.arcgis.com>)

The Extract tools create a new raster with a copy of the cells within some mask area. The Extract By Mask tool lets you use a polygon feature class to extract the raster data.



**Fig. 26.8. New Raster which has been created.** (<http://resources.arcgis.com>)

The Extract Values to Points tool creates a new feature class of points with the values of a single raster at a set of input point features. The Extract By Attributes tool selects cells of a raster based on a logical query. Extract By Polygon and Extract By Rectangle take lists of coordinate values that define an area and output a raster that is either inside or outside the polygon. Extract By Circle takes the centre coordinates and radius of a circle and outputs a raster that is either inside or outside the circle. Extract By Points takes a list of coordinate values that define a set of points and outputs a raster of the cell values at these points (or excluding these points). In all cases, the cells from the original raster that are not part of the Extract area are given No Data values. The 3D Analyst Surface Spot tool extracts elevation values from a surface for a set of point features and adds them to a Spot attribute of the points.

(<http://resources.arcgis.com>)

## 26.5 Application of Raster Calculator

In ArcGIS 10 the Spatial Analyst toolbox includes a Raster Calculator geoprocessing tool in the Map Algebra toolset. This is not the same raster calculator as in previous versions of ArcGIS, so keep reading to find out what it does, how it's improved, and where to find more information.

The Raster Calculator geoprocessing tool in ArcGIS 10 is designed to execute a single-line map algebra expression using multiple tools and operators listed on the tool dialog. When multiple tools or operators from the tool dialog are used in one expression, the performance of this equation will generally be faster than executing each of the operators or tools individually. (<http://blogs.esri.com>)

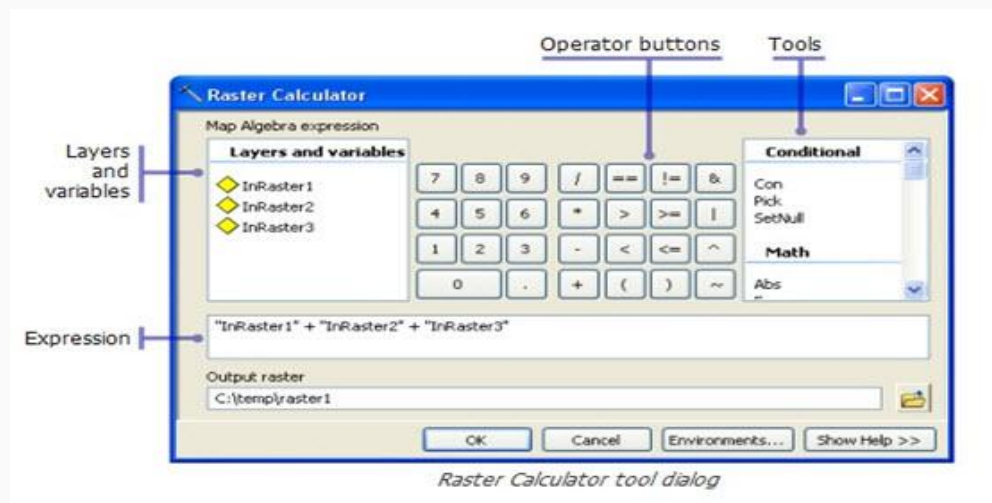


Fig. 26.9. Raster calculator Tool Dialog. (Source: <http://blogs.esri.com>)

The Raster Calculator tool has been designed to replace both the previous Raster Calculator from the Spatial Analyst toolbar and the Single Output Map Algebra geoprocessing tool. The Raster Calculator tool is like all other geoprocessing tools; it honors geoprocessing environment settings, it can be added to Model Builder, and when used in Model Builder it supports variables in the expression. The ability to support variables in the expression makes the new Raster Calculator tool much more powerful and versatile than previous Map Algebra implementations.

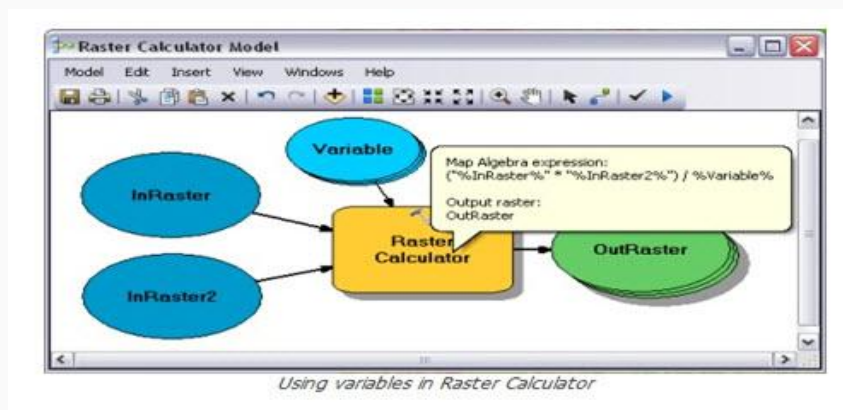


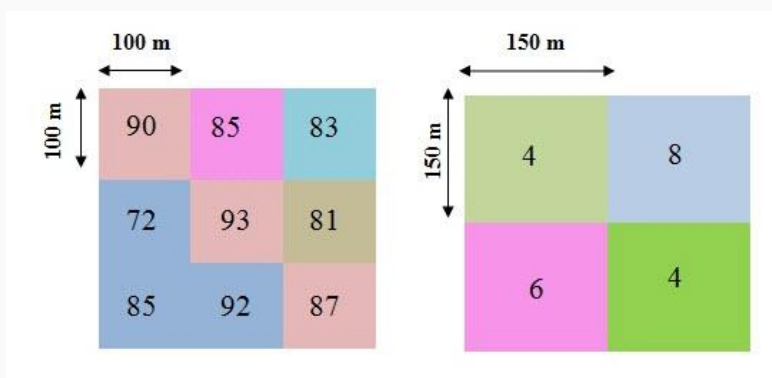
Fig. 26.10. Raster Calculator Model. (Source: <http://blogs.esri.com>)



The Raster Calculator tool is used to execute Map Algebra expressions inside ArcGIS applications. The Raster Calculator is not supported in scripting because in ArcGIS 10 Map Algebra can be accessed directly when using the geoprocessing ArcPy site-package. This seamless integration of Map Algebra into Python extends the capabilities of Map Algebra by taking advantage of Python and third party Python modules and libraries; making Map Algebra far more powerful than it has been in the past. The Map Algebra language in ArcGIS 10 is similar to 9.x Map Algebra with minor syntax changes due to the integration of Python; most notably case sensitivity. (<http://blogs.esri.com>)

## 26.6 Raster Calculation Example

The grids below depict initial snow depth and average temperature over a day for an area.



One way to calculate decrease in snow depth due to melt is to use a temperature index model that uses the formula

$$D_{new} = D_{old} - mT \quad (26.2)$$

Here  $D_{old}$  gives the snow depth at the beginning and end of time step,  $T$  gives the temperature and  $m$  is melt factor  $m=0.5 \text{ cm}/^{\circ}\text{C}/\text{day}$ . Calculate the snow depth at the end of the day.

### Solution:

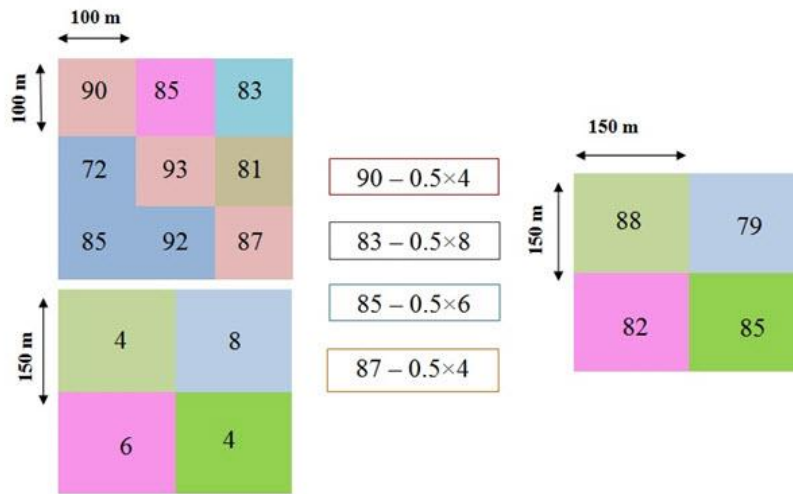
Here using the [Eqn(26.2)] you can write

$$\text{New depth at 100m} = [\text{Snow 100m}] - 0.5 \times [\text{temp at 150m}] \quad (26.3)$$

### Converting the outputs to 150m grid:

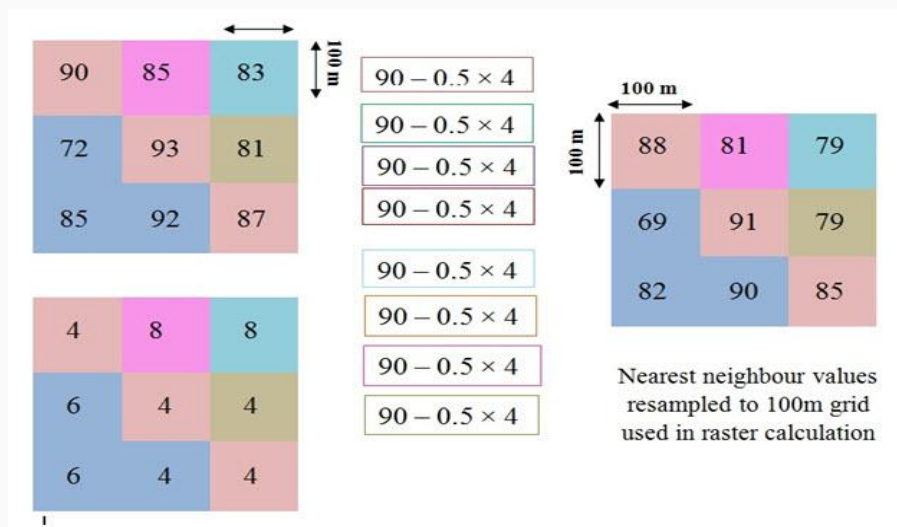
For conversion of 100m grid to 150m grid, it will follow the nearest neighbourhood rule.

First you have to convert the 100m grid to 150m grid. Then you have to choose the nearest grid. And after choosing the nearest grid apply the above formula.



### Converting temperature to 100m grid:

For conversion follow the same nearest neighbour to the East and south for obtaining a 100 temperature grid. Then use the Eqn (26.3).



## 26.7 Characteristics of Vector Data

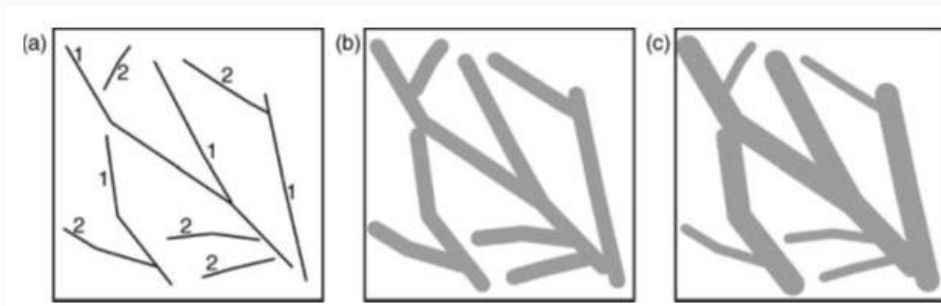
There is no limit to the attribute information which can be stored or linked to a particular feature object. Tabular data represent a special form of vector data which can include almost any kind of data, whether or not they contain a geographic component; tabular data are not necessarily spatial in nature.

A table whose information includes and is referenced by coordinates can be displayed directly on a map. The information which does not must be linked to other spatial data that do have coordinates before it can be displayed on a map. Vector data therefore consist of a series of discrete features described by their coordinate positions than graphically or in any regularly structured way. The vector model could be thought of as the opposite of raster data in this respect, since it does not fill the space it occupies; not every conceivable location is represented, only those where some feature of interest exists.

If we were to choose the vector model to represent some phenomenon that varies continuously and regularly across a region, such that the vector data necessarily become so densely populated as to resemble a raster grid, then we would probably have chosen the wrong data model for those data. (Liu and Mason, 2009).

### 26.7.1 Buffers

A zone calculated as the Euclidean distance from existing vector features, such as roads, is referred to as a buffer. Buffers are calculated at constant distance from the feature or at distances dictated by attribute values, and each zone will be the same width around the feature (see Fig. 26.11).



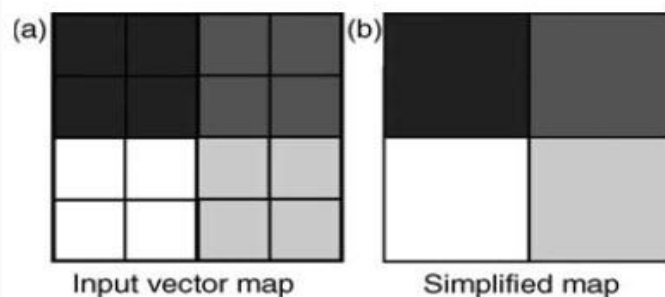
**Fig. 26.11. (a) Simple vector line feature map, labelled with attribute values (1 and 2); (b) output with buffers of constant distance; (c) output map with buffers of distance defined by the attribute values shown in (a).**

(Source: Liu and Mason, 2009).

Features with attribute value 1 having buffers twice the distance of those of features with attribute value 2. No account is taken of the Earth's curvature, so the zones will be at the same width regardless of the coordinate system. Negative distance values can be used, and these will cause a reduction in the size of the input feature. Buffers can also be generated on only one side of input features (should this be appropriate). The input layer in this case is a vector feature but the output may be a polygon file or raster. The same buffering operation can also be applied to raster data by first calculating the Euclidean distance and then reclassifying the output to exclude distances within or beyond specified thresholds; the output will always be a raster in this case. Buffering in this way can be considered as the vector equivalent of conditional logic combined with raster dilation or erosion. (Liu and Mason, 2009).

### 26.7.2 Dissolve

When boundaries exist between adjacent polygon or line features, they could be removed or dissolved because they have the same or similar values for a particular attribute (see Fig. 26.12).

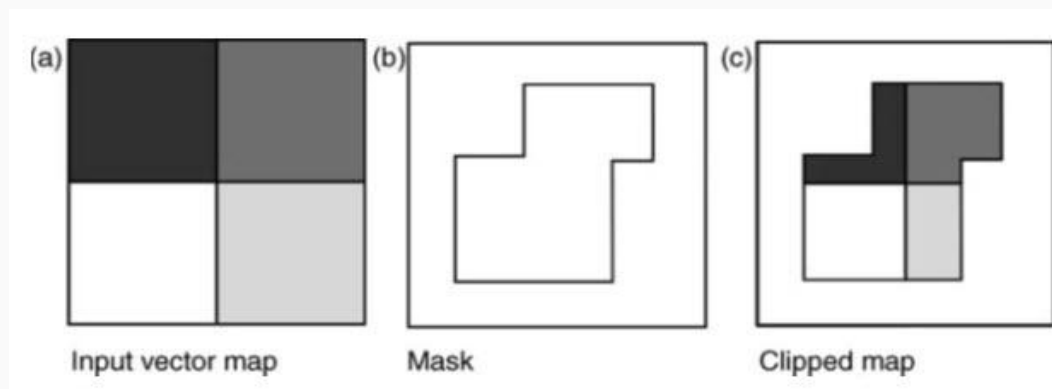


**Fig. 26.12. Vector polygon features (a) and the Dissolved and simplified output map (b).**  
(Source: Liu, and Mason, 2009)

As in a geological map where adjacent litho logical units with similar or identical descriptions can sensibly be joined into one, the boundaries between them are removed by this process and the classes merged intone. Complications in the vector case arise if the features' attribute tables contain other attributes (besides the one of interest being merged) which differ across the boundary; choices must be made about how those other attributes should appear in the output dissolved layer. This is equivalent to merging raster classes through reclassification, or raster generalization/simplification. (Liu, and Mason, 2009).

### 26.7.3 Clipping

The geometry of a feature layer can be used as a mask to extract selectively a portion of another layer; the input layer is thereby clipped to the extent of the mask (see Fig. 26.13). The feature layer to be clipped may contain point, line or polygon features but the feature being used as a mask must have area, i.e. it will always be polygon.



**Fig. 26.13. Vector polygon clipping, using an input vector layer from which an area will be extracted (a), the vector feature whose geometric properties will be used as the mask (b) and (c) the vector output clipped feature.** (Source: Liu and Mason, 2009)

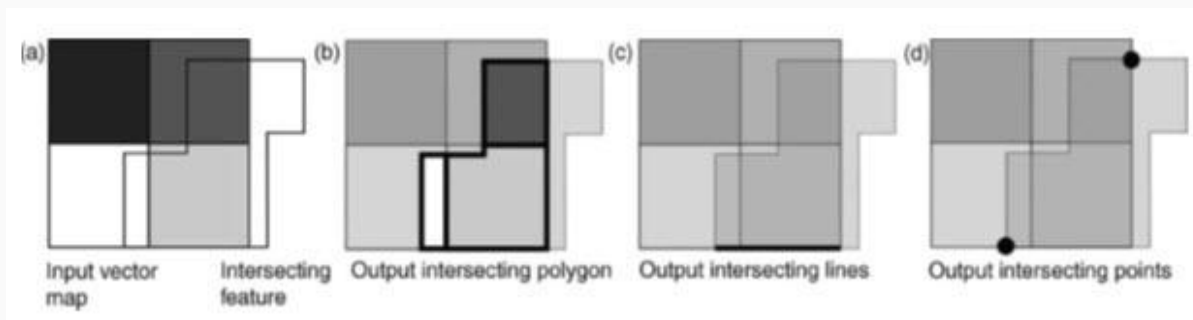
The output feature attribute table will contain only the fields and values of the extracted portion of the input vector map, as the attributes of the mask layer are not combined. Clipping is equivalent to a binary raster zonal operation, where the pixels inside or outside the region are set as null, using a second layer to define the region or mask. (Liu and Mason, 2009).

### 26.7.4 Intersection

If two feature layers are to be integrated while preserving only those features that lie within the spatial extent of both layers, an intersection can be performed (see Figure 26.14).

This is similar to the operation except that the two input layers are not necessarily of the same feature type. The input layers could be point, line and/or polygon, so the output features could also be point, line and/or polygon in nature.

New vertices need to be created to produce the new output polygons, lines endpoints, through a process called cracking.

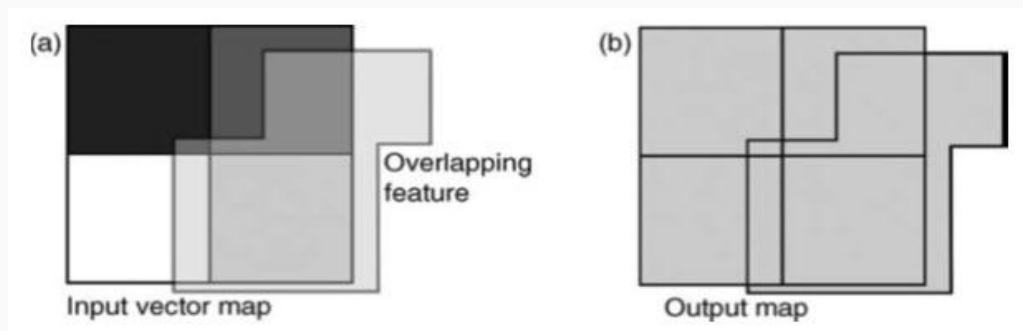


**Fig. 26.14. Intersection operation between two overlapping polygon features (a); the output intersecting polygon (b) which covers the extent and geometry of the area which the two inputs have in common; the intersecting line (c) and points (d) shared by both polygons.**

**The output attribute table contains only those fields and values that exist over the common area, line and points. (Source: Liu and Mason, 2009)**

Unlike the clip operation, the output attribute table contains fields and values from both input layers, over the intersecting feature/area. In the case of two intersecting polygons, intersection is equivalent to a Boolean operation using a logical AND (Min) operator between two overlapping raster images.

When two input overlapping feature layers are required to be integrated such that the new output feature layer contains all the geometric features and attributes of two input layers, the union operation can be used (see Fig. 26.15).



**Fig. 26.15. Vector polygon union operation where two polygon features overlap (a) and the output object (b) covers the extent and geometry of both inputs. The output attribute table also contains the attribute fields and values of both input features. (Source: Liu and Mason, 2009)**

Since vector feature layers can contain only points or only polygons, here the inputs must be of the same type but the number of inputs is not limited to two. Again, new vertices will be created through cracking. This is similar to the intersect operation but the output will have the total extent of the input layers. New, minor polygons are created wherever polygons overlap. The attribute table of the output layer contains attribute fields of both the input layers, though some of the entries may be blank. In the polygon case, it is equivalent to a binary raster operation using logical OR (Max) operator between overlapping images. (Liu and Mason, 2009).



## Module 10: Fundamentals of Global Positioning System

### Lesson 27 GPS

#### 27.1 Components of GPS

GPS is a Satellite Navigation System. It stands for Global Positioning System. GPS is funded by and controlled by the U. S. Department of Defense (DOD). While there are many thousands of civil users of GPS world-wide, the system was designed for and is operated by the U. S. military. GPS provides specially coded satellite signals that can be processed in a GPS receiver, enabling the receiver to compute position, velocity and time. Four GPS satellite signals are used to compute positions in three dimensions and the time offset in the receiver clock. ( <http://www.colorado.edu/geography/gcraft/notes/gps/gps.html#SVData>; November 30.2012).

Here we will mostly focus on the NAVSTAR, the mostly available and used system. GPS consists of mainly three segments, these are

1. Space segment (the satellites )
2. Control segment (the ground stations)
3. User segment (user and their GPS receiver)

##### 27.1.1 Space segment

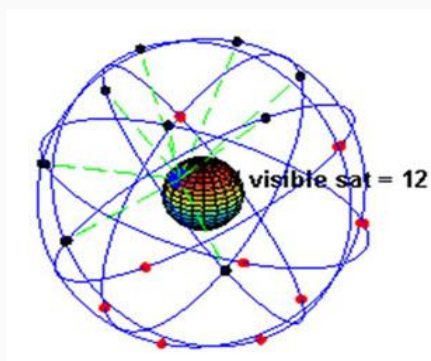
The space segment (SS) is composed of the orbiting GPS satellites, or Space Vehicles (SV). It consists of 24 satellites (21 active plus 3 operating spares), eight each in three circular orbital planes. The satellites are manufactured by Rockwell International, which are launched into space by rockets, from Cape Canaveral, Florida. They are about the size of a car, and weigh about 19,000lbs. This was modified to six planes with four satellites each. The orbital planes are centered on the Earth. The six planes have approximately 55° inclination and are separated by 60° right ascension of the ascending node. This constellation ensures that there will always be at least 4 satellites above the horizon at any location on the surface of the globe.



