Training Manual
on

Remote Sensing &
Geographical Information System (GIS)

For

REWARD
(Rejuvenating Watershed for Agricultural Resilience through Innovative Development)
Project

TABLE OF CONTENTS

1. Introduction

In order to meet the requirement of REWARD (Rejuvenating Watershed for Agricultural Resilience through Innovative Development) project for preparing a user manual for Geographic Information System (GIS), a user friendly Training Manual on GIS has been prepared to cover important aspects. GIS is a tool to efficiently capture, store, update, manipulate, analyze and display all forms of geographically referenced data. GIS is being widely used as spatial analysis tool for effective and efficient means of data generation and management, analysis and display. The true value of GIS can only be realized if the proper tools to collect spatial data and integrate them with attribute data are available. Data used in GIS often come from many types, and are stored in different ways. GIS provides tools and a methods for the integration of different data into a format to be compared and analyzed. Data sources are mainly obtained from remote-sensing satellite imageries, GPS surveys, manual digitization, scanning of aerial photographs, paper maps, and existing digital data sets are promising sources for data input in GIS. Before generation of any spatial data one has to understand the available data types, data formats, data analysis procedures and their capabilities in GIS to get realistic outputs.

Remote sensing (RS), also called earth observation, refers to obtaining information about objects or areas at the Earth's surface without being in direct contact with the object or area. Remote sensing techniques allow taking images of the earth surface in various wavelength region of the electromagnetic spectrum (EMS). One of the major characteristics of a remotely sensed image is the wavelength region it represents in the EMS. Some of the images represent reflected solar radiation in the visible and the near infrared regions of the electromagnetic spectrum, others are the measurements of the energy emitted by the earth surface itself i.e. in the thermal infrared wavelength region.

The manual mainly covers the topics on principles and concepts in GIS, digital image analysis, image registration, digital terrain analysis, assimilation of input datasets in digital terrain analysis, processing of digital elevation datasets in GIS, organization of geospatial datasets in GIS, spatial data integration techniques in GIS and visualization techniques. We would like to state that the training manual does not aim to promote any software, but made an attempt to provide the details on GIS and digital image processing tools to help the project partners and answer their common questions. Keeping the needs of the participants in GIS and digital image processing, every effort has been made to cover all the important aspects, however, if any topics is not cover or partially covered in the training manual, reader's are advised to referred the alternative reference material for detailed information.

2. Fundamentals of Remote Sensing

1.1 What is Remote Sensing?

So, what exactly is **remote sensing**? For the purposes of this tutorial, we will use the following definition:

 "Remote sensing is the science (and to some extent, art) of acquiring information about the Earth's surface without actually being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information."

 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. Note, however that remote sensing also involves the sensing of emitted energy and the use of non-imaging sensors.
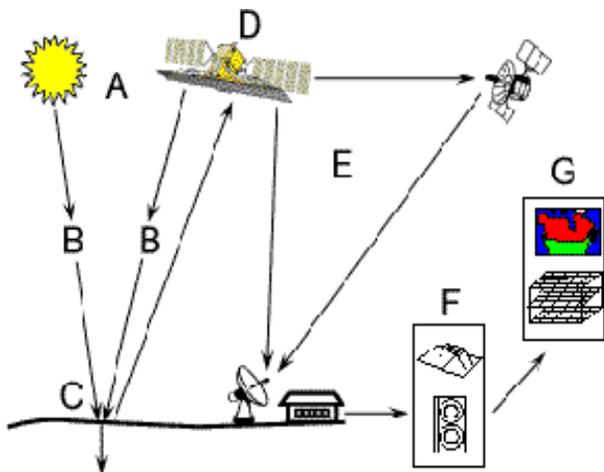
© CCRS / CCT

1. Energy Source or Illumination (A) - the first requirement for remote sensing is to have an energy source which illuminates or provides electromagnetic energy to the target of interest.

2. Radiation and the Atmosphere (B) - as the energy 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.

3. **Interaction with the Target (C)** - 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.

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

5. Transmission, Reception, and Processing (E) - 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).

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

7. Application (G) - the final element of the remote sensing process is achieved when we apply the information we have been able to extract from the imagery about the target in order to better understand it, reveal some new information, or assist in solving a particular problem.

These seven elements comprise the remote sensing process from beginning to end.

1.2 Electromagnetic Radiation
As was noted in the previous section, the first requirement for remote sensing is to have an **energy source to illuminate the target** (unless the sensed energy is being emitted by the target). This energy is in the form of electromagnetic radiation.
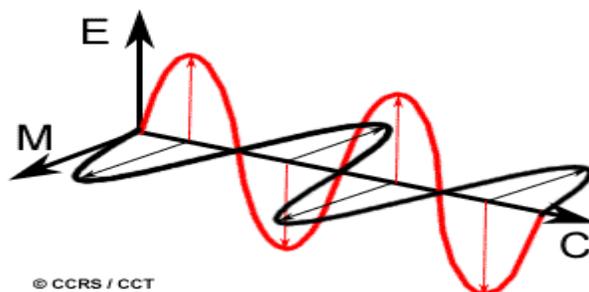
© CCRS / CCT

All electromagnetic radiation has fundamental properties and behaves in predictable ways according to the basics of wave theory. **Electromagnetic radiation** consists of an electrical field(E) which varies in magnitude in a direction perpendicular to the direction in which the radiation is traveling, and a magnetic field (M) oriented at right angles to the electrical field. Both these fields travel at the speed of light (c). Two characteristics of electromagnetic radiation are particularly important for understanding remote sensing. These are the **wavelength and frequency.**

© CCRS / CCT
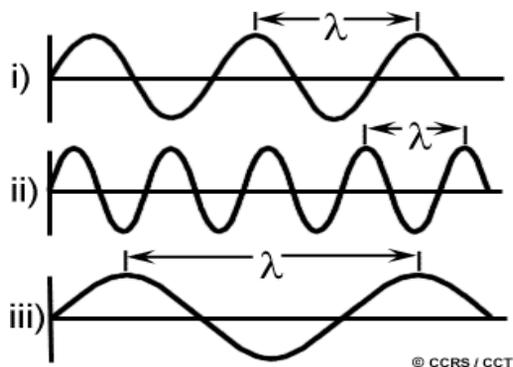
The wavelength is the length of one wave cycle, which can be measured as the distance between successive wave crests. Wavelength is usually represented by the Greek letter lambda (l). Wavelength is measured in meters (m) or some factor of meters such as **nanometers** (nm, 10-9 meters), **micrometers** (mm, 10-6 meters) (mm, 10-6 meters) or centimeters (cm, 10-2 meters). Frequency refers to the number of cycles of a wave passing a fixed point per unit of time. Frequency is normally measured in **hertz** (Hz), equivalent to one cycle per second, and various multiples of hertz.

Wavelength and frequency are related by the following formula:

$$c = \lambda \nu$$

where:

λ = wavelength (m)
ν = frequency (cycles per second, Hz)
c = speed of light (3×10^8 m/s)

Therefore, the two are inversely related to each other. The shorter the wavelength, the higher the frequency. The longer the wavelength, the lower the frequency. Understanding the characteristics of electromagnetic radiation in terms of their wavelength and frequency is crucial to understanding the information to be extracted from remote sensing data. Next we will be examining the way in which we categorize electromagnetic radiation for just that purpose.

1.3 The Electromagnetic Spectrum

The **electromagnetic spectrum** ranges from the shorter wavelengths (including gamma and x-rays) to the longer wavelengths (including microwaves and broadcast radio waves). There are several regions of the electromagnetic spectrum which are useful for remote sensing.

© CCRS / CCT
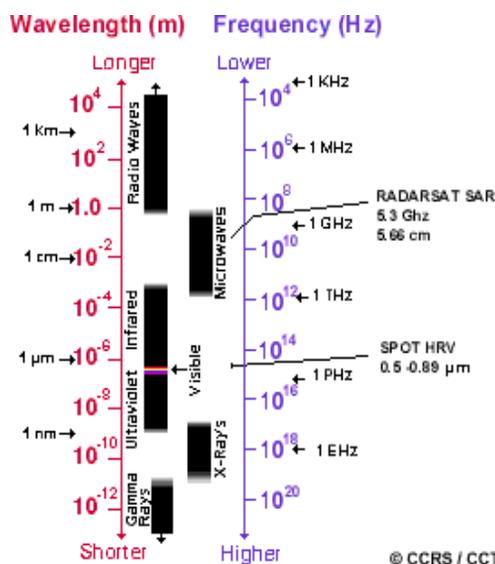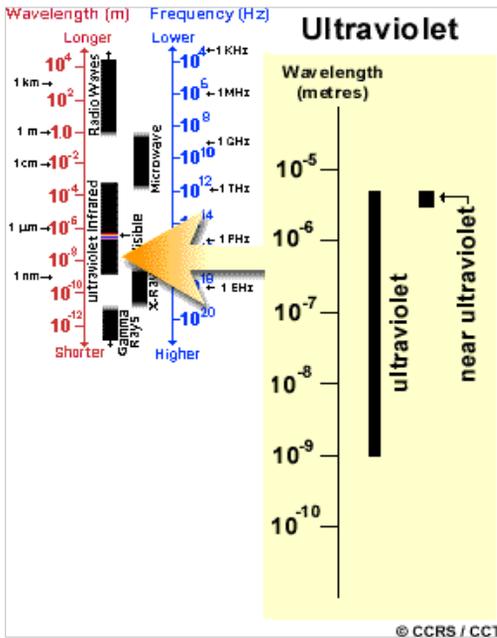
For most purposes, the **ultraviolet or UV** portion of the spectrum has the shortest wavelengths which are practical for remote sensing. This radiation is just beyond the violet portion of the visible wavelengths, hence its name. Some Earth surface materials, primarily rocks and minerals, fluoresce or emit visible light when illuminated by UV radiation.

The light which our eyes - our "remote sensors" - can detect is part of the **visible spectrum**. It is important to recognize how small the visible portion is relative to the rest of the spectrum. There is a lot of radiation around us which is "invisible" to our eyes, but can be detected by other remote sensing instruments and used to our advantage. The visible wavelengths cover a range from approximately 0.4 to 0.7 mm. The longest visible wavelength is red and the shortest is violet. Common wavelengths of what we perceive as particular colors from the visible portion of the spectrum are listed below. It is important to note that this is the only portion of the spectrum we can associate with the concept of **colors**.

- **Violet:** 0.4 - 0.446 mm
- **Blue:** 0.446 - 0.500 mm
- **Green:** 0.500 - 0.578 mm
- **Yellow:** 0.578 - 0.592 mm
- **Orange:** 0.592 - 0.620 mm
- **Red:** 0.620 - 0.7 mm
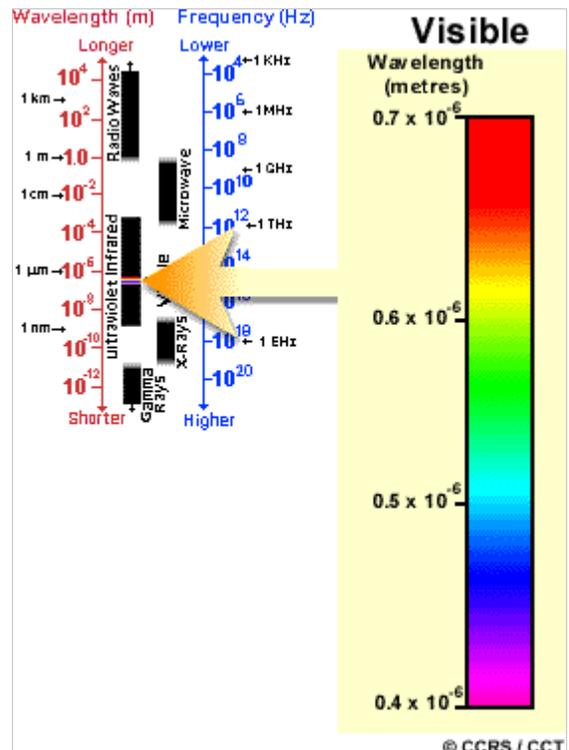
Blue, **green**, and **red** are the **primary colours** or wavelengths of the visible spectrum. They are defined as such because no single primary colour can be created from the other two, but all other colours can be formed by combining blue, green, and red in various proportions. Although we see sunlight as a uniform or homogeneous colour, it is actually composed of various wavelengths of radiation in primarily the ultraviolet, visible and infrared portions of the spectrum. The visible portion of this radiation can be shown in its component colours when sunlight is passed through a **prism**, which bends the light in differing amounts according to wavelength.
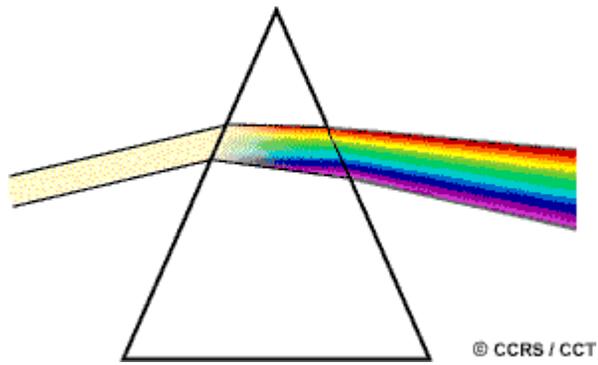
© CCRS / CCT
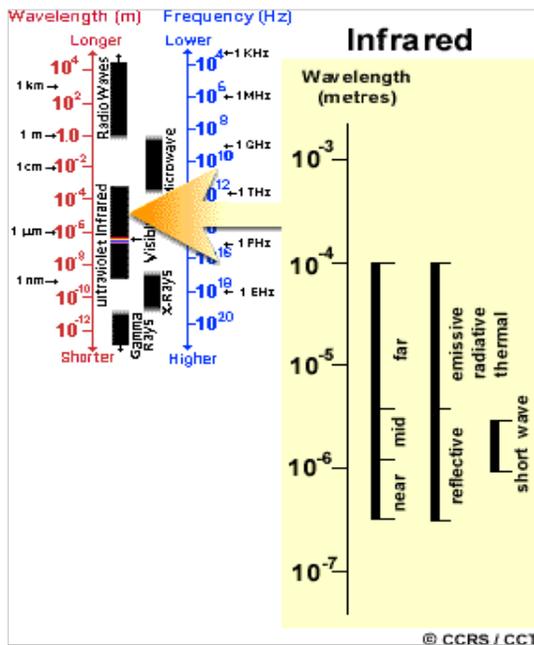
© CCRS / CCT

The next portion of the spectrum of interest is the infrared (IR) region which covers the wavelength range from approximately 0.7 mm to 100 mm - more than 100 times as wide as the visible portion! The infrared region can be divided into two categories based on their radiation properties - the **reflected IR**, and the emitted or **thermal IR**. Radiation in the reflected IR region is used for remote sensing purposes in ways very similar to radiation in the visible portion. The reflected IR covers wavelengths from approximately 0.7 mm to 3.0 mm. The thermal IR region is quite different than the visible and reflected IR portions, as this energy is essentially the radiation that is emitted from the Earth's surface in the form of heat. The thermal IR covers wavelengths from approximately 3.0mm to 100 mm.

The portion of the spectrum of more recent interest to remote sensing is the **microwave region** from about 1 mm to 1 m. This covers the longest wavelengths used for remote sensing. The shorter wavelengths have properties similar to the thermal infrared region while the longer wavelengths approach the wavelengths used for radio broadcasts.
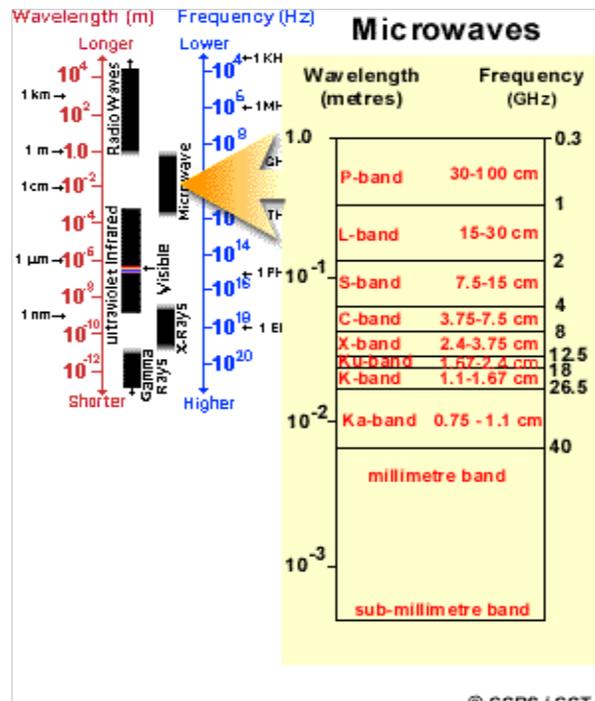
© CCRS / CCT

1.4 Interactions with the Atmosphere

Before radiation used for remote sensing reaches the Earth's surface it has to travel through some distance of the Earth's atmosphere. Particles and gases in the atmosphere can affect the incoming light and radiation. These effects are caused by the mechanisms of **scattering** and **absorption**.

Scattering occurs when particles or large gas molecules present in the atmosphere interact with and cause the electromagnetic radiation to be redirected from its original path. How much scattering takes place depends on several factors including the wavelength of the radiation, the abundance of particles or gases, and the distance the radiation travels through the atmosphere. There are three (3) types of scattering which take place.

© CCRS / CCT
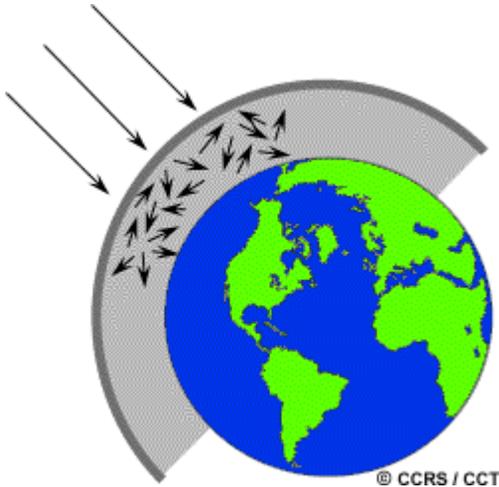
Rayleigh scattering occurs when particles are very small compared to the wavelength of the radiation. These could be particles such as small specks of dust or nitrogen and oxygen molecules. Rayleigh scattering causes shorter wavelengths of energy to be scattered much more than longer wavelengths. Rayleigh scattering is the dominant scattering mechanism in the upper atmosphere. The fact that the sky appears "blue" during the day is because of this phenomenon. As sunlight passes through the atmosphere, the shorter wavelengths (i.e. blue) of the visible spectrum are scattered more than the other (longer) visible wavelengths. At **sunrise and sunset** the light has to travel farther through the atmosphere than at midday and the scattering of the shorter wavelengths is more complete; this leaves a greater proportion of the longer wavelengths to penetrate the atmosphere.

© CCRS / CC
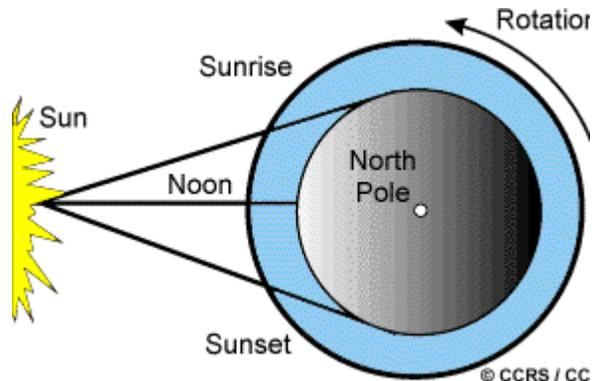
Mie scattering occurs when the particles are just about the same size as the wavelength of the radiation. Dust, pollen, smoke and water vapour are common causes of Mie scattering which tends to affect longer wavelengths than those affected by Rayleigh scattering. Mie scattering occurs mostly in the lower portions of the atmosphere where larger particles are more abundant, and dominates when cloud conditions are overcast.

© CCRS / CCT

The final scattering mechanism of importance is called **nonselective scattering**. This occurs when the particles are much larger than the wavelength of the radiation. Water droplets and large dust particles can cause this type of scattering.
Nonselective scattering gets its name from the fact that all wavelengths are scattered about equally.
This type of scattering causes fog and clouds to appear white to our eyes because blue, green, and red light are all scattered in approximately equal
quantities (blue+green+red light = white light).

is the other main mechanism at work when electromagnetic radiation interacts with the atmosphere. In contrast to scattering, this phenomenon causes molecules in the atmosphere to absorb energy at various wavelengths. Ozone, carbon dioxide, and water vapour are the three main atmospheric constituents which 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.

© CCRS / CCT
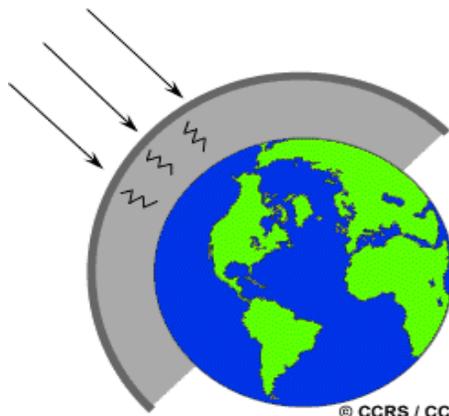
You may have heard **carbon dioxide** referred to as a greenhouse gas. This is because it tends to absorb radiation strongly in the far infrared portion of the spectrum - that area associated with thermal heating - which serves to trap this heat inside the atmosphere. Water vapour in the atmosphere absorbs much of the incoming longwave infrared and shortwave microwave radiation (between 22mm and 1m). The presence of water vapour in the lower atmosphere varies greatly from location to location and at different times of the year. For example, the air mass above a desert would have very little water vapour to absorb energy, while the tropics would have high concentrations of water vapour (i.e. high humidity).
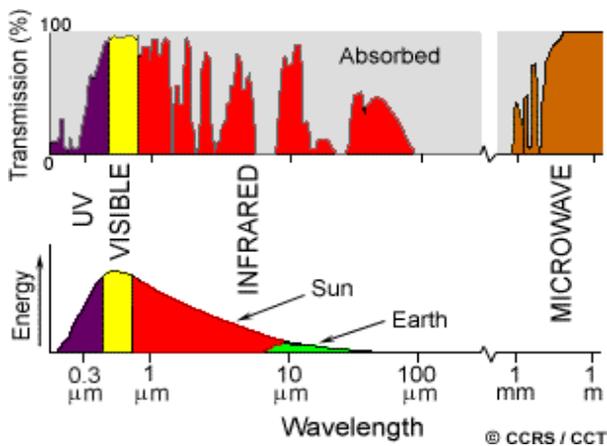
Because these gases absorb electromagnetic energy in very specific regions of the spectrum, they influence where (in the spectrum) we can "look" for remote sensing purposes. Those areas of the spectrum which are not severely influenced by atmospheric absorption and thus, are useful to remote sensors, are called **atmospheric windows**. By comparing the characteristics of the two most common energy/radiation sources (the sun and the earth) with the atmospheric windows available to us, we can define those wavelengths that we can use **most effectively** for remote sensing. The visible portion of the spectrum, to which our eyes are most sensitive, corresponds to both an atmospheric window and the peak energy level of the sun. Note also that heat energy emitted by the Earth corresponds to a window around 10 mm in the thermal IR portion of the spectrum, while the large window at wavelengths beyond 1 mm is associated with the microwave region.

Now that we understand how electromagnetic energy makes its journey from its source to the surface (and it is a difficult journey, as you can see) we will next examine what happens to that radiation when it does arrive at the Earth's surface.

1.5 Radiation - Target Interactions

Radiation that is not absorbed or scattered in the atmosphere can reach and interact with the Earth's surface. There are three (3) forms of interaction that can take place when energy strikes, or is **incident (I)** upon the surface. These are: **absorption (A)**; **transmission (T)**; and **reflection (R)**. The total incident energy will interact with the surface in one or more of these three ways. The proportions of each will depend on the wavelength of the energy and the material and condition of the feature.

Absorption (A) occurs when radiation (energy) is absorbed into the target while transmission (T) occurs when radiation passes through a target. Reflection (R) occurs when radiation "bounces" off the target and is redirected. In remote sensing, we are most interested in measuring the radiation reflected from targets. We refer to two types of reflection, which represent the two extreme ends of the way in which energy is reflected from a target: **specular reflection** and **diffuse reflection**.
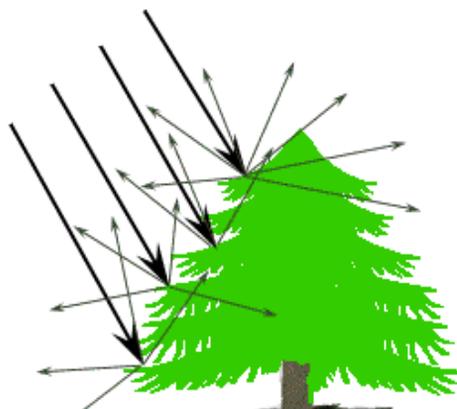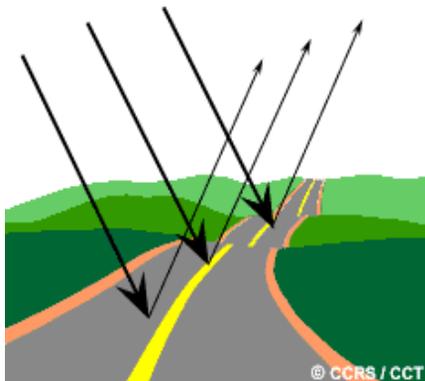
When a surface is smooth we get **specular** or mirror-like reflection where all (or almost all) of the energy is directed away from the surface in a single direction. **Diffuse** reflection occurs when the surface is rough and the energy is reflected almost uniformly in all directions. Most earth surface features lie somewhere between perfectly specular or perfectly diffuse reflectors. 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.

Let's take a look at a couple of examples of targets at the Earth's surface and how energy at the visible and infrared wavelengths interacts with them.
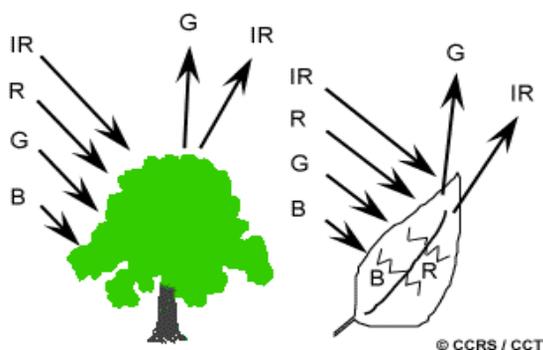
Leaves: A chemical compound in leaves called chlorophyll strongly absorbs radiation in the red and blue wavelengths but reflects green wavelengths. Leaves appear "greenest" to us in the summer, when chlorophyll content is at its maximum. In autumn, there is less chlorophyll in the leaves, so there is less absorption and proportionately more reflection of the red wavelengths, making the leaves appear red or yellow (yellow is a combination of red and green wavelengths). The internal structure of healthy leaves act as excellent diffuse reflectors of near-infrared wavelengths. If our eyes were sensitive to near-infrared, trees would appear extremely bright to us at these wavelengths. In fact, measuring and monitoring the near-IR reflectance is one way that scientists can determine how healthy (or unhealthy) vegetation may be.

Water: Longer wavelength visible and near infrared radiation is absorbed more by water than shorter visible wavelengths. Thus water typically looks blue or blue-green due to stronger reflectance at these shorter wavelengths, and darker if viewed at red or near infrared wavelengths. If there is suspended sediment present in the upper layers of the water body, then this will allow better reflectivity and a brighter appearance of the water. The apparent colour of the water will show a slight shift to longer wavelengths. Suspended sediment (S) can be easily confused with shallow (but clear) water, since these two phenomena appear very similar. Chlorophyll in algae absorbs more of the blue wavelengths and reflects the green, making the water appear more green in colour when algae is present. The topography of the water surface (rough, smooth, floating materials, etc.) can also lead to complications for water-related interpretation due to potential problems of specular reflection and other influences on colour and brightness.
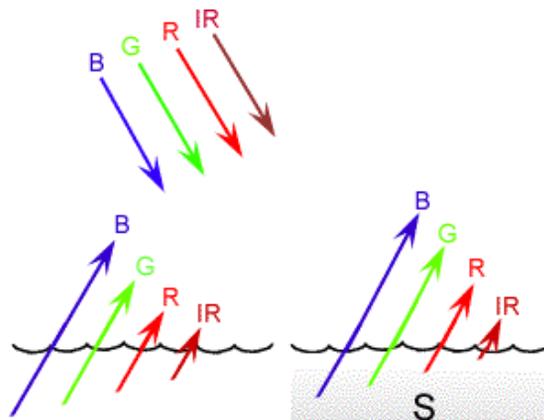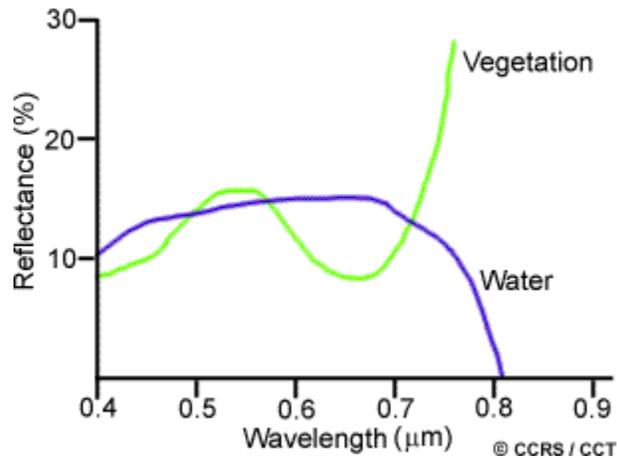
We can see from these examples that, depending on the complex make-up of the target that is being looked at, and the wavelengths of radiation involved, we can observe very different responses to the mechanisms of absorption, transmission, and reflection. By measuring the energy that is reflected (or emitted) by targets on the Earth's surface over a variety of different wavelengths, we can build up a **spectral response** for that object. By comparing the response patterns of different features we may be able to distinguish between them, where we might not be able to, if we only compared them at one wavelength. For example, water and vegetation may reflect somewhat similarly in the visible wavelengths but are almost always separable in the infrared. Spectral response can be quite variable, even for the same target type, and can also vary with time (e.g. "green-ness" of leaves) and location. Knowing where to "look" spectrally and understanding the factors which influence the spectral response of the features of interest are critical to correctly interpreting the interaction of electromagnetic radiation with the surface.

1.6 Passive vs. Active Sensing

So far, throughout this chapter, we have made various references to the sun as a source of energy or radiation. The sun provides a very convenient source of energy for remote sensing. The sun's energy is either **reflected**, as it is for visible wavelengths, or absorbed and then **re-emitted**, as it is for thermal infrared wavelengths. Remote sensing systems which measure energy that is naturally available are called **passive sensors**. Passive sensors can only be used to detect energy when the naturally occurring energy is available. For all reflected energy, this can only take place during the time when the sun is illuminating the Earth. There is no reflected energy available from the sun at night. Energy that is naturally emitted (such as thermal infrared) can be detected day or night, as long as the amount of energy is large enough to be recorded.
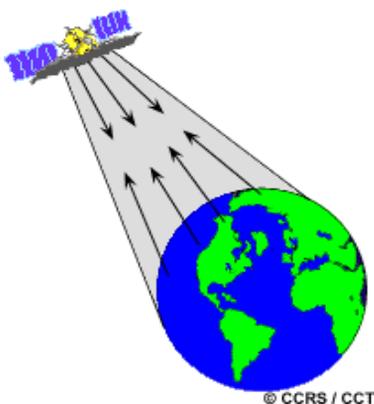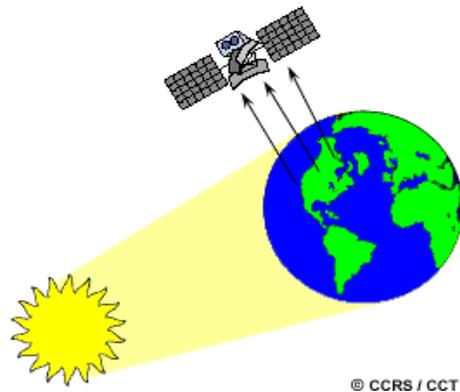
Active sensors, on the other hand, provide their own energy source for illumination. The sensor emits radiation which is directed toward the target to be investigated. The radiation reflected from that target is detected and measured by the sensor. Advantages for active sensors include the ability to obtain measurements anytime, regardless of the time of day or season. Active sensors can be used for examining wavelengths that are not sufficiently provided by the sun, such as microwaves, or to better control the way a target is illuminated. However, active systems require the generation of a fairly large amount of energy to adequately illuminate targets. Some examples of active sensors are a laser fluorosensor and a synthetic

1.7 Characteristics of Images

Before we go on to the next chapter, which looks in more detail at sensors and their characteristics, we need to define and understand a few fundamental terms and concepts associated with remote sensing images.

Electromagnetic energy may be detected either photographically or electronically. The photographic process uses chemical reactions on the surface of light-sensitive film to detect and record energy variations. It is important to distinguish between the terms **images** and **photographs** in remote sensing. An image refers to any pictorial representation, regardless of what wavelengths or remote sensing device has been used to detect and record the electromagnetic energy. A **photograph** refers specifically to images that have been detected as well as recorded on photographic film. The black and white photo to the left, of part of the city of Ottawa, Canada was taken in the visible part of the spectrum. Photos are normally recorded over the wavelength range from 0.3 mm to 0.9 mm - the visible and reflected infrared. Based on these definitions, we can say that all photographs are images, but not all images are photographs. Therefore, unless we are talking specifically about an image recorded photographically, we use the term image.
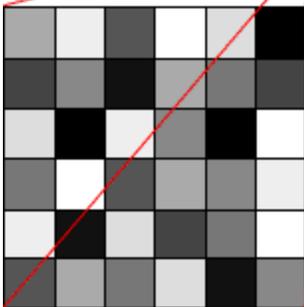
A photograph could also be represented and displayed in a **digital** format by subdividing the image into small equal-sized and shaped areas, called picture elements or **pixels**, and representing the brightness of each area with a numeric value or **digital number**. Indeed, that is exactly what has been done to the photo to the left. In fact, using the definitions we have just discussed, this is actually a **digital image** of the original photograph! The photograph was scanned and subdivided into pixels with each pixel assigned a digital number representing its relative brightness. The computer displays each digital value as different brightness levels. Sensors that record electromagnetic energy, electronically record the energy as an array of numbers in digital format right from the start. These two different ways of representing and displaying remote sensing data, either pictorially or digitally, are interchangeable as they convey the same information (although some detail may be lost when converting back and forth).

In previous sections we described the visible portion of the spectrum and the concept of colours. We see colour because our eyes detect the entire visible range of wavelengths and our brains process the information into separate colours. Can you imagine what the world would look like if we could only see very narrow ranges of wavelengths or colours? That is how many sensors work. The information from a narrow wavelength range is gathered and stored in a **channel**, also sometimes referred to as a **band**. We can combine and display channels of information digitally using the three primary colours (blue, green, and red). The data from each channel is represented as one of the primary colours and, depending on the relative brightness (i.e. the digital value) of each pixel in each channel, the primary colours combine in different proportions to represent different colours.

3. Image Analysis

3.1 Introduction

In order to take advantage of and make good use of remote sensing data, we must be able to extract meaningful information from the imagery. This brings us to the topic of discussion in this chapter - **interpretation and analysis** - the sixth element of the remote sensing process which we defined in Chapter 1.

Interpretation and analysis of remote sensing imagery involves the identification and/or measurement of various targets in an image in order to extract useful information about them. Targets in remote sensing images may be any feature or object which can be observed in an image, and have the following characteristics:

- Targets may be a point, line, or area feature. This means that they can have any form, from a bus in a parking lot or plane on a runway, to a bridge or roadway, to a large expanse of water or a field.
- The target must be distinguishable; it must contrast with other features around it in the image.
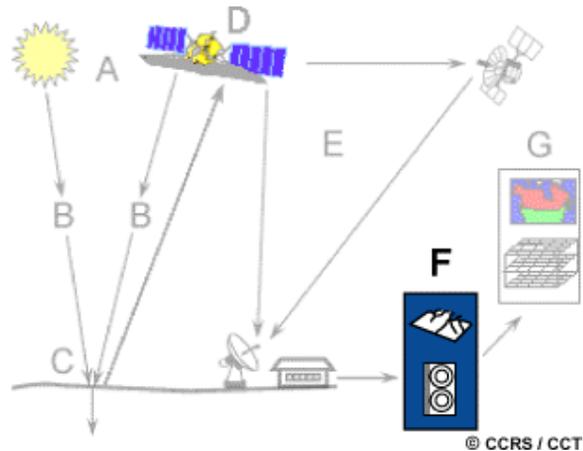
Much interpretation and identification of targets in remote sensing imagery is performed manually or visually, i.e. by a human interpreter. In many cases this is done using imagery displayed in a pictorial or photograph-type format, independent of what type of sensor was used to collect the data and how the data were collected. In this case we refer to the data as being in **analog** format. As we discussed in Chapter 1, remote sensing images can also be represented in a computer as arrays of pixels, with each pixel corresponding to a digital number, representing the brightness level of that pixel in the image. In this case, the data are in a **digital** format. Visual interpretation may also be performed by examining digital imagery displayed on a computer screen. Both analogue and digital imagery can be displayed as black and white (also called monochrome) images, or as colour images (refer back to Chapter 1, Section 1.7) by combining different channels or bands representing different wavelengths.

When remote sensing data are available in digital format, **digital processing and analysis** may be performed using a computer. Digital processing may be used to enhance data as a prelude to visual interpretation. Digital processing and analysis may also be carried out to automatically identify targets and extract information completely without manual intervention by a human interpreter. However, rarely is digital processing and analysis carried out as a complete replacement for manual interpretation. Often, it is done to supplement and assist the human analyst.

Manual interpretation and analysis dates back to the early beginnings of remote sensing for air photo interpretation. Digital processing and analysis is more recent with the advent of digital recording of remote sensing data and the development of computers. Both manual and digital techniques for interpretation of remote sensing data have their respective advantages and disadvantages. Generally, manual interpretation requires little, if any, specialized equipment, while digital analysis requires specialized, and often expensive, equipment.

Manual interpretation is often limited to analyzing only a single channel of data or a single image at a time due to the difficulty in performing visual interpretation with multiple images. The computer environment is more amenable to handling complex images of several or many channels or from several dates. In this sense, digital analysis is useful for simultaneous analysis of many spectral bands and can process large data sets much faster than a human interpreter. Manual interpretation is a subjective process, meaning that the results will vary with different interpreters. Digital analysis is based on the manipulation of digital numbers in a computer and is thus more objective, generally resulting in more consistent results. However, determining the validity and accuracy of the results from digital processing can be difficult.

It is important to reiterate that visual and digital analyses of remote sensing imagery are not mutually exclusive. Both methods have their merits. In most cases, a mix of both methods is usually employed when analyzing imagery. In fact, the ultimate decision of the utility and relevance of the information extracted at the end of the analysis process, still must be made by humans.

3.2 Elements of Visual Interpretation

As we noted in the previous section, analysis of remote sensing imagery involves the identification of various targets in an image, and those targets may be environmental or artificial features which consist of points, lines, or areas. Targets may be defined in terms of the way they reflect or emit radiation. This radiation is measured and recorded by a sensor, and ultimately is depicted as an image product such as an air photo or a satellite image.

What makes interpretation of imagery more difficult than the everyday visual interpretation of our surroundings? For one, we lose our sense of depth when viewing a two-dimensional image, unless we can view it **stereoscopically** so as to simulate the third dimension of height. Indeed, interpretation benefits greatly in many applications when images are viewed in stereo, as visualization (and therefore, recognition) of targets is enhanced dramatically.

Viewing objects from directly above also provides a very different perspective than what we are familiar with. Combining an unfamiliar perspective with a very different scale and lack of recognizable detail can make even the most familiar object unrecognizable in an image.

Finally, we are used to seeing only the visible wavelengths, and the imaging of wavelengths outside of this window is more difficult for us to comprehend.

Recognizing targets is the key to interpretation and information extraction. Observing the differences between targets and their backgrounds involves comparing different targets based on any, or all, of the visual elements of **tone, shape, size, pattern, texture, shadow, and association**. Visual interpretation using these elements is often a part of our daily lives, whether we are conscious of it or not. Examining satellite images on the weather report, or following high speed chases by views from a helicopter are all familiar examples of visual image interpretation. Identifying targets in remotely sensed images based on these visual elements allows us to further interpret and analyze. The nature of each of these interpretation elements is described below, along with an image example of each

Tone refers to the relative brightness or colour of objects in an image. Generally, tone is the fundamental element for distinguishing between different targets or features. Variations in tone also allows the elements of shape, texture, and pattern of objects to be distinguished.

Shape refers to the general form, structure, or outline of individual objects. Shape can be a very distinctive clue for interpretation. Straight edge shapes typically represent urban or agricultural (field) targets, while natural features, such as forest edges, are generally more irregular in shape, except where man has created a road or clear cuts. Farm or crop land irrigated by rotating sprinkler systems would appear as circular shapes.

Size of objects in an image is a function of scale. It is important to assess the size of a target relative to other objects in a scene, as well as the absolute size, to aid in the interpretation of that target. A quick approximation of target size can direct interpretation to an appropriate result more quickly. 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.

Pattern refers to the spatial arrangement of visibly discernible objects. Typically an orderly repetition of similar tones and textures will produce a distinctive and ultimately recognizable pattern. Orchards with evenly spaced trees, and urban streets with regularly spaced houses are good examples of pattern.

Texture refers to the arrangement and frequency of tonal variation in particular areas of an image. Rough textures would consist of a mottled tone where the grey levels change abruptly in a small area, whereas smooth textures would have very little tonal variation. Smooth textures are most often the result of uniform, even surfaces, such as fields, asphalt, or grasslands. A target with a rough surface and irregular structure, such as a forest canopy, results in a rough textured appearance. Texture is one of the most important elements for distinguishing features in radar imagery.

Shadow is also helpful in interpretation as it may provide an idea of the profile and relative height of a target or targets which may make identification easier. However, shadows can also reduce or eliminate interpretation in their area of influence, since targets within shadows are much less (or not at all) discernible from their surroundings. Shadow is also useful for enhancing or identifying topography and landforms, particularly in radar imagery.

Association takes into account the relationship between other recognizable objects or features in proximity to the target of interest. The identification of features that one would expect to associate with other features may provide information to facilitate identification. In the example given above, commercial properties may be associated with proximity to major transportation routes, whereas residential areas would be associated with schools, playgrounds, and sports fields. In our
example, a lake is associated with boats, a marina, and adjacent recreational land.