**Fig. 27.1. Space segment.** (Source: <http://www.aboutcivil.org/components-of-gps-working-mechanism.html>; November 05, 2012)

The space segment transmits a variety of information particularly time and satellite position information. Satellites are orbiting at an altitude of approximately 20, 200 kilometers. The satellites travel at a speed of 7000 miles/h, which allows them circle the earth once in every 12 hours. They are powered by solar energy and are built to last about 10 years. The satellites continuously orient themselves to ensure that their solar panels stay pointed towards the sun, and their antennas point toward the earth. Each satellite carries 4 atomic clocks.

The first GPS satellites are launched into space in 1978. A full consolidation of 24 satellites was achieved in 1994. The satellites are geostationary and non-geostationary. At any given time there are 12 satellites at either side of the hemispheres.



**Fig. 27.2. GPS constellation.** (Source: <http://www.aboutcivil.org/components-of-gps-working-mechanism.html> ;November 05, 2012)

Each satellite consists of three high precision atomic clocks and is constantly transmitting radio signals using its own unique identification code. Each satellite transmits low power radio signals on several frequencies (designed L1, L2 etc.). GPS receivers “listen” on L1 frequency of 1575.42 MHz and L2 frequency of 1227.60 MHz. Civilian GPS receivers can use only the L1 signals. These signals can pass through clouds, grass, plastics, but not pass through solid objects such as buildings and mountains. Each signal consists of two “pseudo random” signals, the protected (P) code and course acquisition (CA) code. C/A is available for civilians and (P) used by US military receivers.

The main purpose of these signals is to calculate the travel time from satellite to the GPS receiver on the earth. The travel time is called time of arrival. The travel time multiplied by speed of light equals the satellite range (distance from the satellite to the GPS receiver). (Bhatta, 2008).

### 27.1.2 Control Segment

The control segment is composed of

1. A master control station (MCS),
2. An alternate master control station,
3. Four dedicated ground antennas and
4. Six dedicated monitor stations



The MCS can also access U.S. Air Force Satellite Control Network (AFSCN) ground antennas (for additional command and control capability) and NGA (National Geospatial-Intelligence Agency) monitor stations.



**Fig. 27.3. Control Segment.** (Source: <http://www.aboutcivil.org/components-of-gps-working-mechanism.html>; November 10, 2012)

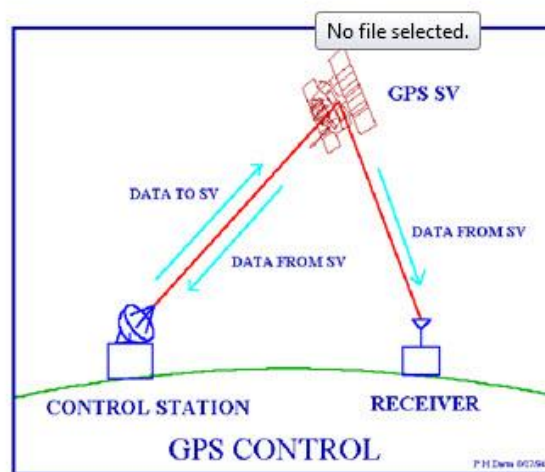
The flight paths of the satellites are tracked by dedicated U.S. Air Force monitoring stations in Hawaii, Kwajalein, Ascension Island, Diego Garcia, Colorado Springs, Colorado and Cape Canaveral, along with shared NGA monitor stations operated in England, Argentina, Ecuador, Bahrain, Australia and Washington DC.

Then 2 SOPS contacts each GPS satellite regularly with a navigational update using dedicated or shared (AFSCN) ground antennas (GPS dedicated ground antennas are located at Kwajalein, Ascension Island, Diego Garcia, and Cape Canaveral).



**Fig. 27.4. Ground monitor station used from 1984 to 2007, on display at the Air Force Space & Missile Museum.**

(Source: [http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System); November 15, 2012)



**Fig. 27.5. GPS Control.**

(Source: [http://www.colorado.edu/geography/gcraft/notes/gps/gps\\_f.html](http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html); December 1, 2012)

These updates synchronize the atomic clocks on board the satellites to within a few nanoseconds of each other, and adjust the ephemeris of each satellite's internal orbital model. The updates are created by a Kalman filter that uses inputs from the ground monitoring stations, space weather information, and various other inputs.

Satellite maneuvers are not precise by GPS standards. So to change the orbit of a satellite, the satellite must be marked unhealthy, so receivers will not use it in their calculation. Then the maneuver can be carried out, and the resulting orbit tracked from the ground. Then the new ephemeris is uploaded and the satellite marked healthy again.

The Operation Control Segment (OCS) currently serves as the control segment of record. It provides the operational capability that supports global GPS users and keeps the GPS system operational and performing within specification. OCS successfully replaced the legacy 1970's-era mainframe computer at Schriever Air Force Base in September 2007. After installation, the system helped enable upgrades and provide a foundation for a new security architecture that supported the U.S. armed forces. OCS will continue to be the ground control system of record until the new segment, Next Generation GPS Operation Control System (OCX), is fully developed and functional.

The new capabilities provided by OCX will be the cornerstone for revolutionizing GPS's mission capabilities, and enabling Air Force Space Command to greatly enhance GPS operational services to U.S. combat forces, civil partners and myriad of domestic and international users.

The GPS OCX program also will reduce cost, schedule and technical risk. It is designed to provide 50% sustainment cost savings through efficient software architecture and Performance-Based Logistics. In addition, GPS OCX expected to cost millions less than the cost to upgrade OCS while providing four times the capability.

The GPS OCX program represents a critical part of GPS modernization and provides significant information assurance improvements over the current GPS OCS program.

- OCX will have the ability to control and manage GPS legacy satellites as well as the next generation of GPS III satellites, while enabling the full array of military signals.
- Built on a flexible architecture that can rapidly adapt to the changing needs of today's and future GPS users allowing immediate access to GPS data and constellations status through secure, accurate and reliable information.
- Empowers the warfighter with more secure, actionable and predictive information to enhance situational awareness.
- Enables new modernized signals (L1C, L2C, and L5) and has M-code capability, which the legacy system is unable to do.
- Provides significant information assurance improvements over the current program including detecting and preventing cyber-attacks, while isolating, containing and operating during such attacks.
- Supports higher volume near real-time command and control capabilities

On September 14, 2011, the U.S. Air Force announced the completion of GPS OCX Preliminary Design Review and confirmed that the OCX program is ready for the next phase of development.

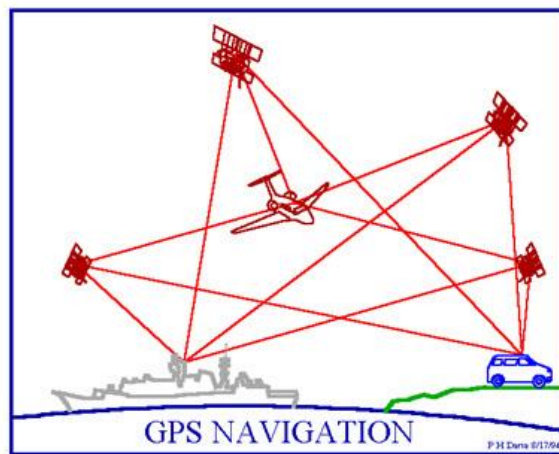
The GPS OCX program has achieved major milestones and is on track to support the GPS IIIA launch in May 2014.

([http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System); October 15, 2012)

### 27.1.3 User segment

The user segment just consists of the user and his GPS receiver. It is composed of hundreds of thousands of U.S. and allied military users of the secure GPS Precise Positioning Service, and tens of millions of civil, commercial and scientific users of the Standard Positioning Service. In general, GPS receivers are composed of an antenna (Internal or external), tuned to the frequencies transmitted by the satellites, receiver-processors, and a highly stable clock (often a crystal oscillator). Generally, they also include a display for providing location and speed information to the user. A receiver is often described by its No. of channels: this signifies signals' from how many satellites it can process simultaneously. Originally limited to 4 or 5, this has progressively increased over the years so that, as of year 2007, receivers typically have between 12 and 20 channels. (Bhatta, 2008)

GPS receivers may include an input for differential corrections, using the RTCM SC-104 format. This is typically in the form of an RS-232 port at 4,800 bit/s speed. Data is actually sent at a much lower rate, which limits the accuracy of the signal sent using RTCM. Receivers with internal DGPS receivers can outperform those using external RTCM data. As of 2006, even low-cost units commonly include Wide Area Augmentation System (WAAS) receivers.



**Fig. 27.6. GPS Navigation.**

(Source: [http://www.colorado.edu/geography/gcraft/notes/gps/gps\\_f.html](http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html); December 1, 2012)

Many GPS receivers can relay position data to a PC or other device using the NMEA 0183 protocol. Although this protocol is officially defined by the National Marine Electronics Association (NMEA), references to this protocol have been compiled from public records, allowing open source tools like `gpsd` to read the protocol without violating intellectual property laws. Other proprietary protocols exist as well, such as the SiRF and MTK protocols. Receivers can interface with other devices using methods including a serial connection, USB, or Bluetooth. [http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System); October 15, 2012.

## 27.2 Working Principles

When the GPS receiver starts its job, it should have information about where the satellites are and how far away they are.

Let us first look at how the GPS receiver knows where the satellites are located in space. This GPS receiver picks up two kinds of coded information from the satellites, almanac and ephemeris. Almanac data is a coarse orbital model for all satellites. Each satellite broadcasts almanac data for all satellites. This almanac data is not very precise and is considered valid for up to several hours. This data is continuously transmitted and stored in the memory of GPS receiver so that it knows the orbits of satellites and where each satellite is supposed to be. The almanac data is periodically updated with new information as the satellites move around.

Ephemeris data by comparison is a very precise orbital and clock correction for each satellite and is necessary for precise positioning. Any satellite can travel slightly out of orbit, so the ground monitor station keeps track of the satellite's orbit, altitude, location, and speed. The ground monitor stations send the monitored orbital data to the master control station, which in turn sends corrected orbital data and clock corrections up to the satellites. This corrected and exact positioned data is called the ephemeris data, which is transmitted as coded information to the GPS receiver. Each satellite broadcasts only its own ephemeris data. Subsets of ephemeris data are broadcast by each satellite continuously, which remains valid for the orbit for a few minutes.

The unit stored data about where the satellites are located at any given time. This data is called the almanac. On the ground, all GPS receivers have an almanac programmed into their computers that tells them where in the sky each satellite is moment by moment. This data helps the receiver to track the satellite faster. Ephemeris data is required to know the exact positions of each visible satellite and clock error for the precise calculation of the receiver location.

Therefore, having received the almanac and ephemeris data, the GPS receiver knows the precise positions of the satellite at all times. (Bhatta, 2008)

### 27.3 Timing and Ranging

Even though the GPS receiver knows the precise location of the satellite in space, it still needs to know how far away the satellites are (the distance), so it can determine its position on earth. There is a simple formula that tells the receiver how far it is from each satellite.

The distance (from us) from a given satellite object equals the velocity of the transmitted signal multiplied by the time taken by the signal to reach us ( $d = vt$ ).

We recollect how in our childhood we tried to find out how far a thunderstorm was from us. When we saw a lightning flash, we counted the number of seconds until we heard the thunder. The longer the count, the farther away the storm was. GPS works on the same principle called time of arrival. We would have noticed that during a thunderstorm, we heard the sound sometime after we saw the light. The reason is that sound waves travel much slower than light waves. We can estimate our distance to the storm by measuring the delay between the time that we see the lightning and the time that we hear it. Multiplying this time delay by the speed of sound gives us our distance to the storm (assuming that the light reaches us almost instantaneously compared to sound). Sound travels about 344 m (1,130 ft) per second in air. So if it takes 2s between the times that we see the lightning and the time we hear it, our distance to the storm is  $2 \times 344 = 688$  m. We calculate the distance to an object by measuring the time that it takes for its signal to reach us.

Using the same basic formula to determine distance, the receiver already knows the velocity. It is the speed of a radio wave-186,000 miles ( $3 \times 10^8$  m) per second (the speed of light), less any delay as the signal travels through the earth's atmosphere.

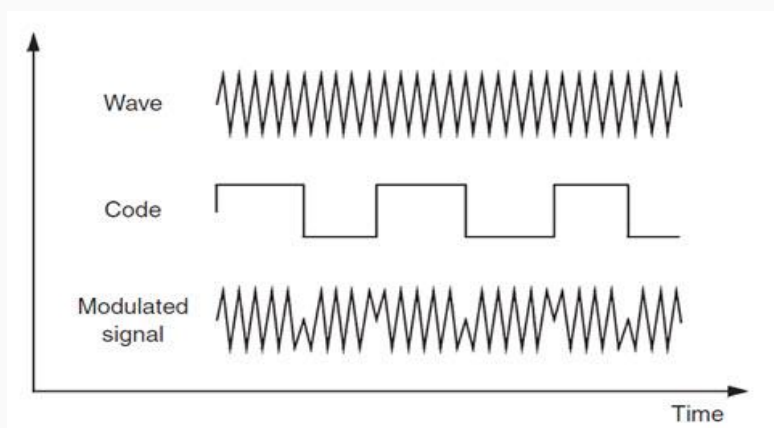


Fig. 27.7. Modulation of carrier wave. (Source: Konecny, 2003)



Now the GPS receiver needs to determine the time part of the formula. The answer lies in the coded signals the satellites transmit. The transmitted code is called 'Pseudo Random Code' (PRC) because it looks like a noise signal. The PRC is a fundamental part of GPS. Physically, it is just a very complicated digital code. The signal is so complicated that it almost looks like random electrical noise and hence the name is 'pseudo-random'. There are several good reasons for its complexity: First, the complex pattern helps make sure that the receiver does not accidentally sync up to some other signal. Since each satellite has its own unique PRC, this complexity also guarantees that the receiver would not accidentally pick up the signal of another satellite. However, the reason that is crucial to make the GPS economical is that the codes make it possible to use 'information theory' to 'amplify' the GPS signal. Because of this reason GPS receivers do not require big satellite dishes to receive the GPS signals.

When a satellite generates the PRC the GPS receiver simultaneously generates the same code and tries to match it up to the satellite's code. The receiver then compares the two codes to determine how much it needs to delay (or shift) its code to match the satellite code. This delay time (shift) is multiplied by the speed of light to get the distance.

Our GPS receiver clock does not keep the time as precisely as the satellite clock. Putting an atomic clock in our GPS receiver would make it much larger and far more expensive. Therefore, each distance measurement needs to be corrected to account for the GPS receiver's internal clock error. For this reason, the range measurement is referred to as a 'pseudo-range'. To determine position using pseudo-range data, a minimum of four satellites must be tracked and the four fixes must be recomputed until the clock error disappears.

**Note:** The delay of time can be understood by the following example:

Let us assume that our friend at the end of a large field repeatedly shouts numbers from 1 to 10 at the rate of one count per second (10 s for a full cycle of 1 to 10 counts). Let us also assume that we do the exact same thing, synchronized with him, at the other end of the field. Synchronization between us and him could have been achieved by both starting at an exact second and observing our watches to count 1 number per second. We assume that both have very accurate watches. Because of the sound travel time, we hear the number patterns of our friend with a delay relative to our patterns. If we hear the friend's count with a delay of one count relative to ours, then our friend must be 344 m away from us ( $344 \text{ m/s}$ , the velocity of sound  $\times 1 \text{ s}$ ). This is because the counts are 1 s apart.

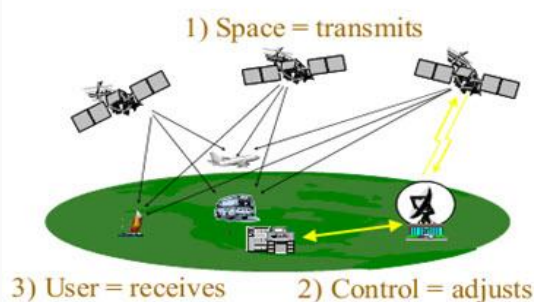


Fig. 27.8. Components of GPS.

(Source: Lecture 16: Global Positioning System.pdf)

## 27.4 How GPS Works

- **Calculating Location**

The basis of GPS is 'triangulation' from satellites also called trilateration. To 'triangulate', a GPS receiver measures distance using the travel time of radio signals. To measure travel time, GPS needs very accurate timing, which is achieved through some tricks. Along with distance, we need to know exactly where the satellites are in space. (Bhatta, 2008)

- **Trilateration**

Trilateration is a process that uses distance from at least three known locations to determine position. GPS receivers calculate the position of objects in two dimensional or three dimensional spaces using a mathematical process called trilateration. Trilateration can be either two dimensional or three dimensional. Let us examine how 2-D and 3-D trilateration work.

<http://www.roseindia.net/technology/gps/what-is-trilateration.html>; November 30.2012.

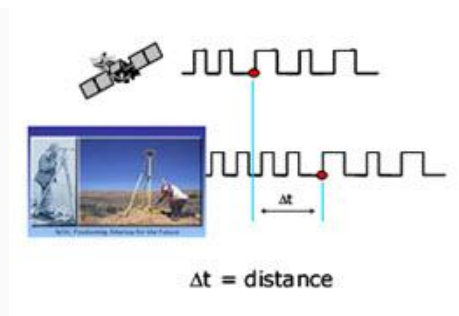
- **2-D Trilateration**

The concept of trilateration is easy to understand through an example. Imagine that you are driving through an unfamiliar country and that you are lost. A road sign indicates that you are 500 km from city A. But this is not of much help, as you could be anywhere in a circle of 500 km radius from the city A. A person you stop by to ask for directions then volunteers that you are 450 km from city B. Now you are in a better position to locate yourself- you are at one of the two intersecting points of the two circles surrounding city A and city B. Now if you could also get your distance from another place say city C, you can locate yourself very precisely, as these three circles can intersect each other at just one point. This is the principle behind 2D trilateration.

<http://www.roseindia.net/technology/gps/what-is-trilateration.html>; November 30.2012.

- **3-D Trilateration**

The fundamental principles are the same for 2D and 3D trilateration, but in 3D trilateration we are dealing with spheres instead of circles. It is a little tricky to visualize. Here, we have to imagine the radii from the previous example going in all directions, that is in three dimensional space, thus forming spheres around the predefined points. Therefore the location of an object has to be defined with reference to the intersecting point of three spheres.



**Fig. 27.9. Travel time determination from shifting of code.**

**(Source: Lecture 16: Global Positioning System.pdf)**

GPS use clocks and trilateration to determine position. The satellite vehicles (SV) and receiving units both contain highly accurate clocks. Part of the information that the SV transmits is a time stamp. When a GPS unit receives the transmission, it compares the time stamp from the satellite to the time it reached the receiver. The difference between the two is multiplied by the speed of the transmission signal  $c$  the distance that the signal traveled.

<http://www.roseindia.net/technology/gps/what-is-trilateration.html>; November 30.2012.

Suppose we measure our distance from a satellite and find it to be 20,000 km, knowing that we are 20,000 km from a particular satellite narrows down all the possible locations we could be in the whole universe to the surface of a sphere that is centered on this satellite and has a radius of 20,000 km.

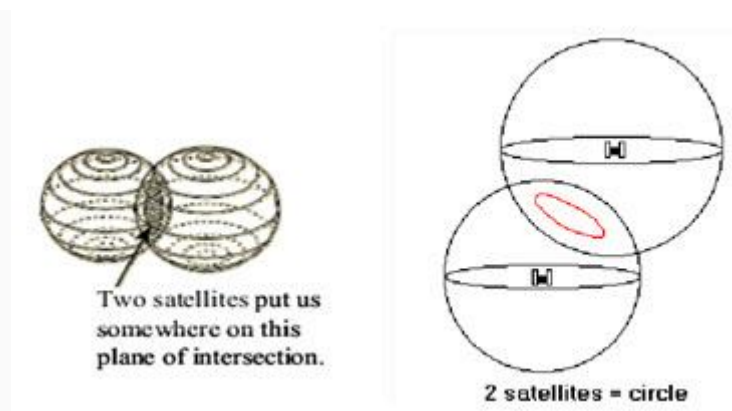


**Fig. 27.10. Distance from a single satellite.**

**(Source: Lecture 16: Global Positioning System.pdf 2<sup>nd</sup> Principles of GPS by Dana m. Sommer.pdf)**

Next, let us say we measure our distance to a second satellite and find out that it is 21,000 km away. This conveys that we are not only on the first sphere but we are also on a sphere that is 21,000 km from the second satellite. Or in other words, we are somewhere on the circle where these two spheres intersecting.

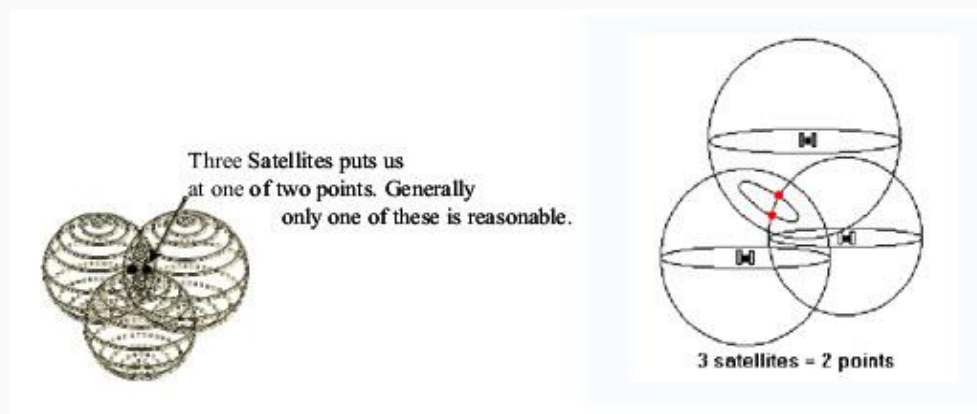




**Fig. 27.11. Distance from two satellites.**

(Source: Lecture 16: Global Positioning System.pdf, 2<sup>nd</sup> Principles of GPS by Dana m. Sommer. pdf)

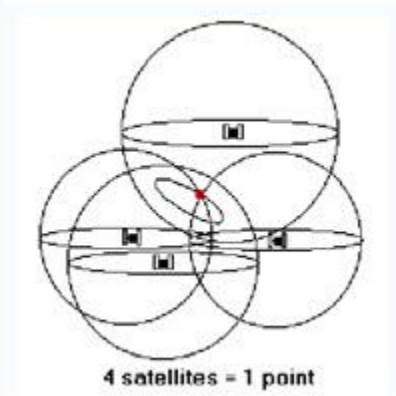
If we then make a measurement from a third satellite and find that we are 22,000 km from that one, which narrows our position down even further, to the two points where the 22,000 km sphere cuts through the circle that is the intersection of the first two spheres. So by ranging from three satellites we can narrow our position to just two points in space.



**Fig. 27.12. Location from three satellites.**

(Source: Lecture 16: Global Positioning System.pdf 2<sup>nd</sup> Principles of GPS by Dana m. Sommer.pdf)

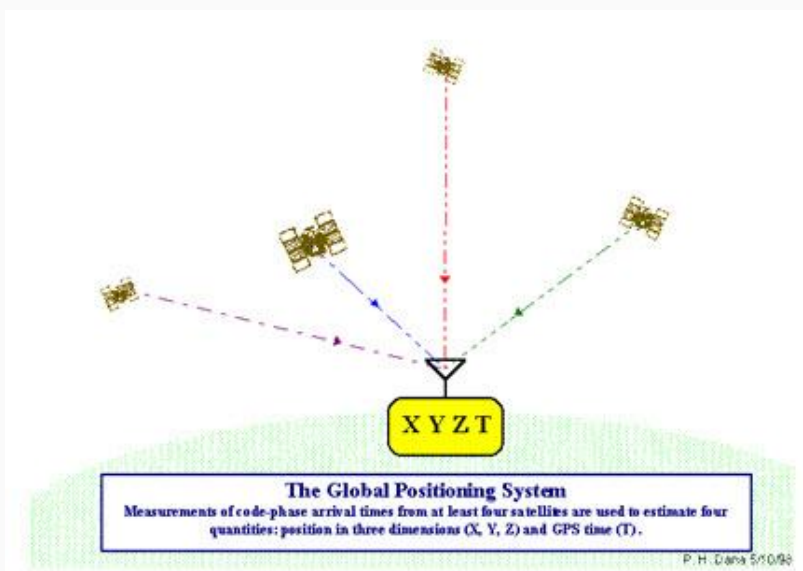
Even though there are two possible positions, they differ greatly in latitude/longitude position and Altitude. However, by adding a fourth satellite, the receiver can determine our three-dimensional position (latitude, longitude, altitude). Let us say our distance from a fourth satellite is 21,000 km. We now have a fourth sphere intersecting the first three spheres at one common point. But usually one of the two points (in case of three satellite references) is a ridiculous answer (either too far from the earth or moving at an impossible velocity) and can be rejected without a measurement, though, a fourth satellite gives us more accurate positional information. (Bhatta, 2008).



**Fig. 27.13. Location from 4 satellites.**

(Source: Lecture 16: Global Positioning System.pdf)

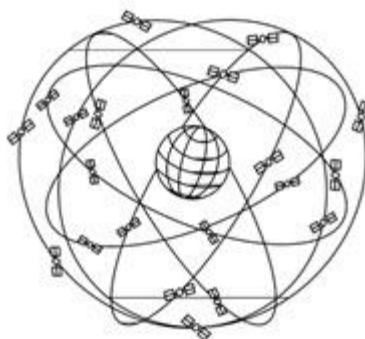
A fourth measurement does come in for another reason, to achieve or measure accurate time in the GPS receiver.



**Fig. 27.14. Measurements of arrival times from four satellites.**

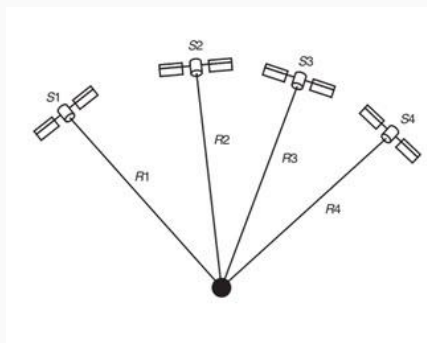
(Source: <http://www.colorado.edu/geography/gcraft/notes/gps/gpsf.html>; December 1, 2012)

On the satellite side, the timing is almost perfect because they have incredibly precise atomic clocks on board, but the receiver does not have precise atomic clock. But both the satellite and the receiver need to be able to precisely synchronize their PRCs to make the system work. The designers of GPS came up with a brilliant little trick that lets us get by with much less accurate clocks in our receivers. This trick is one of the key elements of GPS and as an added side benefit, it means that even GPS receiver is essentially an atomic-accuracy clock. (Bhatta, 2008)



**Fig. 27.15. Global GPS satellite configuration. (Source: Konecny, 2003,)**

If our receiver's clocks were perfect, then all the satellite ranges would intersect at a single point (which is our position). But with imperfect clocks, a fourth measurement, done as a cross-check that will not intersect with the first three. Thus the receiver's computer determines the discrepancy in measurements.



**Fig. 27.16. Shows the determination of the position from the four distance measurements, R1, R2, R3 and R4. (Source: Konecny, 2003)**

Then the receiver looks for a correction factor that it can deduct from all its timing measurements that would cause them all to intersect at a single point. That correction brings the receiver's clock back into synchronise and we have acquired atomic accuracy time right in the palm of our hand. Once that correction is made, it applies to all the rest of its measurements and now we have obtained precise positioning. One consequence of this principle is that, it is necessary for any decent GPS receiver to acquire at least four channels so that it can make the four measurements simultaneously. Since the receiver must solve for its position ( $x_0, y_0, z_0$ ) and the clock error ( $t$ ) (in terms of time-duration), four satellites are required to solve the receiver's position using the following four equations:

$$R_1 = \sqrt{[(X_0 - X_1)^2 + (Y_0 - Y_1)^2 + (Z_0 - Z_1)^2]} + c \cdot t \quad (27.1)$$

$$R_2 = \sqrt{[(X_0 - X_2)^2 + (Y_0 - Y_2)^2 + (Z_0 - Z_2)^2]} + c \cdot t \quad (27.2)$$

$$R_3 = \sqrt{[(X_0 - X_3)^2 + (Y_0 - Y_3)^2 + (Z_0 - Z_3)^2]} + c \cdot t \quad (27.3)$$

$$R_4 = \sqrt{[(X_0 - X_4)^2 + (Y_0 - Y_4)^2 + (Z_0 - Z_4)^2]} + c \cdot t \quad (27.4)$$

where  $(x_1, y_1, z_1)$ ,  $(x_2, y_2, z_2)$ ,  $(x_3, y_3, z_3)$ , and  $(x_4, y_4, z_4)$  stand for the locations of satellites, which are known and  $R_1, R_2, R_3$ , and  $R_4$  are the distances of satellites from the receiver position (Fig. 7.9), which are derived,  $c$  is the velocity of signal (or light). Hence solving the four equations for four unknowns  $x_0, y_0, z_0$  and  $t$ , at least four navigation satellites are necessary. The GPS receiver routinely solves these simultaneous equations using dedicated software installed within it. The coordinates  $(x_0, y_0, z_0)$  are calculated according to the World Geodetic System (WGS 1984) coordinates system. (Bhatta, 2008).

**Note:** A GPS user is likely to have experienced a wrong set-up in his GPS receiver relative to the map he uses. It is essential that the GPS receiver be set up to conform to the map. In particular, the coordinate system and its datum need to be the same. The GPS calculates the position based on World Geodetic System 1984 (WGS 84) spheroid and datum. However, users may encounter maps that use other datums and spheroids, The details for each specific map can be found in the margin information on the map, and must be entered correctly into the GPS receiver if an accurate relationship between the GPS receiver and the map is to be assured.

**Keywords:** Space segment, Control Segment, User segment, Wide Area Augmentation System (WAAS) , GPS, NAVSTAR, OCS, SV, CA code, WGS



## Lesson- 28 GPS Application

### 28.1 GPS Errors

GPS errors are a combination of noise, bias, blunders. GPS measurements are potentially subject to numerous sources of error in addition to clock bias. Among these are uncertainties in the satellite orbits (known as satellite ephemeris errors), errors due to atmospheric conditions (signal velocity depends on time of day, season, and angular direction through the atmosphere), receiver errors (due to such influences as electrical noise and signal matching errors), and multipath errors (reflection of a portion of the transmitted signal from objects not in the straight-line path between the satellite and receiver). (Lillesand and Keiffer, 2004)

#### 28.1.1. Noise Errors

Noise errors are the combined effect of PRN code noise (around 1 meter) and noise within the receiver noise (around 1 meter). Noise and bias errors combine, resulting in typical ranging errors of around fifteen meters for each satellite used in the position solution. (Dana, 1997)

#### 28.1.2. Bias Errors

Bias errors result from Selective Availability and other factors. Selective Availability (SA) is the intentional degradation of the SPS signals by a time varying bias. It is controlled by the DOD to limit accuracy for non-U. S. military and government users. (Dana, 1997)

Other Bias Error sources are discussed in the later part of the chapter.

#### 28.1.3. Blunders

Blunders can result in errors of hundreds of kilometers. Control segment mistakes due to computer or human error can cause errors from one meter to hundreds of kilometers. User mistakes, including incorrect geodetic datum selection, can cause errors from 1 to hundreds of meters. Receiver errors from software or hardware failures can cause blunder errors of any size. (Dana, 1997)

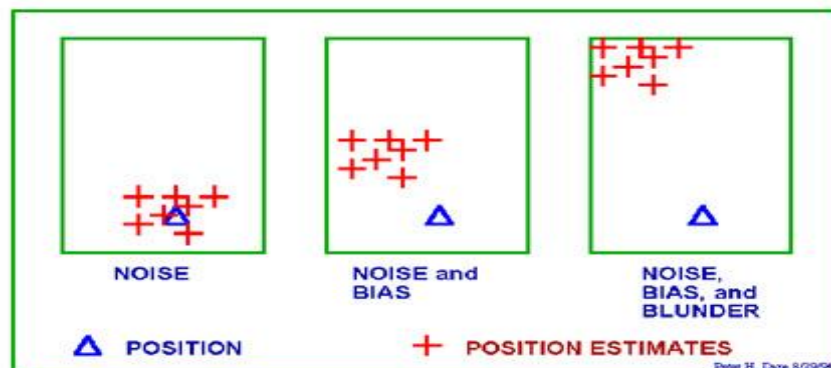


Fig. 28.1. Showing the three types of errors. (Source: Dana, 1997).  
<http://www.ncgia.ucsb.edu/giscc/units/u017/u017.html>, posted August 28, 1997.

The analysis of errors computed using the Global Positioning System is important for understanding how GPS works, and for knowing what magnitude errors should be expected. The Global Positioning System makes corrections for receiver clock errors and other effects but there are still residual errors which are not corrected.

The term user equivalent range error (UERE) refers to the error of a component in the distance from receiver to a satellite. These UERE errors are given as  $\pm$  errors thereby implying that they are unbiased or zero mean errors. These UERE errors are therefore used in computing standard deviations. The standard deviation of the error in receiver position,  $\sigma_{rc}$ , is computed by multiplying PDOP (Position Dilution of Precision) by  $\sigma_R$ , the standard deviation of the user equivalent range errors.  $\sigma_R$  is computed by taking the square root of the sum of the squares of the individual component standard deviations. PDOP is computed as a function of receiver and satellite positions.

User equivalent range errors (UERE) are shown in the Table 28.1. There is also a numerical error with an estimated value,  $\sigma_{num}$ , of about 1 meter. The standard deviations,  $\sigma_R$ , for the coarse/acquisition and precise codes are also shown in the table. These standard deviations are computed by taking the square root of the sum of the squares of the individual components (i.e., RSS for root sum squares). To get the standard deviation of receiver position estimate, these range errors must be multiplied by the appropriate dilution of precision terms and then RSS'ed with the numerical error. Electronics errors are one of several accuracy-degrading effects outlined in the table above. When taken together, autonomous civilian GPS horizontal position fixes are typically accurate to about 15 meters (50 ft). These effects also reduce the more precise P(Y) code's accuracy. However, the advancement of technology means that today, civilian GPS fixes under a clear view of the sky are on average accurate to about 5 meters (16 ft) horizontally.

$\sigma_R$  for the C/A code is given by:

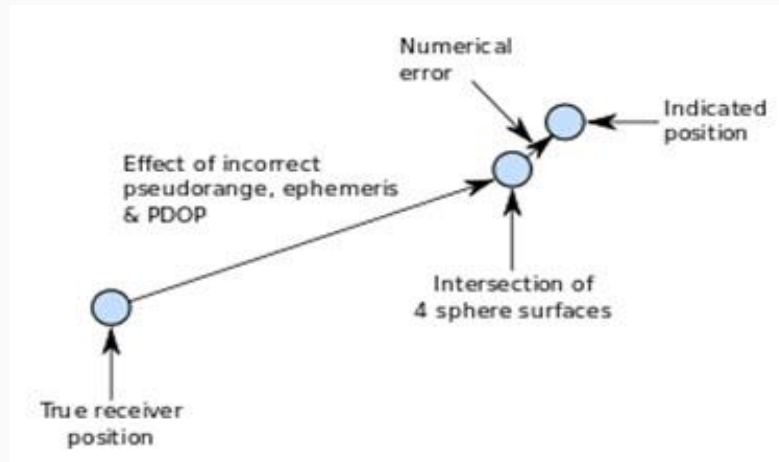
$$\sigma_R = \sqrt{(3^2 + 5^2 + 2.5^2 + 2^2 + 1^2 + 0.5^2)}m = 6.7m$$

The standard deviation of the error in estimated receiver position  $\sigma_{rc}$ , again for the C/A code is given by:

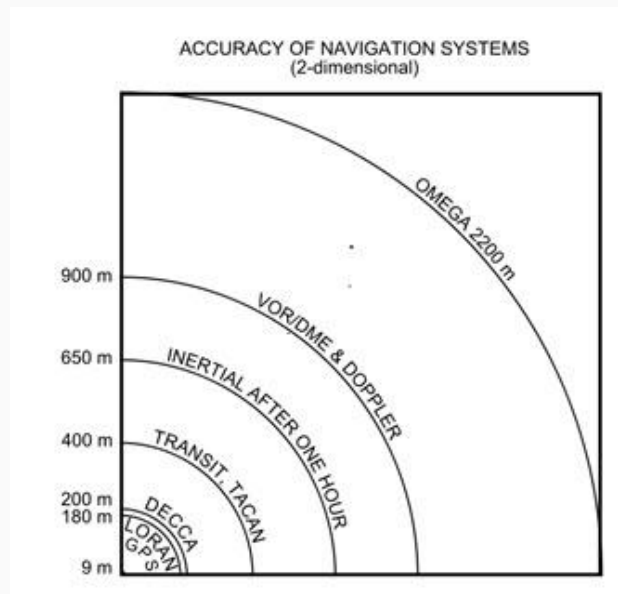
$$\sigma_R = \sqrt{(PDOP^2 + \sigma_R^2 + \sigma_{num}^2)}$$

$$\sigma_R = \sqrt{(PDOP^2 + 6.7^2 + 1^2)}m$$

The error diagram in the Fig. 28.2 shows the inter relationship of indicated receiver position, true receiver position, and the intersection of the four sphere surfaces.



**Fig. 28.2. Geometric Error Diagram Showing Typical Relation of Indicated Receiver Position, Intersection of Sphere Surfaces, and True Receiver Position in Terms of Pseudorange Errors, PDOP, and Numerical Errors. (Source: [http://en.wikipedia.org/wiki/Error\\_analysis\\_for\\_the\\_Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Error_analysis_for_the_Global_Positioning_System))**



**Fig. 28.3. Accuracy of navigation systems. (Source: [http://en.wikipedia.org/wiki/Error\\_analysis\\_for\\_the\\_Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Error_analysis_for_the_Global_Positioning_System))**

**Table 28.1 Sources of user equivalent range errors (UERE)**

Sl. no.	Source	Effect(m)
1.	Signal arrival C/A	$\pm 3$
2.	Signal arrival P(Y)	$\pm 0.3$
3.	Ionospheric effects	$\pm 5$
4.	Ephemeris errors	$\pm 2.5$
5.	Satellite clock errors	$\pm 2$
6.	Multipath distortion	$\pm 1$
7.	Tropospheric effects	$\pm 0.5$
8.	$\sigma_R$ C/A	$\pm 6.7$
9.	$\sigma_R$ P(Y)	$\pm 6$

(Source: [http://en.wikipedia.org/wiki/Error\\_analysis\\_for\\_the\\_Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Error_analysis_for_the_Global_Positioning_System))

## 28.2 Sources of GPS Error

The GPS signals sent from the SVs are subject to a variety of error sources before they are processed into a position and time solution in the receiver. As with most systems these error sources take the form of zero-bias noise, bias errors, and blunders. A number of conditions can reduce the accuracy of a GPS receiver. From a top-down perspective (from orbit down to ground level), the possible sources of trouble look like this:

### 1. Selective Availability

Selective availability is the single largest source of C/A-code error. Y-code capable GPS receivers can remove SA with knowledge of the SA algorithm. SA takes the form of a slowly varying range error for each SV. SA introduces the largest bias errors in the Standard Positioning System accounting for most of the 100 meter (95 percent) error in the SPS. (Dana, 1997)

### 2. Clock and Ephemeris Errors

Ephemeris errors occur when the satellite doesn't correctly transmit its exact position in orbit. Clock and ephemeris data sets represent the difference between the SV clock and GPS time and permit the estimation of SV position at the time of transmission of the tracked codes. A GPS parameter, the User Range Accuracy (URA), is a range error estimate indicative of the "maximum value anticipated during each sub frame fit interval with uniform SA levels invoked" (Anon 1995, 35). The URA is transmitted as an integer power of two. Although the URA is not specified as a definite indicator of SA error magnitude, for a Block II SV affected by SA, a URA of 32 meters is common (Dana, 1997).



### 3. Ionospheric Delays

The ionosphere starts at about 43–50 miles above the Earth and continues for hundreds of miles. Satellite signals traveling through the ionosphere are slowed down because of plasma (a low density gas). Although GPS receivers attempt account for this delay, unexpected plasma activity can cause calculation errors. (Dana, 1997)

A major source of bias error is the delay of the GPS carrier signals as they pass through the layer of charged ions and free electrons known as the ionosphere. (Dana, 1997). Varying in density and thickness as it rises and falls (50 to 500 kilometers) due to solar pressure and geomagnetic effects, the ionosphere can delay the GPS signals by as much as 300 nanoseconds (100 meters) (Klobuchar 1982). The diurnal (24-hour) changes in the ionosphere cause the largest variations in delay. At night the delay is at a minimum and the thinner and higher night-time ionosphere is more easily modeled than the less dense and thicker layer during the day. The signals from SVs at low elevation angles with respect to the local horizon experience the largest delays as the signal passes through more ionosphere than if the SV were directly overhead. Using the P-code, or special codeless (signal-squaring) techniques, the delay through the ionosphere can be computed by a receiver capable of measuring the phase delay difference between the code carried on the L1 and L2 signals (Dana, 1997). These dual frequency methods result in a substantial reduction of the ionospheric bias, making it possible to transfer sub-nanosecond clock offset measurements over thousands of kilometers (Dunn and others 1993, 174). For a single frequency (L1) C/A-code receiver the ionospheric delay can be estimated from the ionospheric delay model broadcast by the SVs. The Master Control station calculates the parameters for delay using a cosine model that computes delay for a given local time-of-day and the elevation angle for the path from the receiver to an SV. Some users compute an ionospheric delay estimate from their own models. Using the broadcast model under normal conditions removes about half of the error (Fees and Stephens 1987) leaving a residual error of around 60-90 nanoseconds during the day and 10 to 20 nanoseconds at night (Knight and Rhoades 1987). Signals from SVs at high elevation angles experience smaller delays, but use of the broadcast model under abnormal conditions can occasionally introduce more error than that caused by the actual delay. (Dana, 1997)

### 4. Tropospheric Delays

The troposphere is the lowest region in the Earth's atmosphere and goes from ground level up to about 11 miles. Variations in temperature, pressure, and humidity all can cause variations in how fast radio waves travel, resulting in relatively small accuracy errors. GPS signal delays through the troposphere, the layer of atmosphere usually associated with changes in weather (from ground level up to 8 to 13 kilometers), are subject to local conditions and are difficult to model. GPS does not broadcast a tropospheric correction model but several such models have been developed. Some receivers make a limited model available that computes tropospheric delay from receiver height and SV elevation angle using nominal atmospheric parameters. (Dana, 1997). Because accurate tropospheric delay models (Turner and others 1986) require local pressure, temperature and humidity (PTH) data as well as receiver height and elevation angle to the SV, these models are difficult to apply in real-time situations. The errors introduced by an unmodeled troposphere may be as much as 100 nanoseconds at low elevation angles (less than 5 degrees), but are more typically

in the 30 nanosecond range (Knight and Rhoades 1987). Residuals after application of a simple, no-PTH, model (Gupta 1980) are in the 10 nanosecond range.

## **5. Multipath**

When a satellite signal bounces off a hard surface (such as a building or canyon wall) before it reaches the receiver, a delay in the travel time occurs, which causes an inaccurate distance calculation. Multipath interference, caused by local reflections of the GPS signal that mix with the desired signal, slowly introduces varying bias errors of one to two nanoseconds for navigation receivers aboard aircraft in flight. For land-based systems, local conditions and exact antenna placement can result in errors of up to 150 nanoseconds. (Dana, 1997). Nominal errors for land-based receivers are in the 30 nanosecond range (Braasch, 1995). Careful attention to antenna placement, antenna design, the use of choke rings, and the use of materials that absorb GPS radio-frequency signals can mitigate much of the potential multipath interference, but these measures must be carefully designed to allow for the different multipath reflections from the constantly changing SV elevations and azimuths. In many applications it is difficult or impossible to completely eliminate multipath errors. (Dana, 1997)

## **6. GPS Signal Noise**

Propagation of the GPS signals from the SV to the receiver introduces noise from galactic sources, ionospheric scintillations, and cross correlation from other GPS SV signals that results in small noise (zero bias) errors in the three nanosecond range. (Dana, 1997)

## **7. Receiver Noise and Delays**

Receiver noise can introduce two to three nanoseconds of zero bias noise in the timing measurements of a GPS receiver. Delays within a receiver can be calibrated by the manufacturer, but if receiver delays change with temperature or change differently between channels of a multi-channel receiver, timing bias errors can result. Antenna cable delays must be recomputed or calibrated if cable lengths change or cables of different materials are used. (Dana, 1997). There have been reports of cable delays being both temperature and signal strength dependent (Lewandowski, Petit and Thomas 1991, 5). Manufacturers can provide cable delays for the equipment they supply. (Dana, 1997)

## **8. Receiver Oscillator Errors**

While precise time standards at the Control and Space Segments of GPS are designed to keep user clock requirements to a minimum, receiver oscillators must provide enough stability to insure that they can be rated properly by GPS receiver software and that they provide a low noise timing reference. This is sometimes difficult to accomplish in high dynamic environments or when the receiver internal temperatures cannot be controlled or compensated for. (Dana, 1997)

## **9. SV clock errors**

The uncorrected by Control Segment can result in one meter errors in position. (Dana, 1997)

## 10. Geometric Dilution of Precision

Geometric Dilution of Precision (GDOP) is a measurement of the sensitivity of a receiver position or time estimate to changes in the geometric relationship between the receiver position and the positions of all of the SVs used to form the position or time estimate. If the SVs used for a navigation solution were all in about the same place in the sky, directly above a receiver position, for instance, the position solution for height would be less sensitive to pseudo-range changes than would the poorly defined (diluted) solution for horizontal position. If the SVs were distributed around the field of view of the receiver, horizontal and vertical positioning would be more equally sensitive to pseudo-range changes. GDOP is a dimensionless multiplier that can be used to estimate the effect of pseudo-range errors on a complete position and time solution. The single GDOP parameter is the square root of the sum of the diagonal terms of the covariance matrix that is formed from the inverse of the matrix of directional derivatives for each of the SV positions and pseudo-ranges used in the position solution. For a specified receiver position and a set of SVs, GDOP can be separated into three-dimensional position (PDOP) or spherical (SDOP) dilution, two-dimensional horizontal (HDOP), or one dimensional vertical (VDOP) or time (TDOP) estimates. These separate components of GDOP are formed from covariance terms and so are not independent of each other. A high TDOP (time dilution of precision) in a navigation receiver will eventually influence position errors as erroneous receiver clock bias estimates are used to correct pseudo-range measurements.