3.3 Digital Image Processing

In today's world of advanced technology where most remote sensing data are recorded in digital format, virtually all image interpretation and analysis involves some element of digital processing. Digital image processing may involve numerous procedures including formatting and correcting of the data, digital enhancement to facilitate better visual interpretation, or even automated classification of targets and features entirely by computer. In order to process remote sensing imagery digitally, the data must be recorded and available in a digital form suitable for storage on a computer tape or disk. Obviously, the other requirement for digital image processing is a computer system, sometimes referred to as an **image analysis system**, with the appropriate hardware and software to process the data. Several commercially available software systems have been developed specifically for remote sensing image processing and analysis.

© CCRS / CCT
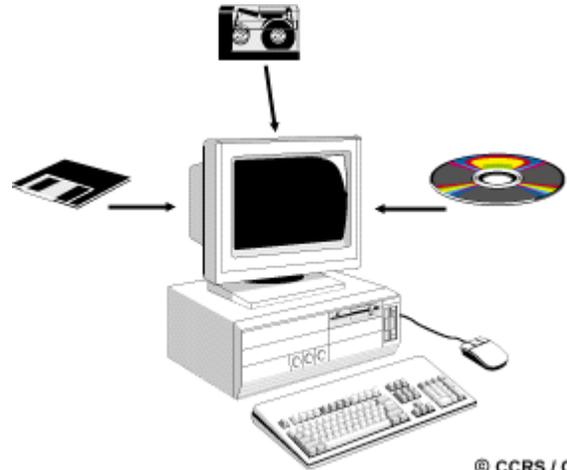
For discussion purposes, most of the common image processing functions available in image analysis systems can be categorized into the following four categories:

- Preprocessing
- Image Enhancement
- Image Transformation
- Image Classification and Analysis

Preprocessing functions involve those operations that are normally required prior to the main data analysis and extraction of information, and are generally grouped **as radiometric or geometric corrections**. Radiometric corrections include correcting the data for sensor irregularities and unwanted sensor or atmospheric noise, and converting the data so they accurately represent the reflected or emitted radiation measured by the sensor. Geometric corrections include correcting for geometric distortions due to sensor-Earth geometry variations, and conversion of the data to real world coordinates (e.g. latitude and longitude) on the Earth's surface.
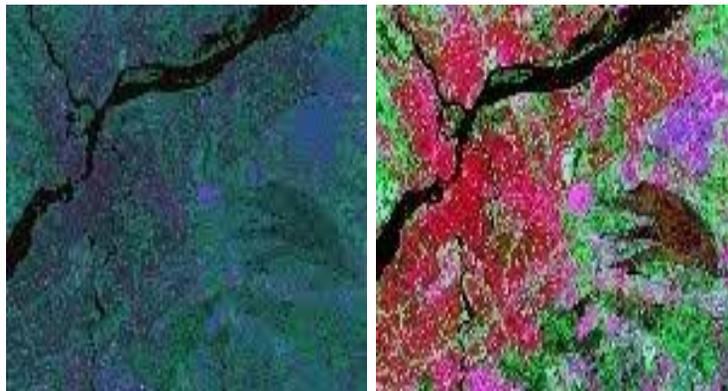
The objective of the second group of image processing functions grouped under the term of **image enhancement**, is solely to **improve the appearance of the imagery** to assist in visual interpretation and analysis. Examples of enhancement functions include contrast stretching to increase the tonal distinction between various features in a scene, and **spatial filtering** to enhance (or suppress) specific spatial patterns in an image.

Image transformations are operations similar in concept to those for image enhancement. However, unlike image enhancement operations which are normally applied only to a single channel of data at a time, image transformations usually involve combined processing of data from multiple spectral bands. Arithmetic operations (i.e. subtraction, addition, multiplication, division) are performed to combine and transform the original bands into "new" images which better display or highlight certain features in the scene. We will look at some of these operations including various methods of **spectral or band** ratioing, and a procedure called **principal components analysis** which is used to more efficiently represent the information in multichannel imagery.

A   B
© CCRS / CCT
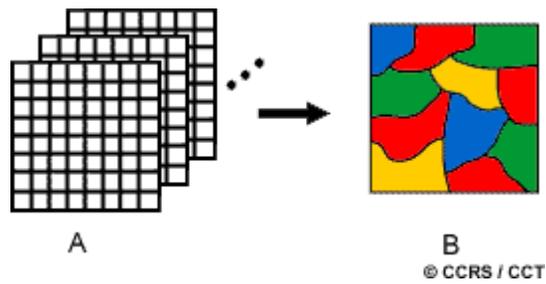
Image classification and analysis operations are used to digitally identify and classify pixels in the data. **Classification** is usually performed on multi-channel data sets (A) and this process assigns each pixel in an image to a particular class or theme (B) based on statistical characteristics of the pixel brightness values. There are a variety of approaches taken to perform digital classification. We will briefly describe the two generic approaches which are used most often, namely **supervised** and **unsupervised** classification.

In the following sections we will describe each of these four categories of digital image processing functions in more detail.

3.4 Pre-processing

Pre-processing operations, sometimes referred to as image restoration and rectification, are intended to correct for sensor- and platform-specific radiometric and geometric distortions of data. Radiometric corrections may be necessary due to variations in scene illumination and viewing geometry, atmospheric conditions, and sensor noise and response. Each of these will vary depending on the specific sensor and platform used to acquire the data and the conditions during data acquisition. Also, it may be desirable to convert and/or calibrate the data to known (absolute) radiation or reflectance units to facilitate comparison between data.

Variations in illumination and viewing geometry between images (for optical sensors) can be corrected by modeling the geometric relationship and distance between the area of the Earth's surface imaged, the sun, and the sensor. This is often required so as to be able to more readily compare images collected by different sensors at different dates or times, or to **mosaic multiple images from a single sensor** while maintaining uniform illumination conditions from scene to scene.

A

B

© CCRS / CCT
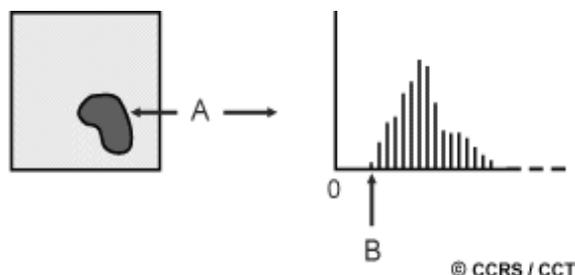
As we learned in Chapter 1, scattering of radiation occurs as it passes through and interacts with the atmosphere. This scattering may reduce, or attenuate, some of the energy illuminating the surface. In addition, the atmosphere will further attenuate the signal propagating from the target to the sensor. Various methods of atmospheric correction can be applied ranging from detailed modeling of the atmospheric conditions during data acquisition, to simple calculations based solely on the image data. An example of the latter method is to **examine the observed brightness values** (digital numbers), in an area of shadow or for a very dark object (such as a large clear lake - A) and determine the minimum value (B). The correction is applied by subtracting the minimum observed value, determined for each specific band, from all pixel values in each respective band. Since scattering is wavelength dependent (Chapter 1), the minimum values will vary from band to band. This method is based on the assumption that the reflectance from these features, if the atmosphere is clear, should be very small, if not zero. If we observe values much greater than zero, then they are considered to have resulted from atmospheric scattering.
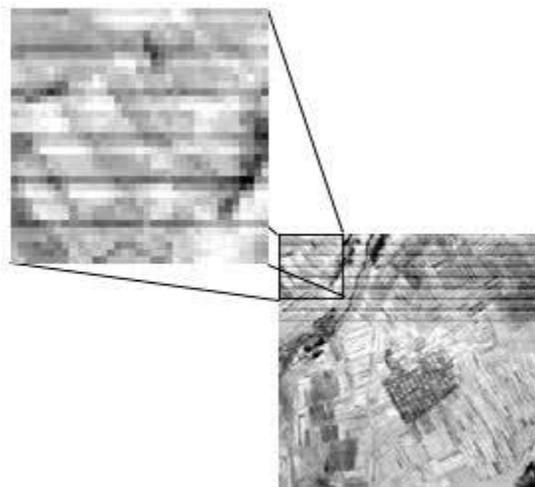
Noise in an image may be due to irregularities or errors that occur in the sensor response and/or data recording and transmission. Common forms of noise include systematic **striping** or banding and **dropped lines**. Both of these effects should be corrected before further enhancement or classification is performed. Striping was common in early Landsat MSS data due to variations and drift in the response over time of the six MSS detectors. The "drift" was different for each of the six detectors, causing the same brightness to be represented differently by each detector. The overall appearance was thus a 'striped' effect. The corrective process made a relative correction among the six sensors to bring their apparent values in line with each other. Dropped lines occur when there are systems errors which result in missing or defective data along a scan line. Dropped lines are normally 'corrected' by replacing the line with the pixel values in the line above or below, or with the average of the two.

For many quantitative applications of remote sensing data, it is necessary to convert the digital numbers to measurements in units which represent the actual reflectance or emittance from the surface. This is done based on detailed knowledge of the sensor response and the way in which the analog signal (i.e. the reflected or emitted radiation) is converted to a digital number, called **analog-to-digital** (A-to-D) conversion. By solving this relationship in the reverse direction, the absolute radiance can be calculated for each pixel, so that comparisons can be accurately made over time and between different sensors.

All remote sensing imagery are inherently subject to geometric distortions. These distortions may be due to several factors, including: the perspective of the sensor optics; the motion of the scanning system; the motion of the platform; the platform altitude, attitude, and velocity; the terrain relief; and, the curvature and rotation of the Earth. Geometric corrections are intended to compensate for these distortions so that the geometric representation of the imagery will be as close as possible to the real world. Many of these variations are **systematic**, or **predictable** in nature and can be accounted for by accurate modeling of the sensor and platform motion and the geometric relationship of the platform with the Earth. Other **unsystematic**, or **random**, errors cannot be modeled and corrected in this way. Therefore, **geometric registration** of the imagery to a known ground coordinate system must be performed.

A   B
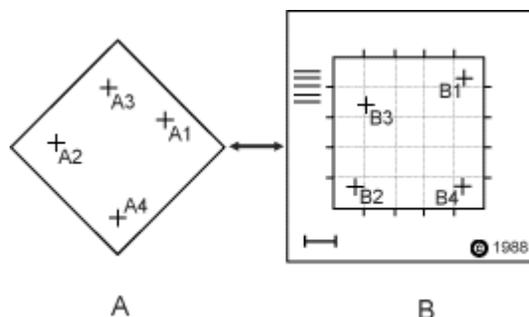
© CCRS / CCT

The **geometric registration process** involves identifying the image coordinates (i.e. row, column) of several clearly discernible points, called **ground control points** (or **GCPs**), in the distorted image (A - A1 to A4), and matching them to their true positions in ground coordinates (e.g. latitude, longitude). The true ground coordinates are typically measured from a map (B - B1 to B4), either in paper or digital format. This is **image-to-map registration**.

Once several well-distributed GCP pairs have been identified, the coordinate information is processed by the computer to determine the proper transformation equations to apply to the original (row and column) image coordinates to map them into their new ground coordinates. Geometric registration may also be performed by registering one (or more) images to another image, instead of to geographic coordinates. This is called image-to-image registration and is often done prior to performing various image transformation procedures, which will be discussed in section 4.6, or for multitemporal image comparison.

Corrected Image

Original Image

© CCRS / CCT
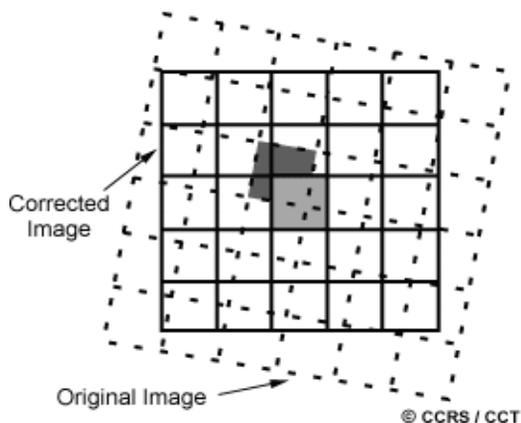
In order to actually geometrically correct the original distorted image, a procedure called **resampling** is used to determine the digital values to place in the new pixel locations of the corrected output image. The resampling process calculates the new pixel values from the original digital pixel values in the uncorrected image. There are three common methods for resampling: **nearest neighbour, bilinear interpolation**, and **cubic convolution**. **Nearest neighbour** resampling uses the digital value from the pixel in the original image which is nearest to the new pixel location in the corrected image. This is the simplest method and does not alter the original values, but may result in some pixel values being duplicated while others are lost. This method also tends to result in a disjointed or blocky image appearance.

Corrected Image

Original Image

© CCRS / CCT
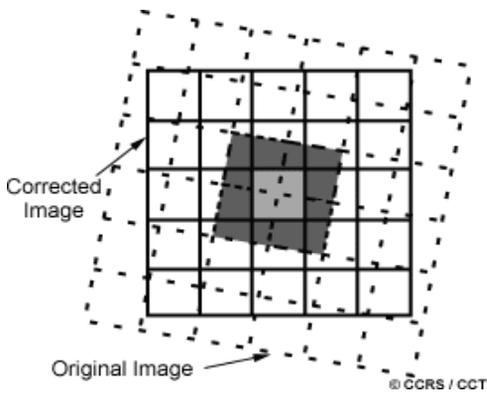
Bilinear interpolation resampling takes a weighted average of four pixels in the original image nearest to the new pixel location. The averaging process alters the original pixel values and creates entirely new digital values in the output image. This may be undesirable if further processing and analysis, such as classification based on spectral response, is to be done. If this is the case, resampling may best be done after the classification process. **Cubic convolution** resampling goes even further to calculate a distance weighted average of a block of sixteen pixels from the original image which surround the new output pixel location. As with bilinear interpolation, this method results in completely new pixel values. However, these two methods both produce images which have a much sharper appearance and avoid the blocky appearance of the nearest neighbour method.

Corrected Image

Original Image

© CCRS / CCT
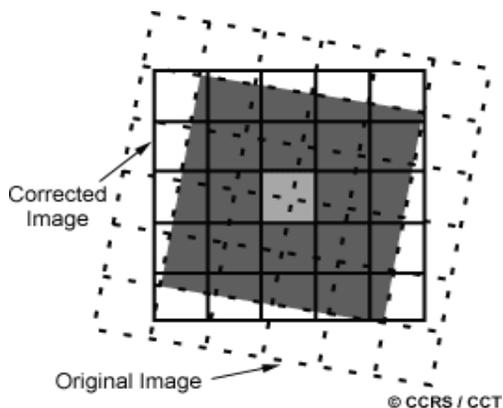
3.5 Image Enhancement

Enhancements are used to make it easier for visual interpretation and understanding of imagery. The advantage of digital imagery is that it allows us to manipulate the digital pixel values in an image. Although radiometric corrections for illumination, atmospheric influences, and sensor characteristics may be done prior to distribution of data to the user, the image may still not be optimized for visual interpretation. Remote sensing devices, particularly those operated from satellite platforms, must be designed to cope with levels of target/background energy which are typical of all conditions likely to be encountered in routine use. With large variations in spectral response from a diverse range of targets (e.g. forest, deserts, snowfields, water, etc.) no generic radiometric correction could optimally account for and display the optimum brightness range and contrast for all targets. Thus, for each application and each image, a custom adjustment of the range and distribution of brightness values is usually necessary.
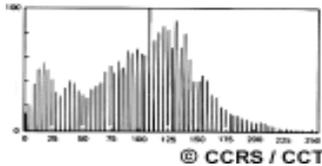
In raw imagery, the useful data often populates only a small portion of the available range of digital values (commonly 8 bits or 256 levels). Contrast enhancement involves changing the original values so that more of the available range is used, thereby increasing the contrast between targets and their backgrounds. The key to understanding contrast enhancements is to understand the concept of an **image histogram**. A histogram is a graphical representation of the brightness values that comprise an image. The brightness values (i.e. 0-255) are displayed along the x-axis of the graph. The frequency of occurrence of each of these values in the image is shown on the y-axis.

By manipulating the range of digital values in an image, graphically represented by its histogram, we can apply various enhancements to the data. There are many different techniques and methods of enhancing contrast and detail in an image; we will cover only a few common ones here. The simplest type of enhancement is a **linear contrast stretch**. This involves identifying lower and upper bounds from the histogram (usually the minimum and maximum brightness values in the image) and applying a transformation to stretch this range to fill the full range. In our example, the minimum value (occupied by actual data) in the histogram is 84 and the maximum value is 153. These 70 levels occupy less than one-third of the full 256 levels available. A linear stretch uniformly expands this small range to cover the full range of values from 0 to 255. This enhances the contrast in the image with light toned areas appearing lighter and dark areas appearing darker, making visual interpretation much easier. This graphic illustrates the increase in contrast in an image before (left) and after (right) a linear contrast stretch.
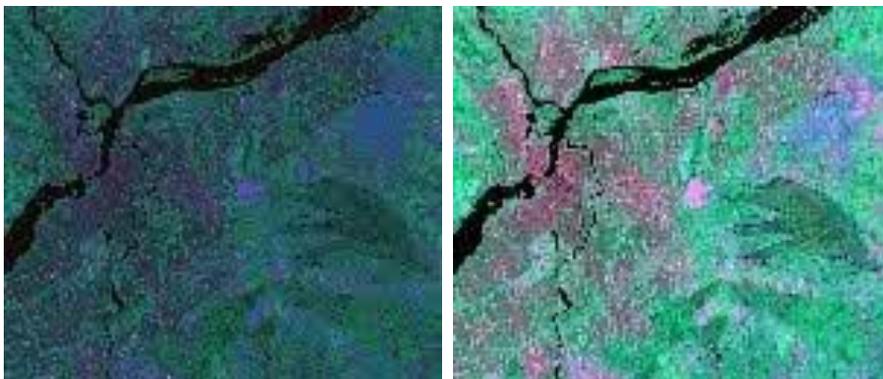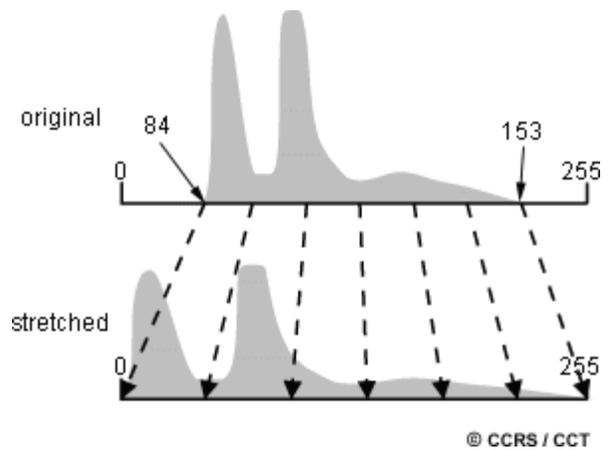
A uniform distribution of the input range of values across the full range may not always be an appropriate enhancement, particularly if the input range is not uniformly distributed. In this case, a **histogram-equalized stretch** may be better. This stretch assigns more display values (range) to the frequently occurring portions of the histogram. In this way, the detail in these areas will be better enhanced relative to those areas of the original histogram where values occur less frequently. In other cases, it may be desirable to enhance the contrast in only a specific portion of the histogram.

For example, suppose we have an image of the mouth of a river, and the water portions of the image occupy the digital values from 40 to 76 out of the entire image histogram. If we wished to enhance the detail in the water, perhaps to see variations in sediment load, we could stretch only that small portion of the histogram represented by the water (40 to 76) to the full grey level range (0 to 255). All pixels below or above these values would be assigned to 0 and 255, respectively, and the detail in these areas would be lost. However, the detail in the water would be greatly enhanced.
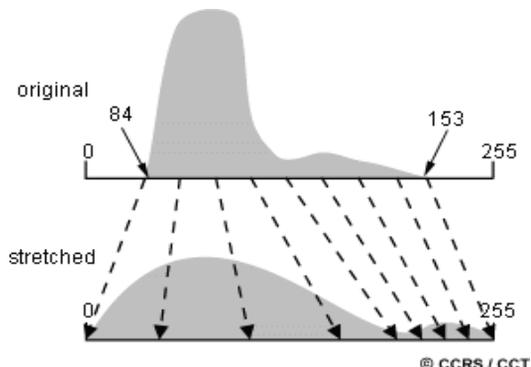
Spatial filtering encompasses another set of digital processing functions which are used to enhance the appearance of an image. Spatial filters are designed to highlight or suppress specific features in an image based on their **spatial frequency**. Spatial frequency is related to the concept of image texture, which we discussed in section 4.2. It refers to the frequency of the variations in tone that appear in an image. "Rough" textured areas of an image, where the changes in tone are abrupt over a small area, have high spatial frequencies, while "smooth" areas with little variation in tone over several pixels, have low spatial frequencies.
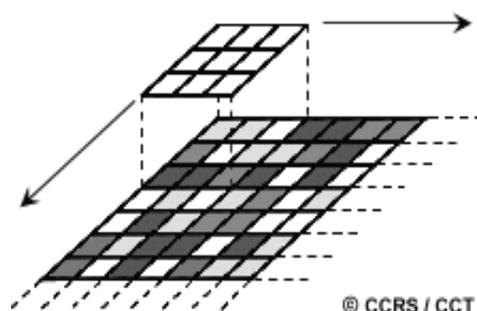
A common **filtering procedure** involves moving a 'window' of a few pixels in dimension (e.g. 3x3, 5x5, etc.) over each pixel in the image, applying a mathematical calculation using the pixel values under that window, and replacing the central pixel with the new value. The window is moved along in both the row and column dimensions one pixel at a time and the calculation is repeated until the entire image has been filtered and a "new" image has been generated. By varying the calculation performed and the weightings of the individual pixels in the filter window, filters can be designed to enhance or suppress different types of features.
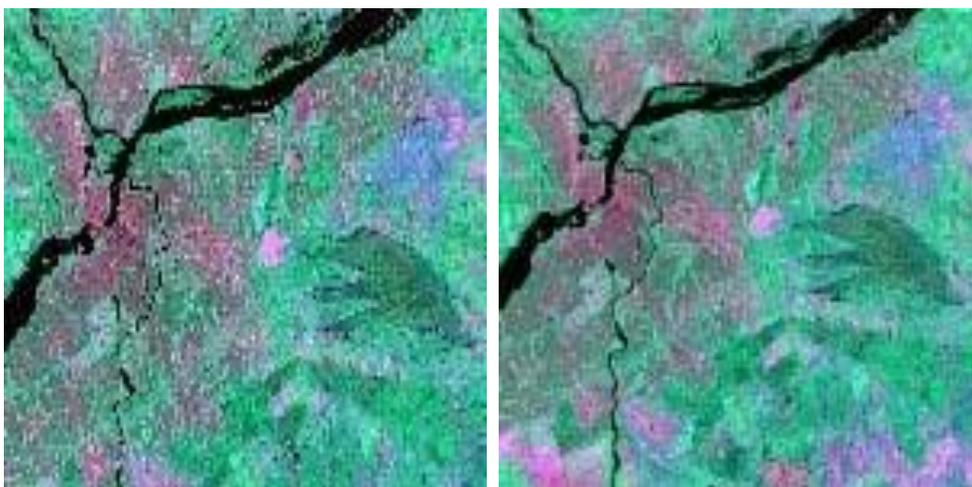
A **low-pass filter** is designed to emphasize larger, homogeneous areas of similar tone and reduce the smaller detail in an image. Thus, low-pass filters generally serve to smooth the appearance of an image. Average and median filters, often used for radar imagery (and described in Chapter 3), are examples of low-pass filters. **High-pass filters** do the opposite and serve to sharpen the appearance of fine detail in an image. One implementation of a high-pass filter first applies a low-pass filter to an image and then subtracts the result from the original, leaving behind only the high spatial frequency information. **Directional, or edge detection filters** are designed to highlight linear features, such as roads or field boundaries. These filters can also be designed to enhance features which are oriented in specific directions. These filters are useful in applications such as geology, for the detection of linear geologic structures.
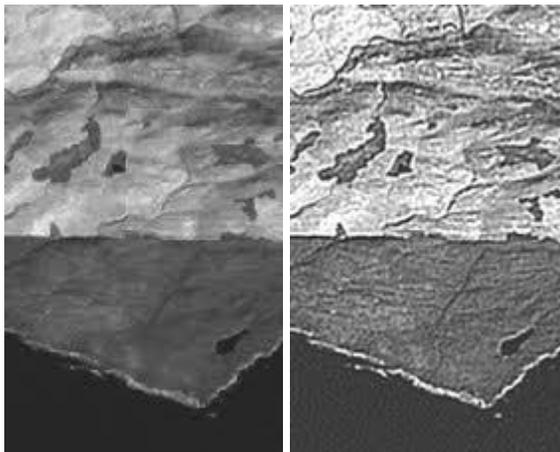
3.6 Image Transformations

Image transformations typically involve the manipulation of multiple bands of data, whether from a single multispectral image or from two or more images of the same area acquired at different times (i.e. multitemporal image data). Either way, image transformations generate "new" images from two or more sources which highlight particular features or properties of interest, better than the original input images.

basic image transformations apply simple arithmetic operations to the image data. **Image subtraction** is often used to identify changes that have occurred between images collected on different dates. Typically, two images which have been geometrically registered (see section 4.4), are used with the pixel (brightness) values in one image (1) being subtracted from the pixel values in the other (2). Scaling the resultant image (3) by adding a constant (127 in this case) to the output values will result in a suitable 'difference' image. In such an image, areas where there has been little or no change (A) between the original images, will have resultant brightness values around 127 (mid-grey tones), while those areas where significant change has occurred (B) will have values higher or lower than 127 - brighter or darker depending on the 'direction' of change in reflectance between the two images . This type of image transform can be useful for mapping changes in urban development around cities and for identifying areas where deforestation is occurring, as in this example.
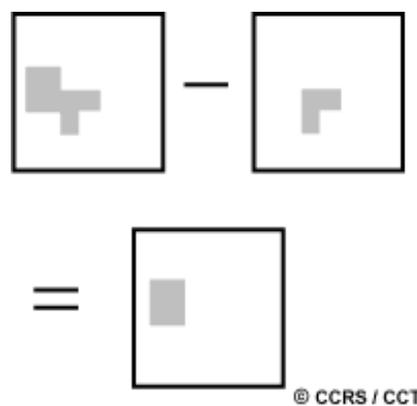
© CCRS / CCT

Image division or **spectral ratioing** is one of the most common transforms applied to image data. Image ratioing serves to highlight subtle variations in the spectral responses of various surface covers. By ratioing the data from two different spectral bands, the resultant image enhances variations in the slopes of the spectral reflectance curves between the two different spectral ranges that may otherwise be masked by the pixel brightness variations in each of the bands. The following example illustrates the concept of spectral ratioing. Healthy vegetation reflects strongly in the near-infrared portion of the spectrum while absorbing strongly in the visible red. Other surface types, such as soil and water, show near equal reflectances in both the near-infrared and red portions. Thus, a ratio image of Landsat MSS Band 7 (Near-Infrared - 0.8 to 1.1 mm) divided by Band 5 (Red - 0.6 to 0.7 mm) would result in ratios much greater than 1.0 for vegetation, and ratios around 1.0 for soil and water. Thus the discrimination of vegetation from other surface cover types is significantly enhanced. Also, we may be better able to identify areas of unhealthy or stressed vegetation, which show low near-infrared reflectance, as the ratios would be lower than for healthy green vegetation.
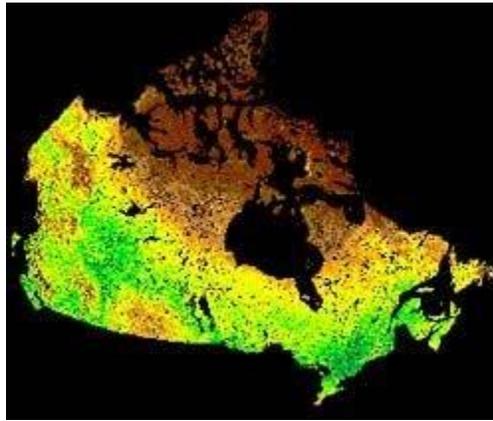
Another benefit of spectral ratioing is that, because we are looking at relative values (i.e. ratios) instead of absolute brightness values, variations in scene illumination as a result of topographic effects are reduced. Thus, although the absolute reflectances for forest covered slopes may vary depending on their orientation relative to the sun's illumination, the ratio of their reflectances between the two bands should always be very similar. More complex ratios involving the sums of and differences between spectral bands for various sensors, have been developed for monitoring vegetation conditions. One widely used image transform is the **Normalized Difference Vegetation Index (NDVI)** which has been used to monitor vegetation conditions on continental and global scales using the Advanced Very High Resolution Radiometer (AVHRR) sensor onboard the NOAA series of satellites (see Chapter 2, section 2.11).
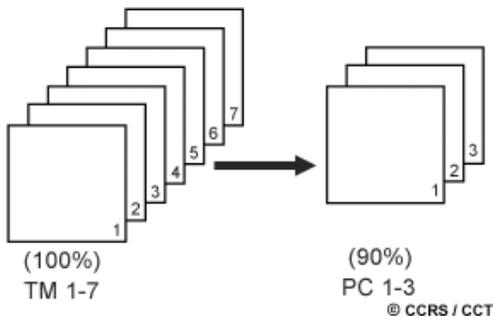
Different bands of multispectral data are often highly correlated and thus contain similar information. For example, Landsat MSS Bands 4 and 5 (green and red, respectively) typically have similar visual appearances since reflectances for the same surface cover types are almost equal. Image transformation techniques based on complex processing of the statistical characteristics of multi-band data sets can be used to reduce this data redundancy and correlation between bands.

One such transform is called **principal components analysis**. The objective of this transformation is to reduce the dimensionality (i.e. the number of bands) in the data, and compress as much of the information in the original bands into fewer bands. The "new" bands that result from this statistical procedure are called components. This process attempts to maximize (statistically) the amount of information (or variance) from the original data into the least number of new components. As an example of the use of principal components analysis, a seven band Thematic Mapper (TM) data set may be transformed such that the first three principal components contain over 90 percent of the information in the original seven bands.

Interpretation and analysis of these three bands of data, combining them either visually or digitally, is simpler and more efficient than trying to use all of the original seven bands. Principal components analysis, and other complex transforms, can be used either as an enhancement technique to improve visual interpretation or to reduce the number of bands to be used as input to digital classification procedures, discussed in the next section.

3.7 Image Classification and Analysis

A human analyst attempting to classify features in an image uses the elements of visual interpretation (discussed in section 4.2) to identify homogeneous groups of pixels which represent various features or land cover classes of interest. **Digital image classification** uses the spectral information represented by the digital numbers in one or more spectral bands, and attempts to classify each individual pixel based on this spectral information.

This type of classification is termed **spectral pattern recognition**. In either case, the objective is to assign all pixels in the image to particular classes or themes (e.g. water, coniferous forest, deciduous forest, corn, wheat, etc.). The resulting classified image is comprised of a mosaic of pixels, each of which belong to a particular theme, and is essentially a thematic "map" of the original image.
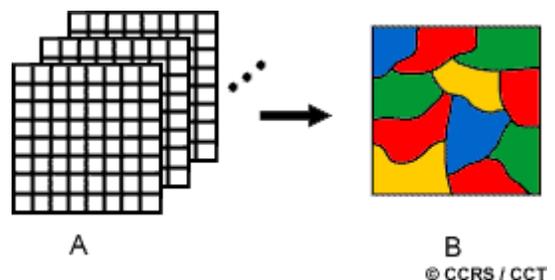
When talking about classes, we need to distinguish between **information classes and spectral classes**. Information classes are those categories of interest that the analyst is actually trying to identify in the imagery, such as different kinds of crops, different forest types or tree species, different geologic units or rock types, etc. Spectral classes are groups of pixels that are uniform (or near-similar) with respect to their brightness values in the different spectral channels of the data. The objective is to match the spectral classes in the data to the information classes of interest. Rarely is there a simple one-to-one match between these two types of classes. Rather, unique spectral classes may appear which do not necessarily correspond to any information class of particular use or interest to the analyst. Alternatively, a broad information class (e.g. forest) may contain a number of spectral **sub-classes** with unique spectral variations. Using the forest example, spectral sub-classes may be due to variations in age, species, and density, or perhaps as a result of shadowing or variations in scene illumination. It is the analyst's job to decide on the utility of the different spectral classes and their correspondence to useful information classes.
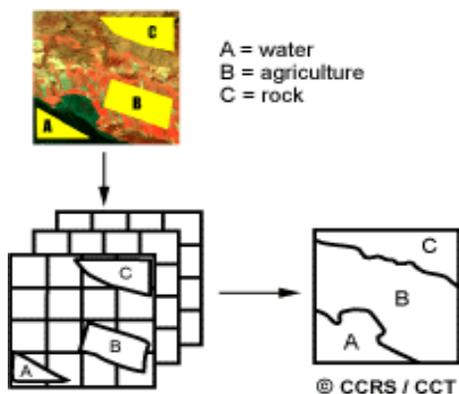
Common classification procedures can be broken down into two broad subdivisions based on the method used: **supervised classification and unsupervised classification**. In a **supervised classification**, the analyst identifies in the imagery homogeneous representative samples of the different surface cover types (information classes) of interest. These samples are referred to as **training areas**. The selection of appropriate training areas is based on the analyst's familiarity with the geographical area and their knowledge of the actual surface cover types present in the image. Thus, the analyst is "supervising" the categorization of a set of specific classes. The numerical information in all spectral bands for the pixels comprising these areas are used to "train" the computer to recognize spectrally similar areas for each class. The computer uses a special program or algorithm (of which there are several variations), to determine the numerical "signatures" for each training class.

Once the computer has determined the signatures for each class, each pixel in the image is compared to these signatures and labeled as the class it most closely "resembles" digitally. Thus, in a supervised classification we are first identifying the information classes which are then used to determine the spectral classes which represent them.
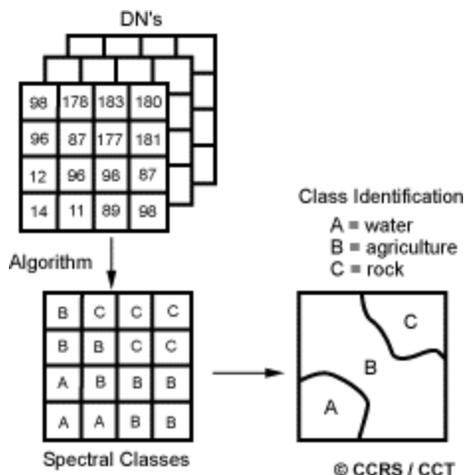
Unsupervised classification in essence reverses the supervised classification process. Spectral classes are grouped first, based solely on the numerical information in the data, and are then matched by the analyst to information classes (if possible). Programs, called **clustering algorithms**, are used to determine the natural (statistical) groupings or structures in the data. Usually, the analyst specifies how many groups or clusters are to be looked for in the data. In addition to specifying the desired number of classes, the analyst may also specify parameters related to the separation distance among the clusters and the variation within each cluster. The final result of this iterative clustering process may result in some clusters that the analyst will want to subsequently combine, or clusters that should be broken down further - each of these requiring a further application of the clustering algorithm. Thus, unsupervised classification is not completely without human intervention. However, it does not start with a pre-determined set of classes as in a supervised classification.

3.8 Data Integration and Analysis

© CCRS / CCT
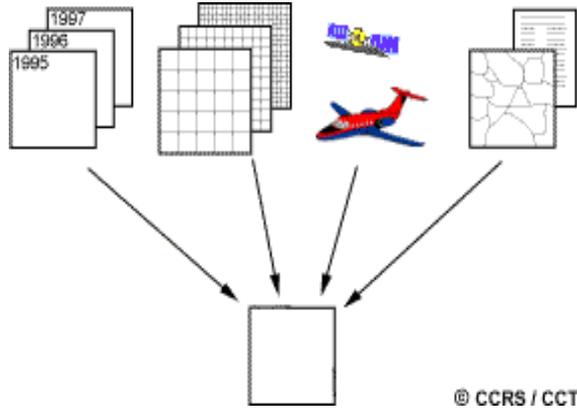
In the early days of analog remote sensing when the only remote sensing data source was aerial photography, the capability for integration of data from different sources was limited. Today, with most data available in digital format from a wide array of sensors, data integration is a common method used for interpretation and analysis. **Data integration** fundamentally involves the combining or merging of data from multiple sources in an effort to extract better and/or more information. This may include data that are multitemporal, multiresolution, multisensor, or multi-data type in nature.

Multitemporal data integration has already been alluded to in section 4.6 when we discussed image subtraction. Imagery collected at different times is integrated to identify areas of change. Multitemporal change detection can be achieved through simple methods such as these, or by other more complex approaches such as multiple classification comparisons or classifications using integrated multitemporal data sets. Multiresolution data merging is useful for a variety of applications. The merging of data of a higher spatial resolution with data of lower resolution can significantly sharpen the spatial detail in an image and enhance the discrimination of features. **SPOT data** are well suited to this approach as the 10 metre panchromatic data can be easily merged with the 20 metre multispectral data. Additionally, the multispectral data serve to retain good spectral resolution while the panchromatic data provide the improved spatial resolution.

Data from different sensors may also be merged, bringing in the concept of multisensor data fusion. An excellent example of this technique is the combination of **multispectral optical data with radar imagery**. These two diverse spectral representations of the surface can provide complementary information. The optical data provide detailed spectral information useful for discriminating between surface cover types, while the radar imagery highlights the structural detail in the image.
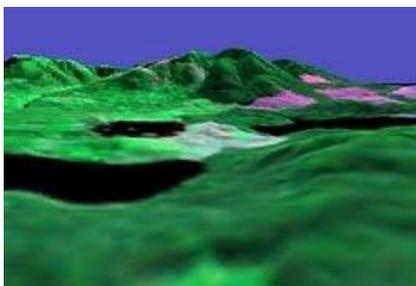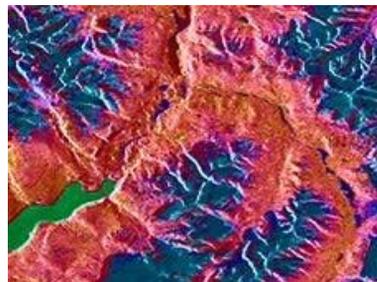
Applications of multisensor data integration generally require that the data be geometrically registered, either to each other or to a common geographic coordinate system or map base. This also allows other **ancillary** (supplementary) data sources to be integrated with the remote sensing data. For example, elevation data in digital form, called **Digital Elevation or Digital Terrain Models (DEMs/DTMs)**, may be combined with remote sensing data for a variety of purposes. DEMs/DTMs may be useful in image classification, as effects due to terrain and slope variability can be corrected, potentially increasing the accuracy of the resultant classification. DEMs/DTMs are also useful for generating **three-dimensional perspective views** by draping remote sensing imagery over the elevation data, enhancing visualization of the area imaged.

Combining data of different types and from different sources, such as we have described above, is the pinnacle of data integration and analysis. 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 **Geographical Information System (GIS)** database. 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.

© CCRS / CCT
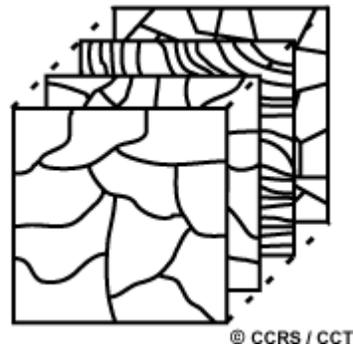
The results from a classification of a remote sensing data set in map format, could also be used in a GIS as another data source to update existing map data. In essence, by analyzing diverse data sets together, it is possible to extract better and more accurate information in a synergistic manner than by using a single data source alone. There are a myriad of potential applications and analyses possible for many applications. In the next and final chapter, we will look at examples of various applications of remote sensing data, many involving the integration of data from different sources.

3.9 Satellite Images used in REWARD Programme

The satellite images planned to be used under the REWARD program are IRS LISS-IV for land use land cover change analysis study, Sentinel-2 for Normalized differential vegetation indices study, Sentinel-1 SAR data for surface soil moisture assessment and correlating hydrological field data of model watershed, and the Worldview-2 data available with KSRSAC for LRI field activities.

4. Fundamentals of Geographic Information System (GIS)

Introduction

A geographic information system (GIS) is a computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface. By relating seemingly unrelated data, GIS can help individuals and organizations better understand spatial patterns and relationships. GIS can use any information that includes location. The location can be expressed in many different ways, such as latitude and longitude, address, or ZIP code.

Functions of GIS

- Data Capture: The input of data into a GIS can be achieved through many different methods of gathering. For example, aerial photography, scanning, digitizing, GPS is just a few of the ways a GIS user could obtain data.

- Storage: Map data stored in a hardcopy, stored on CD or on your hard drive.

- Data Manipulation: The digital geographical data can be edited, this allows for many attribute to be added, edited, or deleted.

- Query And Analysis: GIS used widely in decision making, applying various data retrieving parameters by the users.

- Display: This represents the ability to visualize the data, maps, and information.

Functions and Uses of GIS

Input	Process	Output
❑ Satellite IMAGES ❑ Aerial Photographs ❑ MAPS	Data Capture and Collection	❑ REPORTS ❑ TABLES
❑ Hard Disk ❑ DVD ❑ USB Drive	Storage and Management	❑ Cloud ❑ Online Drive ❑ Update Data
❑ Variables ❑ Factors ❑ Constraints	Data Integration	
	Data Conversion	❑ Raster to Vector ❑ Vector to Raster ❑ Raster to Raster ❑ Vector to Vector
❑ Query Display ❑ Graphs ❑ Tables ❑ Overlay	Data Analysis	
	Data Modeling	❑ Interpolation ❑ Simulation ❑ Mathematical Modeling ❑ Statistical Modeling
❑ Digital Maps ❑ Printed Maps ❑ Tables ❑ Graphs	Display	

Components of GIS

A working GIS integrates five key components: hardware, software, data, people, and methods.

- **Hardware:** Hardware is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

- **Software:** GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are;
  - Tools for the input and manipulation of geographic information.
  - A database management system (DBMS).
  - Tools that support geographic query, analysis, and visualization.
  - A graphical user interface (GUI) for easy access to tools.

- **Data:** Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organizations to organize and maintain their data, to manage spatial data.

- **People:** GIS technology is of limited value without the people who manage the system and develop plans for applying it to real-world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work.

- **Method:** A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

This implies that a GIS user can expect support from the system to enter (geo- referenced) data, to analyze it in various ways, and to produce presentations (including maps and other types) from the data. This would include support for various kinds of coordinate systems and transformations between them, options for analysis of the georeferenced data, and obviously a large degree of freedom of choice in the way this information is presented (such as color scheme, symbol set, and medium used).

Spatial Data and Geoinformation

A subtle difference exists between the terms data and information. Most of the time, we use the two terms almost interchangeably, and without the risk of confusing their meanings. By data, we mean representations that can be operated upon by a computer. More specifically, by spatial data we mean data that contains positional values, such as (x, y) co-ordinates. Sometimes the more precise phrase geospatial data is used as a further refinement, which refers to spatial data that is georeferenced.