The computation of geometric dilution of precision involves many numerical equations. Computations were provided to show how PDOP was used and how it affected the receiver position error standard deviation. When visible GPS satellites are close together in the sky (i.e., small angular separation), the DOP values are high; when far apart, the DOP values are low. Conceptually, satellites that are close together cannot provide as much information as satellites that are widely separated. Low DOP values represent a better GPS positional accuracy due to the wider angular separation between the satellites used to calculate GPS receiver position. HDOP, VDOP, PDOP and TDOP are respectively Horizontal, Vertical, Position (3-D) and Time Dilution of Precision. (Dana, 1997)

## 11. Poor satellite coverage

When a significant part of the sky is blocked, your GPS unit has difficulty receiving satellite data. Unfortunately, you can't say that if 50 percent (or some other percentage) of the sky is blocked, you'll have poor satellite reception; this is because the GPS satellites are constantly moving in orbit. A satellite that provides a good signal one day may provide a poor signal at the exact same location on another day because its position has changed and is now being blocked by a tree. The more open sky you have, the better the chances of not having satellite signals blocked.

Building interiors, streets surrounded by tall buildings, dense tree canopies, canyons, and mountainous areas are typical problem areas. (McNamara, 2004)

### 28.3 Accuracy of GPS

The errors can be compensated for (in great part) using differential GPS measurement methods. In this approach, simultaneous measurements are made by a stationary base station receiver (located over a point of precisely known position) and one (or more) roving receivers moving from point to point. The positional errors measured at the base station are used to refine the position measured by the rover(s) at the same instant in time. This can be done either by bringing the data from the base and rover together in a post-processing mode after the field observations are completed or by instantaneously broadcasting the base station corrections to the rovers. The latter approach is termed real-time differential GPS positioning. (Lillesand and Keiffer, 2004)

In general, each GPS satellite continuously transmits a microwave radio signal composed of two carriers, two codes, and a navigation message. When a GPS receiver is switched on, it will pick up the GPS signal through the receiver antenna.

Once the receiver acquires the GPS signal, it will process it using its built-in software. The partial outcome of the signal processing consists of the distances to the GPS satellites through the digital codes (known as the pseudoranges) and the satellite coordinates through the navigation message.

Theoretically, only three distances to three simultaneously tracked satellites are needed. In this case, the receiver would be located at the intersection of three spheres; each has a radius of one receiver-satellite distance and is centered on that particular satellite. From the practical point of view, however, a fourth satellite is needed to account for the receiver clock offset. The accuracy obtained with this method was limited to 100m for the horizontal component, 156m for the vertical component, and 340 ns for the time component, all at the 95% probability level. This low accuracy level was due to the effect of the so-called selective availability, a technique used to intentionally degrade the autonomous real-time positioning accuracy to unauthorized users. To further improve the GPS positioning accuracy, the so-called differential method, which employs two receivers simultaneously tracking the same GPS satellites, is used. In this case, positioning accuracy level of the order of a sub-centimeter to a few meters can be obtained.

According to the government and GPS receiver manufacturers, the GPS unit is accurate within 49 feet (that's 15 meters for metric-savvy folks). If the GPS reports that we're at a certain location, we can be reasonably sure that we're within 49 feet of that exact set of coordinates. GPS receivers tell us how accurate our position is. Based on the quality of the satellite signals that the unit receives, the screen displays the estimated accuracy in feet or meters.

Accuracy depends on:

- a) Receiver location
- b) Obstructions that block satellite signals

Even if we're not a U.S. government or military GPS user, we can get more accuracy by using a GPS receiver that supports corrected location data. Corrected information is broadcast over radio signals that come from either

- a) Non-GPS satellites
- b) Ground-based beacons

Two common sources of more accurate location data are

- a) Differential GPS (DGPS)
- b) Wide Area Augmentation System (WAAS)

**Source:** McNamara, J., 2004, GPS for Dummies, Wiley Publishing, Inc., Indianapolis, Indiana, pp. 49-68. (McNamara, 2004)

### 28.3.1 Differential Techniques

For both code-phase tracking navigation and carrier-phase tracking survey techniques, bias errors can be removed or mitigated by the use of differential techniques. (Dana, 1997)

- **Post-Processed Precise Ephemerides**

Some GPS position techniques make use of precise ephemeris data that is published by public and private agencies from the measurement of GPS signals at multiple reference locations. These data sets are available from agencies such as the International GPS Service for Geodynamics and the U. S. National Geodetic Survey within a few days or weeks of their reference times. (Dana, 1997). Precise orbital data used in post-processed position solutions can improve the accuracy of both code and carrier-phase derived solutions (Lewandowski, Petit, and Thomas 1991, 3). (Lewandowski, Thomas, 1991)

- **Differential GPS (DGPS)**

Selective Availability errors are correlated to a large extent for receivers within a few hundred kilometers of each other. For code tracking techniques, the ionospheric errors can be considered common to sites separated by a few hundred kilometers. Carrier tracking receivers can resolve differences in integer carrier wavelengths for receivers located within twenty to thirty kilometers of each other. Differential GPS (DGPS) is based on the assumption that bias errors common to two receivers, one a reference receiver at a known location, the other a remote receiver at an unknown location, can be measured at the reference receiver and applied to the remote receiver. DGPS techniques are based on the correction of individual SV pseudo-ranges or SV carrier phase measurements. While it would be possible to apply a simple position correction from the reference receiver to the remote, both receivers would have to be tracking the identical set of SVs with identical GDOP components for the position solution transfer to be effective. While this common-view technique can work for specialized applications where great care is taken to track the same set of SVs over identical time periods, for general-purpose DGPS positioning this technique is not recommended. In most DGPS positioning systems the bias errors in each SV signal are measured at the

reference receiver, which either sends corrections in real time to the remote receiver, or records the corrections for later application in post-processing software. In the remote receiver, or in post-processing software, the pseudo-range corrections are applied to the remote measurements prior to the formation of a position solution. (Dana, 1997)

- **Interferometric Processing**

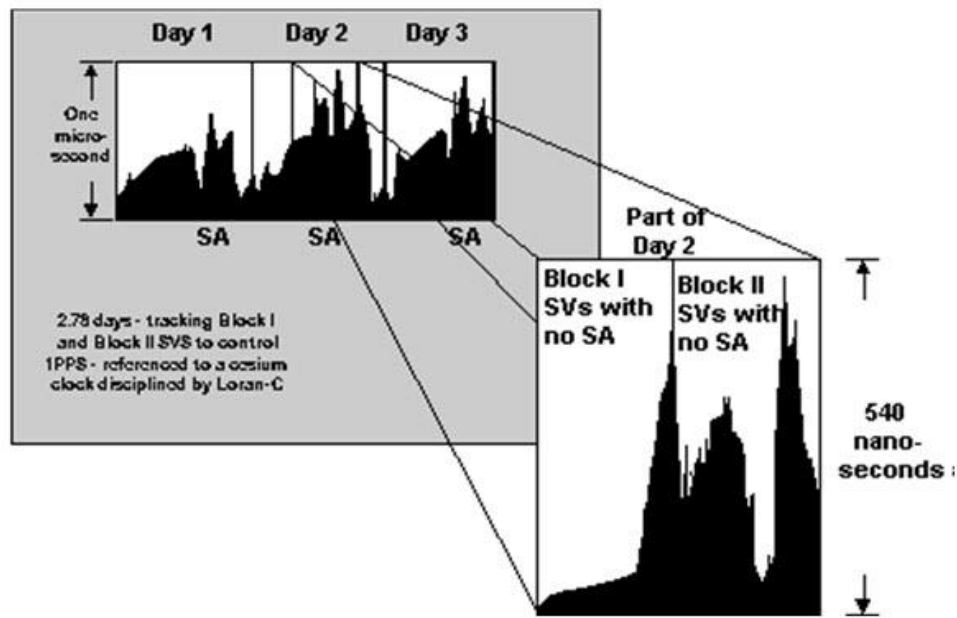
Measurements of crustal movements of the earth, earthquake fault line monitoring, and precise position transfer to isolated islands are possible using GPS interferometric techniques. In these special-purpose differential techniques, recordings of the pseudorandom codes and carrier-phase measurements on the GPS signals from distant sites are correlated and used along with precise ephemeris data in a post-processed mode to achieve position estimates in the centimeter range over thousands of kilometers. Global networks of carrier-phase tracking receivers are used in these processes. (Dana, 1997)

- **Common View**

Not usually suitable for control of real-time systems or for positioning systems, common view measurements are often used to transfer precise time from one location to another. Two receivers, both at known fixed positions, measure signals from a single satellite over the same carefully chosen observation period. Both receivers collect and filter data with the same methods. (Dana, 1997) The clock errors at the location with the reference standard are then transmitted to the other site, allowing the remote site to correct a clock with accuracies that have been obtained in the 8 nanosecond range for 1000 kilometer baselines and 10 nanoseconds for 5000 kilometer baselines (Lewandowski 1993, 138) (Lewandowski 1993, 138).

### **Selective Availability**

GPS has the proven potential to disseminate time and time intervals with accuracies of around 100 nanoseconds. Frequency control can be established globally to accuracies of a few parts in  $10^{-12}$ . The Block I satellites, placed in orbit during the 1970s and 1980s, gave promising results for worldwide time and frequency users. The new Block II satellites being launched now, however, have provisions for the implementation of selective availability (SA). SA is the intentional degradation of GPS signals and navigation messages to limit the full accuracy of the system to authorized users. Implemented last March, SA seriously affects both the time and frequency accuracies of GPS. Early tests indicate that accuracies are degraded from errors of approximately 100 nanoseconds to errors of approximately 500 nanoseconds. Fig. 28.4 shows actual results of selective availability on timing accuracies during a three-day period. Frequency control accuracies are reduced to a few parts in  $10^{-10}$ . (Dana and Bruce, 1990)



**Fig. 28.4. Selective availability effect on 1PPS control.**

(Source: Dana and Bruce, 1990)

Because most authorized users of GPS are in the military or otherwise affiliated with the Department of Defense (DoD), many non DoD time and frequency users are already being denied access to the accurate signals on the Block II satellites. Because the Block I satellites are nearing the end of their useful life and the current DoD policy is to continue SA, many users are turning back to cesium clocks and systems like Loran-C. To date, DoD has denied requests for undegraded signals from certain SVs, which would allow for accurate time and frequency use without compromising the agency's goal of degrading position accuracy. Fortunately, use of common-view, common-mode techniques can considerably reduce SA's effects on time transfer. The technique assumes that SA produces similar errors in both receivers involved in the time transfer. With careful time synchronization of the viewing window at the reference and common-view receivers, and with identical code phase-averaging techniques, the corrections transmitted from the reference station will remove most of the SA errors at the remote receiver site. (Dana and Bruce, 1990)

**Table 28.2. shows the accuracy we can expect from a GPS receiver. These numbers are guidelines; at times, we may get slightly more or less accuracy.**

	GPS Accuracy	
GPS Mode	Distance in Feet	Distance in Meters
GPS without SA	49	15
GPS with DGPS	10–16	3–5
GPS with WAAS	10	3

(Source: McNamara, 2004)

## Signal arrival time measurement

The position calculated by a GPS receiver requires the current time, the position of the satellite and the measured delay of the received signal. The position accuracy is primarily dependent on the satellite position and signal delay.

To measure the delay, the receiver compares the bit sequence received from the satellite with an internally generated version. By comparing the rising and trailing edges of the bit transitions, modern electronics can measure signal offset to within about one percent of a bit pulse width, or approximately 10 nanoseconds for the C/A code. Since GPS signals propagate at the speed of light, this represents an error of about 3 meters.

This component of position accuracy can be improved by a factor of 10 using the higher-chiprate P(Y) signal. Assuming the same one percent of bit pulse width accuracy, the high-frequency P(Y) signal results in an accuracy of or about 30 centimeters.

The U.S. Coast Guard operates the most common differential GPS correction service. This system consists of a network of towers that receive GPS signals and transmit a correction signal using omni-directional beacon transmitters. In order to receive the correction information, users must have a differential beacon receiver and beacon antenna in addition to their GPS. No such additional equipment is necessary with receivers that include Wide Area Augmentation System (WAAS) capability.

([http://en.wikipedia.org/wiki/Error\\_analysis\\_for\\_the\\_Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Error_analysis_for_the_Global_Positioning_System))

- **WAAS**

The WAAS consists of approximately 25 ground reference stations distributed across the United States that continuously monitor GPS satellite data. Two master stations, located on the East and West Coasts, collect the data from the reference stations and create a composed correction message that is location specific. This message is then broadcast through one of two geostationary satellites, satellites occupying a fixed position over the equator. Any WAAS-enabled GPS unit can receive these correction signals. The GPS receiver determines which correction data are appropriate at the current location.

The WAAS signal reception is ideal for open land, aircraft, and marine applications, but the position of the relay satellites over the equator makes it difficult to receive the signals when features such as trees and mountains obstruct the view of the horizon. In such situations, GPS positions can sometimes actually contain more error with WAAS correction than without. However, in unobstructed operating conditions where a strong WAAS signal is available, positions are normally accurate to within 3 m or better.

Paralleling the deployment of the WAAS system in North America are the Japanese Multi-functional Satellite Augmentation System (MSAS) in Asia and the European Geostationary Navigation Overlay Service (EGNOS) in Europe. Eventually, GPS users will have access to these and other compatible systems on a global basis.

Till the year 2002, the U.S. Global Positioning System has only one operational counterpart, the Russian GLONASS system. However, a fully comprehensive European global satellite navigation system, Galileo, is scheduled for operation in the coming years. The future for



these and similar systems is an extremely bright and rapidly changing one. (Lillesand and Keiffer, 2004)

## 28.4 Application of Global Positioning System

There are so many devices made with the implementation of Global Positioning System. Google Earth is the most famous application that uses the signals received by the GPS receivers. It enables public also to access the maps which tell the users about the locations all around the world. 3DEM is freely available software that will create 3D terrain scenes and flyby animations and export GIS terrain data files using any of the following freely available terrain data as a source. People use Global Positioning System for several uses. A research published states that the percentage of uses for each several requirement is as follows.

- Car navigation 37%
- Hand held 26%
- Tracking 10%
- GIS 8%
- Survey 7%
- Manufacturing 7%
- Vessel Voyage 2%
- Military Related 1%

United States and European countries show a rapid growth in using GPS for the car navigations and the number of GPS equipped mobile phone usage. Those facts prove that the Global Positioning System helps many people in many other ways.

Most people who have used a GPS probably can't imagine any limit to its applications. Even if its shortcomings are grievous (it can't be used indoors, nor very well in a forest, or where there are many tall buildings or cliffs), solutions have been developed to these problems. Usually these solutions involve broadcasting radio signals or pseudo-GPS signals that are highly accurate. The configuration of these systems is very complicated and requires large institutional investments. Most are made by governments. For instance, the European Union is developing a high-accuracy network (along the lines of the U.S. WAAS) for navigation purposes. Even if a GPS receiver lacks the ability to use these extra networks, GPS can still be used in a number of applications, some of which are described here.

### 28.4.1 GPS-supported ground surveys

Ground surveys are generally carried out for cadastral purposes of larger areas in terrain, where visibility requirements of the boundaries prohibit the use of photogrammetric techniques. They may also be preferable in countries where photo-adjudication is not accepted due to accuracy concerns. In Europe, cadastral data are already existent. Therefore, a survey cost comparison with photogrammetric methods is not useful there. However, in a

number of development projects, ground survey costs per parcel, including the land registration or land titling aspects, have been established.

In Albania, a new cadaster has been established by European funding at a cost of \$5 per parcel. The procedure used was aerial photography–aerial triangulation–digital elevation model–digital orthophotos generation followed by public photo adjudication process. In Georgia, a large German technical cooperation project was carried out by GPS-supported electronic tacheometers at a cost of \$10 per parcel.

A survey crew is able to measure about fifty parcels per day. In doing so, it has been proved useful to support the ground surveys with aerial photos or orthophotos. For this purpose, ‘digital plane tables’ in the form of large-screen PDAs may be used which record the measured GPS or electronic tacheometer measurements on the screen. These data are superimposed with preprepared (ortho) photographic data on the screen, which helps to identify points to be measured terrestrially. A ‘digital plane table’ costs about \$10 000.

An urban area of 250 km<sup>2</sup> has about 80 000 parcels. The cost of surveying these terrestrially would therefore be \$800 000, which is about the same as the photogrammetric line mapping cost at the scale 1:1000, but about four times as much as the photogrammetric mapping cost at the scale 1:2000. This corresponds to a terrestrial survey cost of \$3200/km<sup>2</sup>.

#### **28.4.2 Mapping and geographic information systems (GIS)**

Most mapping grade GPS receivers use the carrier wave data from only the L1 frequency, but have a precise crystal oscillator which reduces errors related to receiver clock jitter. This allows positioning errors on the order of one meter or less in real-time, with a differential GNSS (Global Navigation Satellite System) signal received using a separate radio receiver. By storing the carrier phase measurements and differentially post-processing the data, positioning errors on the order of 10 centimeters are possible with these receivers.

#### **28.4.3 Geophysics and geology**

High precision measurements of crustal strain can be made with differential GNSS by finding the relative displacement between GNSS sensors. Multiple stations situated around an actively deforming area (such as a volcano or fault zone) can be used to find strain and ground movement. These measurements can then be used to interpret the cause of the deformation, such as a dike or sill beneath the surface of an active volcano.

#### **28.4.4 Archeology**

As archaeologists excavate a site, they generally make a three-dimensional map of the site, detailing where each artifact is found. GPS capable of precise tracking of carrier phases for all or most of available signals in order to bring the accuracy of relative positioning down to cm-level values required by these applications.

#### **28.4.5 GPS Tracking**

In fact, it is this use which represents the simplest form of GPS tracking. The user is able, using a portable GPS device, to keep a track of where they have been, in order to be able to

either retrace their steps, or follow the same path again in the future. When combined with other technologies such as GPS phones, this also gives the possibility for other users of GPS to follow in the footsteps of the initial user; which can be a useful application of GPS tracking for field activities.

Where GPS tracking comes into its own, however, is when it is combined with other broadcast technologies such as radio. GPS watches, for example, can be fitted with a GPS receiver which is capable of calculating its position, whilst also broadcasting that using a miniature radio transmitter. The signal is relayed to a central command center equipped with GPS software systems which can track the position of the wearer, and either store it as a path, or relay that information to a third party. That third party could be an anxious parent, or the police. In fact there are a variety of GPS phones and wristbands which are sold in conjunction with a service which enables third parties to find out where their charges are at any time of the day or night.

#### **28.4.6 GPS Vehicle Tracking**

This is particularly useful when using GPS units attached to vehicles which have distinctive identification such as chassis numbers. The same principle applies as for a GPS tracking device designed to be worn by a human, except that the GPS is integrated within the vehicular electronics. This serves two purposes. On the one hand, it provides the driver with an integrated GPS system, without the necessity to purchase a car navigation system, or a PDA-based GPS system, whilst also offering the possibility to relay that information via a radio or mobile phone transmitter. In fact, these systems have already been tried in the field, primarily as a vehicle locator in the event that the vehicle to which the GPS vehicle tracking system is attached is stolen. The police, once informed, can find out from the control center where the vehicle is, and proceed to track it physically. A useful consequence of being able to use GPS vehicle tracking to locate a vehicle is that the manufacturer can also use the information to alert the driver as to when they near a service center. If, along with the GPS coordinates, the system relays telemetry information such as the status of the engine, time since the last service, or even information not relating to defects, the receiver of this information can make a decision as to what kind of alert to pass on to the driver. More and more people have used GPS-based systems in cars; many more have benefited from the use of GPS in cars, buses, trains, and trucks. The GPS receiver may be hidden in the dashboard, but may be critical for the taxi company to find out which taxi is closest to you when you call for a pickup. A GPS receiver can help a trucking company better organize deliveries to minimize the fuel used. A bus may have a GPS installed to help the bus company indicate to passengers how long they need to wait for the next one. Navigation systems are used for more than vehicles on land. They are also widely used for nautical and aeronautical navigation. They have become for many sailors irreplaceable because they work regardless of the weather and can easily be combined with computerized chart information. Almost all planes use, or will use, GPS. Together with high-precision positional transmission, planes can use GPS-based systems to land in any weather with centimeter precision.

#### **28.4.7 Coordinated Tracking**

This also opens up the possibility to allow for coordinated vehicle tracking, in which GPS tracking is used to share location information between several vehicles, all pursuing the same

end goal. It is an approach that has been used successfully in conjunction with GPS fish-finder units which help fisherman to locate, track and catch schools of fish. These units are more sophisticated than the average GPS unit, having other features such as depth gauges, tide time information and so forth. The basic GPS functionality is the same however, and units can either share that information with each other, or a central point. The central point can also be one of the fishing vessels, and it has on-board computer systems capable of reconciling all the locator information along with a map, thus allowing the different vessels to coordinate their actions. This also has military applications, of course, where units can share, in real time, information about their location, even when line-of-sight is no longer possible. In the past, this was done by relaying often inaccurate map co-ordinate estimations; now the locations can be called in with high absolute accuracy.