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

Data Quality considerations

Key components of spatial data quality include positional accuracy (both horizontal and vertical), temporal accuracy (that the data is up to date), attribute accuracy (e.g. in labelling of features or of classifications), lineage (history of the data including sources), completeness (if the data set represents all related features of reality), and logical consistency (that the data is logically structured). These components play an important role in assessment of data quality for several reasons:

- Even when source data, such as official topographic maps, have been subject to stringent quality control, errors are introduced when these data are input to GIS.
- Unlike a conventional map, which is essentially a single product, a GIS database normally contains data from different sources of varying quality.
- Unlike topographic or cadastral databases, natural resource databases contain data that are inherently uncertain and therefore not suited to conventional quality control procedures.
- Most GIS analysis operations will themselves introduce errors.

If data is to be shared among different users, these users need to know not only what data exists, where and in what format it is held, but also whether the data meets their particular quality requirements. This 'data about data' is known as metadata. Since the real power of GIS lies in their ability to combine and analyze georeferenced data from a range of sources, we must pay attention to the issues of data quality and error, as data from different sources are also likely to contain different kinds of error. This may include mistakes or variation in the measurement of position and/or elevation, in the quantitative measurement of attributes or metadata in the labelling or classification of features. Some degree of error is present in every spatial data set. It is important, however, to distinguish between gross errors (blunders or mistakes), which must be detected and removed before the data is used, variations in the data caused by unavoidable measurement and classification errors.

Data quality parameters

Key components of spatial data quality include positional accuracy (both horizontal and vertical), temporal accuracy (that the data is up to date), attribute accuracy
(e.g. in labelling of features or of classifications), lineage (history of the data including sources), completeness (if the data set represents all related features of reality), and logical consistency (that the data is logically structured).
These components play an important role in assessment of data quality for several reasons:

- Even when source data, such as official topographic maps, have been subject to stringent quality control, errors are introduced when these data are input to GIS.
- Unlike a conventional map, which is essentially a single product, a GIS database normally contains data from different sources of varying quality.
- Unlike topographic or cadastral databases, natural resource databases contain data that are inherently uncertain and therefore not suited to conventional quality control procedures.
- Most GIS analysis operations will themselves introduce errors.

Map and Database

In the GIS environment, the most familiar model is that of a map. A map is a miniature representation of some part of the real world.

Databases are another important class of models. A database can store a considerable amount of data, and also provides various functions to operate on the stored data. The collection of stored data represents some real world phenomena, so it too is a model. Most maps and databases can be considered static models. At any point in time, they represent a single state of affairs. Usually, developments or changes in the real world are not easily recognized in these models. Dynamic models or process models address precisely this issue.

Dynamic models are inherently more complicated than static models, and usually require much more computation. Simulation models are an important class of dynamic models that allow the simulation of real world processes.

Maps

As noted above, maps are perhaps the best known (conventional) models of the real world. Maps have been used for thousands of years to represent information about the real world, and continue to be extremely useful for many applications in various domains.

The map scale determines the spatial resolution of the graphic feature representation. The smaller the scale, the less detail a map can show. The accuracy of the base data, on the other hand, puts limits to the scale in which a map can be sensibly drawn. Hence, the selection of a proper map scale is one of the first and most important steps in map design. With the advent of computer systems, analogue cartography developed into digital cartography, and computers play an integral part in modern cartography.

Databases

A database is a repository for storing large amounts of data. It comes with a number of useful functions:
- o A database can be used by multiple users at the same time - i.e. it allows concurrent use.
- o A database offers a number of techniques for storing data and allows the use of the most efficient one - i.e. it supports storage optimization.
- o A database allows the imposition of rules on the stored data; rules that will be automatically checked after each update to the data - i.e. it supports data integrity.
- o A database offers an easy to use data manipulation language, which allows the execution of all sorts of data extraction and data updates - i.e. it has a query facility.
- o A database will try to execute each query in the data manipulation language in the most efficient way - i.e. it offers query optimization.

Databases can store almost any kind of data. Modern database systems shall organize the stored data in tabular format. A database may have many such tables, each of which stores data of a certain kind. It is not uncommon for a table to have many thousands of data rows, sometimes even hundreds of thousands.

Spatial Databases and Spatial Analysis

A GIS must store its data in some way. For this purpose the previous generation of software was equipped with relatively rudimentary facilities. In more recent years, spatial databases (also known as geodatabases) have emerged. Besides traditional administrative data, they can store representations of real world geographic phenomena for use in a GIS. These databases are special because they use additional techniques different from tables to store these spatial representations.

A geodatabase is not the same thing as a GIS, though both systems share a number of characteristics. These include the functions listed above for databases in general: concurrency, storage, integrity, and querying, specifically, but not only, spatial data.

A GIS, on the other hand, is tailored to operate on spatial data. It 'knows' about spatial reference systems, and supports all kinds of analyses that are inherently geographic in nature, such as distance and area computations and spatial interpolation. This is probably GIS's main strength: providing various ways to combine representations of geographic phenomena. GISs, moreover, built-in tools for map production, of the paper and the digital kind. They operate with an 'embedded understanding' of geographic space. Databases typically lack this kind of understanding.

The phenomena for which we want to store representations in a spatial database may have point, line, area or image characteristics. Different storage techniques exist for each of these kinds of spatial data. These geographic phenomena have various relationships with each other and possess spatial (geometric), thematic and temporal attributes (they exist in space and time). For data management purposes, phenomena are classified into thematic data layers. The purpose of the database is usually described by a description such as cadastral, topographic, land use, or soil database.

Spatial Analysis

Spatial analysis is the generic term for all manipulations of spatial data carried out to improve one's understanding of the geographic phenomena that the data represents. It involves questions about how the data in various layers might relate to each other, and how it varies over space. The aim of spatial analysis is usually to gain a better understanding of geographic phenomena through discovering patterns that were previously unknown to us, or to build arguments on which to base important decisions. It should be noted that some GIS functions for spatial analysis are simple and easy-to-use, others are much more sophisticated, and demand higher levels of analytical and operating skills. Successful spatial analysis requires appropriate software, hardware, and perhaps most importantly, a competent user.

Geographic Information and Spatial Data Types

Geographic Phenomena

A GIS operates under the assumption that the relevant spatial phenomena occur in a two- or three-dimensional Euclidean space, unless otherwise specified. Euclidean space can be informally defined as a model of space in which locations are represented by coordinates—(x, y) in 2D; (x, y, z) in 3D—and distance and direction can defined with geometric formulas. In the 2D case, this is known as the Euclidean plane, which is the most common Euclidean space in GIS use.

In order to be able to represent relevant aspects real world phenomena inside a GIS, we first need to define what it is we are referring to. We might define a geographic phenomenon as a manifestation of an entity or process of interest that:

○ Can be named or described,

○ Can be georeferenced, and

○ Can be assigned a time (interval) at which it is/was present.

The relevant phenomena for a given application depends entirely on one's objectives. For instance, in water management, the objects of study might be river basins, agro-ecologic units, measurements of actual evapotranspiration, meteorological data, ground water levels, irrigation levels, water budgets and measurements of total water use. Note that all of these can be named or described, georeferenced and provided with a time interval at which each exists. In multipurpose cadastral administration, the objects of study are different: houses, land parcels, streets of various types, land use forms, sewage canal forms of urban infrastructure may all play a role. Again, these can be named or described, georeferenced and assigned a time interval of existence.

Types of Geographic Phenomena

Geographic phenomena come in so many different 'flavours', which we will try to categorize below. Before doing so, we must make two further observations. Firstly, In order to be able to represent a phenomenon in a GIS requires us to state what it is, and where it is. We must provide a description or at least a name on the one hand, and a georeference on the other hand. Secondly, some phenomena manifest themselves essentially everywhere in the study area, while others only do so in certain localities. If we define our study area as the equatorial Pacific Ocean, we can say that Sea Surf can be measured anywhere in the study area. Therefore, it is a typical example of a (geographic) field.

A (geographic) field is a geographic phenomenon for which, for every point in the study area, a value can be determined.

Some common examples of geographic fields are air temperature, barometric pressure and elevation. These fields are in fact continuous in nature. Examples of discrete fields are land use and soil classifications. Many other phenomena do not manifest themselves everywhere in the study area, but only in certain localities. The roads are typical examples of (geographic) objects and we know exactly where they are located.

(Geographic) objects populate the study area, and are usually well- distinguished, discrete, and bounded entities. The space between them is potentially 'empty' or undetermined.

A simple rule-of-thumb is that natural geographic phenomena are usually fields, and man-made phenomena are usually objects. Many exceptions to this rule actually exist, so one must be careful in applying it.
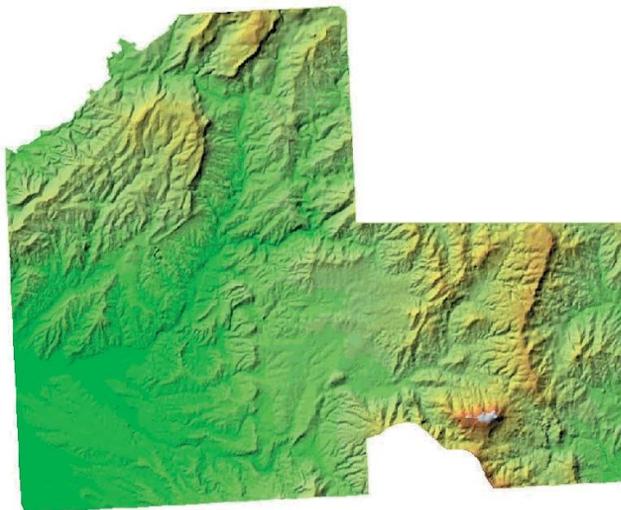
Fig. A continuous field example, namely the *elevation* in a study area
Data source: Department of Earth Systems Analysis (ESA, ITC)

Geographic Fields

A field is a geographic phenomenon that has a value 'everywhere' in the study area. We can therefore think of a field as a mathematical function f that associates a specific value with any position in the study area. Hence if (x, y) is a position in the study area, then f (x, y) stands for the value of the field f at locality (x, y).

Fields can be discrete or continuous. In a continuous field, the underlying function is assumed to be 'mathematically smooth', meaning that the field values along any path through the study area do not change abruptly, but only gradually. Good examples of continuous fields are air temperature, soil salinity and elevation. Continuity means that all changes in field values are gradual.

Discrete fields divide the study space in mutually exclusive, bounded parts, with all locations in one part having the same field value. Typical examples are land classifications, for instance, using either geological classes, soil type, land use type, crop type or natural vegetation type.
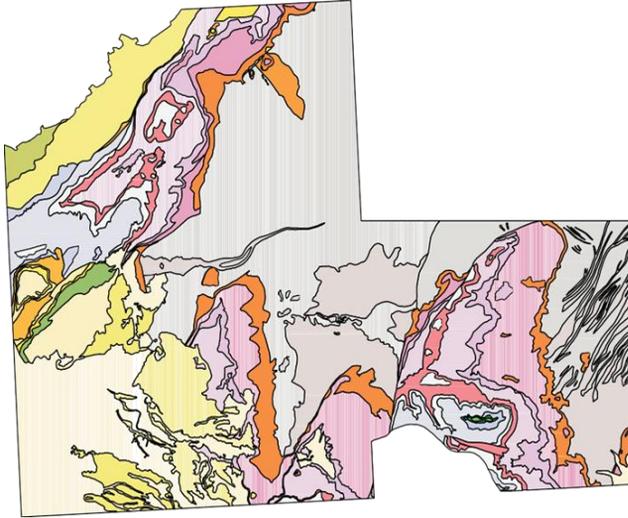
Fig. A discrete field indicating geological units

Geographic objects

When a geographic phenomenon is not present everywhere in the study area, but somehow 'sparsely' populates it, we look at it as a collection of geographic objects. Such objects are usually easily distinguished and named, and their position in space is determined by a combination of one or more of the following parameters:

- o Location (where is it?),
- o Shape (what form is it?),
- o Size (how big is it?), and
- o Orientation (in which direction is it facing?).

How we want to use the information about a geographic object determines which of the four above parameters is required to represent it. For instance, roads are important objects, and for these some notion of location (where does it begin and end), shape (how many lanes does it have), size (how far can one travel on it) and orientation (in which direction can one travel on it) seem to be relevant information components. Shape is usually important because one of its factors is dimension. This relates to whether an object is perceived as a point feature, or a linear, area or volume feature.

Boundaries: Where shape and/or size of contiguous areas matter, the notion of boundary comes into play. This is true for geographic objects but also for the constituents of a discrete geographic field.

GIS Data Types - There are two different types of GIS data, **Vector data** and **Raster data**. Each type of data has its own format.

Vector Data

Vector data is the spatial data most people are familiar with, as it is the format presented in mapping portals such as Open Street Maps and Google Maps. It is also used extensively in computer graphics and computer-aided design (CAD). It consists of points, lines, and polygons.

- o **Point Data** – Point Data typically represents nonadjacent features or distinct data points. Points are zero dimension, so you cannot measure their length or area. Examples of point data would be cities, points of interest, and schools.
- o **Line Data** – Line data is also known as arc data. It represents linear features such as rivers, streets, and trails. Line data has a starting and an ending point, and, since it only has one dimension, can only be used to measure length.

To distinguish arc features from each other, some lines may be solid while others are dashed, and different colours or line thicknesses may be used. For example, a road may be a solid black line, while a river is a dashed blue line.

> o **Polygon Data** – Polygons typically represent areas such as cities, lakes, or forests. Unlike point and line data, polygons are two dimensional and can measure the perimeter or area of a geographic feature. Colour schemes, patterns or gradation colour schemes could be used to identify polygon features.

Vector images are high-quality representations of an image or a shape. They can be enlarged or reduced with no loss of quality.

Vector Data

Points are XY coordinates

Vector points are simply X,Y coordinates. Generally, they are latitude and longitude with a spatial reference frame. When features are too small to be represented as polygons, points are used.

For example, you can't see city boundary lines on a global scale. In this case, maps often use points to display cities .
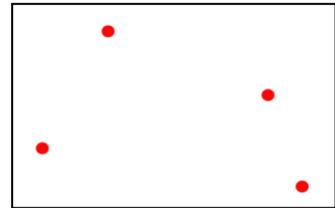
Lines Connect Vertices

Vector lines connect each vertex with paths. Basically, you're connecting the dots in a set order and it becomes a **vector line** with each dot representing a vertex.

Lines usually represent features that are linear in nature. For example, maps show rivers, roads, and pipelines as vector lines. Often, busier highways have thicker lines than abandoned roads.
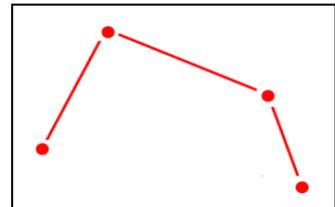
On the other hand, networks are line data sets but they are often considered to be different. This is because linear networks are topologically connected elements. They consist of junctions and turns with connectivity.

If you were to find an optimal route using a traffic line network, it would follow set rules. For example, it can restrict turns and movement on one-way streets.

Polygons Connect Vertices and Close the Path

When you join a set of vertices in a particular order and close it, this is now a **vector polygon** feature. When you create a polygon, the first and last coordinate pairs are the same.

Cartographers use polygons to show boundaries and they all have an area. For example, a building footprint has square footage, and agricultural fields have acreage.
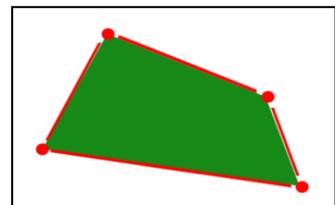
What are the advantages of using vector data?

Because vector data have vertices and paths, this means that the graphical output is generally more aesthetically pleasing. Furthermore, it gives higher geographic accuracy because data isn't dependent on grid size.

Topology rules can help data integrity with vector data models. Not only that, network analysis and proximity operations use vector data structures.

What are the disadvantages of using vector data?

Continuous data is poorly stored and displayed as vectors. If you want to display continuous data as a vector, it would require substantial generalization. Although topology is useful for vector data, it is often processing intensive. Any feature edits require updates on topology. With a lot of features, vector manipulation algorithms are complex.

Raster Data

Raster data, also known as grid data, is made up of pixels, and each pixel has a value. You will typically find raster data on topographic maps, satellite images, and aerial surveys. Raster data is vital for meteorology, disaster management, and industries where analysing risk is essential.

There are two types of raster data, continuous and discrete.

- o **Continuous Data** – Continuous rasters are cells on the grid that gradually change. Some examples would be an aerial photo, elevation and temperature. Continuous raster surfaces come from a fixed registration point. For instance, in digital elevation models, sea level is used as a registration point. Each cell represents a value that is above or below sea level.

- o **Discrete Data** – Discrete rasters have a specific theme or class, and each pixel is assigned to a specific class. Unlike continuous data, discrete data can only take specific values, not values within a range. For example, in a discrete raster land cover/use map, you can see each thematic class, and where it begins and ends is defined.

Unlike vector data, raster data is not scalable. If it is enlarged too much, it will get pixelated, and if stretched too much, it will become distorted. A digital photo is an example of raster data.

Continuous rasters (non-discrete) are grid cells with gradually changing data such as elevation, temperature, or an aerial photograph.
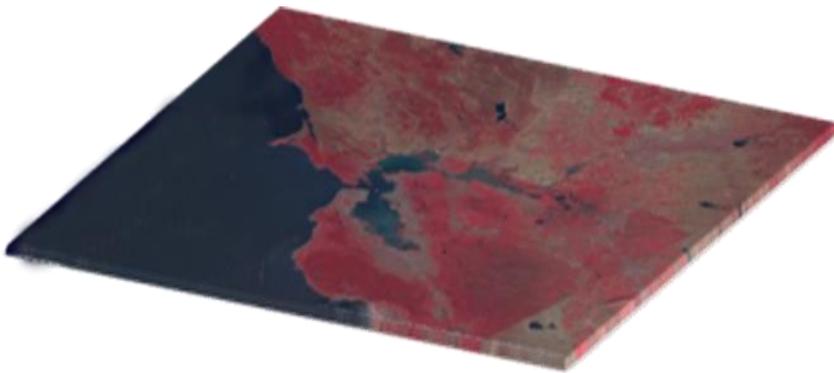
A continuous raster surface can be derived from a **fixed registration point**. For example, digital elevation models use sea level as a registration point.

Each cell represents a value above or below sea level. As another example, aspect cell values have fixed directions such as north, east, south, or west.

Phenomena can gradually vary along a continuous raster from a **specific source**. A raster depicting an oil spill can show how the fluid moves from high concentration to low concentration. At the source of the oil spill, concentration is higher and diffuses outwards with diminishing values as a function of distance.

Discrete Rasters have Distinct Values

Discrete Rasters have distinct themes or categories. For example, one grid cell represents a land cover class or a soil type.
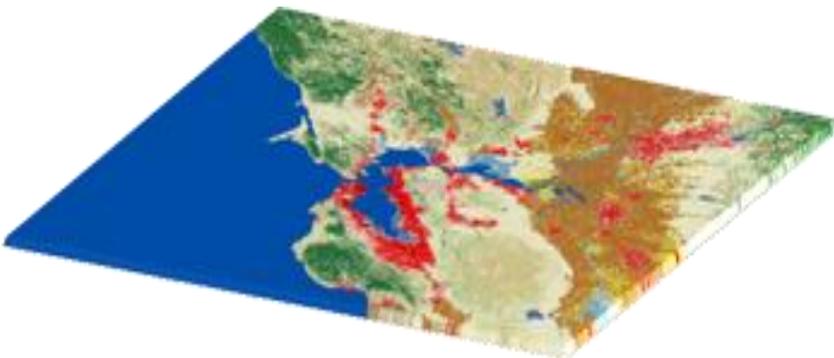
In a discrete raster land use/land cover map, you can distinguish each thematic class. Each class can be discretely defined where it begins and ends.

In other words, each land cover cell is definable and it fills the entire area of the cell.

Discrete data usually consists of integers to represent classes. For example, the value 1 might represent urban areas, the value 2 represents forest, and so on.

What are the advantages of raster data?

A raster grid format is a data model for satellite data and other remote sensing data. For raster positions, it's simple to understand cell size.

Map algebra with raster data is usually quick and easy to perform. Overall, quantitative analysis is intuitive with discrete or continuous raster data.

What are the disadvantages of raster data?

Because cell size contributes to graphic quality, it can have a pixelated look and feel. To illustrate, linear features and paths are difficult to display. You cannot create network datasets or perform topology rules on raster data. Also, you don't have the flexibility with raster data attribute tables.

Raster datasets can become potentially very large because they record values for each cell in an image. As resolution increases, the size of the cell decreases. But this comes at a cost for speed of processing and data storage.
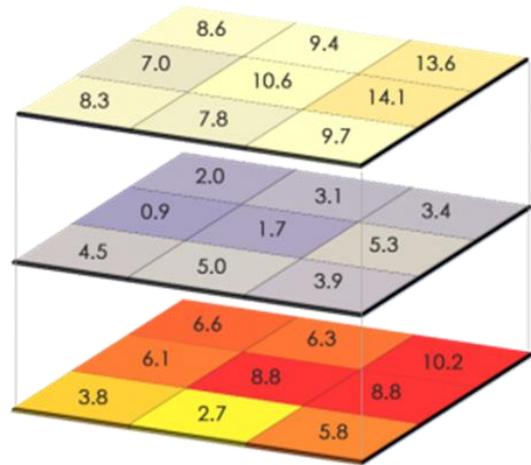
Raster data and vector data are two types of spatial data in GIS. The main difference between Raster and Vector Data is that the raster data represents data as a cell or a grid matrix while vector data represents data using sequential points or vertices.

Geospatial Analysis:

Geospatial analysis deals with the study of the locations and shapes of geographic features and the relationships between them. Spatial analysis is useful when evaluating suitability, when making predictions, and for gaining a better understanding of how geographic features and phenomena are located and distributed.

The following schematic diagram provides the brief information on types of spatial analysis techniques. And the geospatial analysis tools are varied from most simple to complex analysis in the GIS environment.
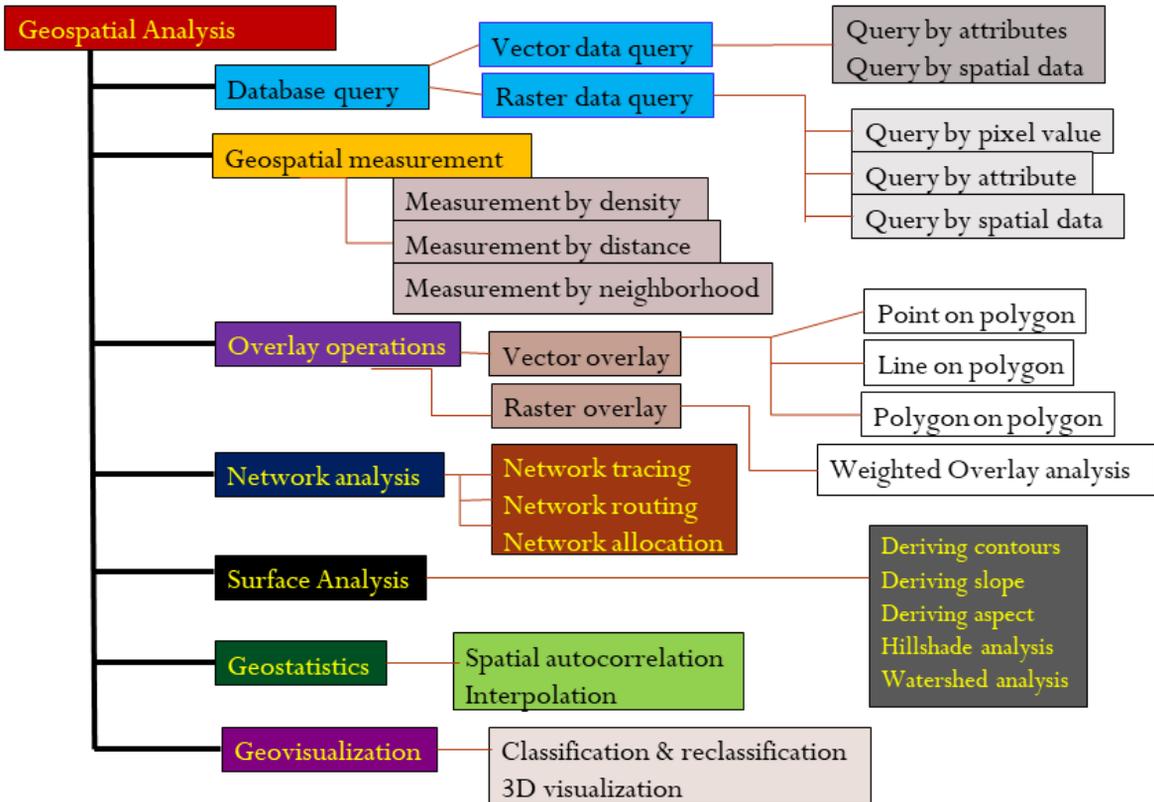
5. Basics of Arc GIS

Introduction

A geographic information system (GIS) is a system that creates, manages, analyses, and maps all types of data. GIS connects data to a map, integrating location data (where things are) with all types of descriptive information (what things are like there). This provides a foundation for mapping and analysis that is used in science and almost every industry. GIS helps users understand patterns, relationships, and geographic context. The benefits include improved communication and efficiency as well as better management and decision making. Geographic Information Systems store information using spatial indices that make it possible to identify the features located in any arbitrary region of a map. In addition to the above capabilities, Maps implements a professional-strength relational database, a feature critical for GIS software. Attribute data may be freely joined to and detached from geographic layers and tables. Relational data manipulation is integrated with robust and powerful geoprocessing for spatial queries, polygon overlay, and other location-based analyses.

How is GIS used?

Hundreds of thousands of organizations in virtually every field are using GIS to make maps that communicate, perform analysis, share information, and solve complex problems around the world. This is changing the way the world works.

- Identify Problems
- Monitor Change
- Perform Forecasting
- Manage & Respond to Events
- Set Priorities
- Understand Trends

ArcGIS provides a scalable framework for implementing GIS for a single user or many users on desktops, in servers, over the Web, and in the field. ArcGIS is an integrated family of GIS software products for building a complete GIS. It consists of several primary frameworks for deploying GIS (ESRI, 2010):

• **ArcGIS Desktop—**An integrated suite of professional GIS applications. Most users recognize this as three products: ArcView, ArcEditor, and ArcInfo.

• **ArcGIS Desktop** is the primary product used by GIS professionals to compile, use, and manage geographic information. It includes comprehensive professional GIS applications that support a number of GIS tasks, including mapping, data compilation, analysis, geodatabase management, and geographic information sharing.

ArcGIS Desktop includes a suite of applications—ArcMap, ArcCatalog, ArcGlobe, ArcScene, ArcToolbox, and ModelBuilder. Using these applications and interfaces, user can perform any GIS tasks from simple to advanced.

ArcGIS Desktop, provide number of elements of GIS information, such as the following:

✓Map documents, globe documents, and layers

✓Geodatabases

✓Geoprocessing toolboxes

✓Other data files such as imagery
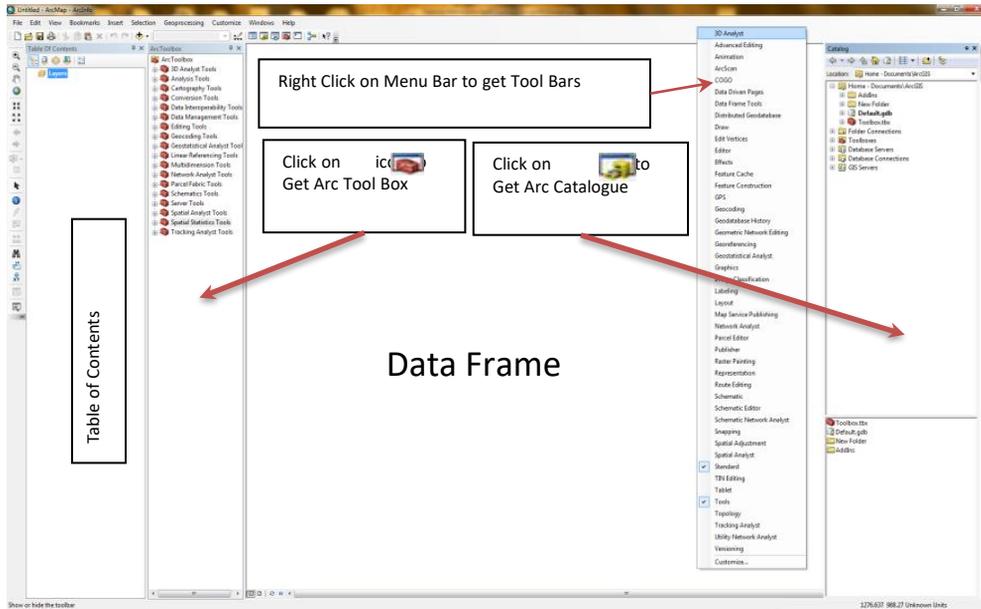
ArcGIS Desktop is scalable and can address the needs of many types of users. It is available at three functional levels:

o **ArcView** focuses on comprehensive data use, mapping, and analysis.

o **ArcEditor** adds advanced geodatabase editing and data creation.

o **ArcInfo** is a complete, professional GIS desktop containing comprehensive GIS functionality and rich geoprocessing tools.

A number of specialized software capabilities are available as add-on extensions to ArcGIS Desktop. Each extension enables user to add capabilities such as raster geoprocessing, 3D GIS, and network analysis.

ArcMap is the primary application used in ArcGIS and is used to perform a wide range of common GIS tasks as well as specialized, user-specific tasks.

2.1 Connect to work space
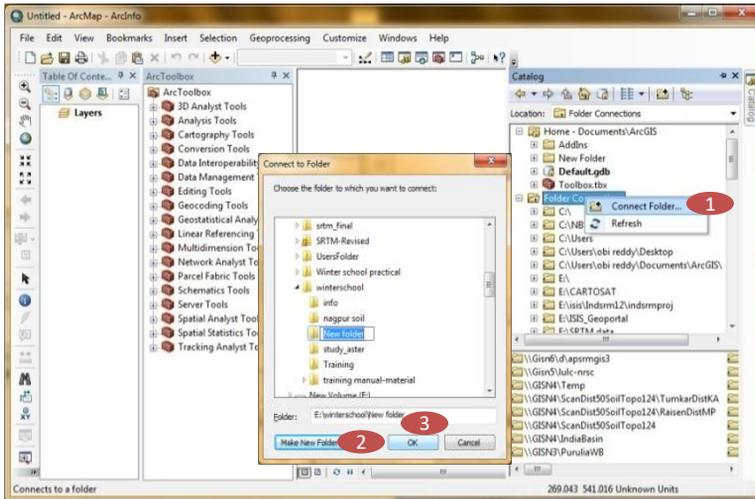
STEP 1: Right Click on **Folder Connections** in **Arc Catalogue** and Select **Connect Folder**

STEP 2: Select the drive or folder of user choice and Click on **Make New Folder** to Create a new folder

STEP 3: Click **OK**

STEP 4: If desired, **Rename** it in the catalogue

Or

Select the folder, if it was already created and click OK

2.2 Creating new vector layer
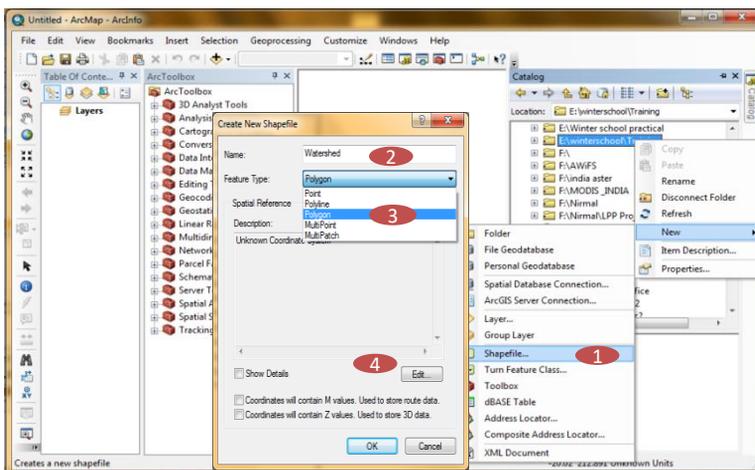
STEP 1: Right Click on created folder and go to **New → Shape File** to open **Create New Shape file** wizard

STEP 2: Define the **Shape file Name**

STEP 3: Select **Feature Type**

STEP 4: Click the **Edit** tab to open **Spatial Reference Properties** wizard
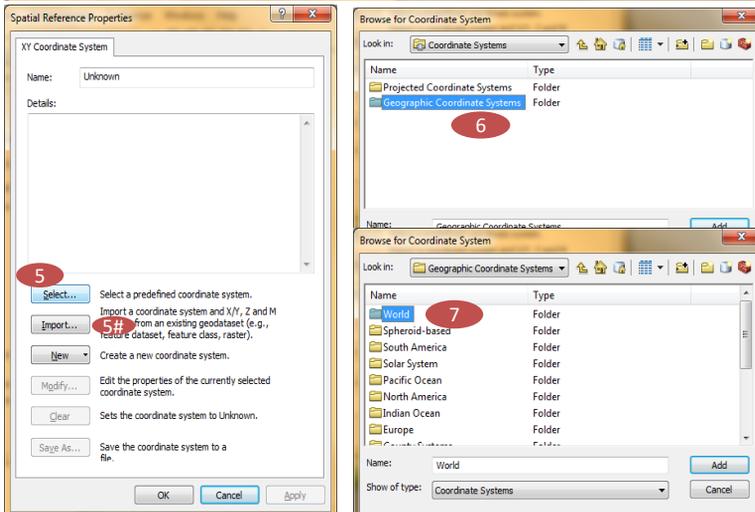
STEP 5: Click **Select** to open **Browse For Coordinate System** window

Or STEP 5#: Click on **Import** tab to import the Spatial reference from already projected image

STEP 6: Double click **Geographic Coordinate System**

STEP 7: Double Click on **World**
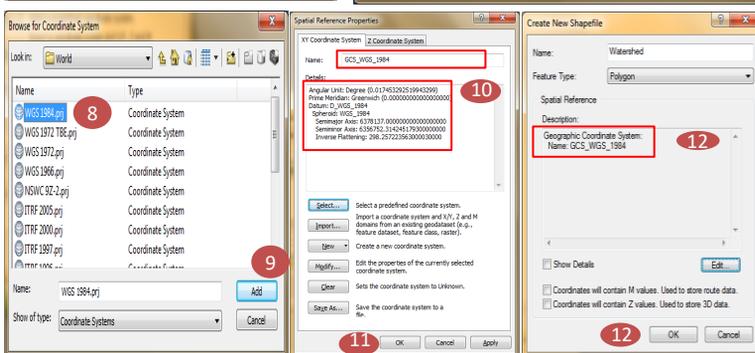
STEP 8: Select **WGS1984.prj**

STEP 9: click **Add**

STEP 10: Check the **Name** of coordinate system and **details**

STEP 11: Click **OK**

STEP 12: Check the **Spatial reference Description** of the new Shape file and **click OK**.

File loading in ArcGIS

In GIS and remote sensing we come across different kinds of raster, vector and database files. Some of the common raster formats are JPEG (Joint Photographic Experts Group) TIFF (Tagged Image File Format) GIF (Graphics Interchange Format) PNG (Portable Network Graphics), etc. Most common non-spatial attribute data formats are .xls, .xlsx, .mdb etc. Different image processing and GIS packages have their own native formats for images, vectors, and tables. For example, different raster formats (.img for ERDAS imagine, .pix for PCI Geomatica, .img for ENVI, .mpr for ILWIS), vector formats (.shp, .evf, .pcx, .grf for ArcGIS and ERDAS, ENVI, PCI Geomatica, and ILWIS, respectively) and table formats (.dbf for ArcGIS, ENVI and ERDAS, .tbt for ILWIS) are available to name a few.

Satellite data providers use different suit of standard file formats for exchange, storage, and transmission of digital-imagery products and image-related products. For example, Landsat data is provided in Fast, Tagged Image File Format (Tiff), Geotiff, National Land Archive Production System (NLAPS), Hierarchical Data Format (HDF, HDF4, and HDF5), Multi-Resolution Land Characteristics (MRLC) format, and Australian Centre for Remote Sensing (ACRES) CCRS format. Similarly IRS satellite data are provided in Fast, Super Structured and Geotiff formats. The GIS and image processing softwares' should be compatible enough to open the most common if not all, formats of the raster, vector, and database files. All leading softwares' are capable to read the most common generic file formats as well as the sensor specific file formats.

A key requirement for working with ArcGIS Desktop is the ability to work with a variety of GIS data formats, DBMS tables, and GIS Web services. In addition to working with geodatabases, ArcView, ArcEditor, and ArcInfo support direct use of a number of GIS and tabular data formats as well as a series of geoprocessing tools that can be used to convert data into supported formats such as a geodatabase. In addition, an optional extension product named the Data Interoperability Extension adds support for over 70 additional data formats that can be used directly within ArcGIS. It also gives the user the ability to define new custom data sources and to define data transformation procedures that help the user perform advanced data transformations between a variety of GIS and tabular data structures. While adding a raster dataset layer to ArcMap, user can choose to display a single band of data or a color composite from three bands of data. If user have data covering the same geographic area but in different coordinate systems, ArcMap uses the coordinate system of the first dataset added to the data frame and transforms any other data on the fly into this coordinate system. This applies to any feature class or raster dataset.

2.3 File loading in Arc GIS

2.3.1 Vector file loading in Arc GIS

STEP 1: Go to **File → Add Data → Add Data** or Click on ic ✛ ▾ the standard Bar

STEP 2: Brows to the folder

STEP 3: Select the Vector layer

STEP 4: Click **Add**.

A Vector file will have icon as:

  For Point

  For Polyline/ Line

  For Polygon

A Table will have icon

Or
Drag the layer from Arc Catalogue and drop it to the Data Frame or Table of Content

2.3.2 Raster file loading in Arc GIS

STEP 1: Go to **File → Add Data → Add Data** or Click on ic ✛ ▾ n the standard Bar

STEP 2: Brows to the folder

STEP 3: Select the Raster layer

STEP 4: Click **Add**.

*A **Raster** file will have icon as:*

*A **TIN** will have icon as:*

By default red Channel will be assigned to Layer 1, green to Layer 2, and Blue to Layer 3

STEP 5: In order to change the band combinations of loaded raster file left click on individual Channel boxes to assign proper Layers to get standard FCC

6. Geo-referencing

Introduction

There is a great deal of geographic data available in formats that can not be immediately integrated with other GIS data. In order to use these types of data in GIS it is necessary to align it with existing geographically referenced data, this process is also called georeferencing. Georeferencing is also a necessary step in the digitizing process. Georeferencing is the process of scaling, rotating, translating and deskewing the image to match a particular size and position. Raster data is commonly obtained by scanning maps or collecting aerial photographs and satellite images. Scanned map datasets don't normally contain spatial reference information. Thus, to use scanned raster datasets in conjunction with other spatial data, user needs to align, or georeference, to a map coordinate system. A map coordinate system is defined using a map projection (a method by which the curved surface of the earth is portrayed on a flat surface). The georeferencing process involves identifying the image coordinates of several clearly discernible points, called ground control points (or GCPs), in the distorted image or a scanned toposheet, and matching them to their true positions in ground coordinates. The true ground coordinates are typically measured from a map, either in paper or digital format. Once several well-distributed GCP pairs have been identified, the coordinate information is processed by the computer to determine the proper transformation equations to apply to the original (row and column) image coordinates to map them into their new ground coordinates. Prior to georeference, user need to carefully examine the digital image to be georeferenced. User should determine in advance what points will be control points. It is important to record these points before georeferencing, which help the user in the georeferencing process.

Datum

Datum is a reference surface used to generate coordinates (i.e. latitude and longitude). Datums have 'nothing to do with' distortions - they have 'everything to do with' location, specifically the coordinate (lat/long) of a feature on the earth's surface. Datum is the choice of ellipsoid to use. As the graticule (latitude/longitude) varies from ellipsoid to ellipsoid, changing datums will change the coordinates of a feature on the earth's surface. Therefore, when features from different maps are combined onto one map (i.e. for an analysis in GIS) it is essential that the maps be of the same datum.

Projections

Projection is the means by which features on the earth's surface are transferred from the globe onto a flat sheet of paper. It is actually accomplished with mathematical transformations. Four properties that can be distorted include area, shape, distance and/or direction. Simply put, a map is not a perfect representation of the features on the surface of the world - the globe is far more accurate. Projections are designed such that one or more of these properties suffers little (no) distortion. Two important projection types are *conformal* and *equal-area*. Each of these two projection types is suited to particular uses. Conformal maintains shape and direction and is well suited to navigation. Equal-area is obviously well suited to uses where area calculations (i.e. absolute size of polygons) or density values (i.e. number of trees per hectare) are important.

Georeferencing

Georeferencing process consists of two basic operations. Identification of geometric relationship between the input pixel coordinates and the associated map coordinates of this same point. The process is called spatial interpolation. The process requires that polynomial equations be fit to the GCP data using least squares criteria to model the corrections directly in the image domain without explicitly identifying the source of the distortion. Depending on the distortion in the imagery, the number of GCPs used, and the degree of topographic relief displacement in the area, higher order polynomial equations may be required. The number of non-correlated control points required for this method must be 3 for a first order, 6 for a second order, and 10 for a third order. The first-order polynomial transformation is commonly used to georeference an image. It is important to determine how well the coefficients derived from the polynomial equations of the initial GCPs account for the geometric distortion in the input image. The most often method involved is calculating RMS error for each GCP. If the RMSE of certain GCP is greater than the user identified threshold, that particular GCP may be deleted and new one may be identified.

The following points are important in selection of control points

- They should be easy to confirm as representing the same geographic location (road intersection, landmark, etc.).
- They should be spread across the image to be registered, one suggestion is to select a control point near each of the corners of the image and few throughout the interior will often work well.
- Good overlap between the two dataset is also important.
- Make sure that clicking should be as close as possible to the same geographic location with zooming the image.

4.1 Geo-referencing in ArcGIS
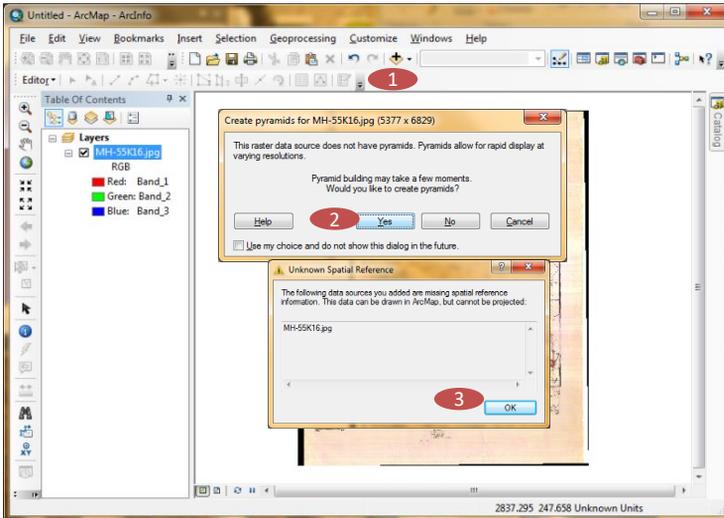
STEP 1: Add the unprojected image

STEP 2: Click **Yes** if prompted for creating Pyramid Layer

STEP 3: Click **OK** in the warning showing **Unknown Spatial Reference**
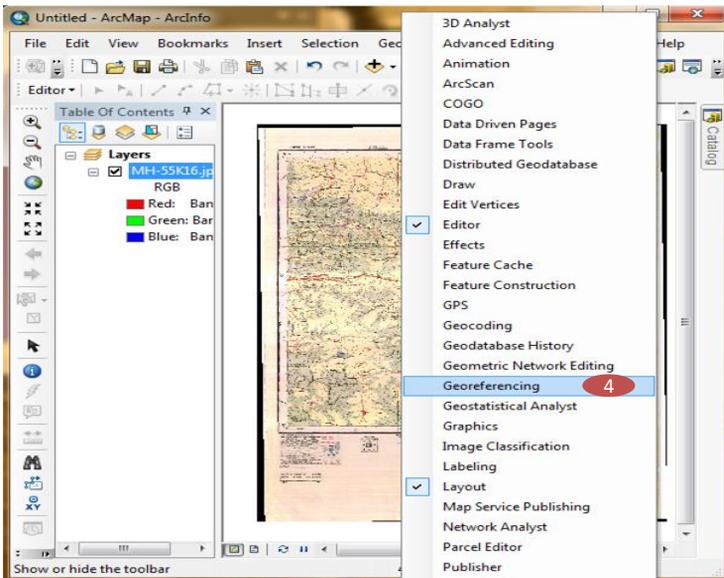
STEP 4: Right click on the Standard bar and check **Georeferencing** tool bar
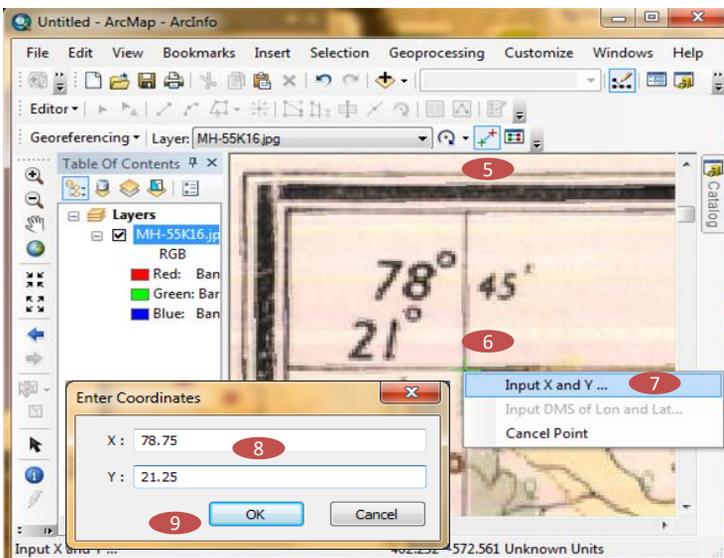
STEP 5: Select the **Add Control Points** Tool under Georeferencing Tool Bar

STEP 6: Click on the point of known Lat/long on the un projected image as a Ground Control Points (GCP)

STEP 7: **Right click** on the same location and select **Add X and Y…**

STEP 8: Enter **Longitude in the X** field and **Latitude in Y** field in **Degree Decimal unit**

STEP 9: Click **OK**.

Add similarly other 3 points (minimum 4 GCPs are required) preferably from corners of the image (if it is toposheet 4 corners can be selected as GCPs)
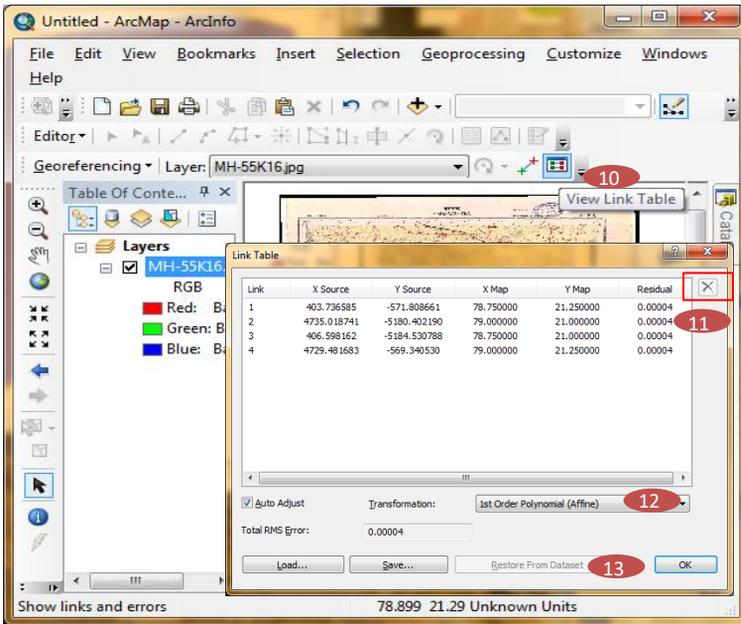
STEP 10: Click on **View Link Table** tool to view the points and **Residual Error**

STEP 11: GCPs showing high **Residual Errors** may be deleted to reduce RMS Error. If required more GCPs can be collected to minimize the RMS Error

STEP 12: Select the Transformation as **1st Degree polynomial (Affine)** as default

STEP 13: Click **OK**
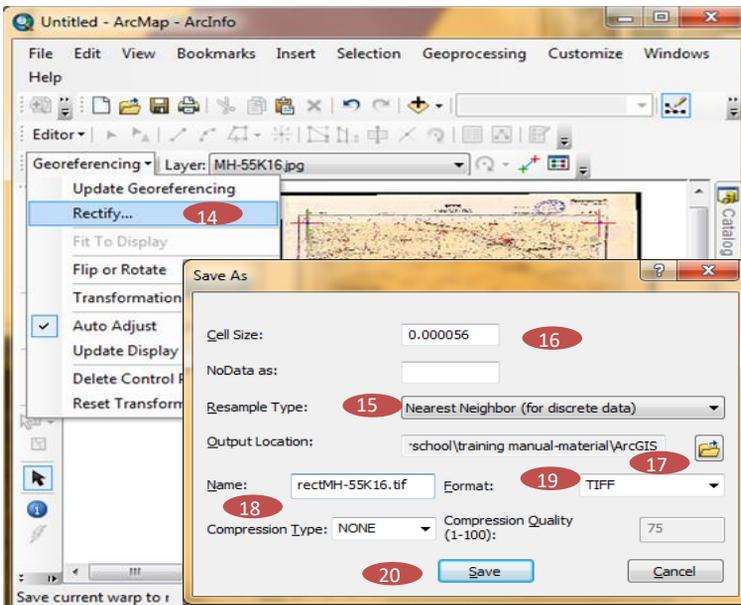
STEP 14: Go to **Georeferncing tool** and select **Rectify**

STEP 15: Select **Resampling Type** as **Nearest Neighbor** which is default.

STEP 16: Keep the **Cell Size** Default

STEP 17: Define **Output Location**

STEP 18: Define **Output Image Name**

STEP 19: Define **Output Image Format**

STEP 20: Click **Save**
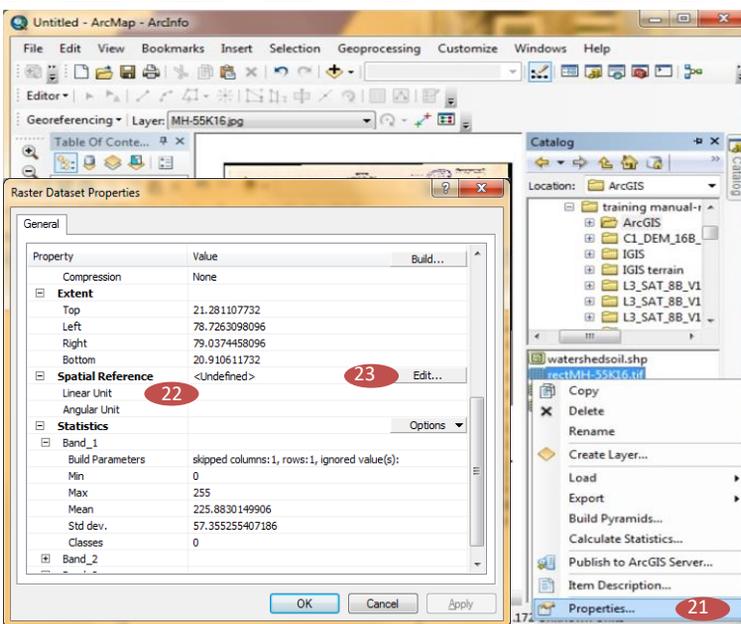
STEP 21: Right click on the rectified image in the **Arc Catalogue** and select **Properties**

STEP 22: Scroll down to **Spatial Reference** which is **Undefined** by default

STEP 23: Click **Edit**
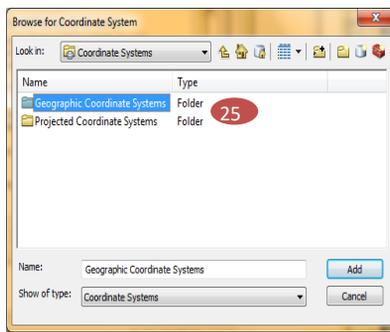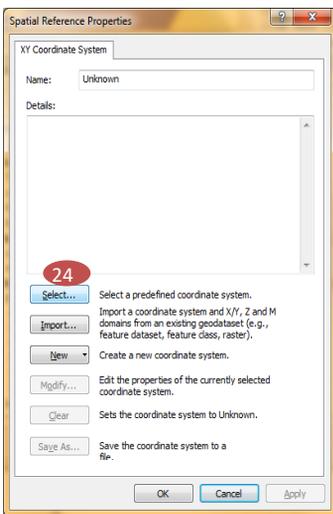
STEP 24: Click **Select**

STEP 25: Open **Geographic Coordinate System**
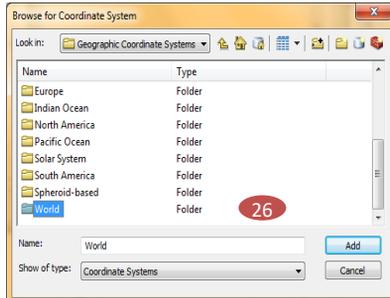
STEP 26: Open **World** Projection System

STEP 27: Select **WGS84.prj**

STEP 28: Click **Add**

STEP 29: Check the **Spatial Reference Name** and **Details**

STEP 30: Click **OK**
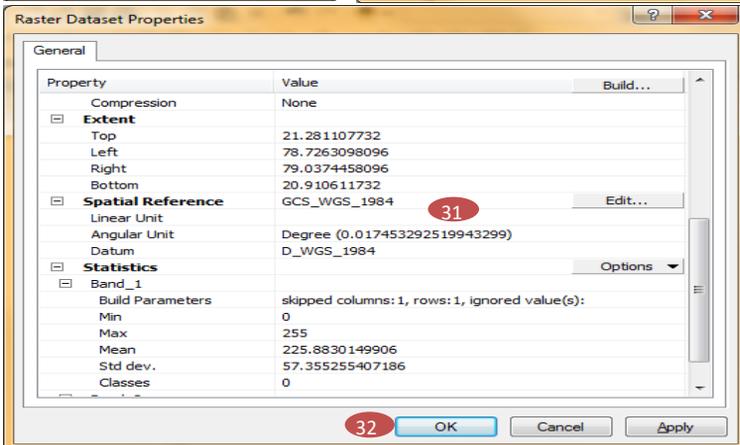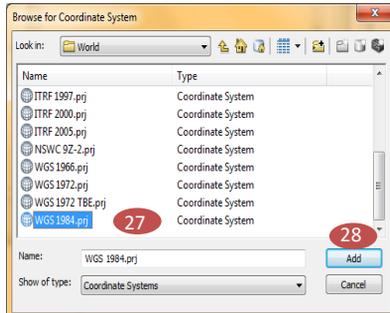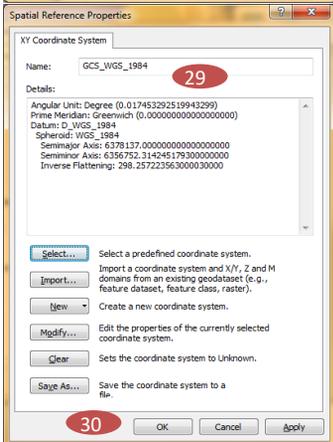
STEP 31: Check the **Spatial Reference** in **Raster Dataset Properties**

STEP 32: Click **OK**
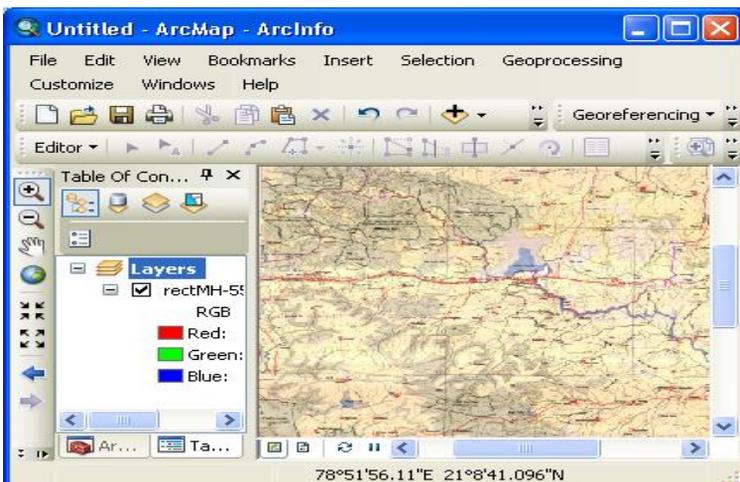
7. Image to Image Registration

Introduction

Raster data is commonly obtained by scanning maps, collecting aerial photographs, and obtaining raw satellite images. Scanned map datasets don't usually contain information (either embedded in the tile or as a separate file) about where the area fits on the surface of the earth. The locational information delivered with raw aerial photographs and satellite imagery is often inadequate to perform analysis or display in proper alignment with other data. Thus, to use these types of raster data in conjunction with spatial data, user often need to align it, or georeference it, to a map coordinate system.

Image to image registration is one of the basic image processing operations in remote sensing. Image registration is the process of matching two images so that corresponding coordinate points in the two images correspond to the same physical region of the scene being imaged. It is a classical problem in several image processing applications where it is necessary to match two or more images of the same scene. The registration process is usually carried out in three steps. The first step consists of selection of features on the images. Next each feature in one image is compared with potential corresponding features in the other image. A pair of features with similar attributes are accepted as matches and are called ground control points (GCPs). Finally the parameters of the best transformation, which models the deformation between the images are estimated using the GCPs obtained in the previous step.

To overlay overlapped images of a scene from different views and dates is frequently required in many application systems of image processing. This process is commonly called image-to-image registration. In many applications of GIS and remote sensing, image-to-image registration is frequently applied to mosaic images. A common operation in desktop photo-grammetry is the registration of new images to existing geo-coded images. Often this entails registration from a previous year's data, or the updating of a database with new overlapping imagery. All remote sensing images are subject to some form of geometric distortions, depending on the manner in which the data are acquired. These distortions may be internal geometric errors or systematic errors e.g., due to the perspective of the sensor optics, the motion of the scanning system, the terrain relief, and the curvature and rotation of the Earth or external geometric errors or nonsystematic errors e.g., due to the motion and (in) stability of the platform – the platform altitude, attitude, and velocity. Most of the commercially available remote sensor data already have much of the systematic errors removed through analysis of sensor characteristics and ephemeris data. Unless otherwise processed, however, the nonsystematic random errors remain in the image, making it non-planimetric.

Once the user has a rectified image or map, the distorted satellite image of the same geographical area may be geometrically corrected based on the rectified image. It is the translation and rotation alignment process by which two images of like geometry and of the same geographic area are positioned coincident with respect to one another so that corresponding elements of the same ground area appear in the same place on the registered images.

Image-to-image registration involves:

- ○ Selecting an appropriate planimetric base map,
- ○ Collecting ground control points (GCP's),
- ○ Determining optimum set of geometric rectification coefficients by iteratively computing the total GCP RMS error, and
- ○ Filling the output matrix using spatial and intensity interpolation resampling methods.

It is often very difficult to locate good GCPs in satellite imageries, especially in rural areas (e.g., forest, wetlands, and water bodies). The optimum method of selecting the GCPs is to have both the rectified base map and the image to be rectified on the screen at the same time. This dual display greatly simplifies GCPs selection. In case of Geomatica and IGIS there are two separate views for both the images. In case of ArcGIS, although there is only one view for registration, it simultaneously shifts the image as the GCPs are added. This makes it very easy to identify good GCPs. All softwares allow knowing the quality of GCPs, i.e., the RMS error. Identified GCPs with high RMS error may be deleted. The total RMS error should not exceed user defined threshold.

After rectification, to judge the proper overlay of the registered image, visualization tools are provided in softwares e.g., swipe, flicker, and transparency.

5.1 Image to image registration in Arc GIS
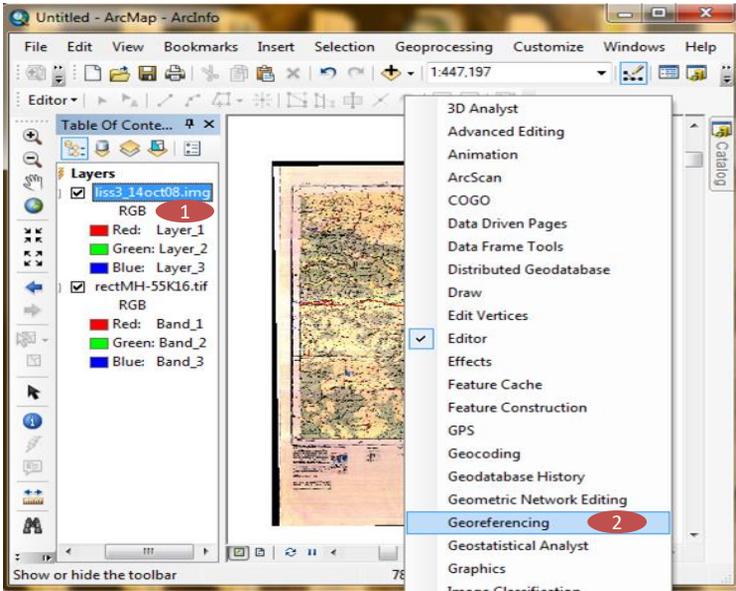
STEP 1: Add both the corrected and uncorrected images with the later one on the top preferably

STEP 2: Add **Georeferencing toolbar** by selecting **Customize** →**Toolbars** → **Georeferencing** or right clicking on Standard bar and select **Georeference**
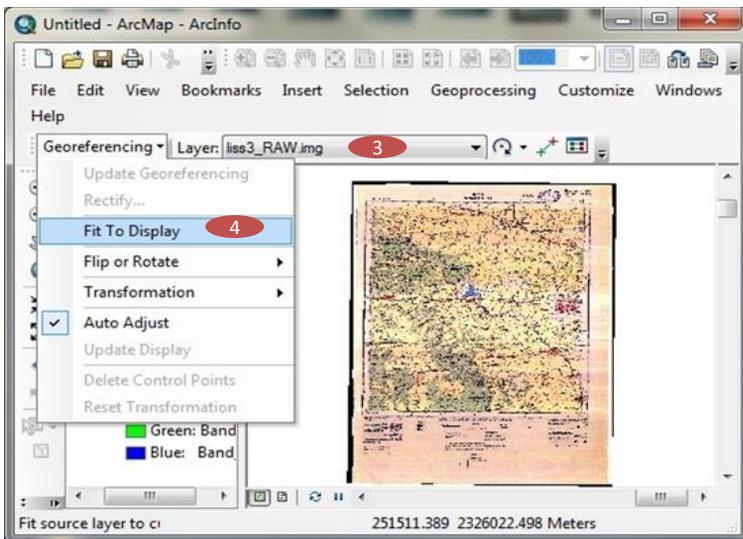
STEP 3: Select the **layer** as the **non-georeferenced** image

STEP 4: Click on **Georeferencing toolbar** and select **Fit To Display** from the drop down
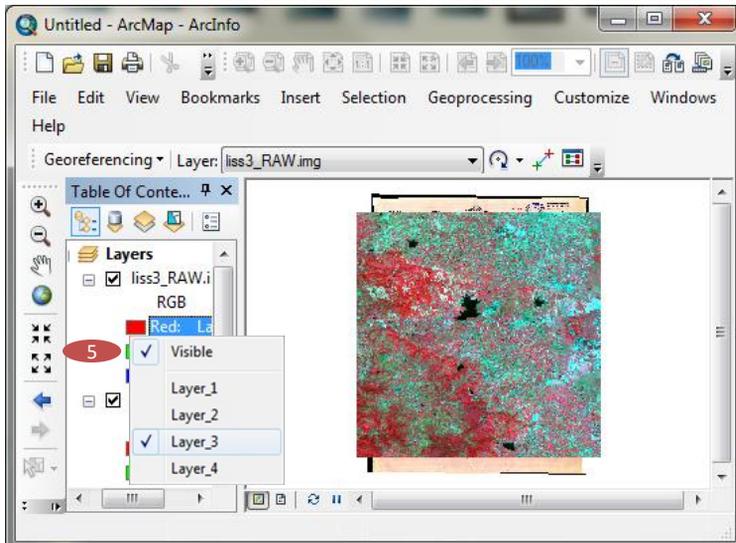
The resulting window looks like this

STEP 5: Set the band combination to view standard FCC if it is satellite data
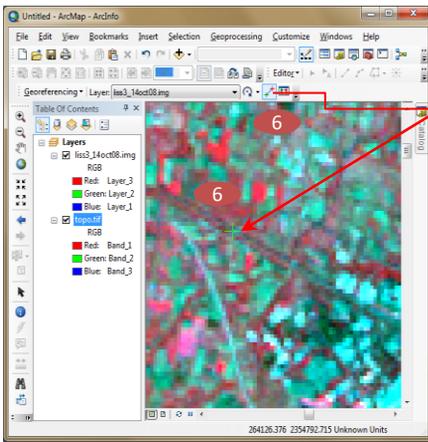
STEP 6: Select **Add Control Point** tool and click a GCP conspicuous in both images

STEP 7: Uncheck the uncorrected image and click at the corresponding point on the corrected image
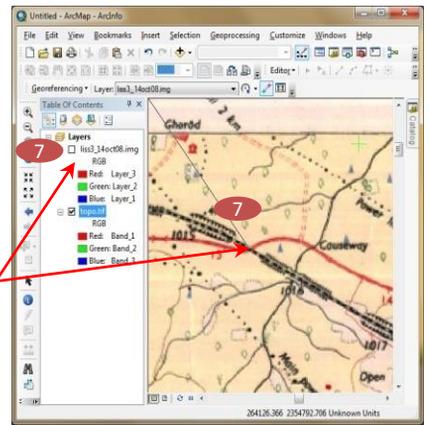
STEP 8: After adding sufficient number of GCPs Click on **View Link Tables** tool for viewing G Link Table

STEP 9: Check for **Total RMS Error**

STEP 10: Set the **Transformation** as **1st Order Polynomial (Affine)**

STEP 11: GCPs of **High RMS Error** can be deleted by selecting the particular GCP and clicking on icon

STEP 12: Click **OK**

Link	X Source	Y Source	X Map	Y Map	Residual
1	637.291898	-58.778730	78.889999	21.240466	0.00030
2	1022.671815	-489.051744	78.975918	21.144010	0.00030
3	990.388215	-801.945634	78.968730	21.073422	0.00015
4	994.449007	-776.646408	78.969748	21.079377	0.00017
5	129.086248	-1046.343034	78.775269	21.018298	0.00009
6	82.837656	-180.380142	78.764998	21.213309	0.00032
7	605.704439	-509.388634	78.882355	21.139055	0.00017
8	1084.476790	-121.360806	78.990272	21.226049	0.00048
9	52.454007	-689.595937	78.758000	21.098326	0.00023
10	929.240775	-955.028634	78.955428	21.039032	0.00038
11	1028.664467	-547.622998	78.977328	21.131109	0.00045

Transformation: 1st Order Polynomial (Affine)

Total RMS Error: 0.00030

STEP 13: Right click on the **top image** and click **Properties** to get **Layer properties** window

STEP 14: Click on **Display** tab

STEP 15: Set **Transparency** as **50 %**

STEP 16: Click **OK**
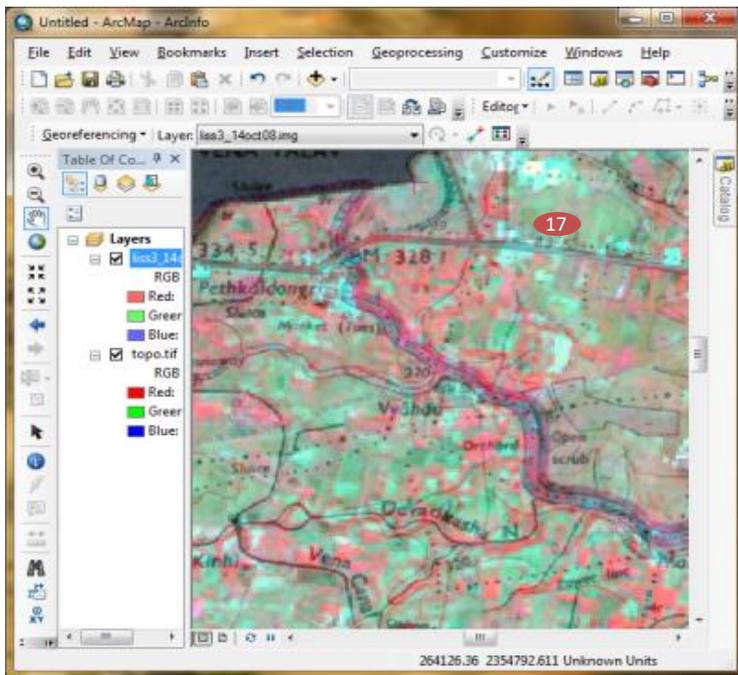
STEP 17: Confirm for proper overlay of features of uncorrected image with reference to corrected image
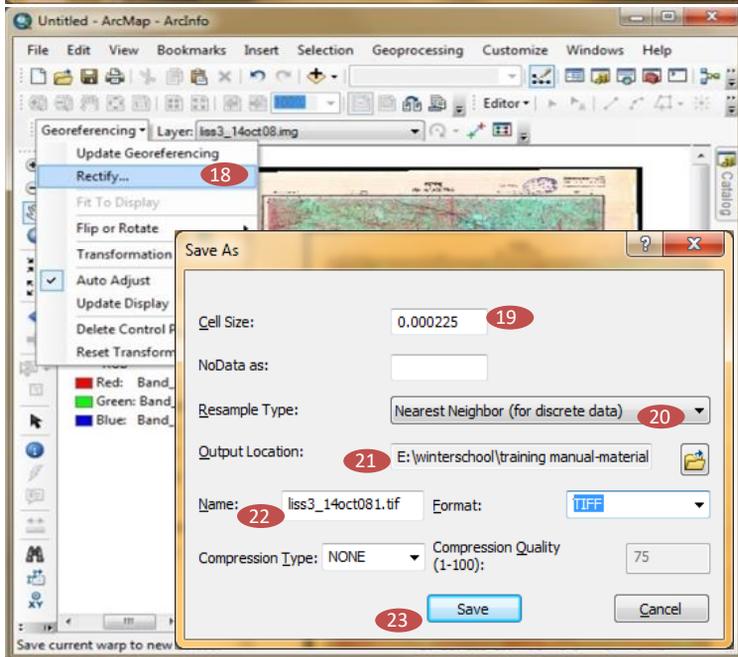
STEP 18: Select **Rectify** from **Georeferencing** drop down

STEP 19: Keep **Default** Cell Size

STEP 20: Keep **Resample Type** as **Nearest Neighbor**

STEP 21: Define **Output location**

STEP 22: Define Output **Name**

STEP 23: Click **Save**
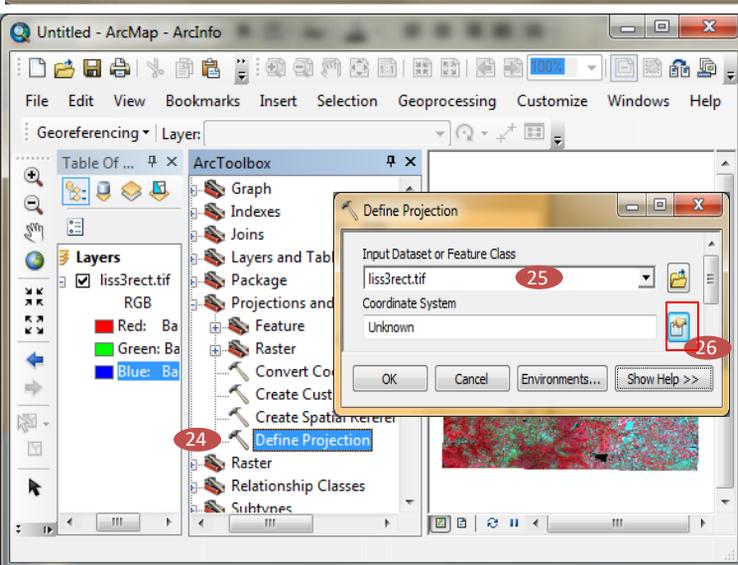
STEP 24: Go to **Arc Toolbox** → **Data Management Tool** → **Projections and Transformations** → **Define Projection**

STEP 25: Select the Rectified image as **Input Dataset**

STEP 26: Click on **Coordinate System**
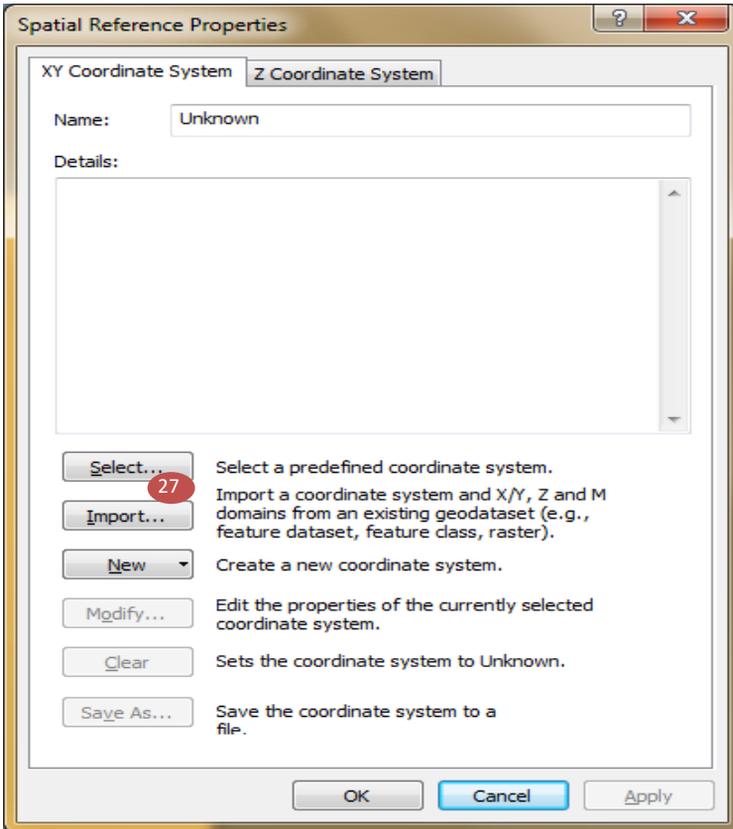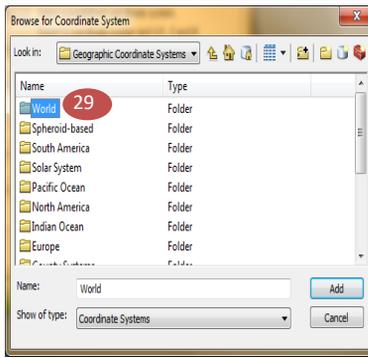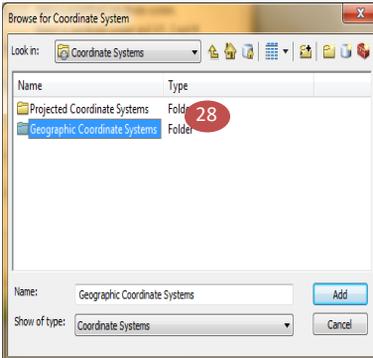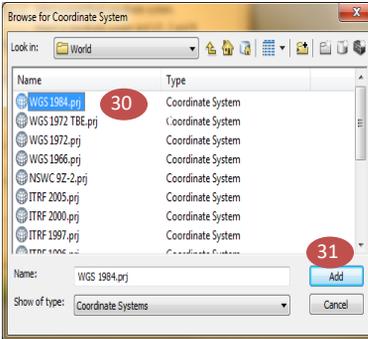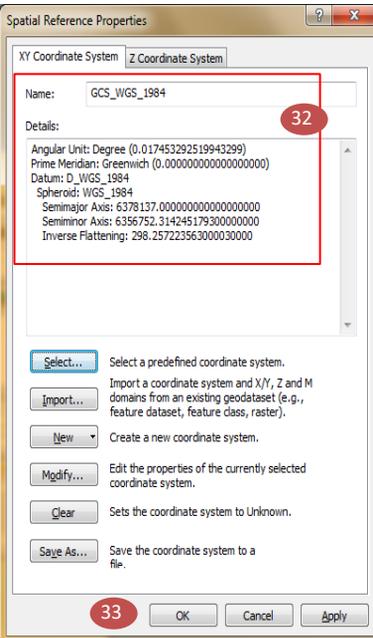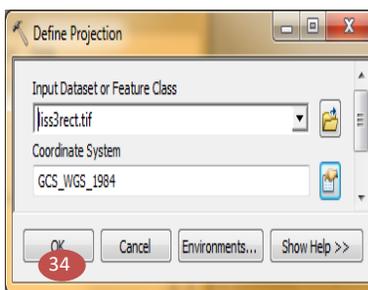
8. Data Model Transformation

Introduction

In GIS, it is common that conversion of spatial dataset from one format to another format to 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. Data may have been digitized in vector form but subsequently needed in raster form for input to some raster based multi-criteria analysis. In such cases it is necessary to convert between models. Understanding the advantages and limitations of particular methods of representing data is key not only to effective storage and functionality but also to produce reliable/accurate result.

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.

Point to raster

There are several ways to assign a point's value to each pixel.

•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.

•Record a value representing merely the presence of a point or points.

•Record the frequency of points found within a pixel.

•Record the sum of the unique identifying numbers of all vector points that fall with the area of the output pixel.

•Record the highest priority value according to the range of values encountered.

•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.

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. 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.

Polygon to raster

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.

Raster to vector conversion (vectorization)

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.

Raster to polyline

This process essentially traces the positions of any non-zero or non-null raster pixels to produce a vector polyline feature. One general requirement is that all the other pixel values should be zero or a constant value. The higher the spatial resolution of the input raster, the more precisely located the vertices will be. The procedure comprises filling, thinning, vectorizing, collapsing, and smoothing.

Raster to polygon

This is the process of vectorizing areas or regions from a raster. 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. 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.

9.1 Data transformation in ArcGIS

9.1.1 Vector to raster transformation in ArcGIS
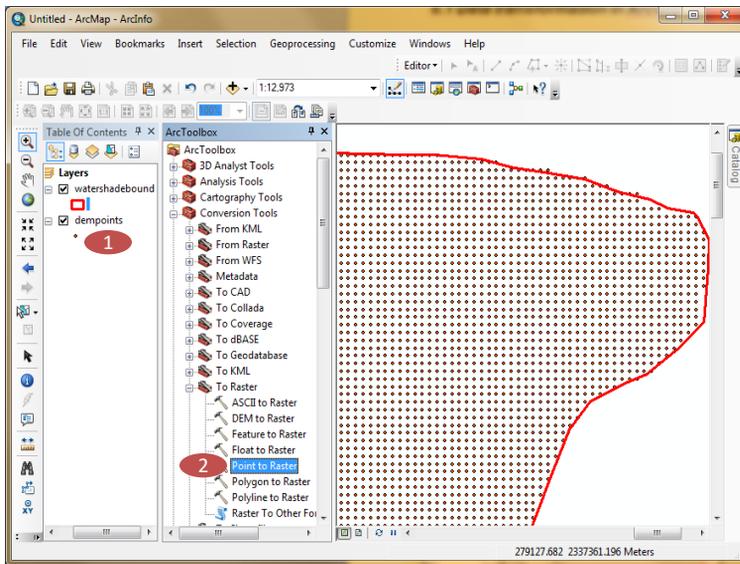
<table>
<tr><td>Points to Raster</td></tr>
</table>

STEP 1: Add the point file to ArcMAP

STEP 2: Go to **Arc tool Box → Conversion Tools → To Raster → Point to Raster**
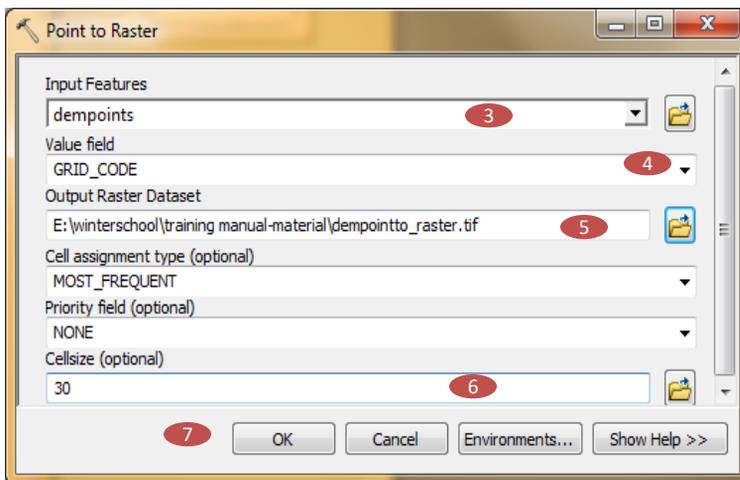
STEP 3: Add the point vector as Input Feature

STEP 4: In **value field** select the column having the attribute to be converted as raster - **GRID_CODE in this case(elevation)**

STEP 5: Define **Output Raster**

STEP 6: Define **Cell Size** keeping in view the resolution of the output image

STEP 7: Click **OK**
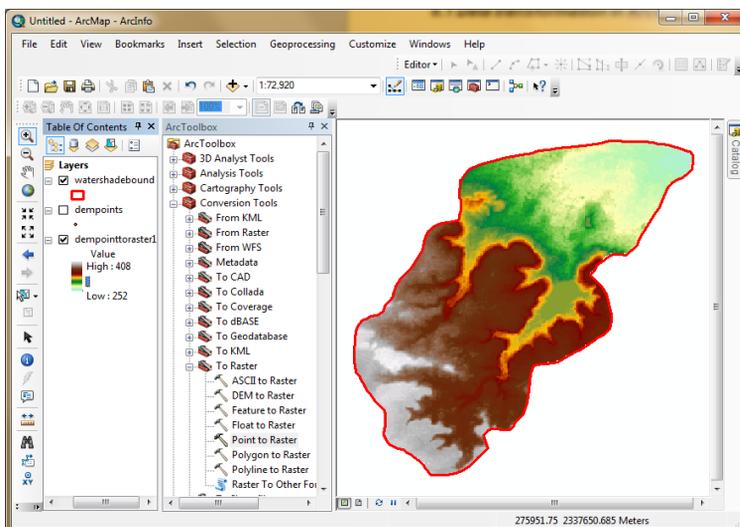
The 30 m elevation raster generated from elevation point layer through vector to Raster data transfer model in ArcGIS
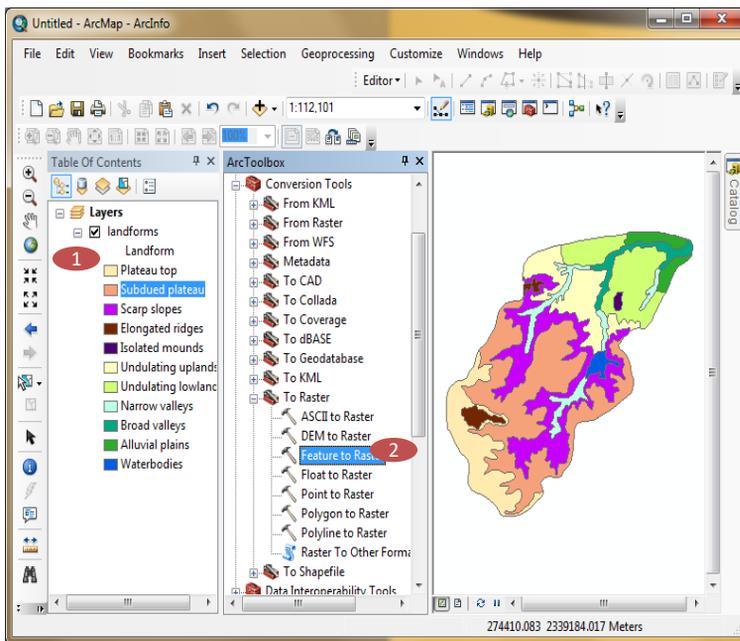
Polygon to Raster

STEP 1: Add the polygon file to ArcMAP

STEP 2: Go to **Arc Tool box →Conversion Tools → To Raster → Feature to Raster**
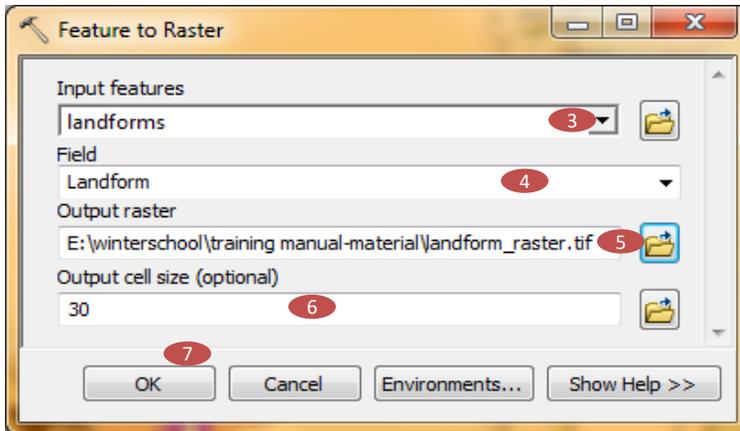
STEP 3: Add the polygon vector as **Input Feature** (in this case landform)

STEP 4: In **field** select the column having the attribute, which is to be transferred as raster (**Landform**)