#### **28.4.8 Consumer GPS Tracking**

Despite its' hi-tech military and commercial fishing applications, as well as use in aviation GPS, the principal application of GPS tracking will be in providing an enabling technology to augment existing systems. These systems will include cell phones and vehicles, usually in conjunction with a central point of service designed to keep track of the location.

#### **28.4.9 Use of GPS to determine well location**

There is typically a lot of good geologic and hydrologic information contained on the well log and drilling report forms. In order for this data to be used for mapping purpose and some regulatory programs, the exact location of well has to be known. GPS have made it possible.

#### **28.4.10 Weather Prediction Improvements**

Measurement of atmospheric bending of GNSS satellite signals by specialized GNSS receivers in orbital satellites can be used to determine atmospheric condition such as air density, temperature, moisture and electron density. Such information from a set of six micro-satellites, launched in April 2006, called the Constellation of Observing System for Meteorology, Ionosphere and Climate COSMIC has been proven to improve the accuracy of weather prediction models.

#### **28.4.11 Photographic Geocoding**

Combining GNSS position data with photographs taken with a (typically digital) camera, allows one to view the photographs on a map or to lookup the locations where they were taken in a gazetteer. It's possible to automatically annotate the photographs with the location they depict by integrating a GNSS device into the camera so that co-ordinates are embedded into photographs as Exif (Exchangeable image file format) metadata. Alternatively, the timestamps of pictures can be correlated with a GNSS track log.

#### **28.4.12 Marketing**

Some market research companies have combined GIS systems and survey based research to help companies to decide where to open new branches, and to target their advertising according to the usage patterns of roads and the socio-demographic attributes of residential zones.

### 28.4.13 Social Networking

Cellular phones equipped with GPS technology, offering the ability to pinpoint friends on custom created maps, along with alerts that inform the user when the party is within a programmed range.

### 28.4.14 Altitude Information

GPS has transformed how altitude at any spot is measured. GPS uses an ellipsoid coordinate system for both its horizontal and vertical datums. An ellipsoid – or flattened sphere – is used to represent the geometric model of the earth.

### 28.4.15 Application to Water Resources

In an effort to protect water resources, GPS is being used to collect the coordinates for well heads as part of the Well Head Protection Program. GPS has also been used to produce coordinates for potable surface water intakes, and reservoir boundaries. To more effectively manage regulatory permits across the various environmental permitting programs, GPS is being used to collect coordinates for facilities that have permits. These include facilities that discharge to surface water, ground water, air, store hazardous waste onsite and/or have underground storage tanks. Environmental monitoring programs are using GPS to generate coordinates for monitoring stations. Water monitoring programs have been determining coordinates of sampling stations for existing water quality monitoring networks. Should a major oil spill occur in their waters, coordinates for the spill location and aerial extent of the plume could be collected. In short order, an effective booming strategy could be developed to protect environmentally sensitive areas in the region of the spill. In the event of a major natural disaster, GPS can be used to assist in the damage assessment and inventory.

### GPS Techniques for Water Stage Measurement and River Slope Calculation Wetland Area

Study of river basin many times involves difficulties of making use of hydrological data such as river stage height due to inaccessibility and political boundaries. The slope of river is very important hydrological data especially from point of view of hydrologic and hydrodynamic models calibration. These specific hydrologic applications need calculation of local changes of water level and slope. Traditionally slope is calculated using data available from water gauge which are always insufficient and the distance between two successive gauge stations varies from few to several kilometers. For hydrodynamic model calibration the water stage determined based on water level measured should be within few hundred meters. In natural river valley such detailed measurements are difficult to perform by use of classical geodetic leveling technique; in case of marginal river wetland it is even impossible, because of harsh measuring condition such as:

- Disturbance of natural vegetation.
- Many oxbow and wetland areas.
- Unstable organic ground.
- Very few network coordination points.

The GPS technique seems to be optimal tool for altitude measurement in wetlands. The average vertical measurement error for DGPS is about 3m which is sufficient for river slope calculation. GPS become not only accurate but also very fast measurement technique. Moreover GPS technique allows performing high accuracy measurement of all three co-ordinates including altitude, easy and fast way. The DGPS definitely can be used for hydrologic application in various water-bodies.

### **Use of GPS receivers as a soil moisture network for water cycle studies**

Soil moisture is fundamental to land surface hydrology, affecting flooding, groundwater recharge, and evapotranspiration [Viterbo and Betts, 1999]. It also influences weather and climate via its influence on turbulent and radiative fluxes between the land surface and atmosphere [Entekhabi and Rodriguez-Iturbe, 1994]. The global distribution and temporal variations of soil moisture are sought both for analyses and modeling purposes. Soil moisture is measured in situ at many locations, both as part of individual studies or as part of monitoring networks.

Measurements of soil moisture, both its global distribution and temporal variations, are required to study the water and carbon cycles. Signals routinely recorded by Global Positioning System (GPS) receivers for precise positioning applications can also be related to surface soil moisture variations. Various studies depicted significant correlation between the result obtain from GPS network and soil moisture fluctuation measured in the top 5 cm of soil with conventional sensors.

### **28.4.16 Application to Agriculture**

Global positioning systems (GPS) are widely available in the agricultural community. Farm uses include:

- Mapping yields (GPS + combine yield monitor)
- Variable rate planting (GPS + variable rate planting system)
- Variable rate lime and fertilizer application (GPS + variable rate controller)
- Field mapping for records and insurance purposes (GPS + mapping software)
- Parallel swathing (GPS + navigation tool).

The Global Positioning System (GPS) provides opportunities for agricultural producers to manage their land and crop production more precisely. Common names for general GPS applications in farming and ranching include precision agriculture, site-specific farming and prescription farming. GPS applications in farming include guidance of equipment such as sprayers, fertilizer applicators and tillage implements to reduce excess overlap and skips. They can also be used to precisely locate soil-sampling sites, map weed, disease and insect infestations in fields and apply variable rate crop inputs, and, in conjunction with yield monitors, record crop yields in fields.

GPS and associated navigation system are used in many types of agricultural operations. These systems are useful particularly in applying pesticides, lime, and fertilizers and in tracking wide planters/drills or large grain-harvesting platforms. GPS navigation tools can replace foam for sprayers and planter/drill-disk markers for making parallel swaths across a field. Navigation systems help operators reduce skips and overlaps, especially when using methods that rely on visual estimation of swath distance and/or counting rows. This technology reduces the chance of misapplication of agrochemicals and has the potential to safeguard water quality. Also, GPS navigation can be used to keep implements in the same traffic pattern year-to-year (controlled traffic), thus minimizing adverse effects of implement traffic.

### **Yield Monitoring Systems**

Yield monitoring systems typically utilize a mass flow sensor for continuous measuring of the harvested weight of the crop. The sensor is normally located at the top of the clean grain elevator. As the grain is conveyed into the grain tank, it strikes the sensor and the amount of force applied to the sensor represents the recorded yield. While this is happening, the grain is being tested for moisture to adjust the yield value accordingly. At the same time, a sensor is detecting header position to determine whether or not yield data should be recorded. Header width is normally entered manually into the monitor and a GPS, radar or a wheel rotation sensor is used to determine travel speed. The data is displayed on a monitor located in the combine cab and stored on a computer card for transfer to an office computer for analysis. Yield monitors require regular calibration to account for varying conditions, crops and test weights. Yield monitoring systems cost approximately \$3,000 to \$4,000, not including the cost of the GPS unit.

### **Field Mapping with GPS and GIS**

GPS technology is used to locate and map regions of fields such as high weed, disease and pest infestations. Rocks, potholes, power lines, tree rows, broken drain tile, poorly drained regions and other landmarks can also be recorded for future reference. GPS is used to locate and map soil-sampling locations, allowing growers to develop contour maps showing fertility variations throughout fields. The various datasets are added as map layers in geographic information system (GIS) computer programs. GIS programs are used to analyze and correlate information between GIS layers.

### **Precision Crop Input Applications**

GPS technology is used to vary crop inputs throughout a field based on GIS maps or real-time sensing of crop conditions. Variable rate technology requires a GPS receiver, a computer controller, and a regulated drive mechanism mounted on the applicator. Crop input equipment such as planters or chemical applicators can be equipped to vary one or several products simultaneously. Variable rate technology is used to vary fertilizer, seed, herbicide, fungicide and insecticide rates and for adjusting irrigation applications.

## Precision Farming

Location coordinate information is needed in precision agriculture to map in-field variability, and to serve as a control input for variable rate application. Differential global positioning system (DGPS) measurement techniques compare with other independent data sources for sample point location and combine yield mapping operations. Sample point location can be determined to within 1m (3ft) 2dRMS using CIA code processing techniques and data from a high-performance GPS receiver. Higher accuracies can be obtained with carrier phase kinematic positioning methods, but this required more time. Data from a DGPS CIA code receiver are accurate enough to provide combine position information in yield mapping. However, distance data from another source, such as a ground-speed radar or shaft speed sensor, needed to provide sufficient accuracy in the travel distance measurements used to calculate yield on an area basis.

Precision farming, sometimes called site-specific agriculture, is a strategic task for agriculture: indeed it has the potential to reduce costs through more efficient and effective applications of crop inputs; it can also reduce environmental impacts by allowing farmers to apply inputs only where they are needed at the appropriate rate. Precision farming requires the use of new technologies, such as GPS, environmental sensors, satellites or aerial images and GIS to assess and understand variations. The various research deals with potentialities and limits of GPS for navigation in agricultural applications. GPS needs for farming applications are:

- Low cost in order to allow farmers to buy GPS technologies;
- High precision in order to reduce the use of pesticides and fertilizers by means of an exact track.

At first, static and kinematic tests needed to be performed, simulating the typical behavior of an agricultural vehicle and using different kinds of GPS receivers and navigation software. The experimental results are presented: particularly, advantages and disadvantages of the popular Kalman filtering on trajectories are discussed. Starting from the analyses of the previous results, and taking into account the typical user requirements, a preliminary design for a new prototype has been done; particularly, both needed instrumentations and their costs and a proposal of a new navigation algorithm will be presented.

Precision farming is a method of crop management by which areas of land within a field may be managed with different levels of input depending upon the yield potential of the crop in that particular area of land. The benefits of so doing are two fold:

- The cost of producing the crop in that area can be reduced.
- The risk of environmental pollution from agrochemicals applied at levels greater than those required by the crop can be reduced.

Precision farming is an integrated agricultural management system incorporating several technologies. The technological tools often include the global positioning system GPS, geographical information system GIS, remote sensing, yield monitor and variable rate technology. The paper talks about the use of GPS to support agricultural vehicle guidance.



Equipment for this purpose consists on a yield monitor installed: the system supports human guide by means of a display mapping with a GIS the exact direction produced by GPS receiver put on vehicle top: the driver follows it to cover in an optimal path the full field. GPS receivers for this applications require, not only an high accuracy to ensure the reduction of input products, but even an easy and immediate way of use for farmers; without forgetting low costs.

Obviously the technology to achieve high precision still exists but it is too expensive and difficult to use for not skilled people. Survey modality usually adopted in agricultural applications is real time kinematic positioning, DGPS RTK, which enable to have a good accuracy by means of corrections received. In this experimentation the aim is to obtain a sub-metric accuracy using low cost receivers, which can provide only point positioning. These receivers have been developed for maritime navigation purposes; our aim is their optimization in order to apply them for land navigation in particular for farming activities. Some tests using these receivers were carried out, but results were not satisfying and probably the reason has to be assigned to the implementation of a Kalman filtering inside the receiver software. This is the starting point for a new project, at the moment still in progress, which aim to develop a new algorithm based on Kalman filter. Its purpose is to improve low cost receiver outputs in order to optimize trajectories and to reach needed accuracy in vehicle positioning during agricultural activities.

#### 28.4.17 Others

- **Hiking:** More and more hikers turn to GPS to help them find out more exactly where they are and to help them to plan a route before they go. GPS may not be reliable in canyons or along steep cliffs, but in most situations and weather it provides accurate positional information. Some map makers have started to change their map designs to make it easier for hikers to use. Some tourist areas offer GPS for people to help them follow a certain tour.
- **Aids for the Visually Impaired:** Combined with acoustic or tactile signaling devices, GPS can be used to help visually impaired people find their way in new settings and navigate places that rapidly change—for example, a state fair or a college campus, as was done by Professor RegGolledge and others at the University of California at Santa Barbara



## ***Module 11: Applications of Remote Sensing for Earth Resources Management***

### **Lesson 29 Agriculture and Forestry**

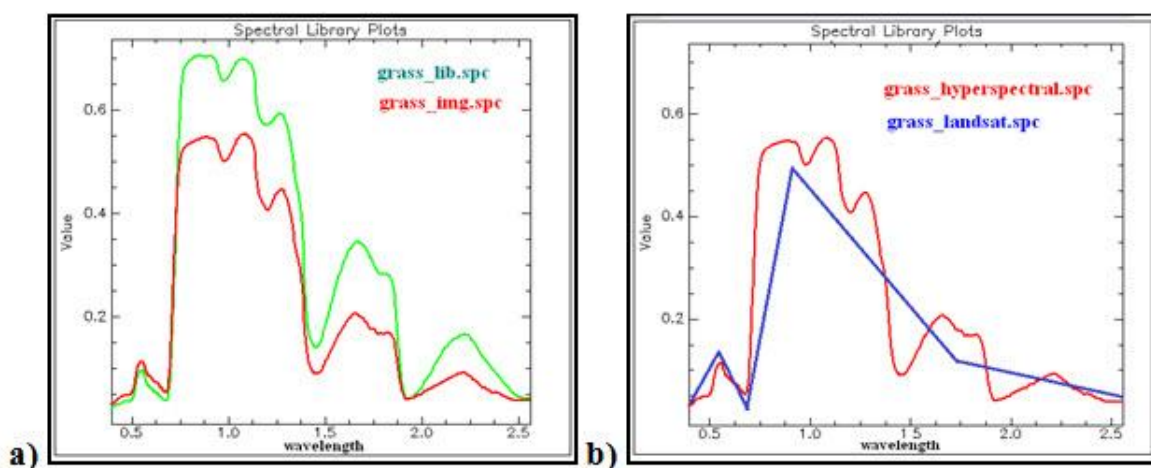
#### **29.1 Crop Type Mapping**

Crop type identification and mapping has a number of important aspects. It can serve for production statistics together with yield prediction, mapping soil productivity, assessment of crop damage and monitoring of farming activities. These include identifying the crop types (winter and spring cereals, rapeseed, sugar beet, potato, maize, grass, etc.) and delineating their parcel extent. Crop type identification is often based on multitemporal, multispectral high resolution imageries, while parcel boundaries delineation is more often based on very-high resolution imageries.

Remote sensing offers an efficient means of collecting the information, in order to map crop type and acreage. Remote sensing can also provide state information about the health of the vegetation. The spectral reflections vary with respect to changes in the phenology and crop health. This can be measured and monitored by multispectral sensors.

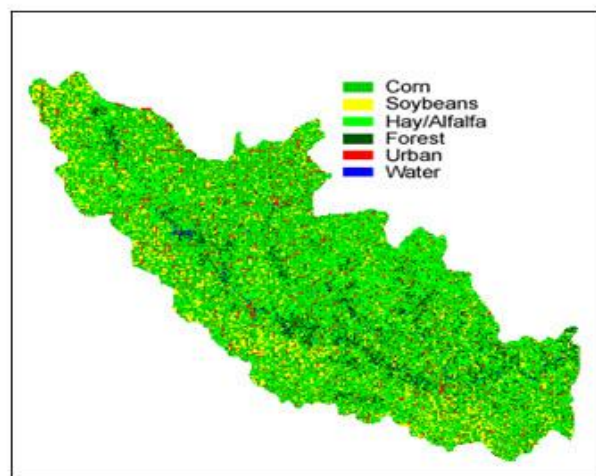
Crop type identification and mapping is based on use of multitemporal imagery to enhance the classification by taking into account changes in reflectance as a function of plant phenology. This in turn requires calibrated sensors and frequent repeat imaging throughout the vegetation season (source: [www.gisat.cz/content/en/applications/agriculture](http://www.gisat.cz/content/en/applications/agriculture); Feb. 12, 2013).

Hyperspectral remote sensing has also helped enhance more detailed analysis of crop classification. Thenkabail et al. (2004) performed rigorous analysis of hyperspectral sensors (from 400 nm to 2500 nm) for crop classification based on data mining techniques consisting of principal components analysis, lambda-lambda models, stepwise Discriminant Analysis and derivative greenness vegetation indices. Hyperspectral sensors images the earth surfaces in hundreds of narrow spaced contiguous bands, from which the spectral signature curves of different objects can be obtained. Spectral signatures are the specific combination of emitted, reflected or absorbed electromagnetic radiation (EM) at varying wavelengths which can uniquely identify an object. These unique spectral signatures of different crops help in their discrimination in the hyperspectral imagery, which can be obtained in multispectral imagery also like Landsat data, but a relatively low accuracy. In comparison to Landsat Enhanced Thematic Mapper data and other broadband sensors, these hyperspectral approaches increased accuracy for crop type mapping (refer Fig. 29.1).



**Fig. 29.1. Comparison of spectral signature curve a) between library spectra (green) and hyperspectral image spectra (red), b) between hyperspectral image spectra (red), Landsat image spectra (blue).**

From the Fig. 29.1. it can be easily inferred that the hyperspectral satellite image data is more suitable than multispectral data for crop type mapping. An classified image is given below showing different crop types identified and mapped.



**Fig. 29.2. Land cover/land use map of the Maquoketa River watershed for 1995.**

(Source: [www.umesc.usgs.gov/reports\\_publications/psrs/psr\\_1997\\_11.html](http://www.umesc.usgs.gov/reports_publications/psrs/psr_1997_11.html))

## 29.2 Crop Monitoring and Damage Assessment

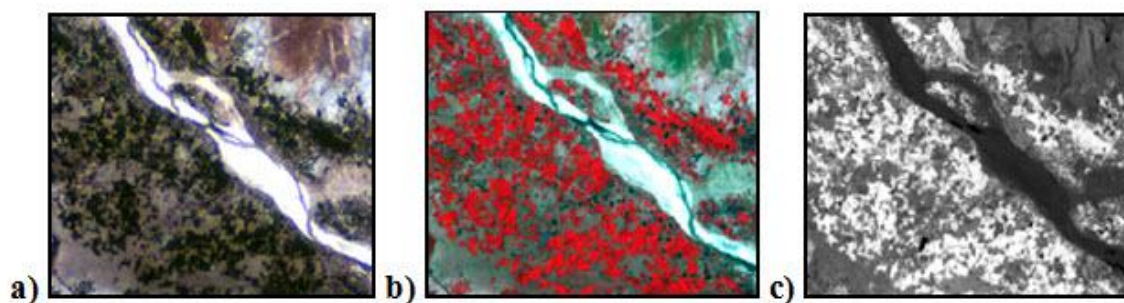
Assessment of the health of a crop, as well as early detection of crop infestations, is critical in ensuring good agricultural productivity. Stress associated with, for example, moisture deficiencies, insects, fungal and weed infestations, must be detected early enough to provide an opportunity for the farmer to mitigate. This process requires that remote sensing imagery be provided on a frequent basis (at a minimum, weekly) and be delivered to the agriculture planner quickly, usually within 2 days.

Remote sensing has a number of attributes that lend themselves to monitoring the crop health. One advantage of optical (VIR) remote sensing is that it can see beyond the visible wavelengths into the infrared, where wavelengths are highly sensitive to crop vigour as well as crop stress and crop damage. Remote sensing imagery also gives the required spatial overview of the land. Remote sensing can aid in identifying crops affected by conditions that are too dry or wet, affected by insect, weed or fungal infestations or **weather related damage**. Images can be obtained throughout the growing season to not only detect problems, but also to monitor the success of the treatment.

Healthy vegetation contains large quantities of chlorophyll, the substance that gives most vegetation its distinctive green colour. In referring to healthy crops, reflectance in the blue and red parts of the spectrum is low since chlorophyll absorbs this energy. In contrast, reflectance in the green and near-infrared spectral regions is high. Stressed or damaged crops experience a decrease in chlorophyll content and changes to the internal leaf structure. The reduction in chlorophyll content results in a decrease in reflectance in the green region and internal leaf damage results in a decrease in near-infrared reflectance. These reductions in green and infrared reflectance provide early detection of crop stress.

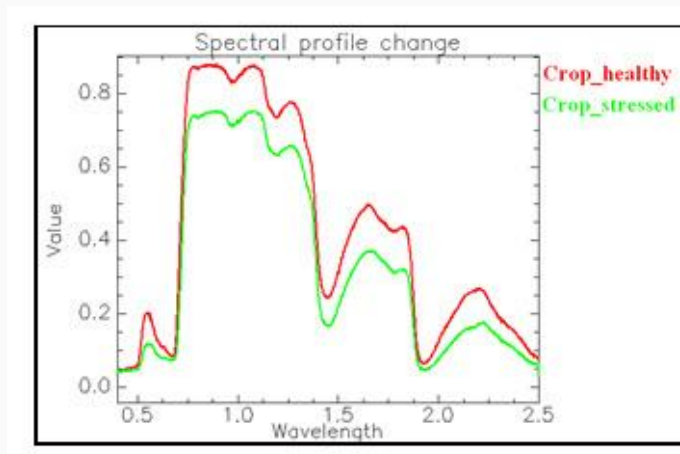
([http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote\\_sensing/fundamentals/1696](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote_sensing/fundamentals/1696); date: 19<sup>th</sup> Feb., 2013).

In case of multispectral data, examining the ratio of reflected infrared to red wavelengths is an excellent measure of vegetation health. The normalized difference vegetation index (NDVI), vegetation condition index (VCI), leaf area index (LAI), General yield unified reference index (GYURI), and temperature crop index (TCI) are all examples of indices uses the multispectral data, used for mapping and monitoring drought and assessment of vegetation health and productivity. An example is the difference between irrigated crops and non-irrigated land is shown in Fig 29.3, in the true color image green areas are cultivated land and grey colored areas are waste lands or lands with very less vegetation. In a FCC (False Color Composite) image, where infrared reflectance is displayed in red, the healthy vegetation appears bright red, while the rangeland remains quite low in reflectance. Healthy plants have a high NDVI value (bright in Fig. 29.3. c)) because of their high reflectance of infrared light, and relatively low reflectance of red light. Phenology and vigour are the main factors in affecting NDVI.