STEP 5: Define **Output Raster**

STEP 6: Define **Cell Size** keeping in view the resolution of the output image

STEP 7: Click **OK**
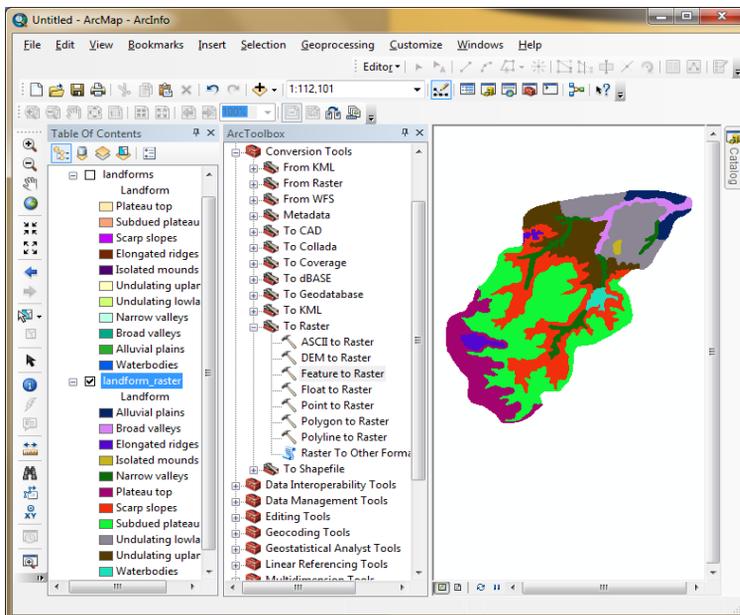
The 30 m landform raster generated from landform polygon layer through vector to Raster data transfer model in ArcGIS

9.1.2 Raster to vector transformation in ArcGIS
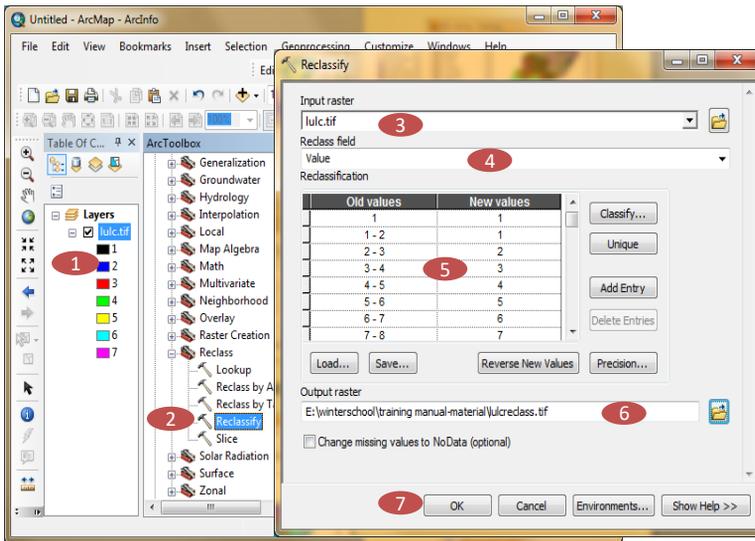
STEP 1: Add the Raster to ArcMap

STEP 2: Go to **ArcTool Box** → **Spatial Analysts Tools** → **Reclass** → **Reclassify**

STEP 3: Select the **Input Raster**

STEP 4: In **Reclass field** select the column to be reclassified

STEP 5: Define **New Values** for **Old Values**

STEP 6: Define **Output raster**

STEP 7: Click **OK**
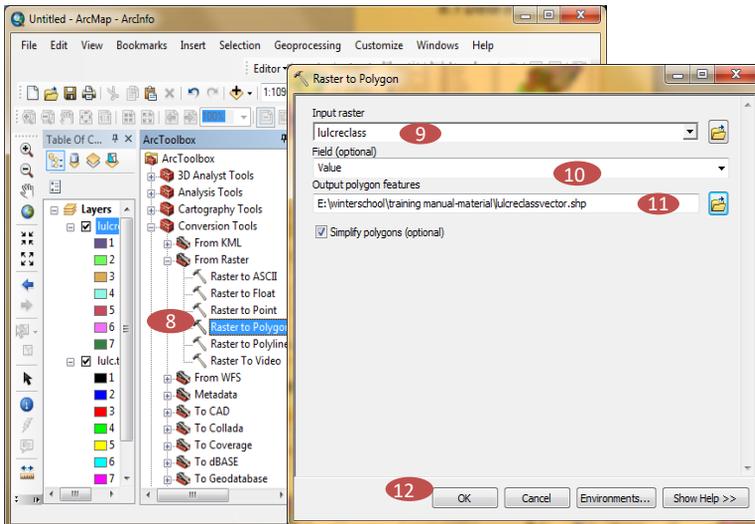
STEP 8: Go to **Arc Tool Box** → **Conversion Tools** → **From Raster** → **Raster to Polygon**

STEP 9: Add the **Reclassified Raster** as **Input raster**

STEP 10: In **value field** select the column having **New values**

STEP 11: Define **Output Polygon Feature**

STEP 12: Click **OK**
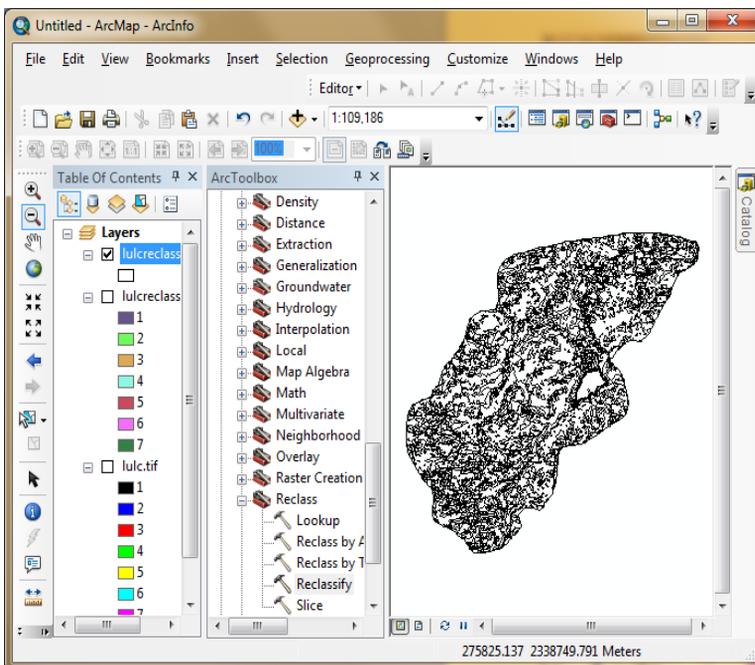
The vector file of landforms generated from landform raster layer through raster to vector data transfer model in ArcGIS

9. Projection Transformation

Introduction

Coordinate systems enable geographic datasets to use common locations for integration. A coordinate system is a reference system used to represent the locations of geographic features, imagery, and observations, such as Global Positioning System (GPS) locations, within a common geographic framework. Location is the most important reference for data in GIS. Coordinate systems (either geographic or projected) provide a framework for defining real-world locations. A coordinate system is a reference system used to represent the locations of geographic features, imagery, and observations such as GPS locations within a common geographic framework. Projection transformation is nothing but the process of converting the coordinates of a map or an image from one system to another, typically by shifting, rotating, scaling, skewing, or projecting them. This functionality concerns the transformation of data in geographic coordinates for an existing map projection to another map projection. Most GIS softwares' require that data layers must be in the same map projection for analysis. Accordingly, if data is acquired in a different projection than the other data layers it must be transformed. Every dataset has a coordinate system, which is used to integrate it with other geographic datasets within a common coordinate framework. Coordinate systems enable user to integrate datasets to perform various integrated analytical operations, such as overlaying data layers from disparate sources and coordinate systems. Typically 20 or more different map projections are supported in a GIS softwares'.

There are two common types of coordinate systems used in GIS:

•A global or spherical coordinate system such as latitude-longitude. These are often referred to as *geographic coordinate systems.*

•A projected coordinate system based on a map projection such as transverse Mercator, Albers equal area, or Robinson, all of which (along with numerous other map projection models) provide various mechanisms to project maps of the earth's spherical surface onto a two-dimensional Cartesian coordinate plane. Projected coordinate systems are sometimes referred to as *map projections*.

Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. However, all map projections representing the earth's surface as a flat map, create distortions in some aspect of distance, area, shape, or direction. The projection systems, which preserve distance, area, shape, or direction are called equidistance (e.g., Sinusoidal), equal area (e.g., Albers conic), conformal (e.g., Lambert conformal canonic), and azimuthal projections, respectively.

When transforming data between projections that are in different datums, a transformation is required. Typically in ArcMap, when data requiring a transformation is added to a data frame, ArcMap provides a list of possible transformations from which a user can select. The selected transformation is then saved as part of the map document. Within ArcGIS Image Server, a single request for an image may involve a number of different projections: the projection of the client request, the projection that is used to define the service, and the one or more projections of the original raster datasets. It is not possible for ArcGIS Image Server to request the exact transformation method that should be used for each of the reprojections from the client. ArcGIS Image Server resolves this by using a table of defined transformations, and when a transformation is required, the table is searched for the first matching transformation. In most cases, the default transformation is most appropriate, but there may be cases where a specific transformation is to be used.

Geographic (datum) transformations

If two datasets are not referenced to the same geographic coordinate system, user may need to perform a geographic (datum) transformation. This is a well-defined mathematical method to convert coordinates between two geographic coordinate systems. As with the coordinate systems, there are several hundred predefined geographic transformations that user can access. It is very important to correctly use a geographic transformation if it is required. When neglected, coordinates can be in the wrong location by up to a few hundred meters. Sometimes no transformation exists, or user has to use a third GCS like the World Geodetic System 1984 (WGS84) and combine two transformations.

Projected coordinate systems

A projected coordinate system (PCS) is defined on a flat, two-dimensional surface. Unlike a GCS, a PCS has constant lengths, angles, and areas across the two dimensions. A PCS is always based on a GCS that is based on a sphere or spheroid. In addition to the GCS, a PCS includes a map projection, a set of projection parameters that customize the map projection for a particular location, and a linear unit of measure.

10.1.1 Projection transformation of vector in ArcGIS
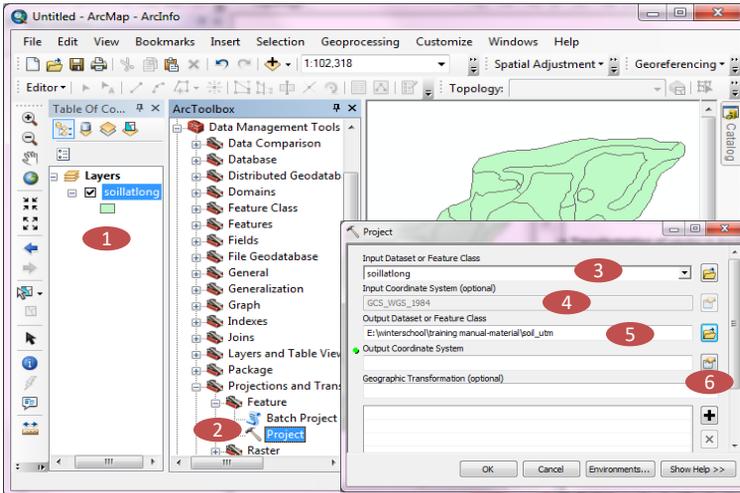
STEP 1: Add the vector to be reprojected

STEP 2: Go to **Arc Tool box → data Management Tool → Projections and Transformations → Feature →Project**

STEP 3: Select **Input Feature Class**

STEP 4: Check the **Input Coordinate System**

STEP 5: Define **Output Feature Class Name**

STEP 6: Click on **Output Coordinate System**
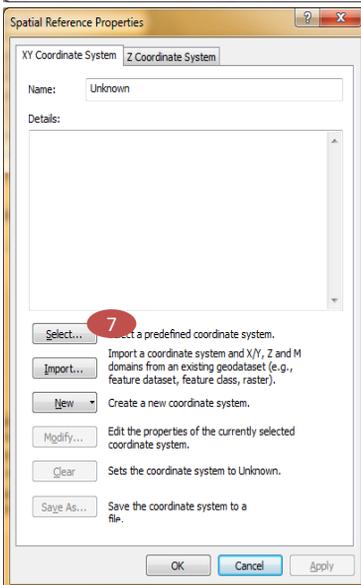
STEP 7: Under **XY Coordinate tab** click on **Select.** Depending upon the desired output projection the projection can be selected

STEP 8: To convert to UTM projection system double click on **Projected Coordinate System** and select **UTM → WGS 1984 → Northern Hemisphere → WGS 1984 UTM Zone 44N** as followed in this case.

STEP 9: Click **Add**
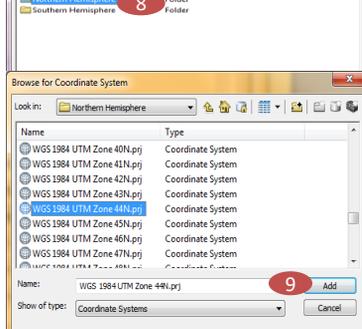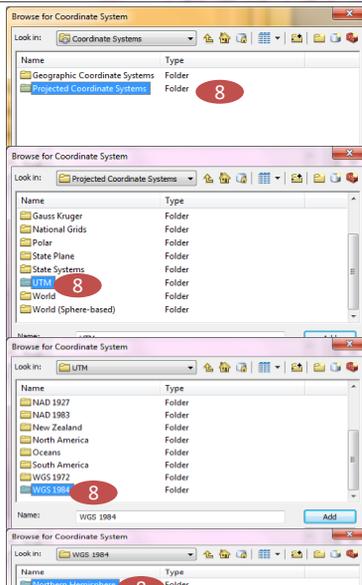
STEP 10: Check the **Coordinate System Name** and **Details**

STEP 11: Click **OK**

STEP 12: Click **OK**
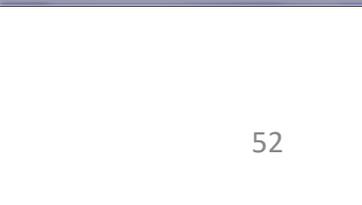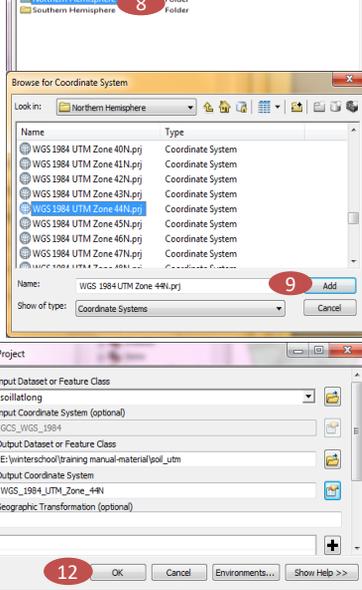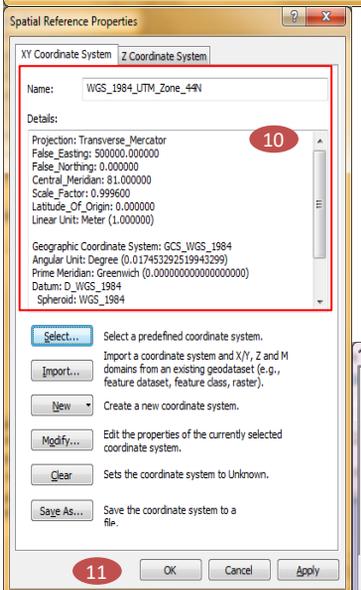
10.1.2 Projection transformation of raster in ArcGIS
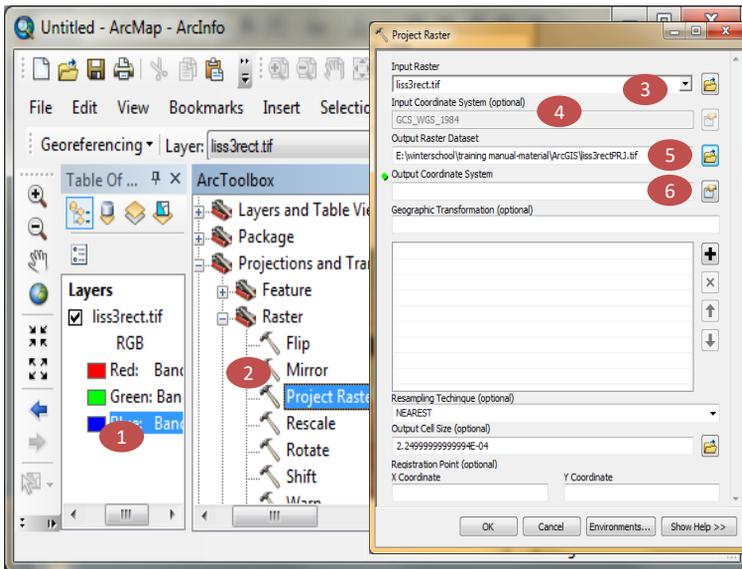
STEP 1: Add the image to be reprojected

STEP 2: Go to **Arc Tool box → data Management Tool → Projections and Transformations → Raster →Project Raster**

STEP 3: Select **Input Raster**

STEP 4: Check the **Input Coordinate System**

STEP 5: Define **Output Image Name**

STEP 6: Click on **Output Coordinate System**
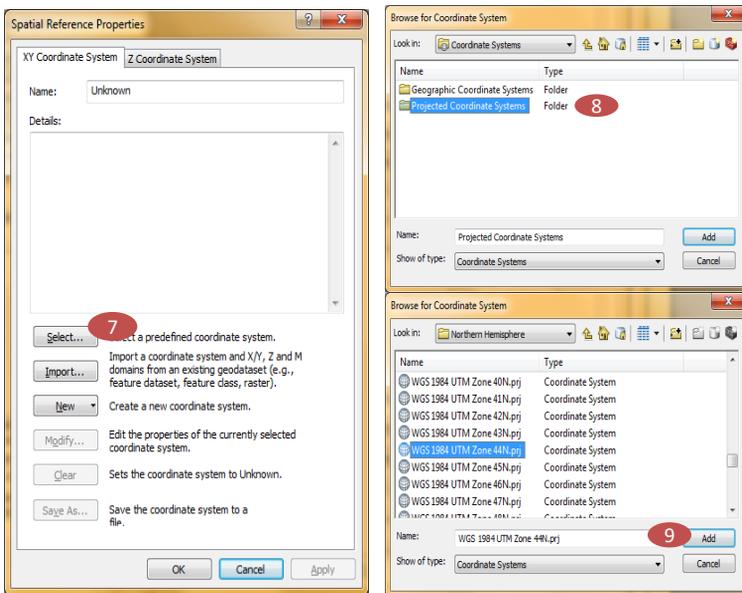
STEP 7: Under **XY Coordinate tab** click on **Select.** Depending upon the desired output projection the projection can be selected

STEP 8: To convert to UTM projection system double click on **Projected Coordinate System** and select **UTM → WGS 1984 → Northern Hemisphere → WGS 1984 UTM Zone 44N** as followed in this case.

STEP 9: Click **Add**
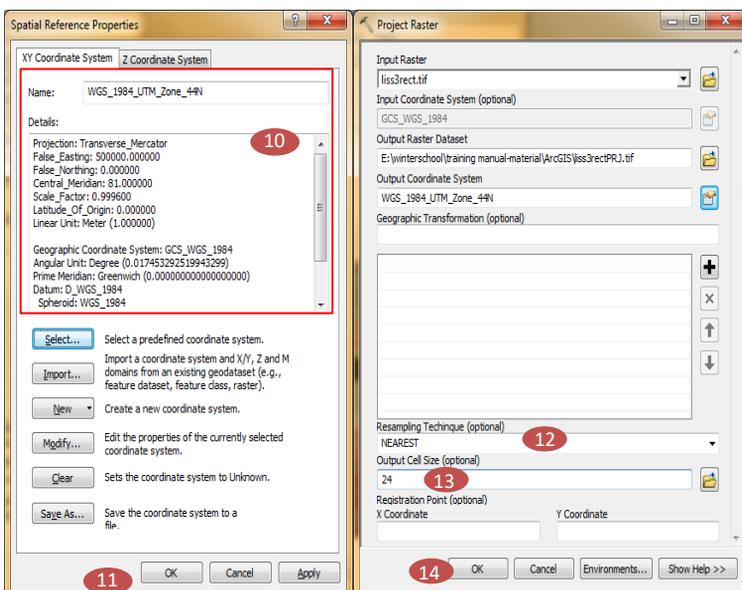
STEP 10: Check the **Coordinate System name** and **Details**

STEP 11: Click **OK**

STEP 12: Select **Resampling technique** as **Nearest**

STEP 13: Assign **Cell Size** as per the desired output resolution (**24 m** in this case)

STEP 14: Click **OK**

10. Spatial Database Management

Introduction

The handling of spatial data usually involves the process of data acquisition, storage, analysis and output. Creation of spatial database in GIS domain has become a very effective tool to aid and facilitate management decision-making. In recent times, GIS is being widely used as spatial analysis tool for effective and efficient means of data generation, management, analysis and display. The true value of GIS can only be realized if the proper tools to collect spatial data and integrate them with attribute data are available. Data used in GIS often come from many types, and are stored in different ways. Remote-sensing satellite imagery and GPS are promising data input sources for GIS. Before generation of any spatial data one has to understand the available data types, data analysis procedures and their capabilities in GIS to get realistic out puts. GIS provides tools and methods for the integration of different data into a format to be compared and analyzed.

Spatial data models/structures

Geo-relational data model

The geo relational approach involves abstracting geographic information into a series of independent layers or coverages, each representing a selected set of closely associated geographic features (e.g., roads, land use, river, settlement, etc). Each layer has the theme of a geographic feature and the database is organized in the thematic layers. With this approach users can combine simple feature sets representing complex relationships in the real world.

Topological data structure

Topology is the spatial relationship between connecting and adjacent coverage features. Topological relationships are built from simple elements into complex elements: points, arcs and areas. Topological data structure, in fact, adds intelligence to the GIS database.

Attribute data management

All data within a GIS are stored within databases. A database is a collection of information about things and their relationships to each other. The principle characteristics of a DBMS are: - Centralized control over the database is possible, allowing for better quality management and operator-defined access to parts of the database; Data can be shared effectively by different applications; The access to the data is much easier, due to the use of a user-interface and the user-views. Data redundancy can be avoided as much as possible.

Relational database

The relational data model is conceived as a series of tables, with no hierarchy nor any predefined relations. The relation between the various tables should be made by the user. This is done by identifying a common field in two tables, which is assigned as the flexibility than in the other two data models. However, accessing the database is slower than with the other two models. Due to its greater flexibility, the relational data model is used by nearly all GIS systems.

Tables

A table is a database component that contains a series of rows and columns, where each row, or record, represents a geographic feature-such as a parcel, power pole, highway, or lake-and each column, or field, describes a particular attribute of the feature-such as length, depth, or cost. Tables are stored in a database-for example, INFO, Access, dBASE, FoxPro, Oracle, or SQL Server.

Working with attributes

From a table, one can identify features with particular attributes and select them on the map. Over time, user might also update the attributes to reflect changes to geographic features, for example, a new subdivision extends your parcel database, or the construction of a dam alters a river network.

Joining data

Tables can also store information related to features such as soil resource inventories, land use/land cover , rainfall etc. By joining this information user can identify new patterns and relationships that were not readily apparent.

Database query

The ability to selectively retrieve information from GIS is an important facility. Basically there are two types of query most general GIS allow: viz., Q*uery by attribute and Query by geometry.* The attribute database, in general, is stored in a table with a unique code linked to the geometric data. This database can be searched with specific characteristics. However, more complex queries can be made with the help of SQL. GIS can carry out a number of geometric queries. There are five forms of primitive geometric query: viz., *Query by point, Query by rectangle, Query by circle, Query by line and Query by polygon.* A more complex query still is one that uses both geometric and attributes search criteria together.

11.1 Design of spatial database in ArcGIS
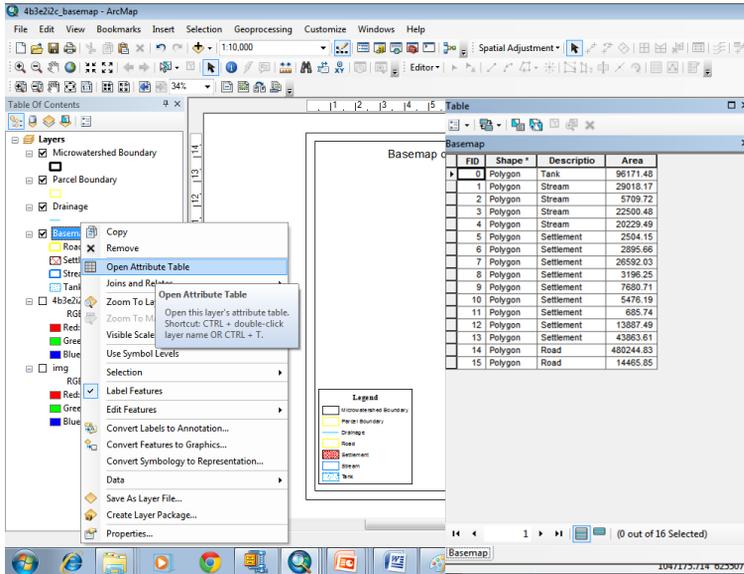
Adding Fields

STEP 1: Right click on the layer and select **Open Attribute Table**
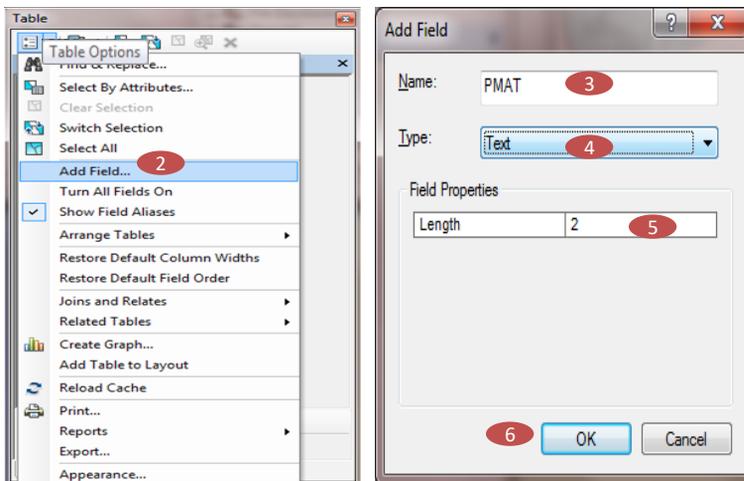
STEP 2: Go to **Table Options → Add Field**

STEP 3: Enter **Name** of the field

STEP 4: Select **Type** of the field (**Text in this case)**

STEP 5: Define **Length** of the filed

STEP 6: Click **OK**
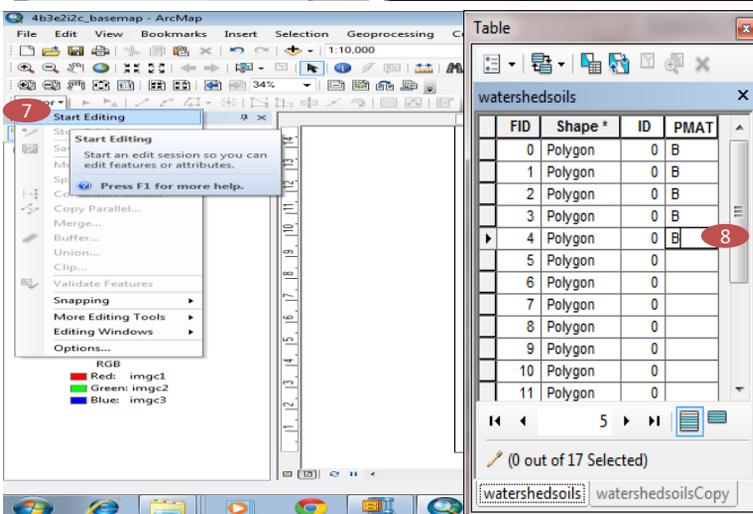
STEP 7: Start Editing by selecting **Editor → Start Editing**

STEP 8: Enter the values in the field
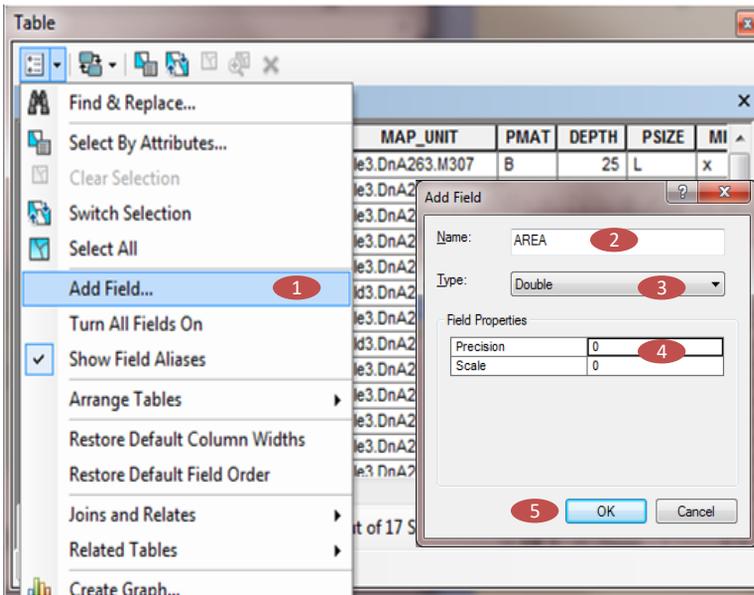
Calculating Geometry (AREA and PERIMETER)

STEP 1: Go to **Table Options** → **Add Field**

STEP 2: Enter **Name** of the field as Area

STEP 3: Select **Type** of the field as **Float** or **Double**

STEP 4: Define **Precision and scale**

STEP 5: Click **OK**
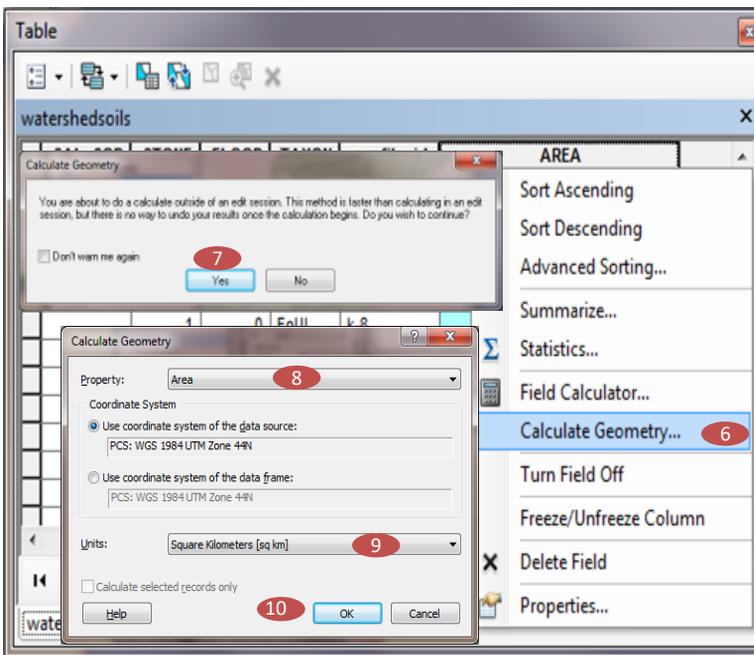
STEP 6: Right click on the field label and select **Calculate geometry**

STEP 7: Click **Yes** if prompted whether user wish to calculate geometry outside edit session

STEP 8: select **Area** in **Property**

STEP 9: Select **Unit**

STEP 10: Click **OK**
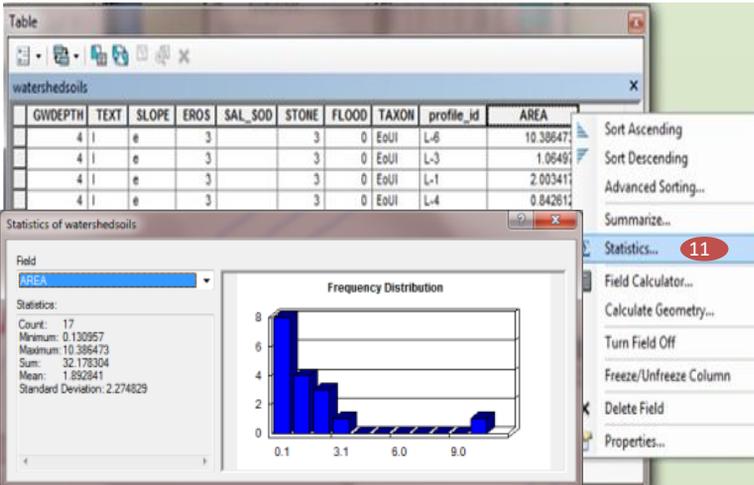
STEP 11: Right click on the field label and select **Statistics** to see basic descriptive statistics of area

Perimeter can also be generated in the same way

Or,

STEP 1: Go to **Arc Tool Box** → **Data Management Tool** → **Fields** → **Delete Fields**
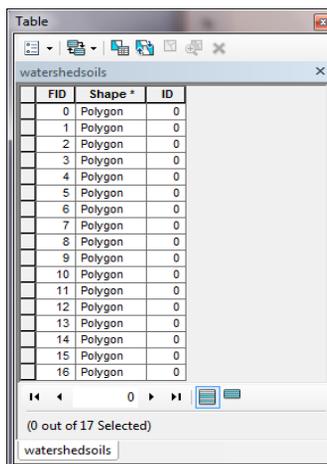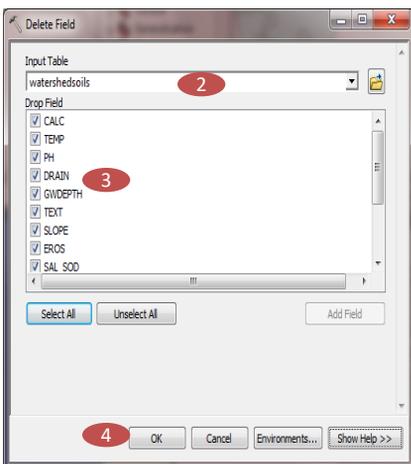
STEP 2: Select the layer in **Input Table**

STEP 3: Select the fields to be deleted

STEP 4: Click **OK**
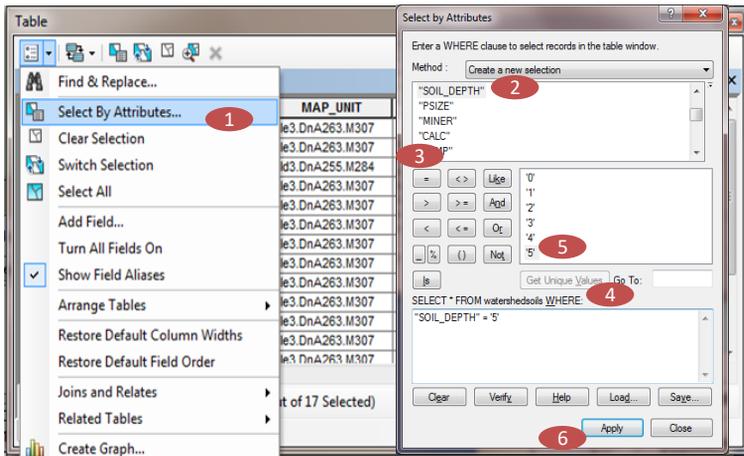
Spatial Data Query-Select by Attribute

Condition: Select the soil polygons with **soil depth** as **deep (code:5)**

STEP 1: Go to **Table Options** → **Select By Attributes**

STEP 2: Double click on the field user want to query upon to add it in the query (Soil Depth in this case)

STEP 3: Click on **Get Unique Value** to see the available unique values for that field

STEP 4: Click on suitable operator **(= in this case)**

STEP 5: Double click on the value user want to select (**5** in this case)

STEP 6: Click **Apply** for the desired polygons to be selected

---
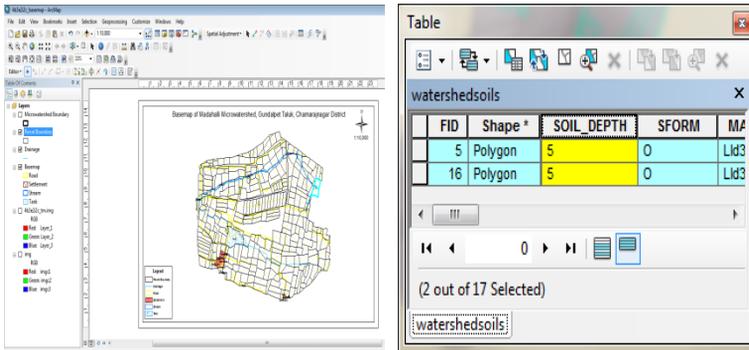
The selected polygons with attribution value 5 are shown in the figure

---

Use of AND operator:

In order to query the database based on two conditions AND operator can be used

Condition: Select the soil polygons with **soil depth** as **Very Shallow (code:1)** and **Slope** as **Moderately Sloping (Code: d)**

STEP 7: Add the 2 queries by **AND**

STEP 8: Click **Apply**
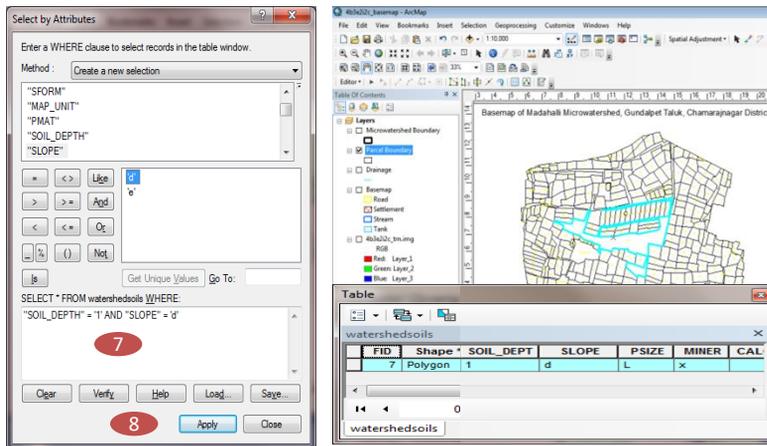
The selected polygons with two conditions are shown in the figure

---

Use of OR operator:

In order to select both the attributes OR operator can be used

Condition: Select the soil polygons with **soil depth** as **Moderately Deep (code:4)** and **Deep (Code: 5)**

STEP 9: Add the 2 queries by **OR**

STEP 10: Click **Apply**
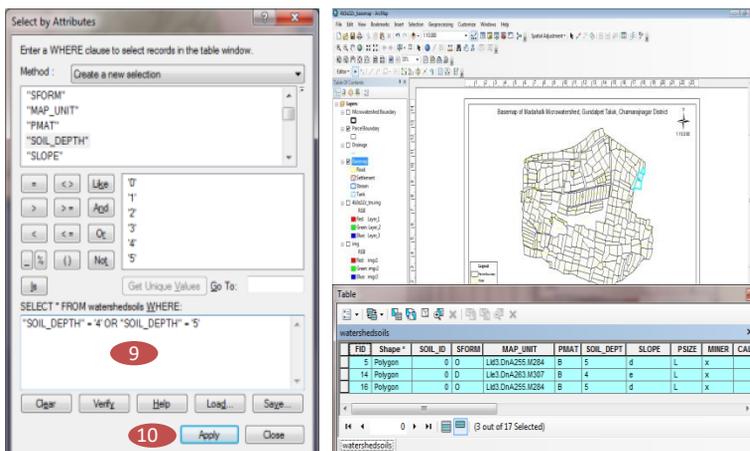
The selected polygons with desired attributes are shown in the figure

11. Mosaic Generation

Introduction

Each scene of a remote sensing image, obtained from one orbit, is performed on a subset of the total area. The mosaicking process is employed to obtain the whole area. The mosaicking process can be broken down into a number of logical steps: pre-processing, selection of tie points, renavigation, resampling to output space and compositing and mosaicing with radiometric blending. It is possible to mosaic unrecified individual frames or flight lines of remotely sensed data. However, it is more common to mosaic multiple images that have already been rectified to a standard map projection and datum. To be mosaicked the pixel size of the individual images should be same. One of the images to be mosaicked is designated as the base image. The adjacent images will normally overlap a certain amount. Interest points are found and paired between overlapping images to generate potential tie point conjugates. A representative geographic area in the overlapping region in the base image is contrast stretched according to the user specification. The histogram of the base image is then applied to the adjacent images using a histogram matching algorithm. This causes the two images to have approximately same grayscale characteristics.

Mosaic generation

A mosaic dataset allows user to store, manage, view, and query small to vast collections of raster and image data. It is a data model within the geodatabase used to manage a collection of raster datasets (images) stored as a catalog and viewed as a mosaicked image. Some software allows the users to select histogram of which image to be applied in the output or which should remain unchanged. It is possible that pixel values in one scene are dominating the pixel values in the overlapping scene. This can result in noticeable seams in the final output. Therefore, it is common to blend the seams between mosaicked images using feathering. Some digital image processing systems allow the user to specify a feathering buffer distance (e.g., 100 pixels) wherein 0% of the base image is used in the blending at the edge and 100% of second image is used to make the output image. At the specified distance (e.g., 100 pixels) in from the edge, 100% of the base image is used to make the output image and 0% of the second image is used. At 50 pixels in from the edge, 50% of each image is used to make the output. Sometimes analysts prefer to use a linear feature such as a river or road to subdue the edge between adjacent mosaicked images. In this case, the user has to identify a polyline in the image and then to specify a buffer distance where the feathering will take place. ArcGIS provides the option for appending the second image to the first or to make a separate mosaicked image. Other two softwares' create separate mosaicked files.

Mosaic datasets are designed to handle data with varying resolutions - spectral, spatial, temporal, and radiometric. The raster types and functions in a mosaic dataset play a strong role in how all this data is handled and displayed. Additionally, the mosaic dataset is particularly aware of the spatial and temporal information as attributes of the raster data. Based on the cell sizes, the mosaic dataset will display the imagery at the most appropriate scales. With some additional display control properties, called mosaic methods, a user can control the temporal information, allowing them to view the images for the dates they require. User can use mosaic methods in a mosaic dataset to control what raster data is presented each time a mosaic (from the mosaic dataset) is displayed. By default, the mosaic is generated by displaying the raster dataset that is the closest to the centre of the image. Another mosaic method user can define a query based on attributes such as acquisition date or cloud cover. These mosaic methods and querying capabilities allow users to access every raster dataset within the mosaic dataset, even when there is overlap.

There is no pixel data loss or metadata loss when using mosaic datasets, as the source pixels are never altered or converted, and the files are never moved; therefore, any metadata files remain in their location. Because the mosaic dataset does not alter the source data or its location, the pixel values are not altered. Additionally, the mosaicking performed by the mosaic dataset occurs dynamically when the mosaic dataset accessed. Users have access to the mosaicked image as well as the source data; therefore, there is no data loss occurring for overlapping datasets. Overviews, which are like raster pyramids, can be generated for a mosaic dataset. Overviews are reduced resolution datasets that are generated to improve the speed at which the mosaic is displayed. User can allow the default overviews to be generated over the entire mosaic dataset. Alternatively, user can control how they're created by defining the down sampling ratio, the extent, a specific spatial resolution, and so on.

Mosiac generation has been widely used in change detection, image fusion and other related areas. In order to integrate different kinds of sensor data and different temporal data, image registration is an indispensable pre-processing tool in integrating multi-source and multi-temporal images. In change detection process, the image registration accuracy directly influences the accuracy of change detection result.

8.1 Mosiac generation in ArcGIS
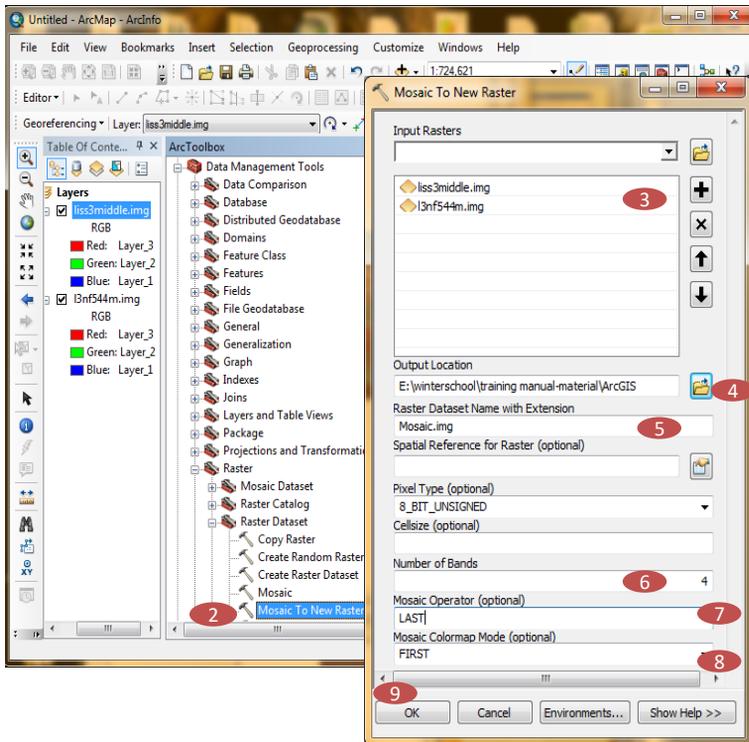
STEP 1: Add both Images to be mosaicked

STEP 2: Go to **Arc Tool box** → **data Management Tool** → **Raster** → **Raster Dataset** →**Mosaic To New Raster**

STEP 3: Add **Input Rasters**

STEP 4: Set **Output Location**

STEP 5: Define **Output Raster Dataset Name**

STEP 6: Enter **Number of Bands 4 (in case of IRS P 6 LISS III)**

STEP 7: The Mosaic Operator signifies the method used to mosaic overlapping areas.

The default method is LAST which indicates that the output cell value of the overlapping areas will be the value from the last raster dataset mosaicked into that location. The other mosaic operators like FIRST, BLEND, MEAN, MINIMUM and MAXIMUM are also available.

STEP 8: The Mosaic Color Map Mode signifies the method used to choose which colormap from the input rasters will be applied to the mosaic output.

The default method is FIRST which indicates the colormap from the first raster dataset in the list will be applied to the output raster mosaic. Other methods like LAST, MATCH, and REJECT are also available

STEP 9: Click **OK**
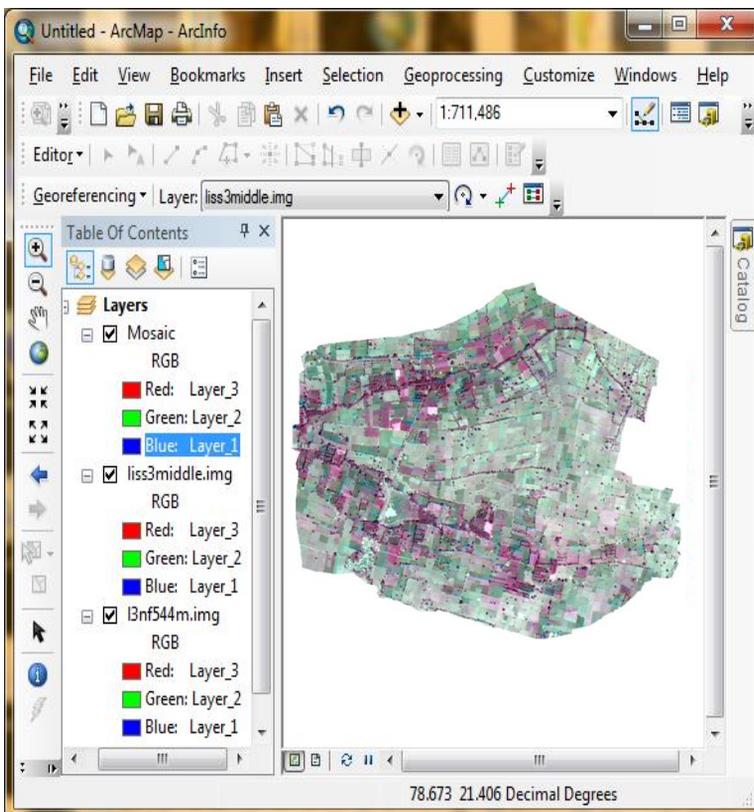
The view of the generated mosaic image of IRS P-6 LISS III

12. FCC Generation and Image Enhancement

Introduction

Multi-spectral sensors from earth observation satellites like Landsat or IRS series, collect information from different wavelengths of light in a digital camera but there are two major differences. The first is that instead of limiting itself to the visible wavelengths (what we can see with the human eye - red, green, and blue) a much broader range of wavelengths, such as, infrared and thermal wavelengths are detected. The second major difference is that instead of automatically combining the information from the different wavelengths to form a picture, the information for each specific wavelength range is stored as a separate image, commonly called a band. When viewed, this single band image is similar to a black and white digital photograph. User can create color images on a computer screen by assigning red, green, and blue light – the three primary colors, to illuminate any band combinations. When the three colors are assigned to their respective bands it forms the natural color image. This means that the colors of an object in an image appear to a human observer the same way as if this observer were to directly view the object. When, the color of a target in the displayed image does not have any resemblance to its actual color, the resulting product is known as a false color composite image. There are many possible schemes of producing false color composite images. However, some schemes may be more suitable for detecting certain objects in the image.

Enhancements are used to make it easier for visual interpretation and understanding of imagery. In raw imagery, the useful data often populates only a small portion of the available range of digital values (commonly 8 bits or 256 levels). Contrast enhancement involves changing the original values so that more of the available range is used, thereby increasing the contrast between targets and their backgrounds. The advantage of digital imagery is that it allows user to manipulate the digital pixel values in an image.

Generation of false color composite (FCC)

Assigning red color to NIR band, green to Red band, and blue to Green band makes the standard false color composite image. This false color composite scheme allows vegetation to be detected readily in the image. In this type of false color composite images, vegetation appears in different shades of red depending on the types and conditions of the vegetation, since it has a high reflectance in the NIR band, clear water appears dark-bluish (higher green band reflectance), while turbid water appears cyan (higher red reflectance due to sediments) compared to clear water. Bare soils, roads and buildings may appear in various shades of blue, yellow or grey, depending on their composition. Assigning red color to SWIR band, green to NIR band, and blue to Red band makes a false color composite image which allows active fire to be detected readily. In this display scheme, vegetation appears in shades of green. Bare soils and clear cut areas appear purplish or magenta. Location of active fires appear bright red and smoke plume originating from the active fire site appears faint bluish in color.

Image contrast enhancement

There can be linear and nonlinear digital contrast enhancement techniques. The simplest type of linear enhancement is a **minimum– maximum contrast stretch**. This involves identifying lower and upper bounds from the histogram (the minimum and maximum brightness values in the image) and applying a transformation to stretch this range to fill the full range. Linear contrast enhancement is best applied to images with Gaussian or near- Gaussian histograms. If the lower and upper bounds lay a certain percentage of pixels from the mean of the histogram, the enhancement is called a **percent linear contrast stretch**. When the percentage coincides with a standard deviation (SD) percentage (68% for one SD, 95.4% for two SD, and 99.73 for three SD), then it is called a **standard deviation contrast stretch**. When the histogram of the image is not Gaussian (i.e., bimodal or trimodal, etc.), a **piecewise linear contrast stretch** can be applied, which means applying separate linear stretches to expand the brightness range for each mode in the histogram. In effect, this corresponds to setting up a series of lower and upper bounds within user selected regions of the histogram. A uniform distribution of the input range of values across the full range may not always be an appropriate enhancement, particularly if the input range is not uniformly distributed. In this case, nonlinear contrast enhancement like, a **histogram-equalized stretch** may be better. This stretch assigns more display values (range) to the frequently occurring portions of the histogram. This way, the detail in these areas will be better enhanced relative to those areas of the original histogram, where values occur less frequently. Another type of nonlinear contrast stretch, the **logarithmic contrast stretch** involves scaling the input data logarithmically. This enhances greatly the darker parts of the histogram. The **square root contrast stretch** has the same effect. These methods are useful for reducing the gray level range of images with a large dynamic range and a histogram skewed towards lower gray levels. To enhance the brighter part of the histogram more, the **square contrast stretch** method can be applied, which enlarges the gray level range of images with a small dynamic range and with a histogram skewed to higher gray levels.

3.1.1 FCC generation in ArcGIS
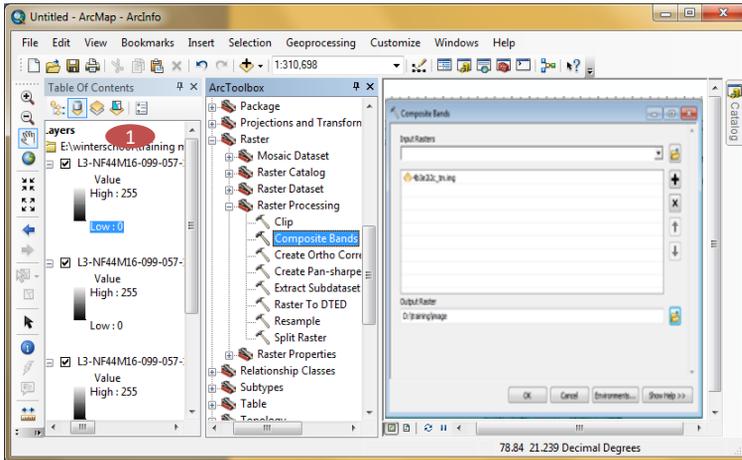
STEP 1: Add all the bands in the **Table of Contents**

STEP 2: Go to **ArcToolBox** → **Data management Tools** → **Raster** → **Raster Processing** → **Composite Bands**

STEP 3: Add all the bands in order

STEP 4: The order of bands can be changed by using **up** and **down arrows**

STEP 5: Define **Output Raster** name

STEP 6: Click **OK**
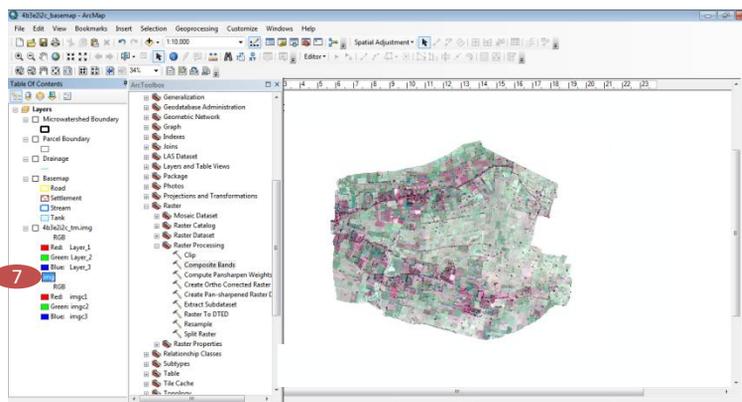
By default red Channel will be assigned to Layer 1, green to Layer 2, and Blue to Layer 3

STEP 7: Left click on individual Channel boxes to assign proper Layers to get standard FCC

3.1.2 Image enhancement in ArcGIS
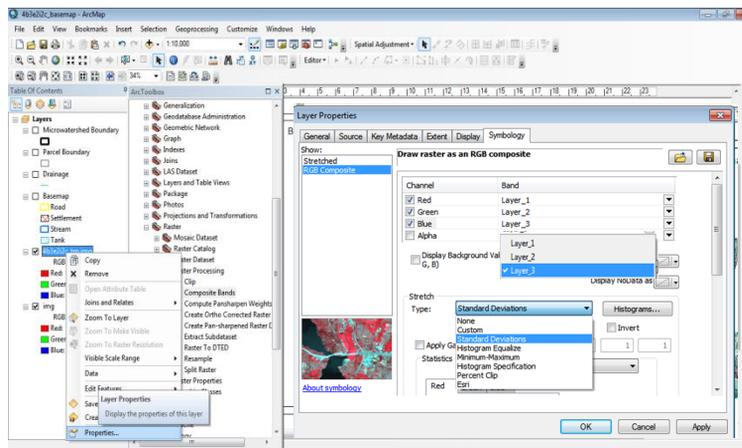
STEP 1: Right click on the Image in **Table of Contents** and go to **Properties**

STEP 2: Click on **Symbology** Tab

STEP 3: Select **RGB Composite**

STEP 4: Assign Red channel to NIR band, Green Channel to Red band, and Blue Channel to Green Band

STEP 5: In the **Stretch Type** select desired enhancement

STEP 6: Click **OK**
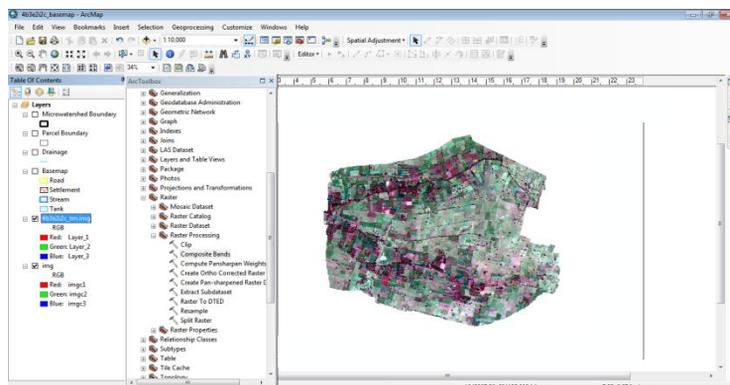
The output of enhanced IRS P 6 - LISS III image generated with Standard Deviation Enhancement Technique

13. Preparation of Soil Map

Introduction

        In REWARD (Rejuvenating Watershed for Agricultural Resilience through Innovative Development) the Imagery and cadastral maps are provided by KSRSAC. Both are geo-referenced. We have to prepare all other layers like Soils, Hydrology etc. to match the existing Imagery and cadastral maps. Soil boundaries are drawn on the base map provided by KSRSAC.
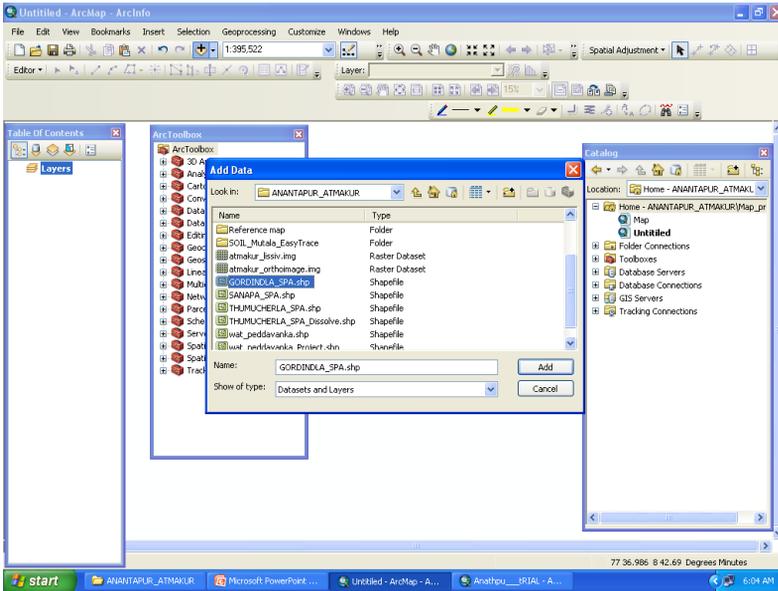
STEP 1: Add the shape file and the LISS IV through the **Add data** *A Vector file will have icon as:*

 For Point

 For Polyline/ Line

 For Polygon

STEP 2: *Add* **Raster** *file* through the **Add data** *will have icon as:*

 Raster *(LISS IV + Cortasat)*

Dissolve
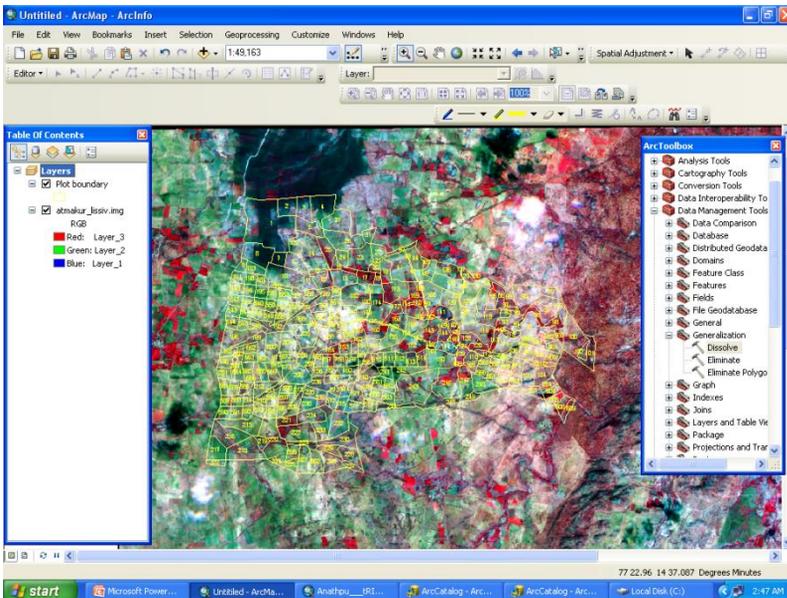
STEP 3: The watershed boundary with survey numbers or plot numbers (viz., xxxx_Amos.shp given by KSRSAC) and also the image is also seen.

STEP 4: Open the tool box. For preparing external boundary.

NOTE: Here we are making the soil external boundary using the xxx_Amos.shp (given by KSRSAC); Otherwise, there is chances of the soil outer boundary not matching exactly with plot outer boundary.
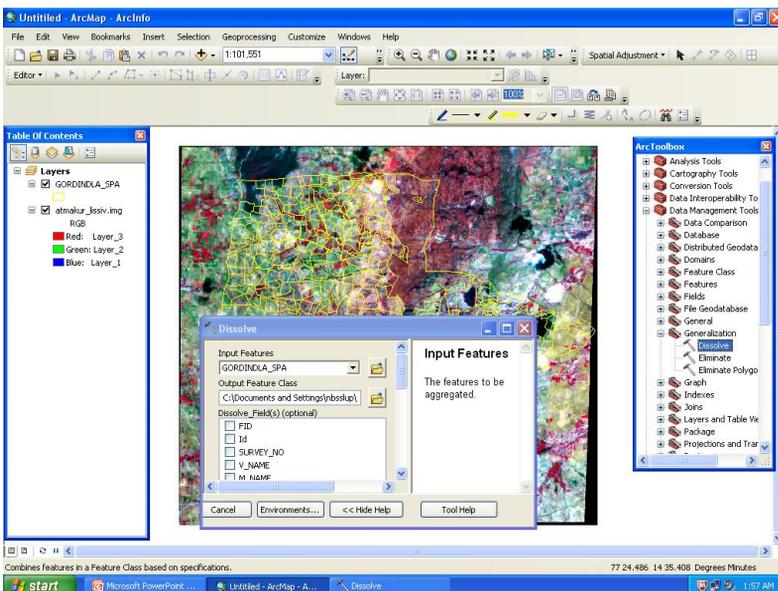
STEP 5: Open the tool box - Data management tools - Generalization – **Dissolve**.

STEP 6: Input feature is the shape file (xxx_Amos.shp or any other given by KSRSAC) and in the Output feature class the name of shape file and source has to be shown.

STEP 7: Click **OK**

STEP 8: The new external watershed boundary is created.

STEP 9: Open a new arc map, add already geo-referenced shape file or external boundary.

STEP 10: *Add* **Raster or** the scanned image of the soil boundary through the **Add data.** *The file will have icon as:*
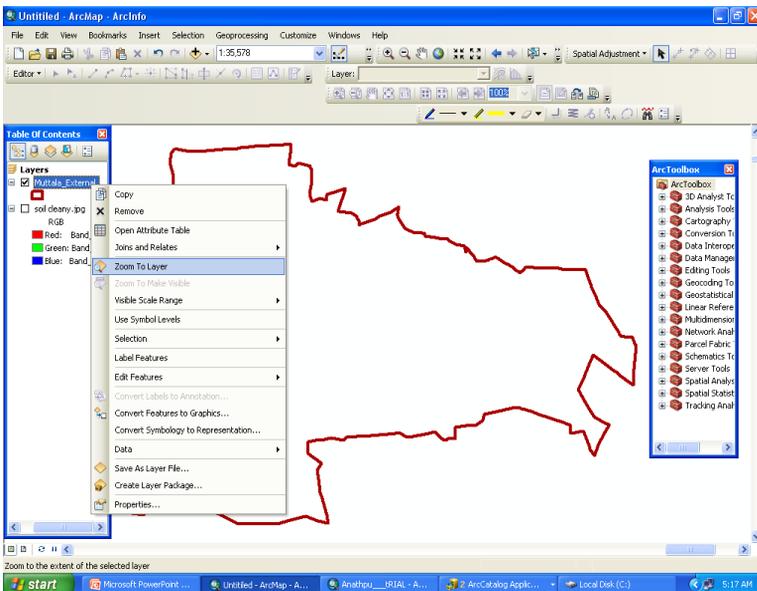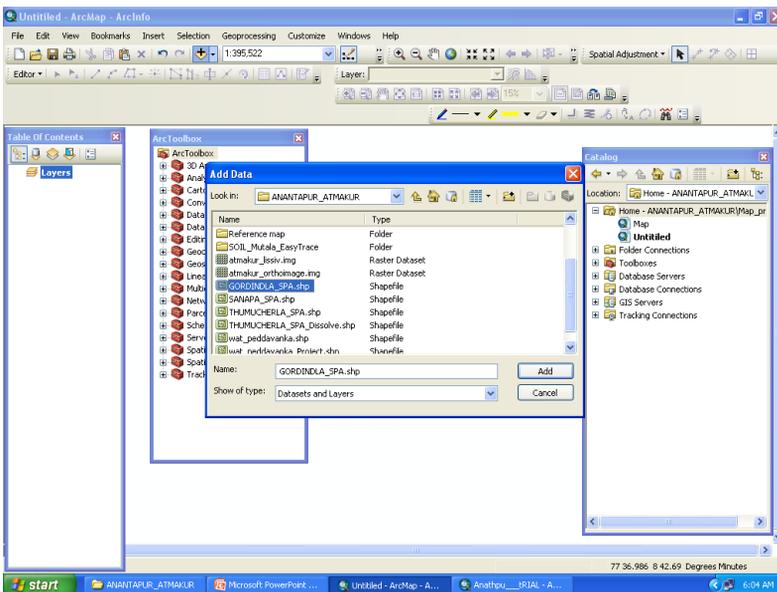
STEP 11: Right click on the vector layer name, then click on ⮡ **Zoom to layer**.

STEP 12: Go to Geo-referencing tool ⮡ **Fit to display**. The Vector and Raster image is in the same map interface but they are in different places as Shape file is Geo-referenced & the scanned soil base(image) is not.

NOTE: If it is not showing in the same interface then try Zoom to layer and Fit to display twice.
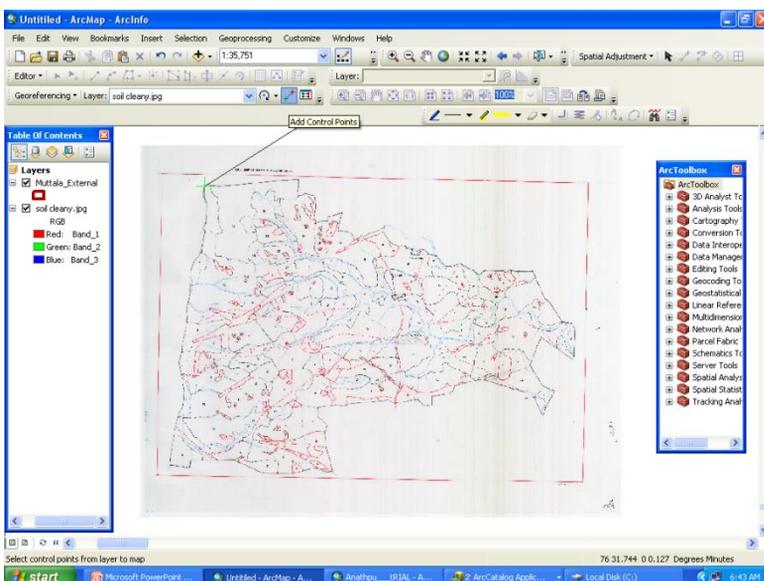
STEP 13: Go to the Geo-referencing tool & select **Add control points tool.** Right click on the Scanned image Zoom to layer Select a point (zoom to max.) and mark there.
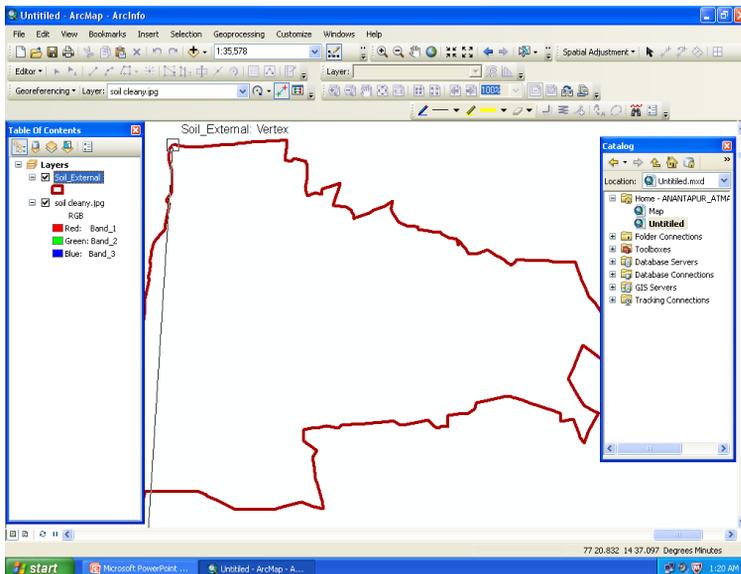
STEP 14: Go to the soil external or .shp file ▯ Right click then **Zoom to layer** ,then go to the same place (where you had started the point in **Step 13** on scanned image) and end the marker. So we have done with one point.

STEP 15: Again go to the scanned image of the soil boundary ▯ Right click then **Zoom to layer** & same procedure has to be repeated till it sits properly. Minimum of 4-6 points covering full external boundary.
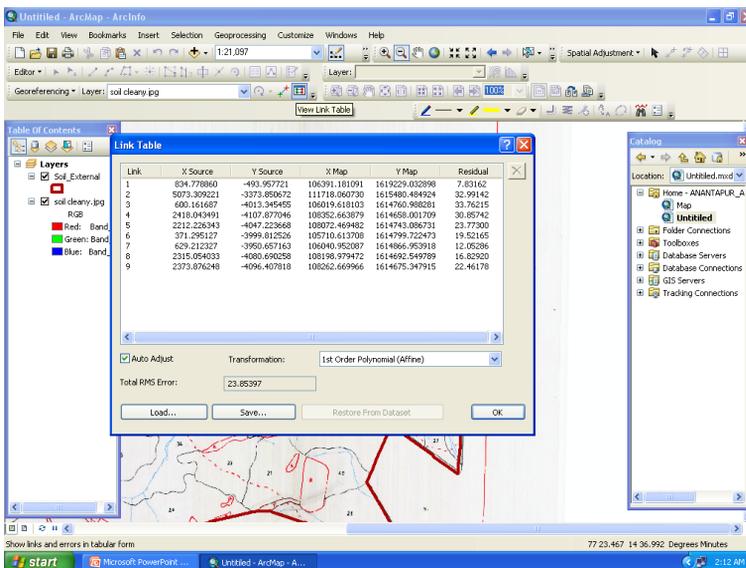
STEP 16: Now we can see that the soil image is almost sitting on the soil external shape file.

STEP 17: Go to *Link table (*adjacent to *Add control points). Can view all the points here can delete the points if found not necessary.*

STEP 18: In *Geo-referencing tool* drop down and click on "**Rectify**" and show the output source to save your required geo-referenced image.

OR

activated "**Update Geo-referencing**" to update in the same image.
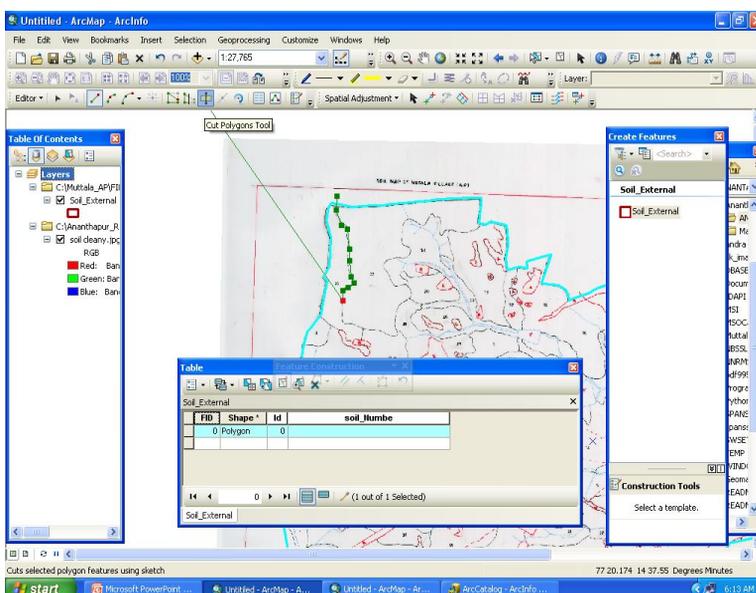
Preparation of Soil boundary

STEP 19: We can use soil external and prepare the soil boundary by using *Cut polygon tool* (Editor tool)*.*

STEP 20: In *Editor tool* dropdown Click Start editing then cut polygons tool gets activated.

STEP 21: Using the *edit tool* select the soil external boundary then go to the *Cut polygon tool* and start to cut and end the cut with a double click. (NOTE; Starting from outside the selection & ending outside the selected boundary)
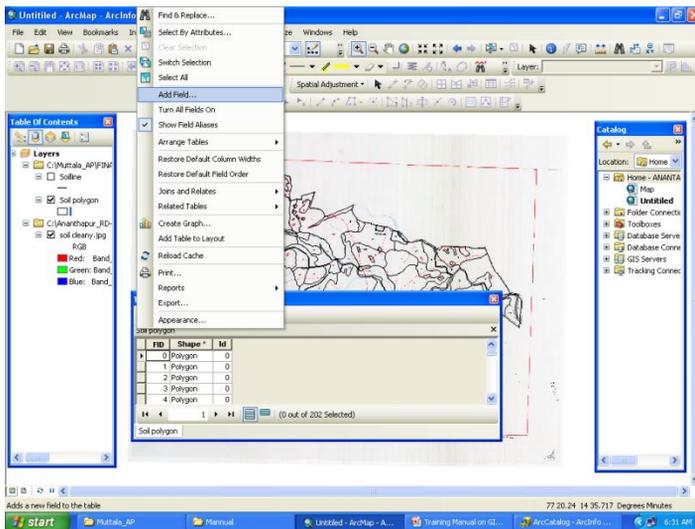
STEP 22: Soil polygon is soil boundary can be made hollow so that the soil phases/units details (name or number) can be visible.

STEP 23: Make attribute table ready by using:

Right click on the soil polygon → Open Attribute layer → Open Table options → Add field → Give the field name as "Soil_Leg"

Now the whole map vectorisation is complete**.**
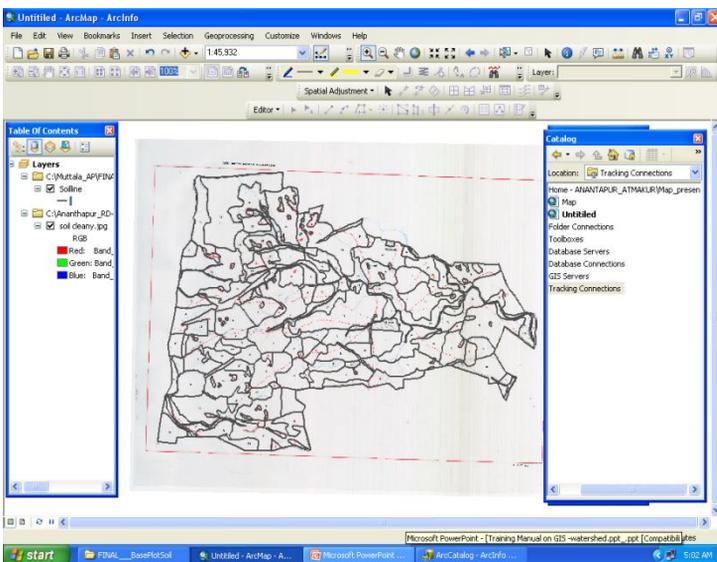
STEP 24: Enter the details of the entries appearing on the scanned image and finalize the "Soil_Leg" field.

NOTE: In the soil layer which occupies area (space) is taken in like Stream, tank, lake, burial ground, habitation, highway road and railway line on a small watershed.
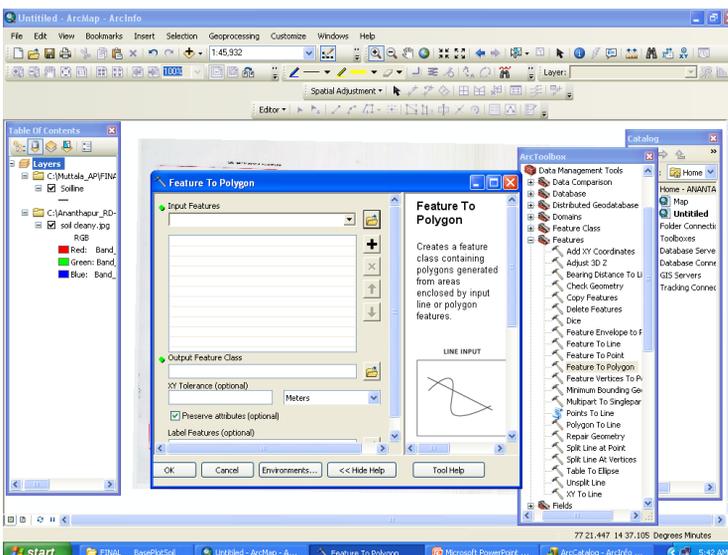
STEP 25: Take Habitation and Waterbody as "Others*" and simplify the "Soil_Leg" field as "Soil_Leg1"

STEP 26: The finalized layer appears on the table of contents.

CREATING FILE GEODATABASE

Follow these steps to create a file geo-database from the Catalog tree in ArcGIS for Desktop:
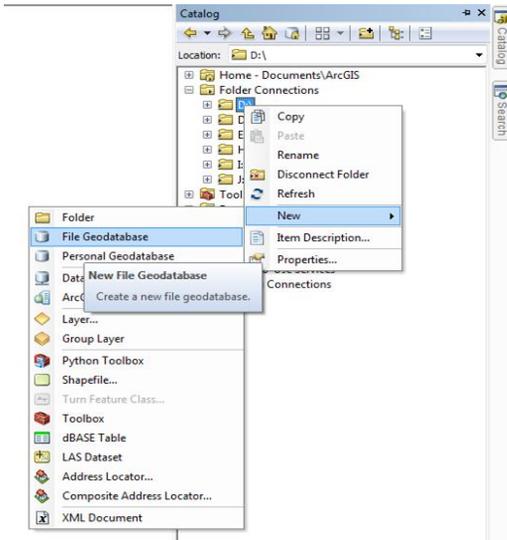
✓ Start ArcCatalog or ArcMap and open the Catalog window.

✓ Right-click the file folder in the Catalog tree where you want to create the file geodatabase.

✓ Point to **New**.

✓ Click File Geodatabase.

✓ A new file geodatabase is created in the location you selected.

✓ Rename the file geodatabase by right-clicking the file geodatabase name, choosing **Rename**, then typing a new name.

OR
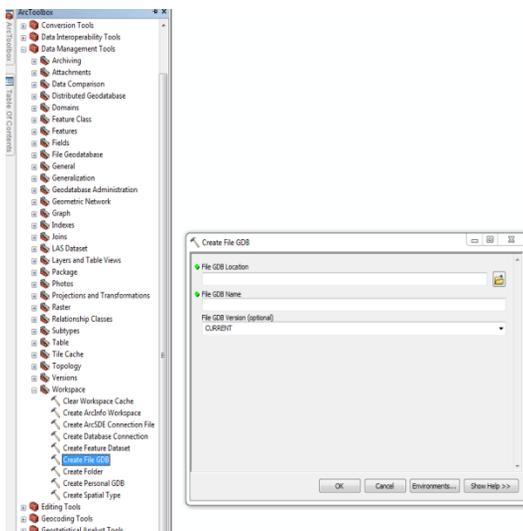
STEPS

✓ Use the tool Create File GDB in ArcToolbox to specify the path name to the file folder location and name of the new geo-database you want to create.

✓ The Create File GDB tool is located in the Data Management toolbox in the Workspace toolset (Data Management > Workspace).

✓ Create File GDB.

NAMING OF FILE GDB

Taluk name and Batch number of the district should be given as folder inside the particular system where the shape files are saved.

For eg., If we are doing **Batch_5** work of **Harve** Sub watershed under **Chamarajanagara** district, **Chamarajanagara** taluk then we have to proceed following way:
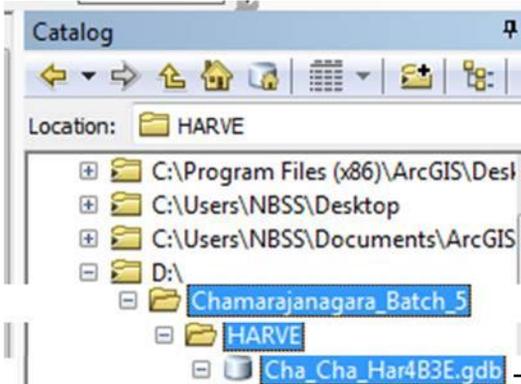
Chamarajanagar Taluk name given to File GDB

✔ Create one folder with the Taluk name with Batch number in the source system
> D:Chamarajanagara_Batch_5

✔ Inside the above folder create one more folder with the Sub watershed name (ie., in Caps)
> D:Chamarajanagara_Batch_5\HARVE

✔ Naming the File GDB in the following way:

First three letters of district name followed by underscore (_), first three letters of taluk name followed by underscore (_) and first three letters of Subwatershed name followed by first four letters of sub-watershed code should be given inside the earlier folder as given below:
> D:Chamarajanagara_Batch_5\HARVE\Cha_Cha_Har4B3E

CREATING FEATURE DATASET

First, create a new feature dataset.

Choosing the co-ordinate system (WGS 1984 UTM Zone 43 N)

Add the set of feature classes that will be held in your feature dataset and populate their contents (i.e., add features)

1. Creating a New Feature Dataset in Arc Catalog

Here are the steps to create a feature dataset:

- ○ In the Arc Catalog tree, right-click the File Geodatabase in which you want to create a new feature dataset.
- ○ Point to New and click Feature Dataset.
- ○ Type a name for the feature dataset.

Naming of Feature Dataset should be given as

First three letters of Micro-watershed name (ie., in Caps) followed by underscore(_), then last four letters of micro-watershed code.

> For example.
>> If we take Kethahalli Micro-watershed - 4B3E1G1a;
>> So, its feature dataset should create as KET_1G1a

Navigate to the spatial reference you want to use. Alternatively, click Import and navigate to the feature class or feature dataset that uses the spatial reference you want to use as a template.

        Click **OK**.
        X ,Y, Z, tolerance as default
        Click **Next** ..
        Click **Finish**.
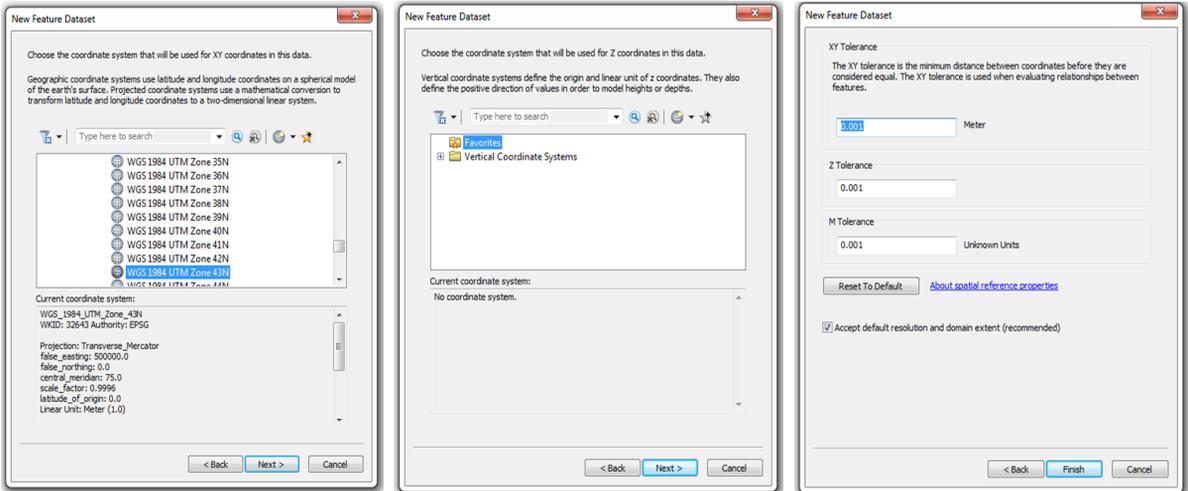
Now feature dataset is formed.
You will receive File GDB in ArcCatalog as follows:
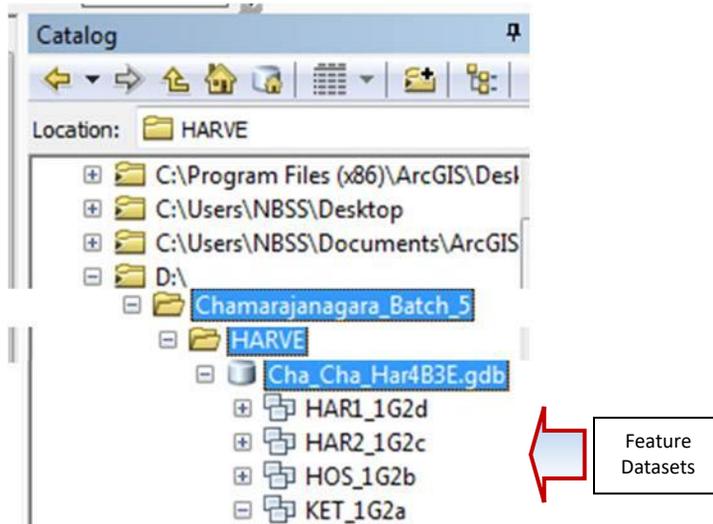
IMPORTING FINALIZED SHAPE FILES INTO DATASET

NOTE: Only the person who created the feature dataset can add feature classes to it.
NOTE: All feature classes added to the geo database must have unique names, even if they are grouped into a feature dataset.