**Fig. 29.3. a) True colored image of an area, b) standard false colored image of the same area, c) NDVI output image of the same area.**

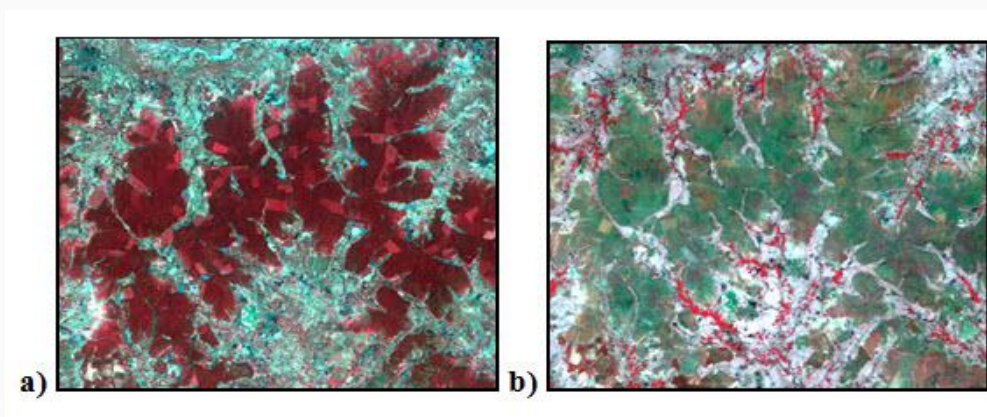
In case of hyperspectral remote sensing in crop monitoring, there will be change in the spectral profile of the crops (Fig. 29.4).



**Fig. 29.4 Change in spectral profile of crop due to infection in crop.**

Remote sensing doesn't replace the field work performed by the agriculture planner to monitor their fields, but it does direct them or provide enough information for the areas in need of immediate attention.

Fig. 29.5 provides scenario of an area before and after long time drought condition.



**Fig. 29.5. a) Before incidence of drought, b) just after occurrence of drought.**

### 29.3 Mapping of Forest Resources and Species Identification

The utility of the remote sensing data for forest type, canopy density mapping and monitoring on various scales is well established. Multispectral data such as that from IRS LISS-4 with 5.8m, *IKONOS* with 1m, and *QuickBird* with 0.61m spatial resolutions and hyperspectral data such as hyperion with 242 spectral resolutions provides an unprecedented opportunity to identify and monitor forest. Barring single species dominated forests and forest plantations, majority of the Indian forests are highly heterogeneous. This makes their differentiation, delineation and mapping a difficult task. Problem gets further compounded in case of forests that are located in hilly and mountainous regions on account of topographic effects.



Although passive satellite sensors offer routine and repeated assessments at very high spatial and spectral resolution, but this technology has difficulty in capturing reflectance beyond upper canopy layers and is better suited for mapping horizontal structure rather than vertical structure. In humid tropical areas, forest resource assessments and measurements are difficult to obtain because of cloudy conditions hindering conventional remote sensing efforts, and difficult terrain impeding ground surveys. Active remote sensing technologies offer great potential to spatially map a forest three-dimensional (3D) structure at various scales from landscape, stand and individual tree levels. Active satellite systems based on interferometric synthetic aperture radar (InSAR) can provide measures of horizontal and vertical structure of vegetation at regional scales, but this technology does not provide the spatial resolution required in detailed forest studies. However, LIDAR (Light Detection and Ranging System) data provides improved capabilities for the estimation of forest canopy dimensions at the individual tree level.

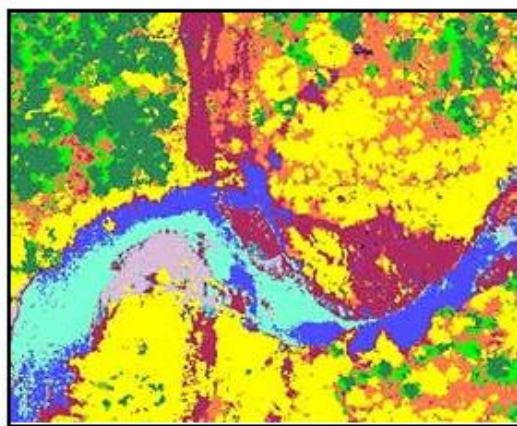
**Table 29.1 Indexing algorithms for forest mapping using multispectral remote sensing data**

Vegetation index	Formulae (source)
Simple Ratio (SR)	$= \text{NIR} / \text{R}$
Normalized Difference Vegetation Index (NDVI)	$= (\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$
Soil Adjusted Vegetation Index (SAVI)	$= (1 + L) * (\text{NIR} - \text{R}) / (\text{NIR} + \text{R} + L)$ L = correction factor, its value is dependent on the vegetation cover.
Atmospherically Resistant Vegetation Index (ARVI)	$= (\text{NIR} - \text{R}) / (\text{NIR} + \text{QRB})$ , $\text{QRB} = \text{R} - \gamma (\text{B} - \text{R})$ , $\gamma = 1$ (calibration factor) (Kaufman and Tanré 1992)
Modified Soil Adjusted MSAVI	$= (1 + L) * (\text{NIR} - \text{R}) / (\text{NIR} + \text{R} + L)$ Vegetation Index $L = 1 - 2a * \text{NDVI} * \text{WDVI}$ (Qi et al. 1994)
Enhanced Vegetation Index (EVI)	$= G * (\text{NIR} - \text{R}) / (\text{NIR} + \text{C1} * \text{R} - \text{C2} * \text{B} + L)$ $G = 2.5$ , $\text{C1} = 6$ , $\text{C2} = 7.5$ , $L = 1$ (Liu and Huete 1995)
Modified Simple Ratio MR	$= (\text{NIR} / \text{R} - 1) / ((\text{NIR} / \text{R})^{1/2} + 1)$ (Chen 1996)

(Source: [cdn.intechopen.com/pdfs/30811/InTech-Using\\_remotely\\_sensed\\_imagery\\_for\\_forest\\_resource\\_assessment\\_and\\_inventory.pdf](https://cdn.intechopen.com/pdfs/30811/InTech-Using_remotely_sensed_imagery_for_forest_resource_assessment_and_inventory.pdf))

Where, R, NIR and B are red, near-infrared, and blue bands, respectively. For Modified Soil Adjusted Vegetation Index,  $\text{WDVI} = \text{NIR} - a\text{R}$  ( $a = 0.08$ , slope of the soil line). For Enhanced Vegetation Index, G, C1, C2 and L are coefficients to correct for aerosol scattering, absorption, and background brightness.

Hyperspectral imageries provide a very high spectral resolution which can be used to generate signatures of vegetation species and certain stresses (e.g. infestations) on trees. Hyperspectral data offers a unique view of the forest cover, available only through remote sensing technology. An output map generated from hyperspectral data is given in Fig. 29.6.



**Fig. 29.6. Output classified map using hyperspectral data, where: Dark green: conifers, Green: lower branches, Light purple: gravel, Yellow: deciduous, Orange: dry ground cover, Red: wet ground cover, Blue (light): water, Blue (dark): deep or clear water**

(Source: [www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1258](http://www.nrcan.gc.ca/earth-sciences/geography-boundary/remote-sensing/fundamentals/1258))

## 29.4 Forest Fire Mapping

Land and forest degradation process is considered to be one of the major environmental problems, which has lead to a variety of environmental disasters that adversely affect human life. Traditional land use practices and changes in weather patterns have affected the incidence of fires over a period of time. In tropical deciduous forests, fire is a natural factor due to high levels of water stress during summer. Frequent occurrence of fire is one of the reasons for the degradation of forests in India also. Annual fires may decrease the growth of the grasses, shrubs and forests, which may result in increased soil erosion.

Fire is the most spectacular natural disturbance that affects the forest ecosystem composition and diversity. Fire has a devastating effect on the landscape and its impact is felt at every level of the ecosystem. Majority of induced fires in the Indian forests are mainly caused for timber harvesting, land conversion, agriculture, cattle grazing, fodder and creating access to forest interiors. Fires set on farmlands, to clear agricultural residues, sometimes spread into the forests.

Satellite remote sensing provides the only practical means of monitoring and acquiring information about the spatial distribution of fire scars and fire activity. Understanding the behavior of forest fires, the factors that contribute to making an environment fire prone and the factors that influence fire behavior is essential for forest fire risk zone mapping.

(Source: [www.jeb.co.in/journal\\_issues/201011\\_nov10/paper\\_13.pdf](http://www.jeb.co.in/journal_issues/201011_nov10/paper_13.pdf)).

The severity of a burn for post-fire ecological effects can be assessed using some indexing method over satellite image data, such as normalized difference vegetation index (NDVI), composite burn index (CBI) and differenced Normalized Burn Ratio (dNBR) etc. The algorithms of these indexes are as follows:

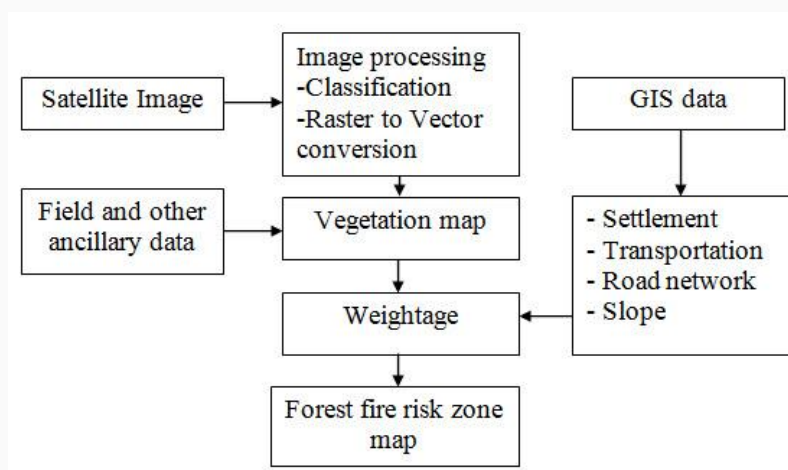
$$NDVI = \frac{NIR - RED}{NIR + RED} \quad NBR = \frac{NIR - TIR}{NIR + TIR}$$

$$dNBR = NBR_{pre-fire} - NBR_{post-fire}$$

NIR-Near Infrared band & RED- Red band of a multispectral data set.

Geographical Information System (GIS) along with satellite remote sensing imageries aid more accuracy in forest fire detection, monitoring, assessing the forest fire risk zone mapping and management.

Flowchart (Fig 29.7) for preparing a forest risk map is shown below taken from a study reported for Bhadra Wildlife Sanctuary, Karnataka.



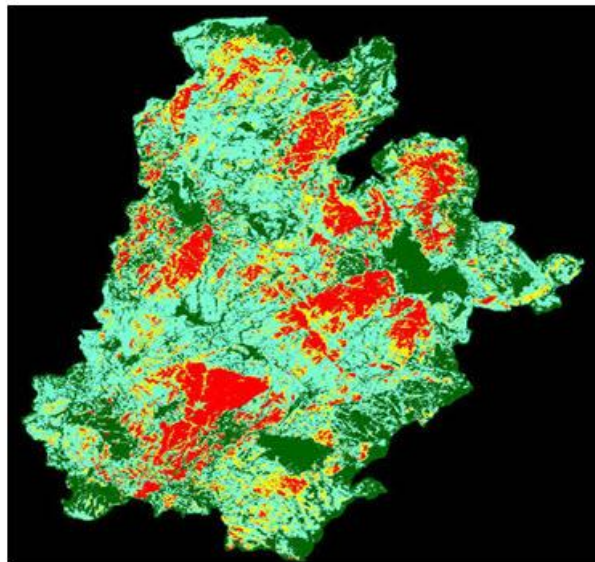
**Fig. 29.7. Flow chart of the procedure followed for preparing forest risk map in Bhadra Wildlife Sanctuary, Karnataka**

(Source: [www.jeb.co.in/journal\\_issues/201011\\_nov10/paper\\_13.pdf](http://www.jeb.co.in/journal_issues/201011_nov10/paper_13.pdf))

One change detection process can also be used to show the change of forest cover before and after of a forest fire hazard. Depending on the purpose different kinds of procedure can be taken to get satisfied output results. Satellite image interpretation of a forest during fire hazard can be used to monitor real time situation, and will help the disaster management personnel to take necessary action in this regard.



Fig. 29.8 shows the burnt and unburnt areas generated from bNBR data.



**Fig. 29.8.** The image generated from the dNBR data. The green colors indicate unburned or very lightly burnt areas; red areas illustrate highly burnt land. Credit: US Forest Service, Remote Sensing Applications Center (RSAC)

(Source: [www.nasa.gov/mission\\_pages/fires/main/post-fire2011.html](http://www.nasa.gov/mission_pages/fires/main/post-fire2011.html))

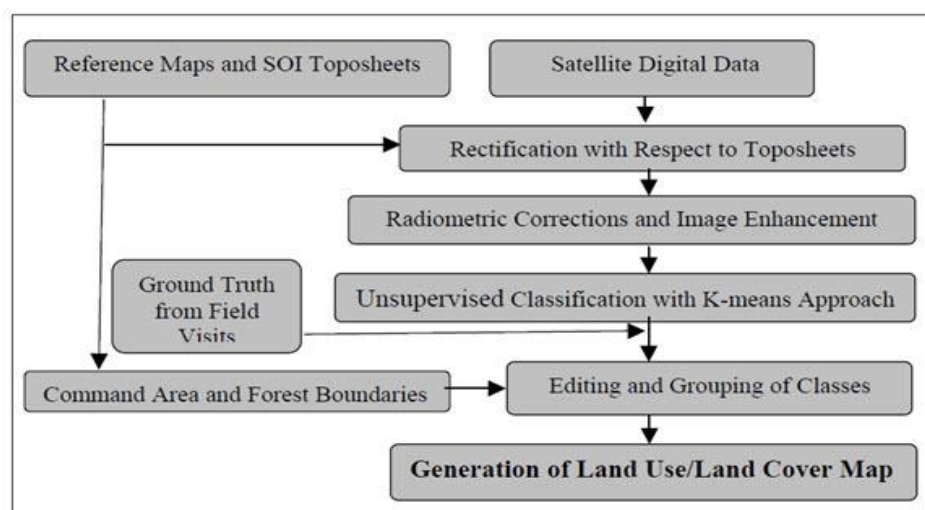
## *Lesson 30 Land Use and Land Cover Mapping and Change Detection*

### 30.1 Land Cover Mapping

Land cover refers how much of a region of the earth surface is covered by forests, wetlands, impervious surfaces, agriculture, water body etc. Water body includes wetlands or open water. Whereas land use shows how people use the land – whether for social-economic development, conservation, or mixed uses.

Land cover can be identified and mapped using satellite remote sensing data in a short time over a large area at a time. These maps help planners and managers to understand about the current landscape which helps them to evaluate past management decisions as well as gain insight into the possible effects of their current decisions before they are implemented. As an example, coastal managers use land cover data and maps to understand the impacts of natural phenomena and human use of the landscape. Maps can help managers to assess urban growth, model water quality issues, predict and assess impacts from floods and storm surges, track wetland losses and potential impacts from sea level rise, prioritize areas for conservation efforts, and compare land cover changes with effects in the environment or to connections in socio-economic changes such as increasing population.

A simple land cover land use map can be prepared by supervised or unsupervised techniques or combination of both. The output map depends on the spectral, spatial and radiometric resolution of the image data set. The basic need of an image classification is a general knowledge about the landscape of the study area, post classification field verification and accuracy assessment. The steps to generate land use land cover map are shown through flowchart in Fig 30.1.



**Fig. 30.1. Flowchart generation of land use land cover map.**

(Source: Gontia, 2005)

The study has been carried out for irrigation planning using remote sensing and GIS for wheat crop grown in Tarafeni South Main Canal command (TSMC), (Gontia, 2005). A land use land cover map classification generated for TSMC is shown in Fig 30.2. The corresponding error matrix and accuracy assessment carried out are given in the Table 30.1 and Table 30.2 respectively.

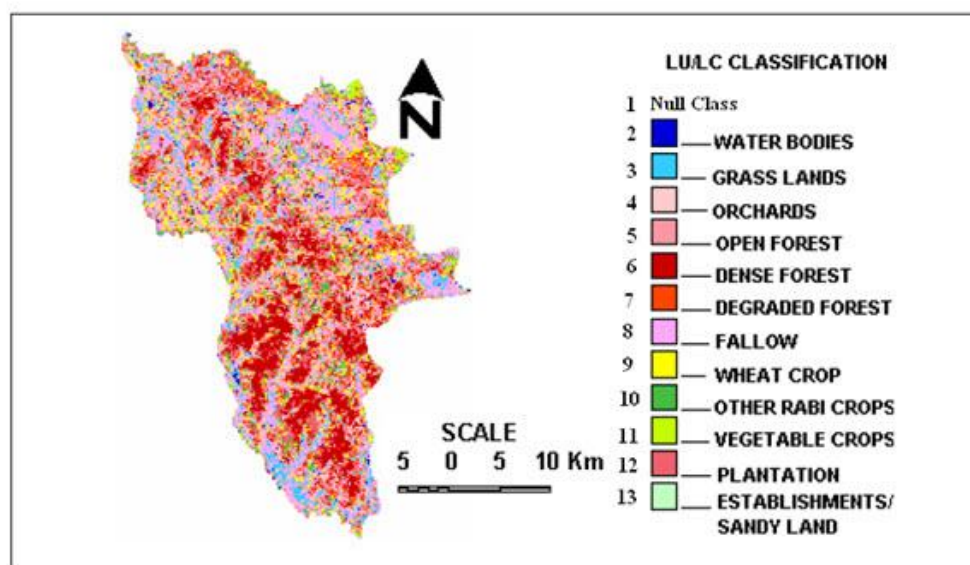


Fig. 30.2. Land use land cover classification map of an area.

(Source: Gontia, 2005)

Table 30.1 Error matrix of the classified image

Classified Data	Reference Data													Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	
1	166	0	0	0	0	0	0	1	0	0	1	1	0	169
2	0	19	8	0	0	0	0	0	0	0	1	0	0	28
3	0	5	21	0	1	0	0	2	3	0	0	0	0	32
4	0	0	0	12	1	2	0	0	0	0	0	0	1	16
5	0	0	1	2	16	0	1	1	2	0	0	0	1	24
6	0	0	0	1	0	35	2	0	0	1	0	0	0	39
7	0	0	0	3	2	2	13	0	0	1	1	0	2	24
8	0	0	0	0	0	0	1	17	3	1	0	0	1	23
9	0	0	3	0	1	1	0	1	51	2	2	1	1	63
10	0	0	1	0	0	0	0	1	2	36	3	1	0	44
11	0	0	1	0	0	0	0	1	1	1	19	3	0	26
12	0	0	1	0	0	2	1	0	1	1	2	18	0	26
13	0	1	0	2	1	0	4	1	2	0	3	0	8	22
Total	166	25	36	20	22	42	22	25	65	43	32	24	14	536

(Source: Gontia, 2005)

**Table 30.2 Accuracy statistics of land use land cover classification**

Class Name	Producer's Accuracy, %	User's Accuracy, %
Null	100.00	98.22
Water Bodies	76.00	67.86
Grass Lands	58.33	65.63
Orchards	60.00	75.00
Open Forest	72.73	66.67
Dense Forest	83.33	89.74
Degraded Forest	59.09	54.17
Plantation	68.00	73.91
Fallow	78.46	80.95
Wheat Crop	83.72	81.82
Other Rabi Crops	59.38	73.08
Vegetables	75.00	69.23
Establishments/Sandy Lands	57.14	36.36

Overall accuracy = 80.41 per cent and Overall Kappa statistics ( $\hat{k}$ ) = 0.7734.

(Source: Gontia, 2005)

### 30.2 Land use/Land Cover Change Detection

Land use and land cover is an important component in understanding the interactions of the human activities with the environment and thus it is necessary to simulate changes. Land use and land cover are dynamic in nature. Land use and land cover (LULC) change is a major issue of global environment change. Changes may be natural or manmade for land use land cover modification. Land degradation also takes place due to increased human activity without proper management practices. Changes in land use can be due to urban expansion and the loss of agriculture land, changes in river regimes, the effects of shifting cultivation, the spread of erosion and desertification and so on. This, therefore, requires not only the identification of features but also the comparison of subsequent data in order to recognize when valid change has taken place. The land use change has a direct bearing on the hydrologic cycle. Various hydrologic processes such as interception, infiltration, evapotranspiration, soil moisture, runoff and ground water recharge are influenced by land use land cover characteristics of a catchment. Application of remotely sensed data made possible to study the changes in land cover in less time, at low cost and with better accuracy in association with Geographical Information System (GIS) that provide suitable platform for data analysis, update and retrieval helps in monitoring, mapping and management of natural resources.

A number of techniques for accomplishing change detection using satellite imagery can be used, which are broadly grouped into two general types: (1) those based on spectral classification of the input data such as post-classification comparison and direct two-date classification; and (2) those based on radiometric change between different acquisition dates, including (a) image algebra methods such as band differencing, ratioing, and vegetation indices; (b) regression analysis; (c) principal component analysis; and (d) change-vector analysis (CVA).

Chen. et al. (2003) stated the methodology of change-vector analysis (CVA) which is shown through Fig. 30.3. The output image generated using CVA method is given in Fig 30.4.