✔ To create empty feature classes in ArcCatalog. A new feature class can be created as part of a feature dataset or as a stand-alone feature class in a geodatabase. They can also be contained within the following structures: shapefile, coverage, CAD, SDC, and VPF.

✔ When creating a feature class, you choose whether to create one that stores polygons, lines, points, annotation, multipoint, multipatch or dimension features.

✔ All feature classes in a feature dataset must use the same **spatial reference**, which is defined when the feature dataset is created.
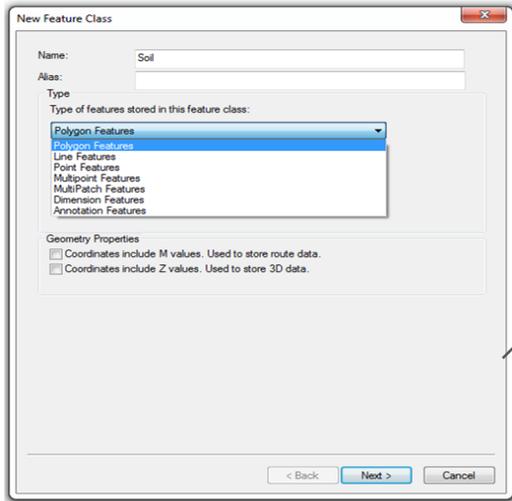
Select the feature type Polygon, Linear Point.
Click Next.
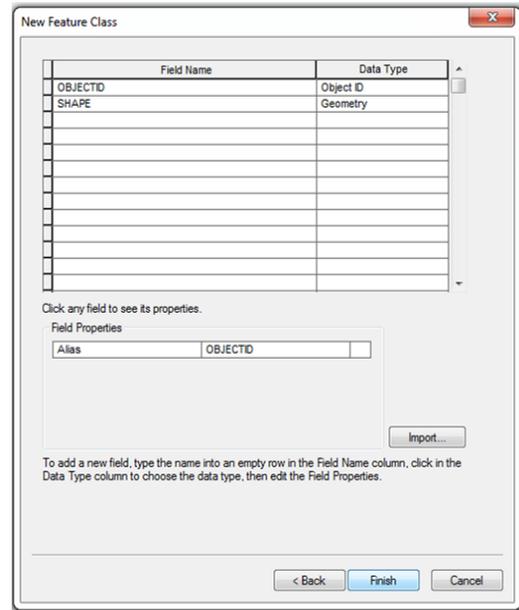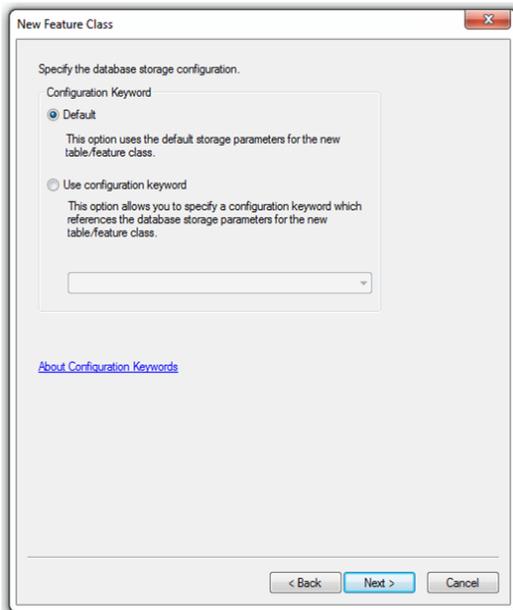
Click **Finish.**

Importing Feature Class

Importing a feature class (single)

In ArcCatalog, right-click the geodatabase or feature dataset to which you want to import a dataset.
Navigate to the Import pull-right menu and choose Feature Class (single). This opens the Feature Class To Feature Class tool from the Conversion toolbox.
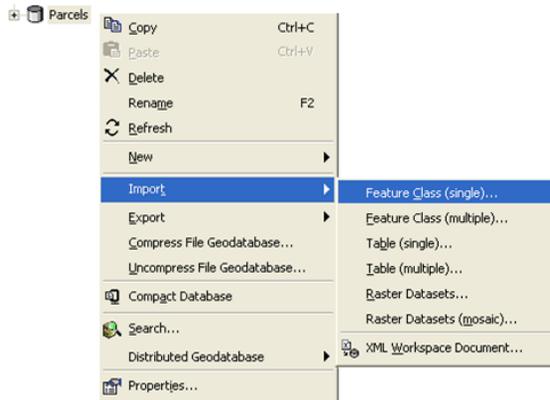
For this tool, you will have to fill out the following parameters:
- o Input Features—Browse to the dataset you want to import.
- o Output Location—The location in which the output feature class will be created.
- o This can be either a geodatabase, feature dataset, or a folder. If the Output Location is a folder, the output will be a shapefile.
- o Output Feature Class—The name of the output feature class.

Importing a feature class (multiple)

In ArcCatalog, right-click the geodatabase or feature dataset into which you want to import multiple datasets.
Click Import and choose Feature Class (multiple) from the pull-right menu. This opens the Feature Class To Geodatabase (multiple) tool from the Conversion toolbox.

Click **OK**

Examples of Feature Datasets with different feature classes done in **File GDB**
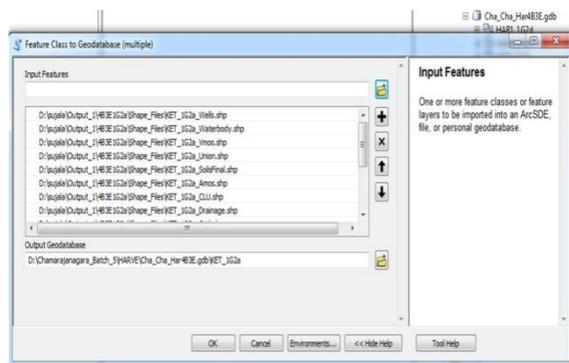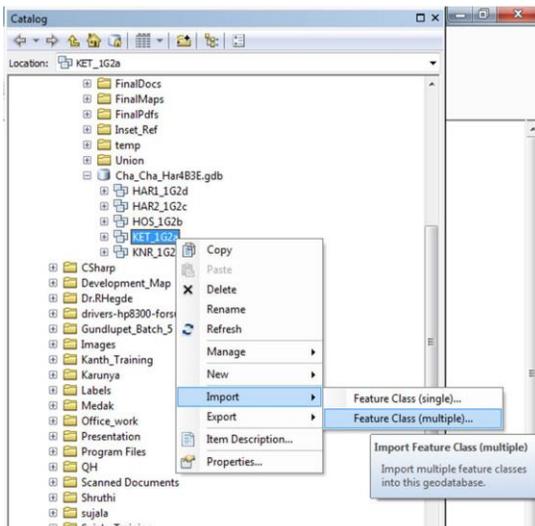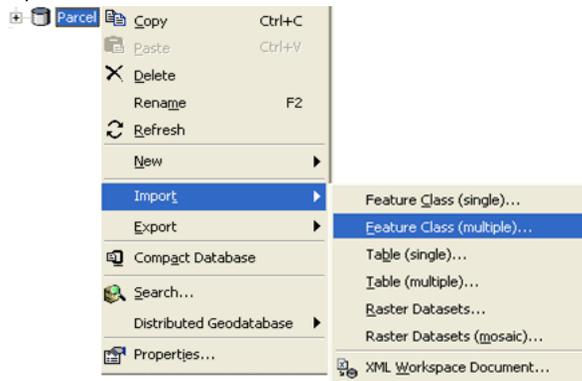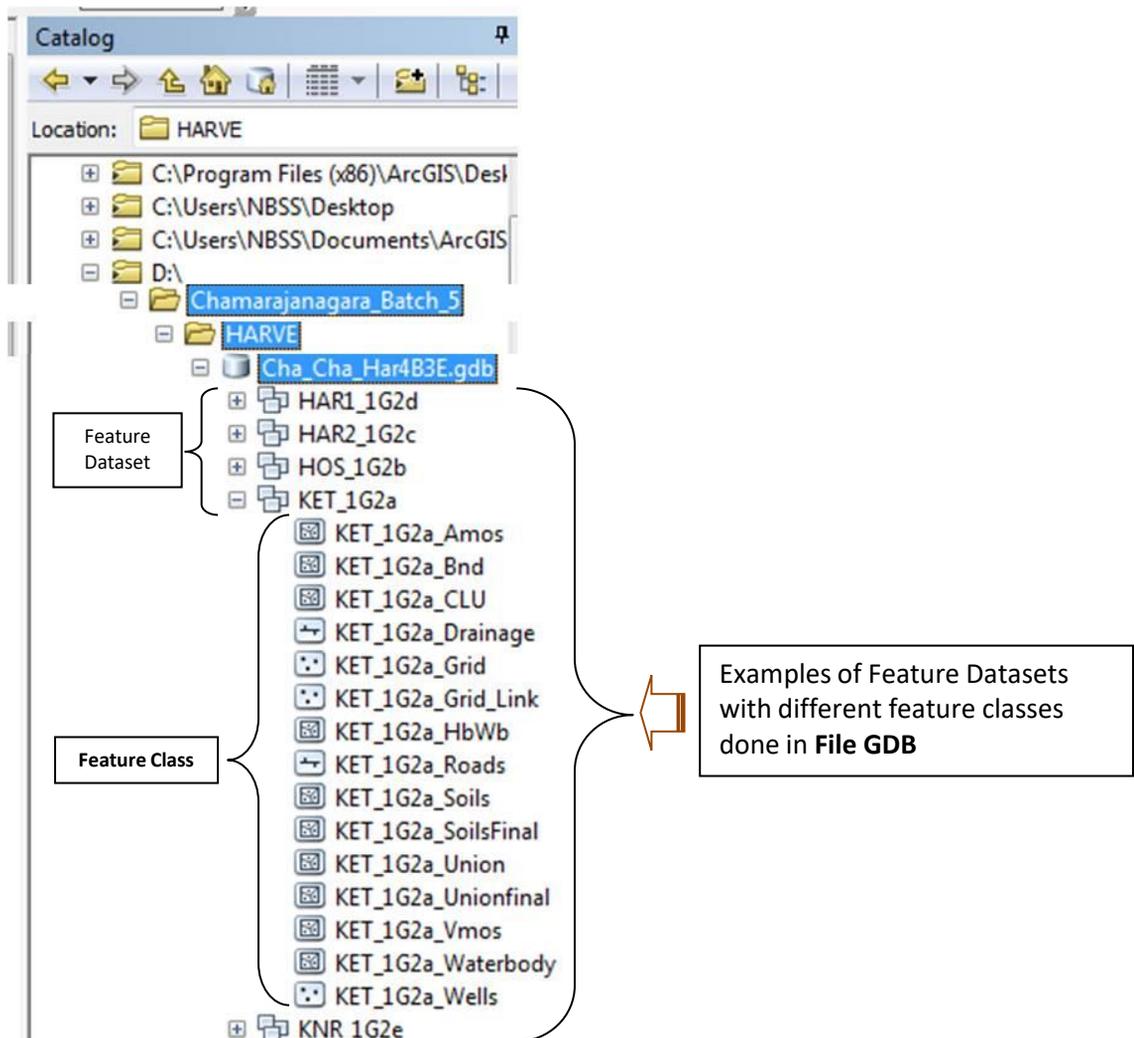
We can see feature classes added in feature dataset.

CREATING AND IMPORTING RASTER DATASET

Steps:

In *ArcCatalog* or the *Catalog* window, right-click the geodatabase, and click **Import** > **Raster Datasets**.

- ○ Click the **Input Rasters** browse button to navigate to the raster datasets you want to import.
- ○ Select the raster dataset you want to import and click **Add**.
- ○ You can use the SHIFT or CTRL keys to select more than one raster dataset.
- ○ Repeat steps 2 and 3 to add raster datasets from different locations.
- ○ Click the **Output Geo-database** browse button to set the output geodatabase if it has not already been specified.
- ○ Optionally, type a configuration keyword.
- ○ Click **OK**.
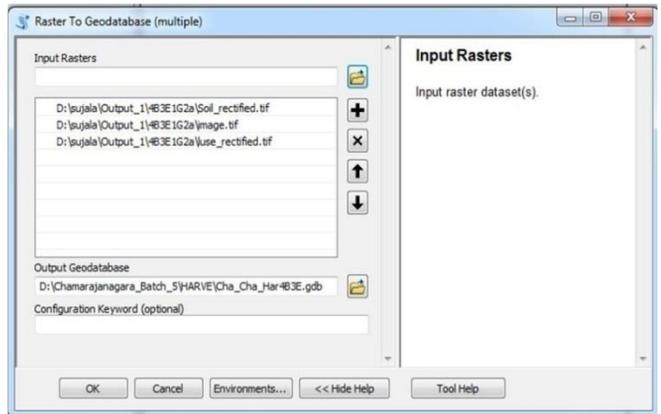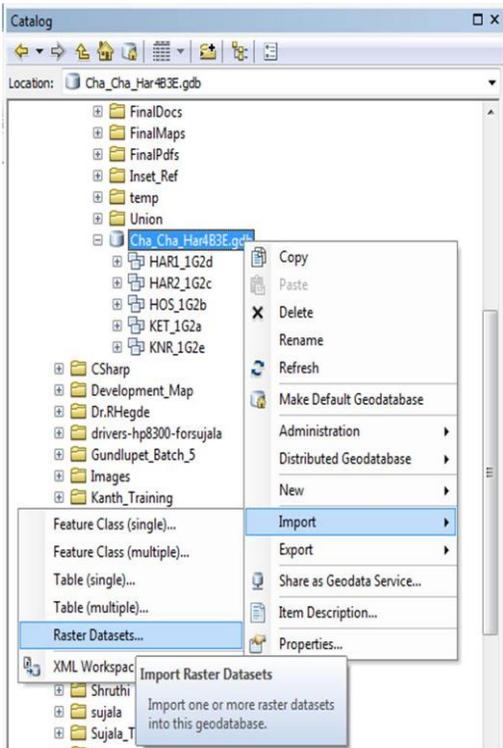
Click **OK**

Raster Dataset in File GDB
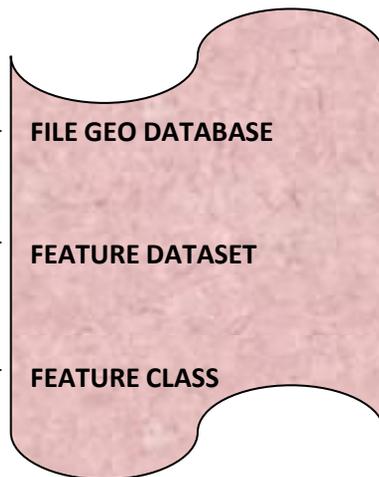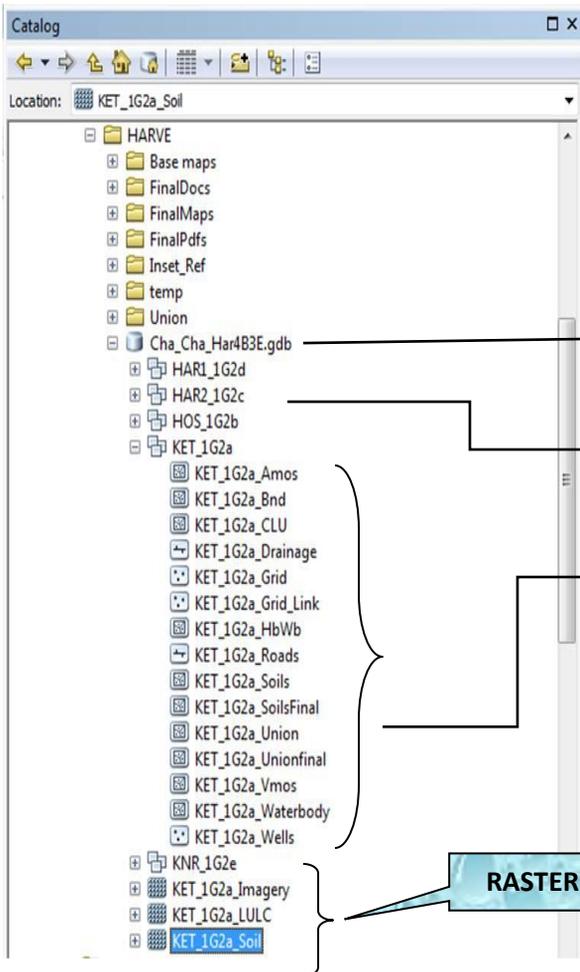
GENERATION OF MAPS

NOTE: Before starting the generation of maps cross check finalized feature classes

      Take care of Fonts and its sizes, types has to maintain uniformly in all the maps

Generate maps using following steps:

○ Create folders (maintain uniformity while giving the folder names) under the sub watershed

    For eg., If we are doing **Kethahalli** Micro-watershed (**4B3E1G2a**) maps then create folder with name **Final Maps** under Sub-watershed **HARVE** and then create one more folder **KET_1G2a** folder to save all .mxd format final maps. It appears as follows:
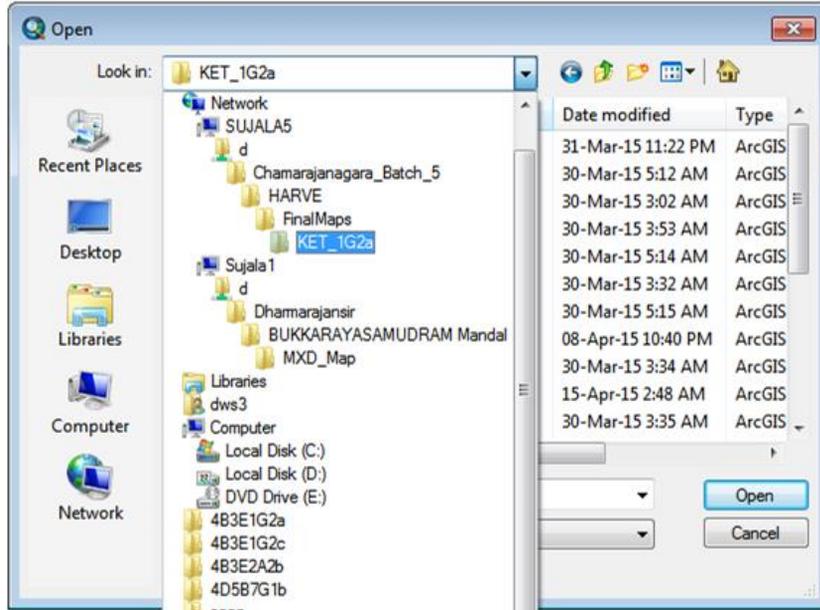
○ Whatever the needed folders you want that you have to create in the above manner like...

    Base Maps, Final Pdfs, **Final Docs**, *etc.,* for each and every sub-watershed.

<table>
<tr><td>Rectified
images</td><td>Exported
maps</td><td>Summarized/
required documents
(ie., excel/word etc.)</td></tr>
</table>
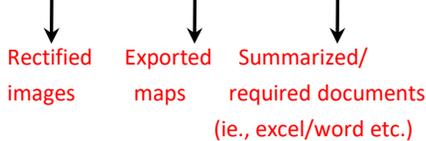
○ Inside these folders maintain separately Micro-watershed information.

Instructions before starting map composition:

a) First of all take care of attribute table with uniform column names in each and every micro-watershed.

b) Take care of "Soil_Leg" and "Soil_Leg1" for future reference (Separate Habitation & Waterbody from "Soil_Leg" column and show it as "Others*" in attribute table by taking a new column "Soil_Leg1" without disturbing.

c) Consider Railway line, Grass Land, Gullied Region, Forest, Paddy Area, Eroded Area, Mining/Industrial Area, Rock Crushing Area, Rockout crops and Miscellaneous are common features and they should be include in "Legend"

d) Roads should be polyline feature class except National Highways and include in "References".
    If particularly "National Highways" are there then it should be shown in "Legend"

Generation of Soil-Site Characteristics Map:

To generate soil site characteristics go through the following steps.

○ Locate your required features in the **layout window in A3-Paper size print set up and Set the data source to File GDB** is must to all feature classes.

○ Provide **surrounding villages** in a label (Arial - 13.00 - Italic - Black) format to micro-watershed and give **plot numbers** overlay on map by taking "DXF_TEXT" column from Amos feature class's attribute table and take care while text is appearing on the map and if it is there remove text and overlay (Arial - 8.00 - Regular - Gray 70%)

NOTE: While giving <u>labels</u> only numeric should be visible and ignore text wherever it appears

For eg., If "123_TANK" is there then you have to ignore the word TANK label

The output should be:
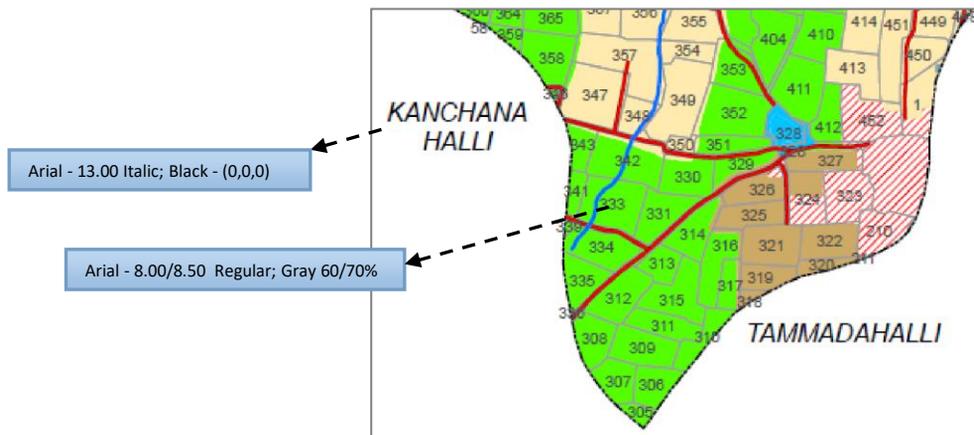
○ Give the titles of the micro-watershed which you are working

For eg., if you consider Kethahalli Micro-watershed should be in the following color combinations and with specific font sizes depend on the map size:
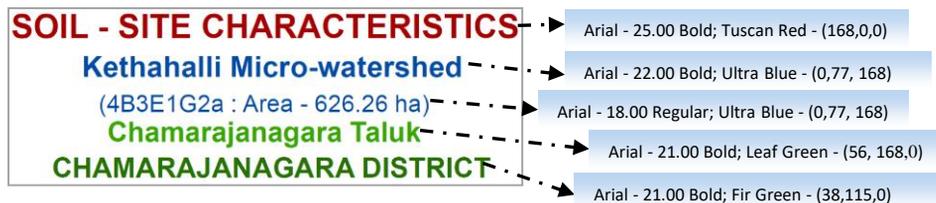
○ Give "**Alternating Scale Bar1**" from the scale bar selector ie which is shown below and made corrections for required scale bar to get scale in Kilometers.

Steps followed to get required format are:

1. Select "Alternating Scale Bar1" from Scale Bar Selector and give Ok
2. Go to properties of Scale Bar and select " **Scale and Units**" tab
3. Change number of divisions and subdivisions
4. Change scale units from Miles to "Kilometers"
5. Give unit label as "km" and change the required label position
6. Go to very next tab "Numbers and Marks "and give the required gap between division and subdivisions
7. Change divisions and subdivisions label position to "Below bar"
8. Apply and give OK
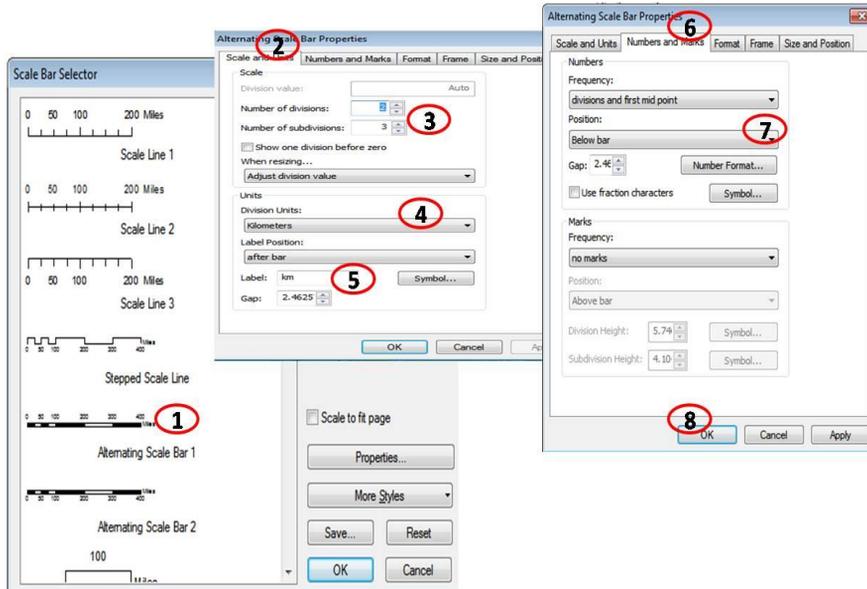
Finally you will get **Scale Bar** as given below:
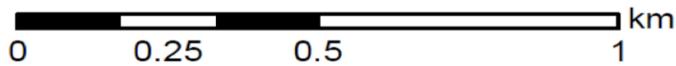
○ Select North Arrow from "North Arrow Selector" and select "**ESRI North 3**" and arrange in the map
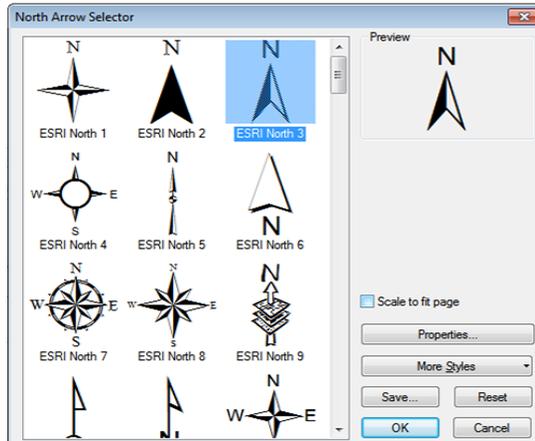
The output should be as follows:

○ Prepare "**INSET**" map separately as .Jpeg format and then Insert as picture into the current map.
The inset map should prepare as given here: ------------▯
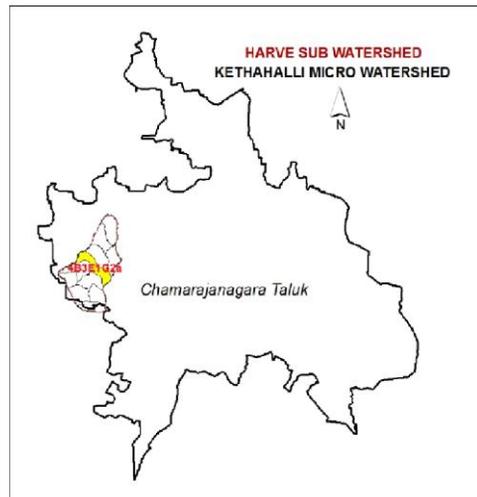
○ Finalize "**Legend**" including common features with area in hectares and percentage (%) by giving heading as

"**Phases   Area in ha. (%)**"

and include " **Others***" in the legend

For eg.,

Phases	Area in ha.(%)
bB1g1	21 (3.30)
bB1g2	107 (17.09)
cA1g1	39 (6.31)
.....	
.....	
.....	
.....	
Others*	22 (3.47)

Arial - 15..50 Bold; Black - (0,0,0)

Arial - 15.20 Regular; Black - (0,0,0)
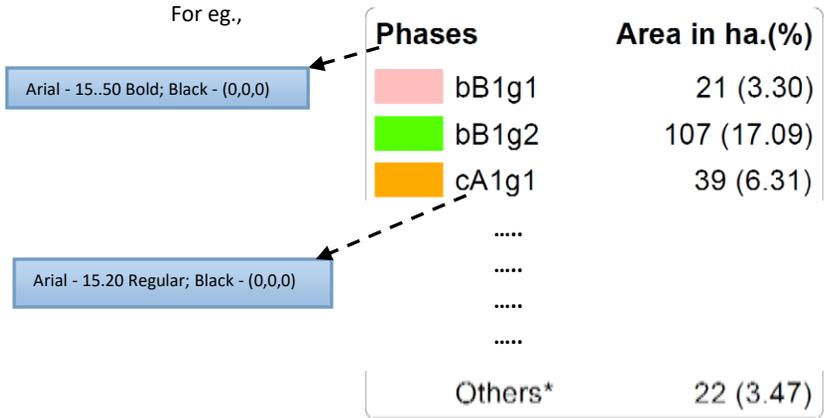
NOTE: Font size and adjustment of values varies with the size of map. Ultimately you have to highlight your legend.

Here we have to specify/explain simply what is **Others*** in the map as follows:

* - Habitation & Waterbody

Arial – 14.00 Regular; Black - (0,0,0)
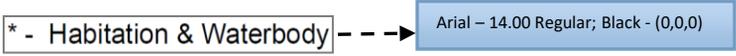
○ We have to explain the things whatever is appearing in the legend by giving heading as "**KEY**" is as follows:

KEY

Arial – 14.00 Bold; Tuscan Red - (168,0,0)

Arial – 13.00 Bold; Black - (0,0,0)

TEXTURE
b – Loamy sand
c – Sandy loam

SLOPE
A – Nearly level (0-1%)
B – Very gently sloping (1-3%)

EROSION
1 – Slight

GRAVELLINESS
g1 – Gravelly (15-35 %)
g2 – Very gravelly (35-60 %)

Arial – 14.00 Regular; Black - (0,0,0)
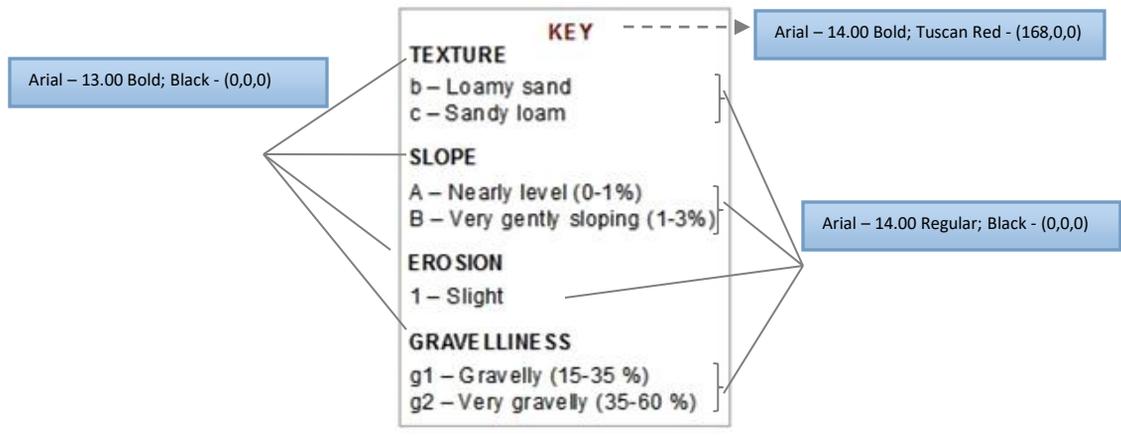
○ Show the **indication** in the following smart manner by showing any one of the phase in the legend.

Texture   Gravelliness

b B 1 g1

Slope   Erosion

Arial – 10.00 Regular; Black - (0,0,0)

Arial – 10.00 Regullar; Black - (0,0,0)
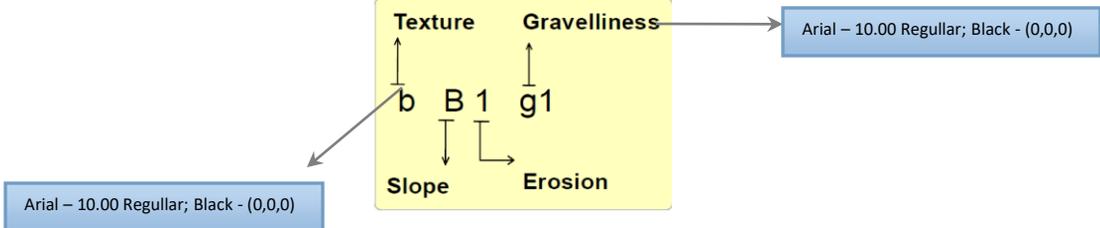
○ You must take care of all the feature classes required for "**References**". They are:

Profile locations
Wells (Bore or Open)   _____ Point feature

Drainages
Roads/Cart track   _____ Polyline features

Habitation
Waterbody
Land parcels with No's   ___ Polygon features
Micro-watershed boundary
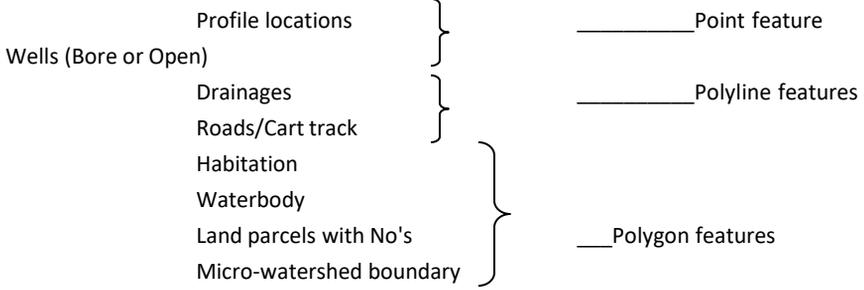
The final "**References**" legend appears in the following way:
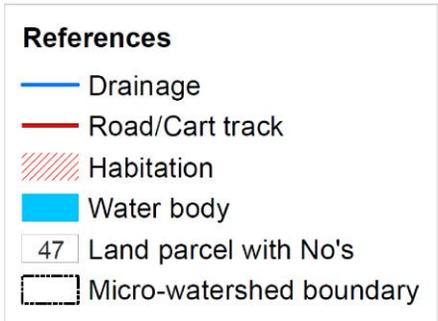
If we follow the above steps final output looks like in the below format:
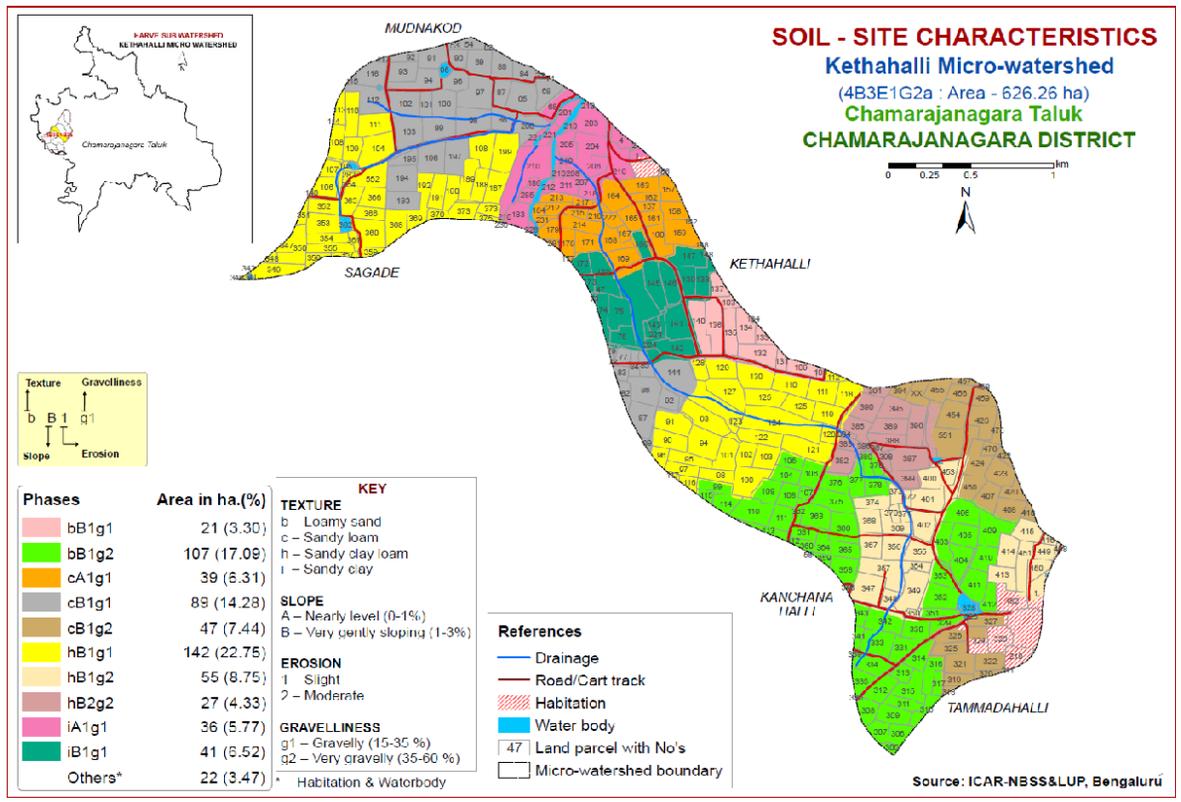
14. Spatial Interpolation

Introduction

Spatial interpolation tools in GIS create a continuous (or prediction) surface from sampled point values. Visiting every location in a study area to measure the height, concentration, or magnitude of a phenomenon is usually difficult or expensive. Instead, one can measure the phenomenon at strategically dispersed sample locations, and predicted values can be assigned to all other locations. Input points can be either randomly or regularly spaced or based on a sampling scheme. The continuous surface representation of a raster dataset represents some measure, such as the height, concentration, or magnitude. Surface interpolation tools make predictions from sample measurements for all locations in an output raster dataset, whether or not a measurement has been taken at the location. There are a variety of ways to derive a prediction for each location; each method is referred to as a model. With each model, there are different assumptions made of the data, and certain models are more applicable for specific data. Each model produces predictions using different calculations. The interpolation tools are generally divided into deterministic and geostatistical methods. The deterministic interpolation methods assign values to locations based on the surrounding measured values and on specified mathematical formulas that determine the smoothness of the resulting surface. The deterministic methods include inverse distance weighting (IDW), Natural Neighbor, Trend, and Spline. The geostatistical methods are based on statistical models that include autocorrelation (the statistical relationship among the measured points). Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions. Kriging (Isaaks and Srivastava, 1989) is a group of geostatistical techniques to interpolate the value of a random field (e.g., the elevation, z, of the landscape as a function of the geographic location) at an unobserved location from observations of its value at nearby locations. Spatial interpolation is the procedure of estimating the value of properties at unsampled sites within the area covered by existing observations. As per the first law of Geography "Everything is related to everything else, but near things are more related than distant things (Tobler, 1970), this is the basic premise behind interpolation, and near points generally receive higher weights than far away points.

Triangular irregular networks (TIN)

Triangular irregular networks (TIN) have been used by the GIS community for many years and are a digital means to represent surface morphology. TINs are a form of vector-based digital geographic data and are constructed by triangulating a set of vertices (points). The vertices are connected with a series of edges to form a network of triangles. There are different methods of interpolation to form these triangles, such as Delaunay triangulation or distance ordering. The input features used to create a TIN remain in the same position as the nodes or edges in the TIN.

Spline

Spline can be divided into Regularized and Tension. The Regularised option of Spline type usually produces smoother surfaces than those created with the Tension option. With the Regularized option, higher values used for the weight parameter produce smoother surfaces. The values entered for this parameter must be equal to or greater than zero. With the Tension option, higher values entered for the weight parameter result in somewhat coarser surfaces, but surfaces that closely conform to the control points. The values entered must be equal to or greater than zero. The greater the value of Number of Points, the smoother the surface of the output raster.

Inverse Distance Weighting (IDW)

The IDW is simple and intuitive deterministic interpolation method based on principle that sample values closer to the prediction location have more influence on prediction value than sample values farther apart. Using higher power assigns more weight to closer points resulting in less smoother surface. On the other hand, lower power assigns low weight to closer points resulting in smoother surface. The output value for a cell using inverse distance weighting (IDW) is limited to the range of the values used to interpolate. Because IDW is a weighted distance average, the average cannot be greater than the highest or less than the lowest input. The best results from IDW are obtained when sampling is sufficiently dense with regard to the local variation to simulate the phenomena. If the sampling of input points is sparse or uneven, the results may not sufficiently represent the desired surface.

Kriging

The basis of this technique is the rate at which the variance between points changes over space. This is expressed in the variogram, which shows how the average difference between values at points changes with distance between points. Depending on the stochastic properties of the random field different types of kriging apply. The type of kriging determines the linear constraint on the weights w_i implied by the unbiasedness condition; i.e. the linear constraint, and hence the method for calculating the weights, depends upon the type of kriging. There are a number of kriging algorithms, and each is distinguished by how the mean value is determined and used during the estimation process. The four most commonly used methods are Ordinary kriging, indicator kriging, universal kriging and gassuain kriging.