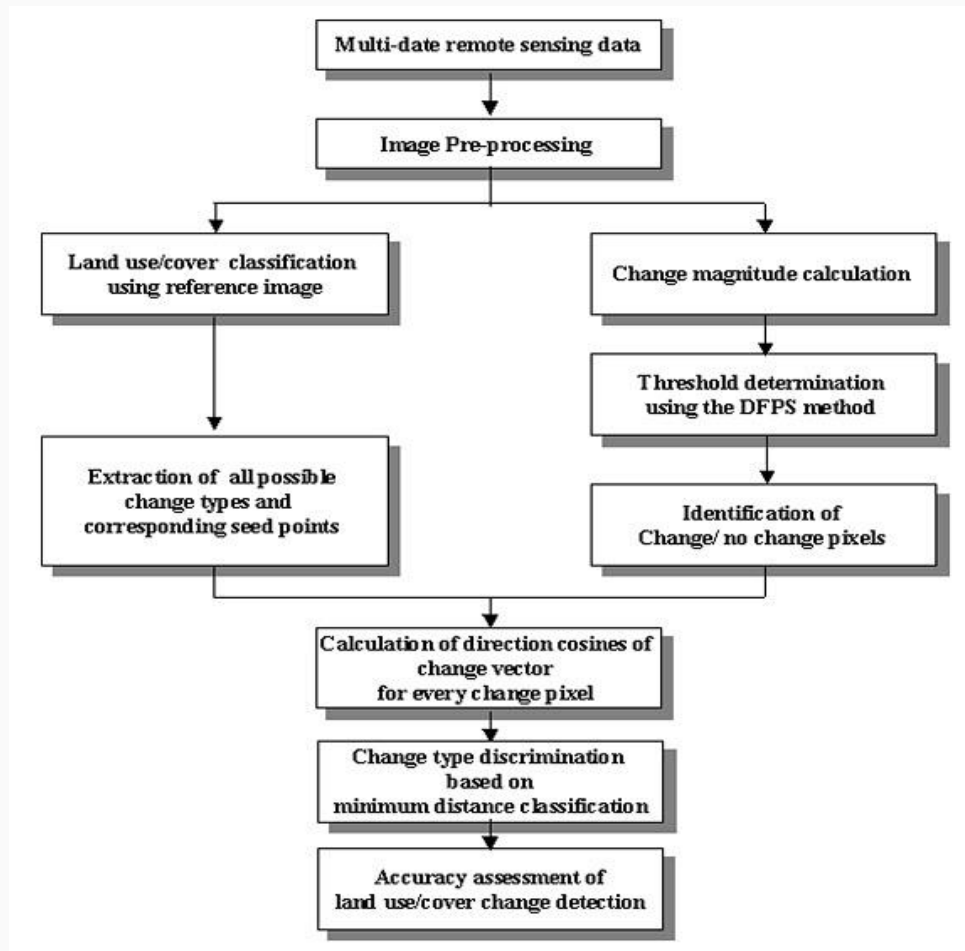


Fig. 30.3. Flowchart of land-use/land-cover change detection based on improved change vector analysis.(Source: Chen, et al., 2003)

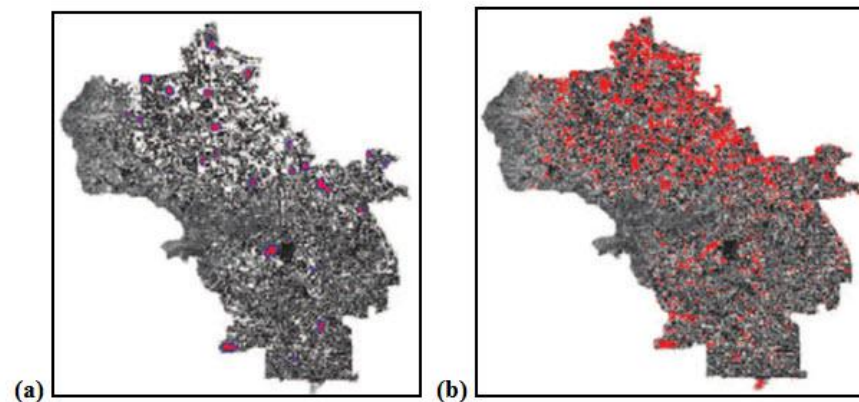


Fig. 30.4. The selected change sample areas and results of change/no-change detection. (a) Typical change sample areas (red) with outer no-change buffer boundary (blue). (b) Change pixels detected by the improved CVA method.

(Source: Chen, et al., 2003)



Another study on land use land cover change detection has been carried out for Tarafeni catchment of West Bengal by Jena (2002). The output change detection map and statistics of obtained through analysis of satellite image of Tarafeni catchment are given in Fig. 30.5 and Table 30.3.

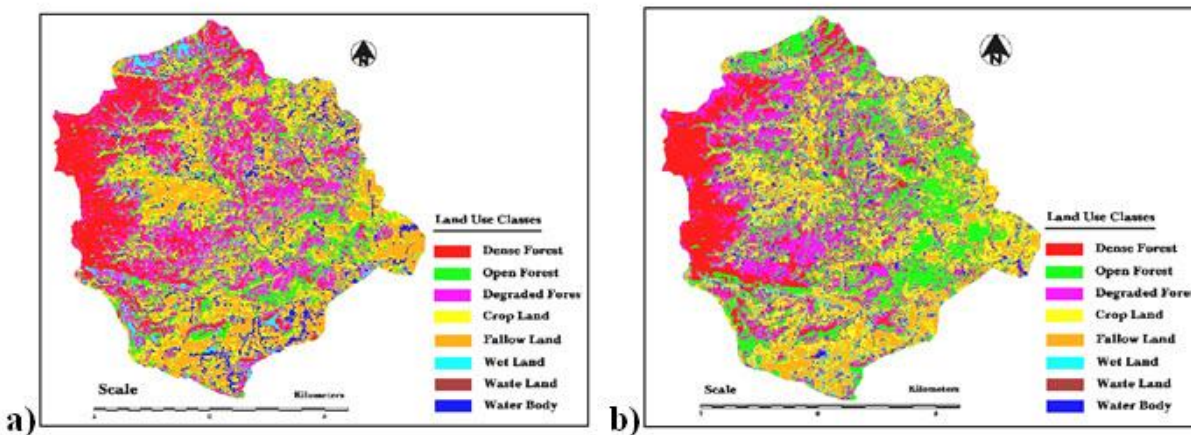


Fig. 30.5. Land cover land use map of Tarafeni catchment of year (a)1989, (b)2000.

(Image Source: Jena, 2002)

Table 30.3 Land use land cover statistics of Tarafeni catchment

Land use land cover Type	Area (ha.)	
	Year 1989	Year 2000
Dense forest	2715.0	1874.9
Open forest	2109.6	3046.6
Degraded forest	3337.0	2932.7
Cropland	1961.6	2207.0
Fallow land	4029.8	3815.4
Wet land	605.07	639.40
Waste land	0.00	367.9
Water bodies	1048.1	922.13

(Source: Jena, 2002)

From the Table 30.3 it can be observed that in 11 years open forest, crop land, wet land and waste land have increased, where as dense forest, degraded forest, fallow land, water bodies have decreased. Such information can be used for planning purpose.

## **Lesson 31 Soil Mapping and Soil Erosion Estimation**

### **31.1 Soil Mapping**

There are several thousand types of soil throughout the world, a fact that is not surprising when bearing in mind the differences there are worldwide in the agents responsible for the building and forming of soil (landscape, climate, geology, vegetation, time and man). We know that there are at least these numbers of soil types because they have been mapped. In the past 50 years many countries of the world have been involved in making maps of their soils to determine the range of soil types in their territory, types of soils occur and how they can be used. Soil mapping involves locating and identifying the different soils that occur, collecting information about their location, nature, properties and potential use, and recording this information on maps and in supporting documents to show the spatial distribution of every soil. In order to map and identify different types of soil it is necessary to have a system of soil classification.

Traditional soil mapping is conducted with an auger and spade at intervals throughout the landscape. The intervals between inspections can be according to a pre-determined grid (grid-survey) or, more often, are based upon the judgment of the surveyor who uses their knowledge of the inter-relationship between soil type and landscape, geology, vegetation, etc. to determine where to make inspections. Auger borings are supplemented by excavated profile pits at determined points in the landscape. These profile pits are used to demonstrate lateral changes in the soil as well as vertical ones, and are important for the full description of soil type and for the taking of soil samples for chemical, physical and less commonly, biological laboratory analysis. In this way a picture is built up of the soil in a region and its relationship to the landscape in which it lies, (Source: [www.soil-net.com/legacy/advanced/soil\\_mapping.htm](http://www.soil-net.com/legacy/advanced/soil_mapping.htm)).

Soils can be mapped at a range of scales from very detailed at 1:1,250 to 1:5,000 by which the pattern of soils in individual fields can be identified, through to scales of 1:500,000 to 1:500,000,000 which provide only a very generalized picture of the soils of a country or continent. Sometimes soils are mapped with a specific aim in mind, such as the suitability of soils for a particular crop, suitability for irrigation, erosion risk and many other specific needs or environmental threats. Most organized soil surveys in the past have been general purpose surveys. These have the advantage that they provide the basis for many different uses, some of which may yet not be known.

### **Soil Mapping of Upper DVC Catchment**

The Damodar a comparatively small river running a distance of 540 km originates from the Khamarpat at an altitude of 1067m in the district of Jharkhand. It has a drainage area of 24,235 km<sup>2</sup>. The area shown in Fig 31.1 lies between 23°34' to 24°9' N latitude and 85°00' to 87°00' E Longitude. Mean annual rainfall in the basin is of the order of 1,100 mm and about 80% of rain precipitates during the summer monsoon (June to September). The soil mapping

of this study area was done for estimation of soil erosion using USLE. The details of study area is available in Mishra (2009).

The soil maps of the study area obtained from AASSLU in the scale of 1:250,000 were traced, scanned and exported to Erdas Imagine 8.5. There were total five maps, four comprised of soils of Bihar (recently renamed as Jharkhand) soils and one map was with soils information of West Bengal soil information. The scanned maps were loaded in ERDAS and georeferenced. Then boundaries of different soil textures were digitized carefully and the polygons representing various soil categories were assigned with different colours for identification. Using ArcView 3.3, digitized soil map was clipped with micro-watershed boundaries and all required data like soil texture, bulk density etc. were extracted for each micro-watershed. The soil is classified mainly on the percentage of available sand into various texture groups such as sandy loam (80.16%), sandy clay loam (9.97%), and loamy sand (9.88%). Fig. 31.1. shows the soil map for Damodar valley catchment. Soil data on texture, structure, permeability and organic matter content by analyzing soil samples were used to derive soil erodibility factor (K).

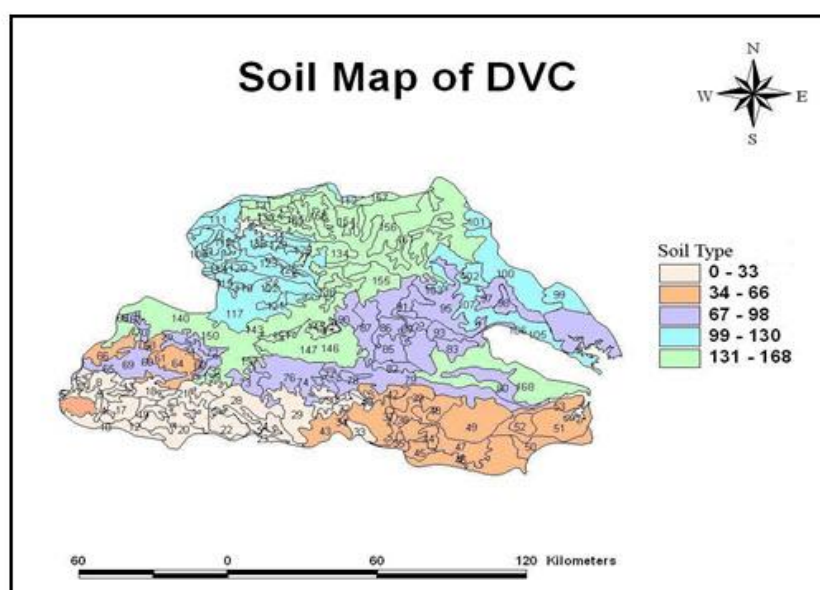


Fig. 31.1. Soil map of DVC. (Source: Shinde, 2009)

## 31.2 Soil Erosion Estimation and Best Management Practices

### 31.2.1 Soil Erosion Estimation using USLE

For assessing soil erosion from the watershed, several empirical models based on the geomorphological parameters were developed in the past to quantify the sediment yield. Several other methods such as Sediment Yield Index method and Universal Soil Loss Equation (USLE) are extensively used for prioritization of the watersheds. The USLE formula is given in equation 31.1, has been widely applied at a watershed scale on the basis of lumped approach to catchment scale. The USLE model application in the grid form in GIS environment allows us to analyze soil erosion in much more detail and in more meaningful scene. It is more reasonable to use the USLE on physical basis than to apply it to an entire watershed as a lumped model. Although, GIS permits more effective and accurate



application of the USLE model for small watershed, most GIS-model applications are subject to data limitations. The universal soil loss equation (USLE) has widely been used for predicting annual average soil loss for specific areas. In this equation, soil loss ( $A$ ) is defined as a product of rainfall erosivity factor  $R$ , soil erodibility factor  $K$ , slope steepness factor  $S$ , slope length factor  $L$ , cover management factor  $C$ , and support practice factor  $P$ .

$$A = R K L S C P \quad (31.1)$$

The Soil Conservation Department of DVC, Hazaribagh, Jharkhand (India) has established network of stream gauging stations and sediment observation posts for monitoring hydrological parameters. The time series data on rainfall, runoff, and sediment yield are available through the network of gauging stations. It is possible to analyze the cause-effective relationship between the various factors that affect the hydrologic response of a watershed. Information obtained from such analysis can be used to develop appropriate management plans for soil conservation and water resource development to increase the agricultural productivity of the region on sustainable basis.

### 31.2.1.1 Development of Model Database for USLE

#### Rainfall Erosivity ( $R$ ) Factor

The rainfall erosivity factor ( $R$ ) map is prepared using daily rainfall data from six stations located in the Upper DVC. The  $R$  factor map (Fig. 31.2) is based on 9 year-average rainfall data which is used to calculate annual average  $R$  factor values. All the storms do not produce runoff and hence storms more than 12.5 mm were only used in computation. Eq.31.2, is used for computation of  $R$  factor using the daily rainfall amount, (Source: Panigrahi et al., 1996 and Wischmeier, 1959).

$$R = P^2(0.00364 \log_{10} P - 0.000062) \quad (31.2)$$

where  $P$  = storm depth in mm.

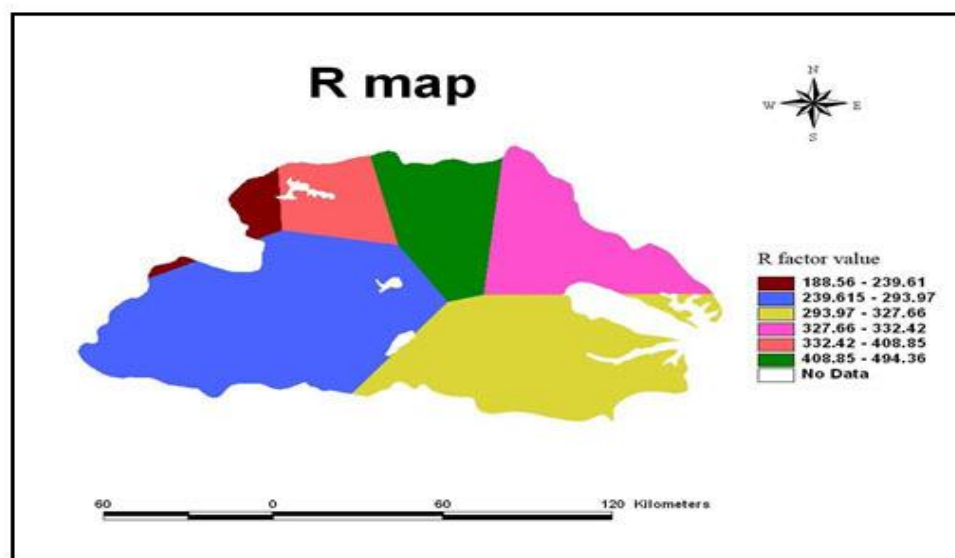


Fig. 31.2. Rainfall erosivity ( $R$ ) factor map of DVC. (Source: Shinde, 2009)

## Soil Erodibility Factor (K)

Soil erodibility factor (K) is a measure of the total effect of a particular combination of soil properties. Some of these properties influence the soil's capacity to infiltrate rain, and therefore, help to determine the amount of rate of runoff; some influence its capacity to resist detachment by the erosive forces of falling raindrops and flowing water and thereby determine soil content of the runoff. The inter-relation of these variables is highly complex. The K factor in USLE model relates to the rate at which different soils erode. The soil erodibility factor (K) was computed using field and laboratory estimated physico-chemical properties of the surface soils. The laboratory soil analysis was carried out to determine soil texture, structure, permeability and organic matter content for 179 soil samples covering various soil group of the area. The K-factor is computed for each soil type based on the various soil properties using equation 31.3. K factor map generated using GIS is shown in Fig. 31.3, (Source: Wischmeier *et al.*, 1971)

$$100K = 2.1M^{1.14}(10^{-4}) (12 - a) + 3.25 (b - 2) + 2.5 (c - 3) \quad (31.3)$$

Where, K = soil erodibility factor, M = percentage silt, very fine sand and sand > 0.10 mm, a = organic matter content, b = structure of the soil, c = permeability of the soil.

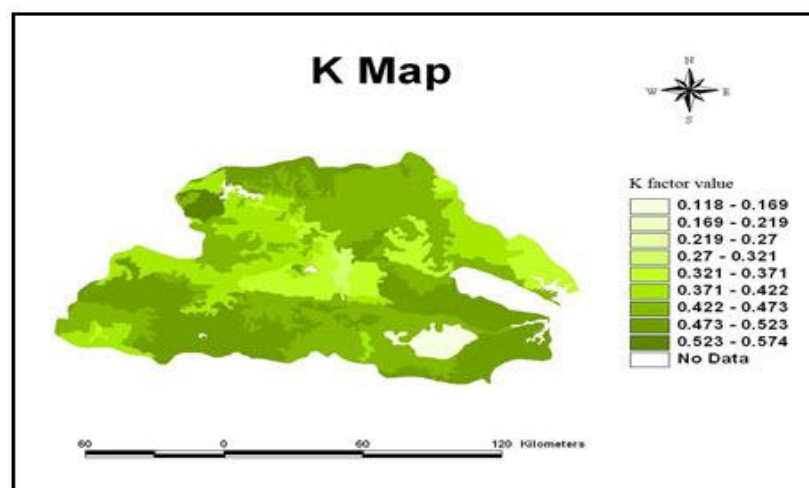


Fig. 31.3. Soil erodibility (K) factor map of DVC. (Source: Shinde, 2009)

## Topographic factor (LS)

Topographic factor (LS) is the expected ratio of soil loss per unit area from a field slope to that from a 22.13 m length of uniform 9% slope under otherwise identical conditions. It is this quantity that is significant in rain dependent processes taking place at soil surfaces. Hillslope gradient (S) and length (L) factors are sometimes combined into a topographic factor (LS) while estimating soil erosion. Slope Length factor: The relationship between the slope steepness in percentages (Sp) and slope length in meters (L) was used to generate slope length map. It is given by

$$L = 0.4 * Sp + 40$$

By applying this equation the resultant map was prepared for slope length.

Although L and S factors were determined separately, the procedure has been further simplified by combining the L and S factors together and considering the two as a single topographic factor (LS). The combined LS factor layer was generated and shown in Fig. 31.4.

The slope length modified by Wischmeier and Smith (1965) is used to estimate as given by-

$$LS = (L / 22.1)^m * (65.41 \sin^2 \theta + 4.56 \sin \theta + 0.065) \quad (31.4)$$

Where, LS is the slope length and gradient factor,  $\theta$  is angle of the slope and L is slope length in meters and  $m = 0.5$  if the percent slope is 5 or more, 0.4 on slopes of 3.5 to 4.5 percent, 0.3 on slopes of 1 to 3 percent, and 0.2 on uniform gradients of less than 1 percent.

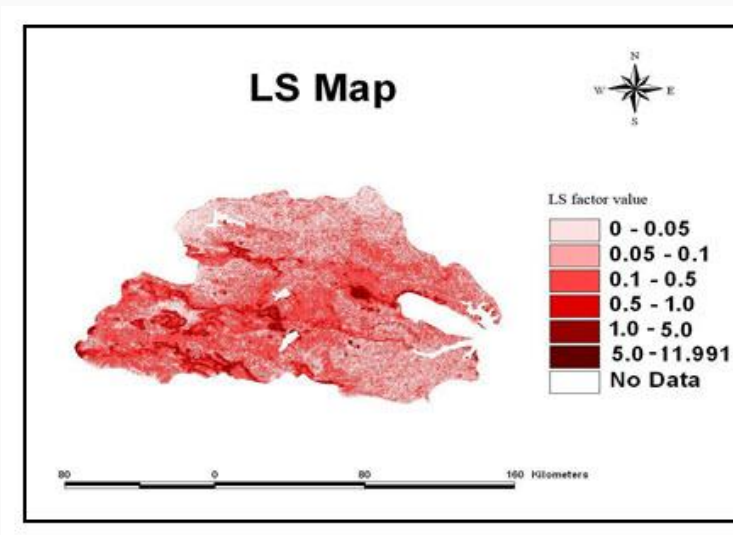


Fig. 31.4. Topographic factor (LS) map of DVC. (Source: Shinde, 2009)

### Crop Management Factor (C)

The crop management factor (C) reflects the combined effect of cover, crop sequence, productivity level, length of growing season, tillage practices, residue management and the expected time distribution of erosive rainstorm with respect to seeding and harvesting date in the locality. Actual loss from the cropped field is usually much less than the amount of soil loss for a field kept continuously in fallow conditions. This reduction in soil loss depends on the particular combination of cover, crop sequence and management practices.

### Conservation Practice Factor (P)

Presence of soil conservation practices in the region is duly considered in the USLE by including support conservation practice factor (P). Depending upon the LU/LC map, the C factor values are assigned to all classes as shown in Table 31.1. The prepared C factor map is shown in Fig 31.5. In the present study, the land use/land cover map was derived from the supervised classification of LANDSAT ETM images which served as a guiding tool in the allocation of C factor for different land use classes. Conservation practice factor (P) depends on land use/cover information of the study area. If no conservation measures are established, then P factor value of 1 is considered.

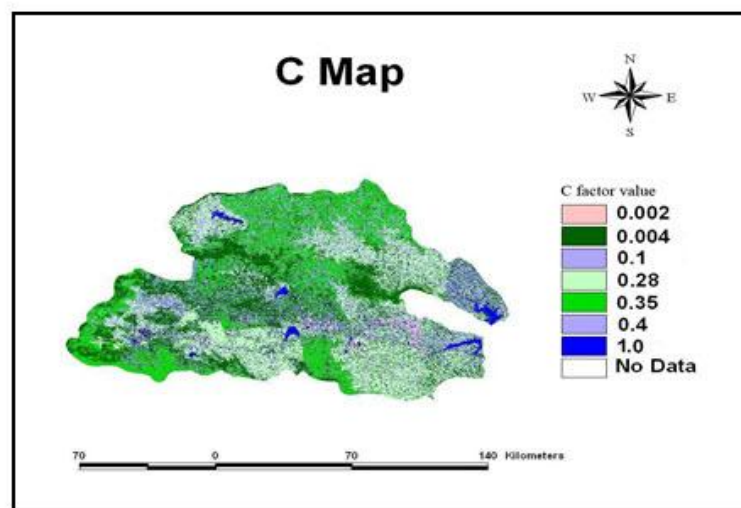


Fig. 31.5. C factor map of DVC. (Source: Shinde, 2009)

Table 31.1. Land use/land cover class and C factor value

Land use/land cover	C value
Forest	0.004
Range	0.1
Water body	1
Urban	0.002
Wetland	0.4
Corn	0.35
Paddy	0.28

(Source: Shinde, 2009)

### 31.2.1.2 Average Annual Soil Loss using USLE

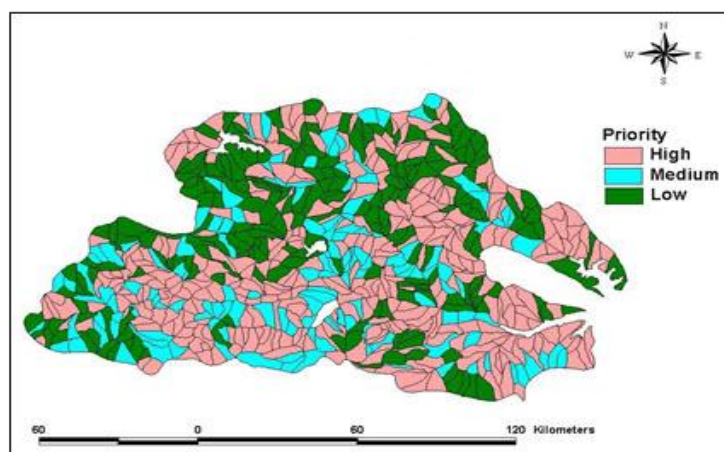
Annual soil loss based on 9-year average rainfall erosivity factor is termed as average annual soil loss. The annual soil loss for all the micro-watersheds was calculated by using annual average R (based on annual average rainfall data of 1993-2001) and K, LS, C and P factors. The soil erosion rate (t/ha/yr) was estimated as total soil loss of a micro-watershed (t/yr) divided by geographical area of particular micro-watershed (ha). The classification of erosion rate has been given three categories of soil loss. Fig. 31.6 indicates distribution of the micro-watersheds (MWs) of DVC according to soil loss categories. All the layers viz. R, K, LS, C and P were generated in GIS and were overlayed to obtain the product, which gives annual soil loss (A) for the upper DVC. These values gave annual soil loss per hectare per year at pixel level. These values are converted to the loss per pixel in  $m^2$  (i.e.  $84 \times 84$  m) and all values are added in GIS domain to obtain total annual soil loss per hectare for the upper DVC as a

whole. A surface of annual soil loss was overlaid with the micro-watershed map of DVC which contains 705 micro-watersheds to get micro-watershed wise soil loss. All values for each micro-watershed were summed using 'summarized zone' option in ArcView GIS 3.3 to obtain soil loss in each micro-watershed. The total loss thus obtained per micro-watershed is then converted to the soil loss/ha/year for the upper Damodar Valley catchment. Micro-watersheds are then classified in different soil loss classes as given in Table 31.2.

**Table 31.2. Soil loss categories according to annual average soil loss**

Soil loss (t/ha/yr)	Class	Number of Micro-watersheds	Area (km <sup>2</sup> )
>20	High	345	8503.25
5.0 to 20.0	Medium	159	3981.95
0.0 to 5.0	Low	201	5271.77
Total		705	17756.98

(Source: Shinde, 2009)



**Fig. 31.6. Priority map of DVC Using USLE. (Source: Shinde, 2009)**

### 31.3 Best Management Practices

Adoption of Best Management Practices (BMPs) is required to reduce sediment yield, runoff rate after the deciding of priority of micro-watershed. The various types of BMPs can be suggested to control soil loss for the prioritization soil conservation works in micro-watersheds (Shinde, et al, 2011). BMPs are not exactly same for all micro-watersheds. It varies according to geomorphology of micro-watersheds. The BMPs like afforestation, trenching, bunding, vegetative barriers, sediment basin, bend way weir and different kinds of check dams are selected in different studies. It was found from available treatment data from DVC that, for the completion of treatment in any micro-watershed 3 to 4 years are required. In present study treatment factors (P) were found as 0.7, 0.6 and 0.5 after 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> year of treatment respectively. While calculating micro-watershed wise soil erosion, P factor applied

was 0.5 to those micro-watershed in which treatment has been completed (i.e. after 3<sup>rd</sup> year of treatment).



www.AgrIMoon.Com

## Lesson 32 Water Resources and Flood Damage Assessment

### 32.1 Water Resources and Water Quality Detection

Water is a precious commodity which is rapidly becoming scarce. The demands for water made by agriculture, industry and for human and cattle consumption is increasing rapidly, particularly with the global increase in population. Therefore, the optimum use of water resources is absolutely essential and this can be done by adopting suitable management practices and preventing wastage caused by uncontrolled runoff. According to Central Water Commission report, after allowing for evaporation losses etc. the countries estimated water potential from surface flow is 1,800 cubic km and all this is carried by the various river systems in the four or five months in a year of the monsoon season.

Remote sensing & GIS applications have been operationalized in most of the natural resources themes and at present that trend is on integrated approach to arrive at sustainable development packages. With the specific problems such as high runoff, shallow ground water, soil erosion etc from hilly areas. The use of remote sensing becomes highly relevant for the development and management of water resources of such regions. Remote sensing & GIS techniques have been applied in several locations for identifying the suitable locations of water harvesting structures. *Padmavathy et al, 1992*, used GIS for identifying check dams sites in Alur taluka of Hasan district in Karnataka. They selected second and third order streams from the drainage map and superimposed over lineament (fracture) map to avoid fracture zones. Based on the different themes and different weightages were assigned for water spread area and immediate downstream area. Based on the normalized cumulative weighted index the sites were selected after verifying through field checks.

Ground water is a dynamic renewable natural resource, its availability with good quality and quantity in appropriate time and space is of significant importance. As a result of population explosion, urbanization and deforestation, there is continuous pressure on it. Advent of satellite Remote Sensing and Geographic Information System (GIS) has added a new dimension in the field of ground water investigations. Although ground water investigation through satellite data and aerial photographs are indirect approach and involve complex job, when it is integrated with field data (e.g. thickness of weathered zone, surface fractures, faults, depth to water level, seasonal water level fluctuations, water quality and tubewell discharge) in a GIS environment, one can get a fairly accurate idea about ground water of the area by preparing "Hydromorphological (HMG)" maps (*Roy 1993, Jugran 1995, Rao et al, 1992; Shah et al, 1992*)

Locating and delineating with remote sensing data are done most easily in NIR wavelength because of the absorption property. However, various conditions of water bodies manifest themselves primarily in visible wavelengths. Clear water absorbs little energy having wavelengths less than 0.6  $\mu\text{m}$ . High transmittance typifies these wavelengths with a maximum in blue-green portion of the spectrum. However, as the turbidity of the water changes (because of the presence of organic or inorganic materials), transmittance and



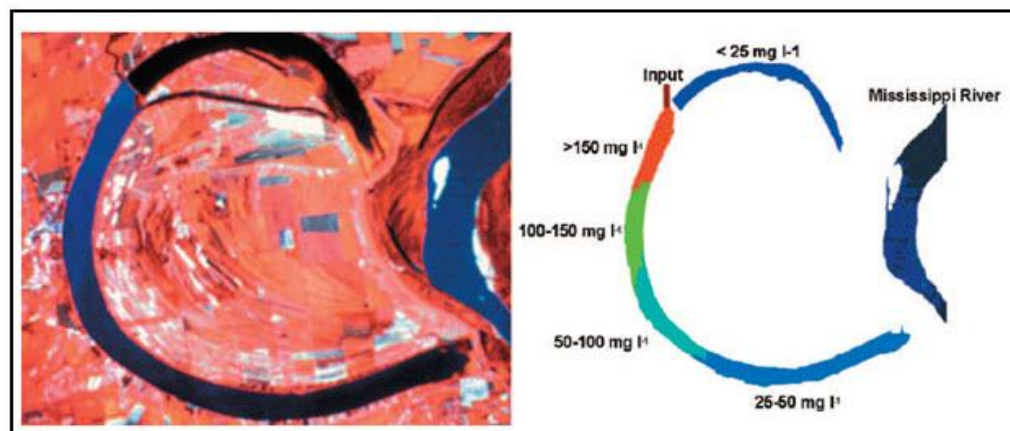
therefore reflectance changes dramatically. For example, waters containing large quantities of suspended sediments resulting from soil erosion normally have much higher visible reflectance than other clear waters of the same geographical area. Similarly, reflectance of water changes with the chlorophyll concentration involved. Similarly, concentration of algae has been monitored using remote sensing data. The use of remote sensing and GIS in monitoring water quality parameter such as suspended matter, phytoplankton, turbidity, and dissolved organic matter has been long recognized. The spectral characteristics of the signal received from water are a function of hydrological, biological and chemical characteristics of water, and other interference. Suspended sediments increase the emergent radiance from surface waters in the visible and near infrared proportion of the electromagnetic spectrum. Materials that form films on the water surface, such as oil films, can also be detected through the use of satellite images.

As an example Landsat imageries are also very useful in water quality measurement. The first four bands (TM1, TM2, TM3 and TM4) and thermal band (TM7) are correlated with some water quality parameters, temporal variations of water bodies. It found that atmospheric correction was essential to water quality assessment using satellite remotely sensed imagery because it improved significantly the water reflectance. Band rationing and regression analysis can be utilized for water quality monitoring. The ratio of TM1/TM3 is the most effective in estimating chlorophyll  $\alpha$ . In other studies it has been found that TM Band 2, Band 3/Band 2, Band 4/Band 3, and a multivariate regression analysis using Bands 3 and 2 are best in turbidity mapping.

Newly developed hyperspectral imaging, which can simultaneously record in 100s of spectral bands, is a much more powerful tool. Hyperspectral imaging has greater potential because of its simultaneous collection of images covering many narrow, contiguous wavelength bands that allow various aspects of water quality to be measured and monitored. Each water quality parameter such as suspended matter, phytoplankton concentration, turbidity, and dissolved organic matter has their own reflectance value within the range of 400-850  $\mu\text{m}$ .

Spectral signatures of suspended sediments in the lake water have been correlated with the satellite image data. Different concentration of suspended sediments in the lake water is shown in the Landsat TM image data (Fig. 32.1.).





**Fig. 32.1. Landsat TM image of Lake Chicot, Arkansas (left) and a derived image (right) showing categories of suspended sediments mapped in Lake Chicot based on the radiance in the TM image.**

(Source: [www.asprs.org/a/publications/pers/2003journal/june/2003\\_jun\\_695-704.pdf](http://www.asprs.org/a/publications/pers/2003journal/june/2003_jun_695-704.pdf))

## 32.2 Flood Inundation Delineation and Damage Assessment

Several flood events turned out to be the most devastating catastrophes for India's population, economy and environment during the past decades. Identification of flood affected areas is an important input for taking up flood management schemes for alleviating the problems affected by floods. Mapping the extent of flood by using optical, radar, digital elevation model (DEM), and river gauge data, supported and verified by ground observations, has received increased attention because of the availability of these data sets and the effectiveness of using the data for emergency responses. Analysis of satellite remote sensing is a powerful tool for identification and quantification of water spread area. The growing availability of multi-temporal satellite data has increased opportunities for monitoring large rivers from space, (Manushree & Rao, 2010). A variety of passive and active sensors operating in the visible and microwave range image data can be used to estimate inundation area and delineate flood boundaries. It also appears to be possible to obtain estimates of river discharge from space, using ground measurements and satellite data to construct empirical curves that relate water surface area to discharge. Extrapolation of these curves to ungauged sites may be possible for the special case of braided rivers.

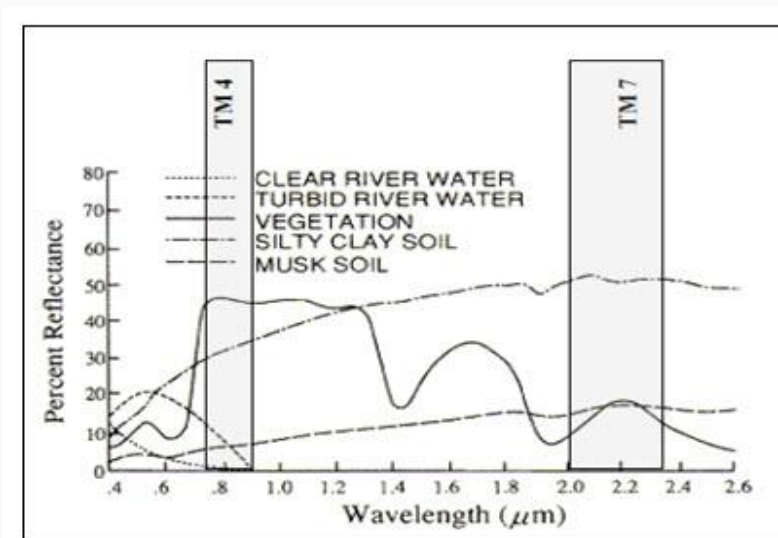
VIR sensors provide good delineation of inundated areas, where clouds, trees and floating vegetation do not obscure the water surface. Whereas the Synthetic Aperture Radar (SAR) sensors emitted microwave signal can penetrate clouds and can also detect standing water through emergent aquatic plants and forest canopies. Cloud penetration is particularly important for monitoring flood events, as they commonly occur during periods of extended rainfall. However, interpretation of synthetic aperture radar (SAR) imagery is less straightforward than for the visible/infrared range.

### 32.2.1 Flood inundation delineation using VIR data

The spectral reflectance curves of water, vegetation and soil are shown in Fig. 32.2. From the Fig 32.2, it should be noted that water has almost no reflectance in the infrared region, when

vegetation, soil reflection continues. The spectral signature of water and other features can be used in flood inundation delineation. In a Landsat image data, the TM 4 band is very useful in identifying land and water boundaries. But confusion arises between the reflectance of water with asphalt areas i.e. road pavement and rooftops of building as they reflect little back to sensor and appeared black on the TM 4 image. It was found that on the TM 7 band the reflectance of water, paved roof surfaces and rooftops are different and on TM 5, the differences are slightly smaller than those are in TM 7. Hence the addition of TM 4 and TM 7 (i.e. TM 4+TM 7) will be useful for determining water verses non water area. If the reflectance of a pixel is low in TM 4+TM 7 image the pixel is considered as water, otherwise it will be represented as non water features. However aerial photographs, DEM, water level measurements and high water marks after flood events are required for the aid of analysis and interpretation.

Generally for flood mapping two sets of remotely sensed data are required: one set consisting of data acquired before the flood event and the other acquired during the flood occurrence. The image before the flood usually used as the reference. One basic way of flood inundation delineation using passive multispectral data is by computing the threshold DN values, creating map and then creating a change detection map between pre flood and during flood scenarios. A methodology for flood mapping is shown in Fig 32.3. The output image generated after application of the methodology in the city of Greenville, Pit County located in the coastal flood plain of North Carolina, USA is shown through Fig. 32.4.



**Fig. 32.2. Spectral reflectance curve of water, vegetation and soil.**

(Source: [www.ualberta.ca/~mdzahidu/flood\\_mapping.pdf](http://www.ualberta.ca/~mdzahidu/flood_mapping.pdf))

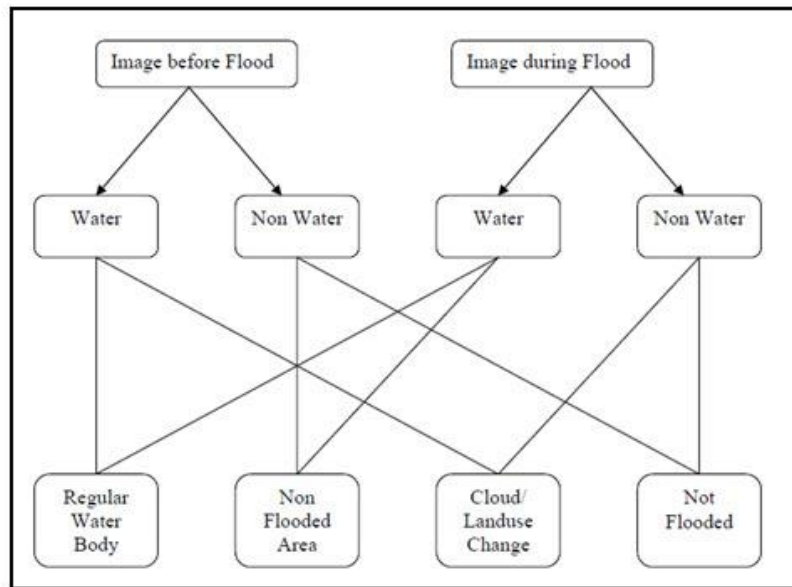


Fig. 32.3. Determination of flooded and non flooded area.

(Source: [www.ualberta.ca/~mdzahidu/flood\\_mapping.pdf](http://www.ualberta.ca/~mdzahidu/flood_mapping.pdf))

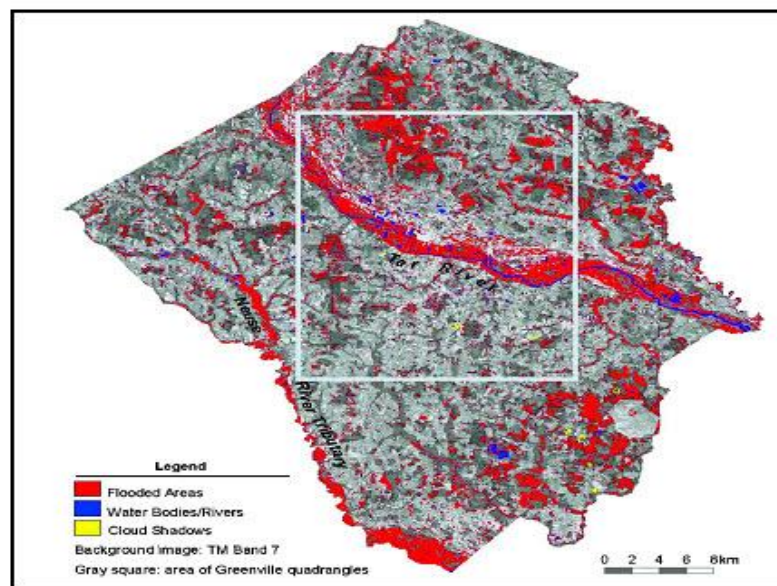


Fig. 32.4. Flood mapping output map generated through landsat data.

(Source: [www.ualberta.ca/~mdzahidu/flood\\_mapping.pdf](http://www.ualberta.ca/~mdzahidu/flood_mapping.pdf))

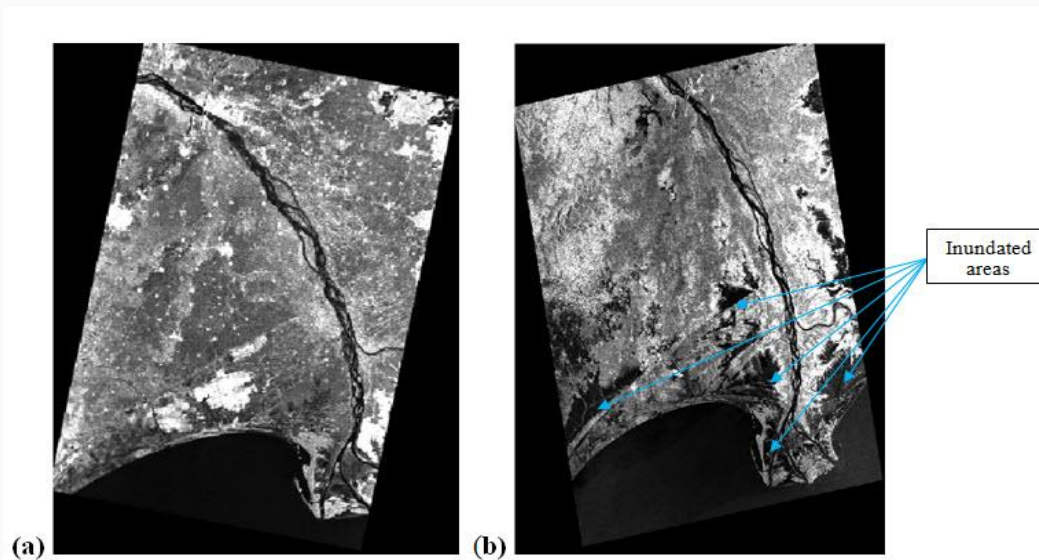
### 32.2.2 Flood inundation delineation using microwave data

Active remote sensing data such as Synthetic Aperture Radar (SAR) data is also useful for mapping floods because of their all weather functionality, their independency from sun as the illumination source and ability of penetration through forest canopy at a certain frequencies and polarization. In SAR system microwave radiation is produced which transmitted to the target object or area. The amount of microwave energy returned to the sensor is heavily dependent on the surface roughness and the dielectric constant of the

elements. The wet and rough ground surface yield strong backscatter than the dry surface. As open water usually exhibit strong specular reflection, so they produce low backscatter and appear dark.

Thresholding method is a very effective method to delineate flooded areas using SAR data. Commonly, a threshold value of radar back scatter is set in decibel (dB) and a binary algorithm can be used to determine whether a given raster cell is 'flooded' or not. A major problem is associated with the relation between radar wavelength and roughness of the terrain and water body. Normally calm water acts as a specular reflector to the radar signals, therefore appears in dark. But during floods, due to windy condition wind induced ripples appear in the water surface makes the water rough which appears brighter tone in the SAR image.

A case study carried out by *Manushree and Rao, 2010* in a coastal floodplain of Krishna river in Andhra Pradesh using ENVISAT data. The methodology of the study was as follows- first the SAR images (before and after cyclone) were processed by methods which include calibration, multi looking, speckle filtering, slant range to ground range conversion, Geo-coding. A pixel-based classification was used to classify both scenes (i.e., before and after cyclone) into different categories. A Decision tree model was prepared based on backscattering coefficients of different classes, which was used to identify the area inundated by the flood. The pre and post-flood SAR images are shown through Fig. 32.5. (a) and (b).



**Fig. 32.5. (a) Geo coded image of 4th April, 2006 (pre flood), (b) Geo coded image of 4th Nov, 2006 (post flood). (Source: Manushree and Rao, 2010)**

Areas inundated due to flood are shown in Fig. 32.5. (b). areal extent of flood and corresponding damage can be assessed. The recurrence of flood can be avoided through proper planning using these images.



**This Book Download From e-course of ICAR**  
**Visit for Other Agriculture books, News,**  
**Recruitment, Information, and Events at**  
**[WWW.AGRIMOON.COM](http://WWW.AGRIMOON.COM)**

Give FeedBack & Suggestion at [info@agrimoon.com](mailto:info@agrimoon.com)

**Send a Massage for daily Update of Agriculture on WhatsApp**  
**+91-7900 900 676**

**DISCLAIMER:**

The information on this website does not warrant or assume any legal liability or responsibility for the accuracy, completeness or usefulness of the courseware contents.

The contents are provided free for noncommercial purpose such as teaching, training, research, extension and self learning.



*Connect With Us:*