12.1 Spatial interpolation in ArcGIS
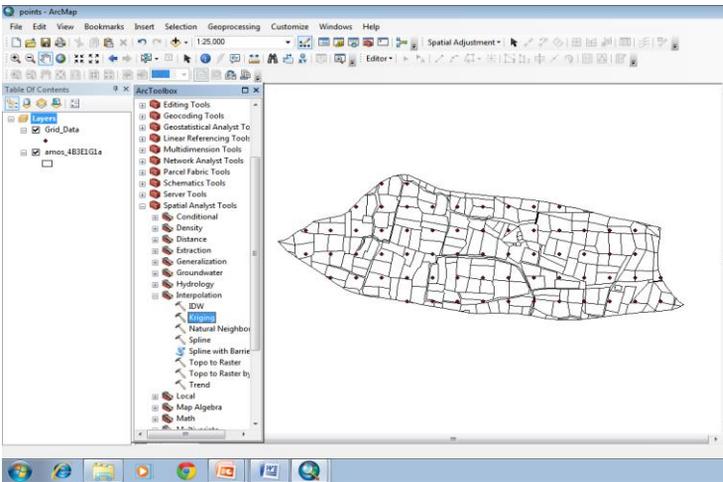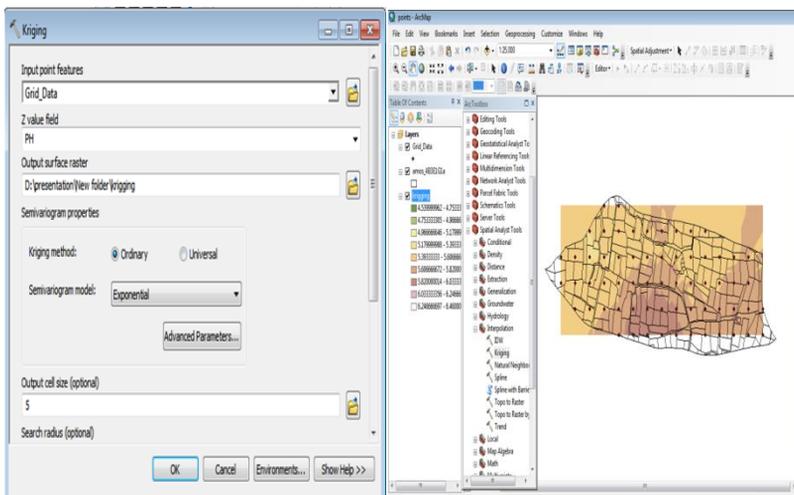
Interpolated raster by Kriging is shown in the figure
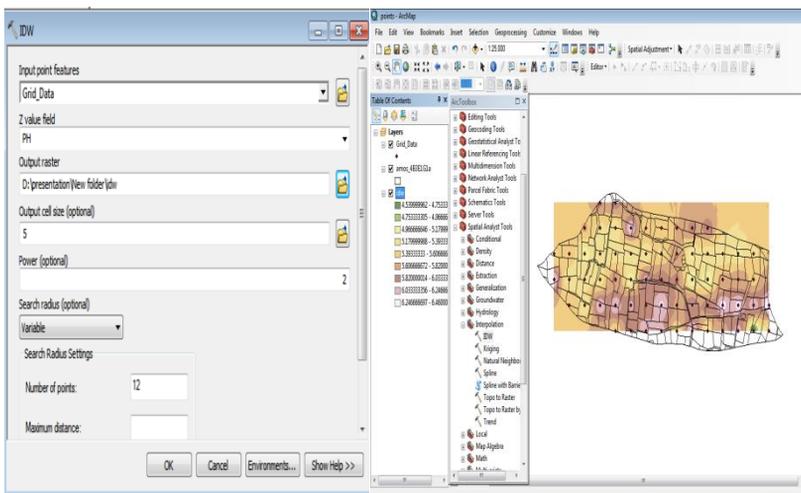
Interpolated raster by IDW is shown in the figure

Interpolation Techniques following under REWARD
(Rejuvenating Watershed for Agricultural Resilience through Innovative Development)
Project

Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data, such as elevation, rainfall, chemical concentrations and so on.

In the following graphic, each symbol in the point layer represents a location where the samples has been measured. By interpolating, the values between these input points will be predicted
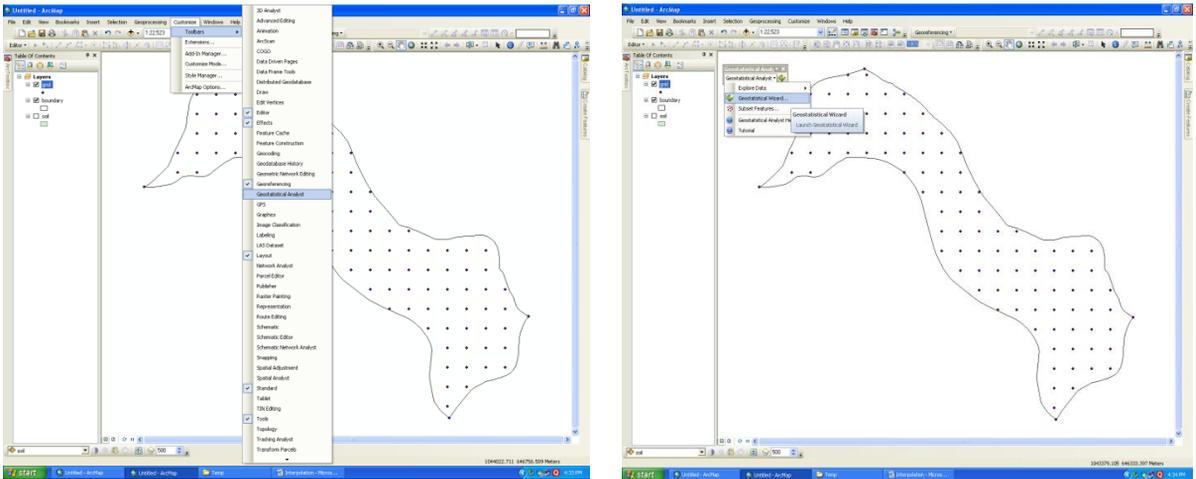
Geo-wizard Kriging method for Interpolate your data: Select Customize option which is in Menu Bar, and select Geo-statistical Analyst tool in drop down window. Select Geo statistical Wizard in geo statistical Analyst tool.
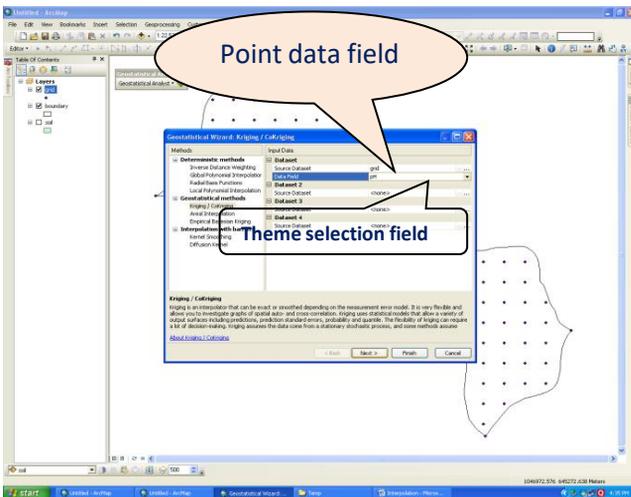
Theme selection field

In newly opened window select Kriging option. In the Data set field select your point data which is having all your data and in the Data field Colum select the theme, to which one you want interpolate (Ex. Boron, Zinc, Soil Reaction, etc.) and click on Next option

In newly opened page select "**Ordinary**" in Kriging type and select "**Prediction**" in Surface output type after that click on "**Next**"
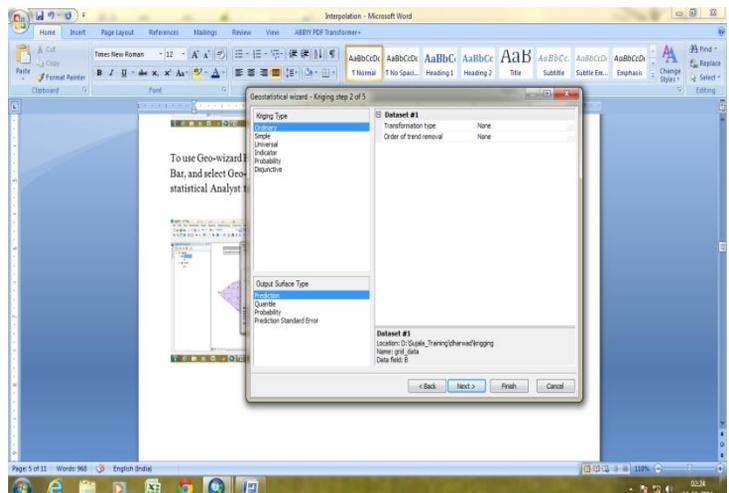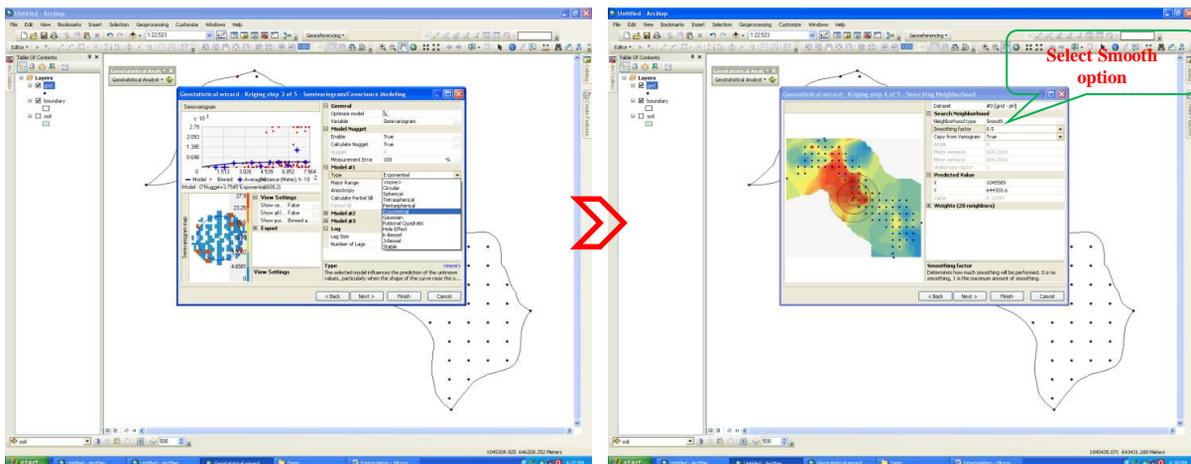
In the place model selection select "**Exponential**" or "**Spherical**". In the next step select "**Smooth**" option in Neighbourhood type tool and below that you can select '**smoothing factor**' how much you want.
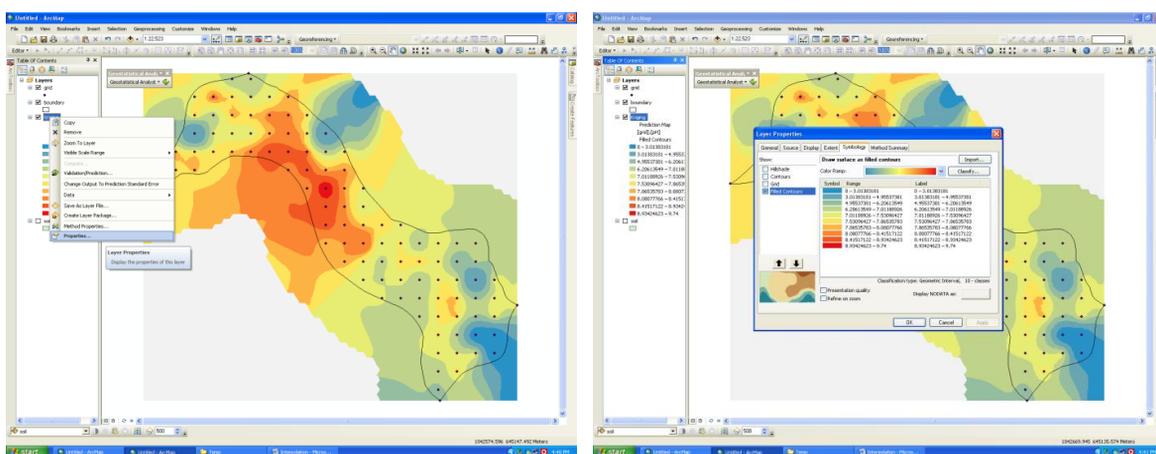
Once you got the result go to properties of the interpolated layer. One new window will open that window select Symbology option in that select classify inside of that window select Method as Manual and below that select classes according to your classifications.

After the classification the values we need to take data layer as vector file. Right click on the classified layer go to Data-Export to vector. One new window will open in that window select "Filled contour" in Contour type option than mention your output location and give ok.

To the final output vector layer you need clip your common features (Waterbody, Habitation, Railway, Forest, Mining/Industrial, Rock out crop, Gullied Region and Eroded Area) from Soil shape file of your study area.

NOTE: Take the instructions from your officials before clipping the common features.

From the soil shape file copy all polygons of "Water body" & "Habitation" (Before copy the shape file that should be in the **Start Editing** mode).

Copied polygon features paste on to interpolated layer. After you pasted the feature you need to Clip that using clip option which is in Editor Tool as shown in adjacent image.

After clipping the feature you can modify your attribute table and you can calculate the area.
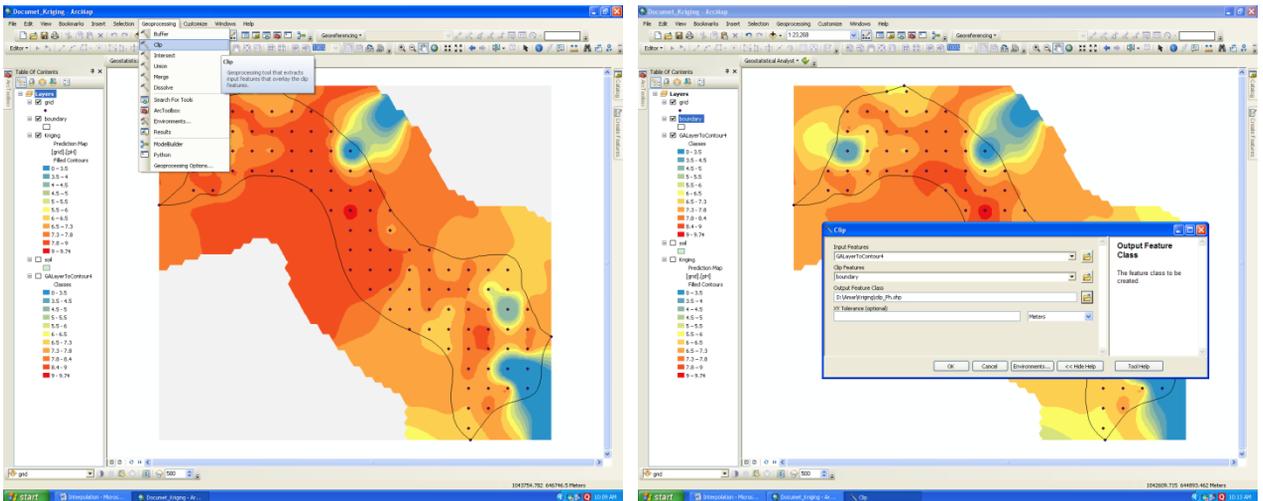
Extract the required study area by using "**clip**" option which is there in the **Geo processing** option. In the option of input feature provide your Interpolated vector file and in the next step provide your boundary (study area) in "clipping feature" and show the output source before giving **Ok**. So that it will clip the required extent of study area.
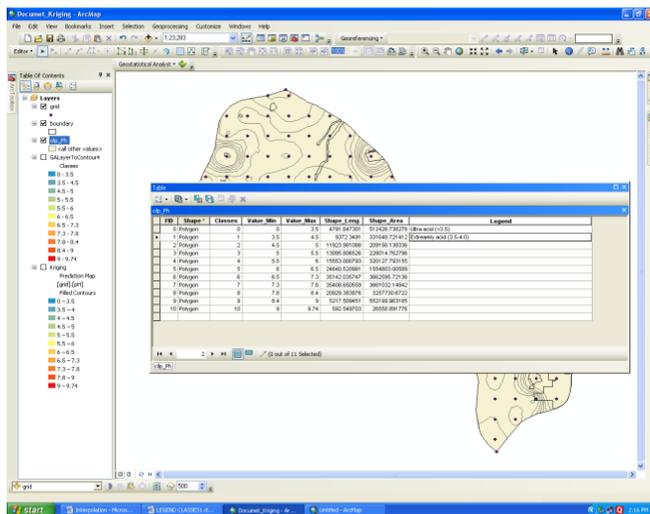
Attributes has to be edit according to your classification. After editing all the required fertility maps can be prepared as shown below :

SOIL REACTION STATUS
Kethahalli Micro-watershed
(4B3E1G2a : Area - 626.26 ha)
Chamarajanagara Taluk
CHAMARAJANAGARA DISTRICT

pH Status	Area in ha.(%)
Strongly acid (5.5-6.0)	0.2 (0.03)
Slightly acid (6.0-6.5)	40 (6.40)
Neutral (6.5-7.3)	175 (27.94)
Slightly alkaline (7.3-7.8)	190 (30.35)
Moderately alkaline (7.8-8.4)	150 (23.90)
Strongly alkaline (8.4-9.0)	48 (7.64)
Very strongly alkaline (>9.0)	2 (0.27)
Others*	22 (3.47)

References
— Drainage
— Road/Cart track
Habitation
Water body
47 Land parcel with No's
Micro-Watershed boundary

* - Habitation & Waterbody

Source: ICAR-NBSS&LUP, Bengalurú

15. Overlay Analysis

Introduction

The heart of GIS is the analytical capabilities of the system. What distinguish the GIS system from other information system are its spatial analysis functions. The analysis functions use the spatial and non-spatial attributes in the database to answer questions about the real world. Geographic analysis facilitates the study of real-world processes by developing and applying models. Such models illuminate the underlying trends in geographic data and thus make new information available. In data analysis the most common operations carried out by GIS are database query, map algebra, and distance-and context-related analysis. The objective of geographic analysis is to transform data into useful information to satisfy the requirements or objectives of decision-makers. Data integration and conversion are only a part of the input phase of GIS. What is required next is the ability to interpret and to analyze the collected information quantitatively and qualitatively.

Overlay analysis

The overlay function creates composite maps by combining diverse data sets. The overlay function can perform simple operations such as laying a road map over a map of soils, or more sophisticated operations such as multiplying and adding map attributes of different value to determine averages and co-occurrences. In GIS, there are basically two different types of overlay operations depending upon data structures.

Simple overlay: The layers were overlaid without assigning any weightage either for layer or classes. Weighted overlay is a technique for applying a common scale of values to diverse and dissimilar input to create an integrated analysis. Geographic problems often require the analysis of many different factors.

Weighted overlay: This process reclassifies values in the input rasters onto a common evaluation scale of suitability or preference, risk, or some similarly unifying scale. The input rasters are weighted by importance and added to produce an output raster. The Weighted Overlay process accepts only discrete rasters (integer values) as input. Continuous rasters should and must be reclassified to discrete rasters before they can be used.

Topological overlays and their functions

An analysis procedure for determining the spatial coincidence of geographic features.

- **Union:** A topological overlay of two polygon coverages, which preserves features that fall within the spatial extent of either input datasets.
- **Identity:** The topological overlay of a line or polygon coverage with another polygon coverage that computes the geometric intersection of the two coverages.
- **Intersect:** The topological integration of two spatial datasets that preserves features that fall within the spatial extent common to both input datasets.

Vector overlay

The vector overlay, however is far more difficult and complex and involves more processing. In simple vector overlay, the layers were overlaid without assigning any weightage either for layer or classes. Weighted overlay is a technique for applying a common scale of values to diverse and dissimilar input to create an integrated analysis. Geographic problems often require the analysis of many different factors. The weighted overlay process reclassifies values in the input rasters onto a common evaluation scale of suitability or preference, risk, or some similarly unifying scale. The input rasters are weighted by importance and added to produce an output raster. Continuous rasters should and must be reclassified to discrete rasters before they can be used.

Raster overlay

It is a relatively straightforward operation and often many data sets can be combined and displayed at once. In raster overlay, the pixel or grid cell values in each map are combined using arithmetic and Boolean operators to produce a new value in the composite map. The maps can be treated as arithmetical variables and perform complex algebraic functions and it is called as map algebra.

Logical Operators

The logical operators are Boolean functions. There are basically four types of Boolean Operators: viz., OR, AND, NOT, and XOR. During vector overlay, map features and the associated attributes are integrated to produce new composite maps. Logical rules can be applied to how the maps are combined. Vector overlay can be performed on different types of map features: viz., Polygon-on-polygon overlay, Line-in-polygon overlay and Point-on-polygon overlay. During the process of overlay, the attribute data associated with each feature type id merged. The resulting table will contain both the attribute data. The process of overlay will depend upon the modelling approach the user needs.

13.1.1 Overlay analysis in ArcGIS: Simple overlay
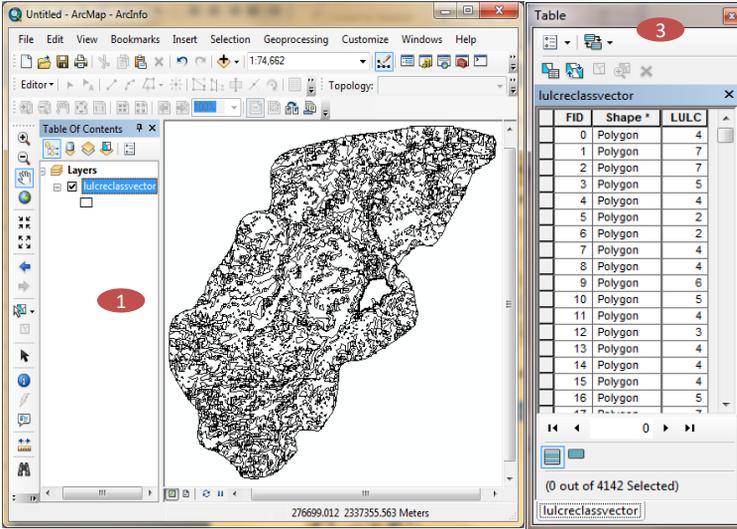
Vector Overlay

STEP 1: Open 1st vector file to be used in Overlay Analysis (Land use/Land Cover in this case)
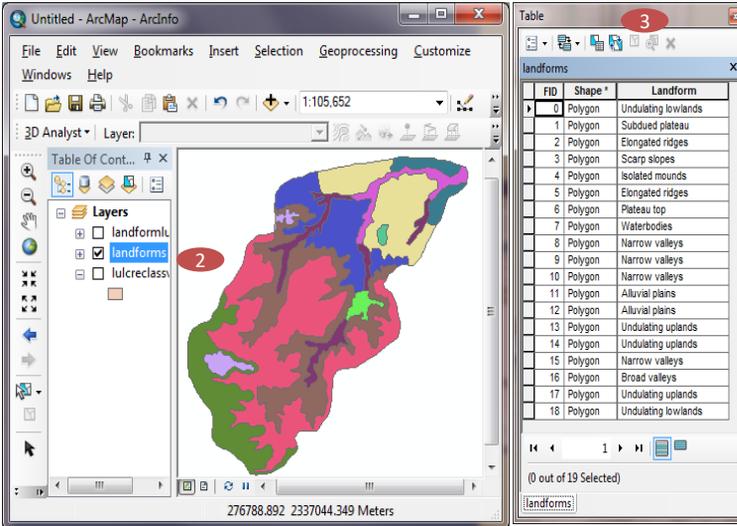
STEP 2: Open 2nd vector file to be used in Overlay Analysis (Landform in this case)

STEP 3: Analyze the attribute table of the two vectors
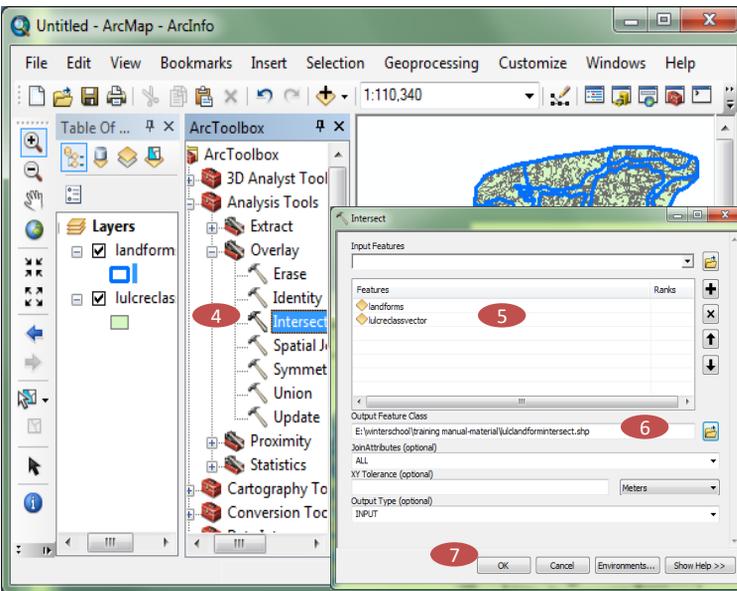
STEP 4: Go to **Arc Tool Box** → **Analysis Tools** → **Overlay** → Select suitable overlay technique (**Intersect in this case**)

STEP 5: Add both **Input Features** to be overlaid

STEP 6: Define **Output Vector**

STEP 7: Click **OK**
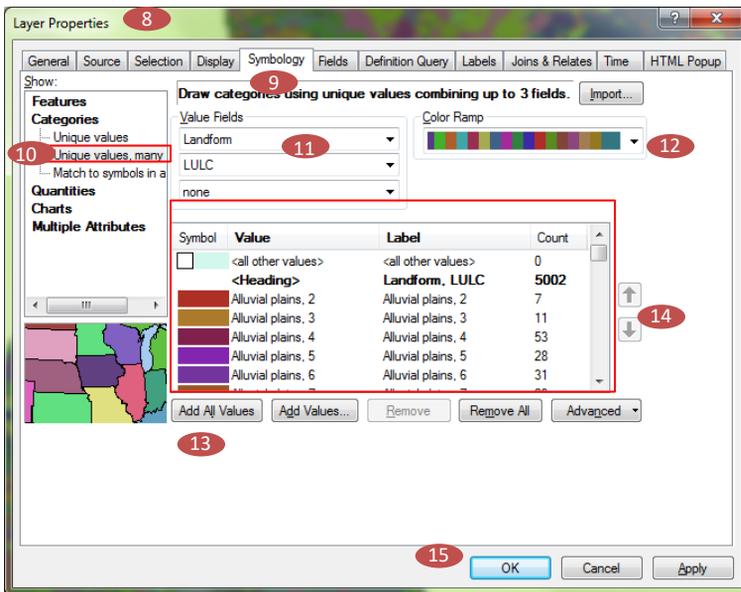
STEP 8: Open **Layer Properties** (by right clicking on the layer and selecting Properties)

STEP 9: Go to **Symbology** Tab

STEP 10: Under Show Table go to **Categories → Unique values, many fields**

STEP 11: Under **Value Fields** Select two fields in drop down

STEP 12: Specify the **Color Ramp**

STEP 13: Click **Add All Values**

STEP 14: See the combined Labels of the two fields

STEP 15: Click **OK**
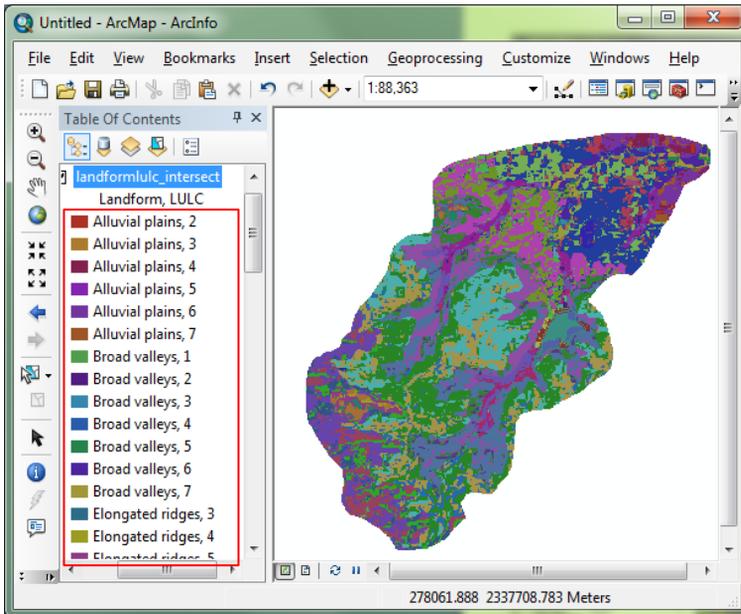
The overlay output with two labels combined is shown in figure

The attribute table of overlaid vectors of landform and land use / land cover after overlay is shown in table

13.1.2 Overlay analysis in ArcGIS: Weighted Overlay

16. Digital Elevation Model (DEM)

Introduction

The most common digital data of the shape of the earth's surface is cell-based digital elevation models (DEMs). This data is used as input to quantify the characteristics of the land surface. A DEM is a raster representation of a continuous surface, usually referencing the surface of the earth. The accuracy of this data is determined primarily by the resolution (the distance between sample points). Other factors affecting accuracy are data type (integer or floating point) and the actual sampling of the surface when creating the original DEM. Errors in DEMs are usually classified as either sinks or peaks. A sink is an area surrounded by higher elevation values and is also referred to as a depression or pit. This is an area of internal drainage. Some of these may be natural, particularly in glacial or karst areas, although many sinks are imperfections in the DEM. Likewise, a spike, or peak, is an area surrounded by cells of lower value.

DEM Readiness:

A Digital Elevation Model is a raster whose pixel values represent elevations of a surface, most commonly the elevations of the ground. Elevation models are a data type that can be derived from satellite and aerial imagery. There are two main types of elevation models: DEMs, which are raster datasets depicting the earth's topography as a regularly spaced grid, and triangular irregular networks (TINs), which connect irregularly spaced elevation points with triangular surfaces. There are many variations of DEMs, with a wide range of applications for many disciplines.

Contours

Contours are lines that connect locations of equal value in a raster dataset that represents continuous phenomena such as elevation, temperature, precipitation, pollution, or atmospheric pressure. The line features connect cells of a constant value in the input. Contour lines are often generally referred to as isolines but can also have specific terms depending on what is being measured. Some examples are isobars for pressure, isotherms for temperature, and isohyets for precipitation. This is useful in identifying suitable locations for projects, such as estimating water storage capacity or approximating the volume of excavation work. In ArcGIS Pro, existing tools and functions can be used to create contour lines, depending on the input and the desired output data.

Slope

The Slope tool identifies the steepness at each cell of a raster surface. The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. The Slope tool uses a 3 by 3 window of cells to compute the value, while the Surface Parameters tool allows window sizes from 3 by 3 to 15 by 15 cells. Larger window sizes are useful with high resolution elevation data to capture land surface processes at an appropriate scale. Surface Parameters also provides an adaptive window option that evaluates the local variability of the terrain and identifies the largest appropriate neighborhood size for each cell. This can be useful with gradual homogeneous terrain interrupted by streams, roads, or sharp breaks in slope.

15.1.1 Creating DEM, Contours & Slope in ArcGIS
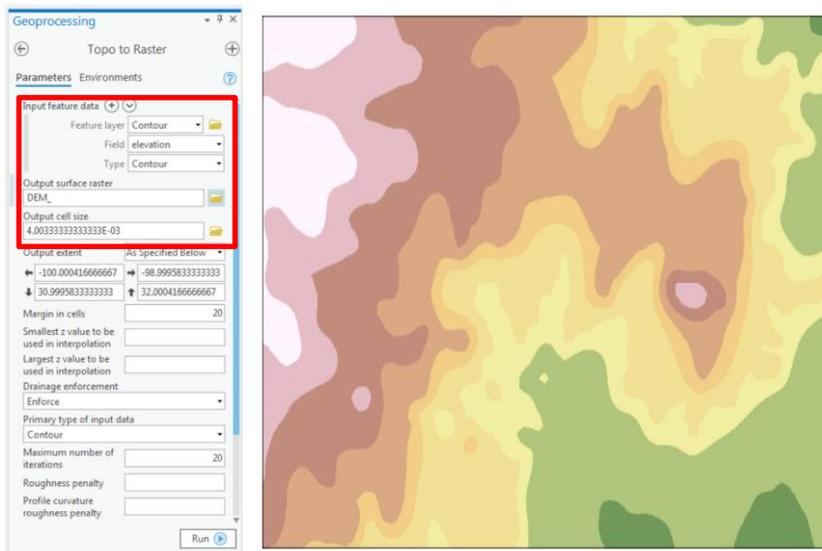
Create DEM

- STEP 1: For Input feature data, select the contour layer for Feature layer, select the field containing the elevation values for Field, and select Contour for Type.
- STEP 2: Set a name and location for Output surface raster.
- STEP 3: Type the desired Output cell size or accept the default.
- STEP 4: Set other parameters as desired and click Run.
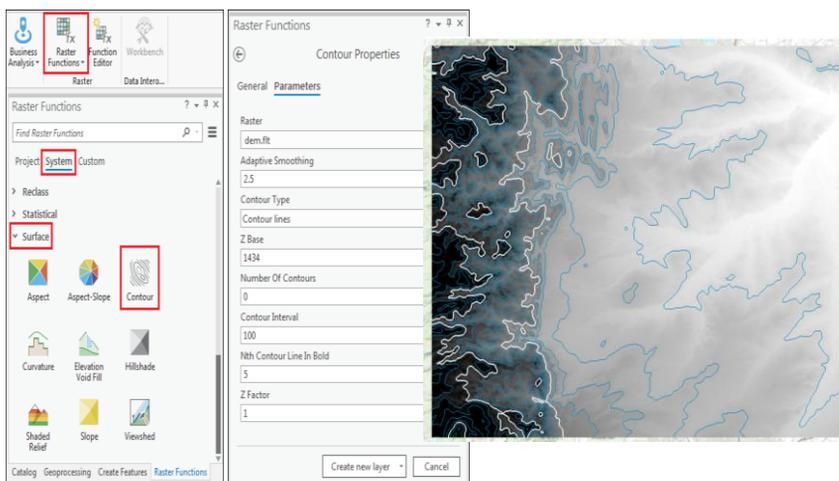
Create Contours

- STEP 1: In ArcGIS Pro, select either the Analysis or the Imagery tab. Click Raster Functions.
- STEP 2: In the Raster Functions pane, click System > Surface > Contour.
- STEP 3: Configure the Contour Properties pane. For Raster, specify the desired single-band raster layer. In this example, it is dem.flt. For Contour Type, ensure the default Contour lines option is selected.
- STEP 4: Specify either the Number Of Contours or Contour Interval field to determine the space between two contour lines. In this example, Contour Interval is specified as 100.
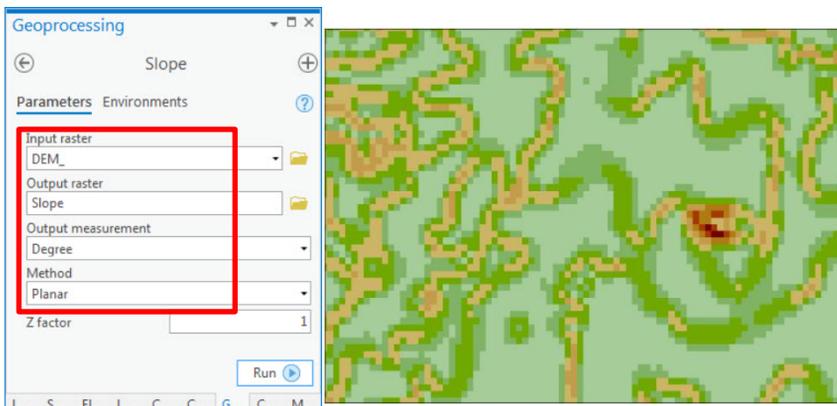
Create Slope

- STEP 1: Select the created DEM layer for Input raster.
- STEP 2: Set a name and location for Output raster.
- STEP 3: Select the desired Output measurement. In this example, Degree is selected.
- STEP 4: Set other parameters as desired and click Run.

17. Drainage Analysis

Introduction

Hydrologic process and water resource issues are commonly investigated by use of distributed watershed models. These watershed models require physiographic information such as configuration of the channel network, location of drainage divides, channel length and slope, and sub catchment geometric properties. Traditionally, these parameters are obtained from maps or field surveys. Over the last two decades this information has been increasingly derived directly from digital representations of the topography. The digital representation of the topography is called a Digital Elevation Model (DEM). The automated derivation of topographic watershed data from DEMs is faster, less subjective and provides more reproducible measurements than traditional manual techniques applied to topographic maps. Digital data generated by this approach also have the advantage that they can be readily imported and analyzed by Geographic Information Systems (GIS). The technological advances provided by GIS and the increasing availability and quality of DEMs have greatly expanded the application potential of DEMs to many hydrologic, hydraulic, water resources and environmental investigations.

DEM

A Digital Elevation Model is a raster whose pixel values represent elevations of a surface, most commonly the elevations of the ground. Elevation models are a data type that can be derived from satellite and aerial imagery.

Fill

Sinks are often errors due to the resolution of the data or rounding of elevations to the nearest integer value. Sinks should be filled to ensure proper delineation of basins and streams. If the sinks are not filled, a derived drainage network may be discontinuous. The Fill tool uses the equivalents of several tools, such as Focal Flow, Flow Direction, Sink, Watershed, and Zonal Fill, to locate and fill sinks. The tool iterates until all sinks within the specified z limit are filled. As sinks are filled, others can be created at the boundaries of the filled areas, which are removed in the next iteration.

Flow Accumulation

The Flow Accumulation tool calculates accumulated flow as the accumulated weight of all cells flowing into each downslope cell in the output raster. If no weight raster is provided, a weight of 1 is applied to each cell, and the value of cells in the output raster is the number of cells that flow into each cell. Flow Accumulation tool with an input weight raster might be to determine how much rain has fallen within a given watershed. In such a case, the weight raster may be a continuous raster representing average rainfall during a given storm. The output from the tool would then represent the amount of rain that would flow through each cell, assuming that all rain became runoff and there was no interception, evapotranspiration, or loss to groundwater. This could also be viewed as the amount of rain that fell on the surface, upslope from each cell.

Watershed tool

A watershed is the upslope area that contributes flow (generally water) to a common outlet as concentrated drainage. It can be part of a larger watershed and can also contain smaller watersheds, called subbasins. The boundaries between watersheds are termed drainage divides. The outlet, or pour point, is the point on the surface at which water flows out of an area. It is the lowest point along the boundary of a watershed. Watersheds can be delineated from a DEM by computing the flow direction and using it in the Watershed tool. Delineation of watersheds can take place at different spatial scales. A large watershed may cover an entire stream system and, within the watershed, there may be smaller watersheds, one for each tributary in the stream system. Watersheds can be as small as a footprint or large enough to encompass all the land that drains water into rivers that drain into Bay or sea or ocean.

16.1.1 Creating Watershed in ArcGIS
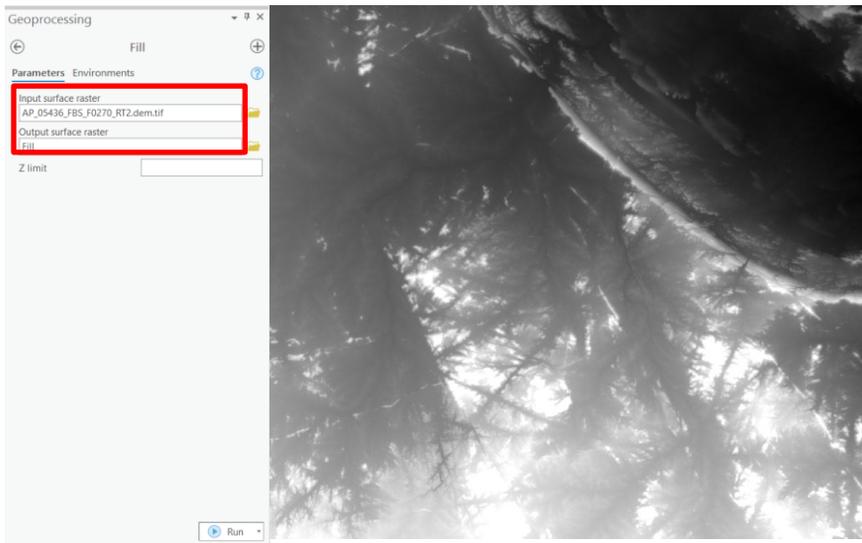
Create Fill

- STEP 1: In Catalog Pane, navigate to Toolboxes > System Toolboxes > Spatial Analyst Tools > Hydrology > Fill.

- STEP 2: Use a digital elevation model (DEM) as the 'Input surface raster'.

- STEP 3: Verify the path name for the 'Output surface raster'.

- STEP 4: Input the Z-limit, if necessary and Click OK.
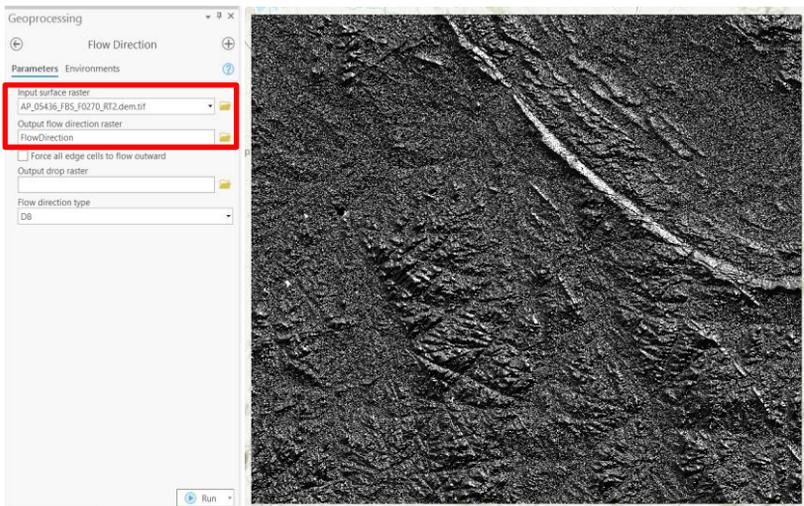
Create Flow Direction

- STEP 1: In Catalog Pane, navigate to Toolboxes > System Toolboxes > Spatial Analyst Tools > Hydrology > Flow Direction.

- STEP 2: Use the DEM output from Step 1 as the 'Input surface raster'.

- STEP 3: Verify the path name for the 'Output flow direction raster'and Click OK.
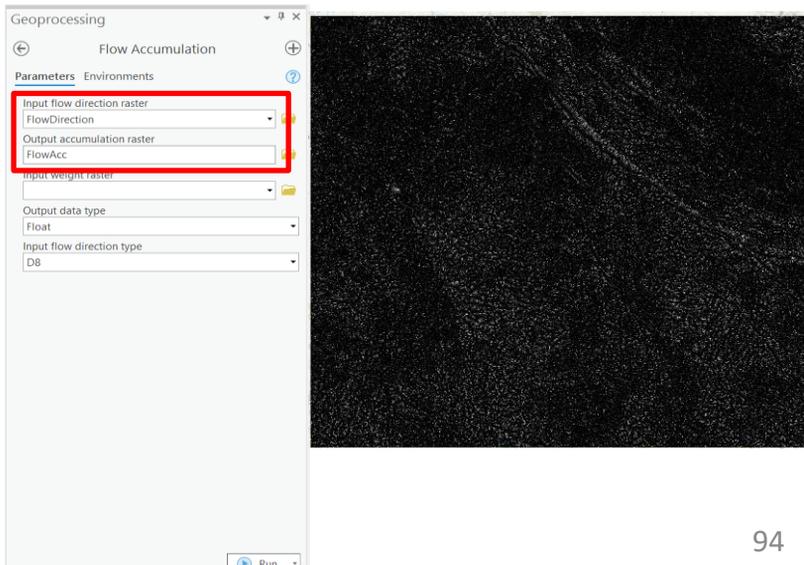
Create Flow Accumulation

- STEP 1: In Catalog Pane, navigate to Toolboxes > System Toolboxes > Spatial Analyst Tools > Hydrology > Flow Accumulation.

- STEP 2: Use the output raster from flow direction as the 'Input flow direction raster'.

- STEP 3: Verify the path name for the 'Output accumulation raster' and Click OK.

16.1.1 Creating Watershed in ArcGIS
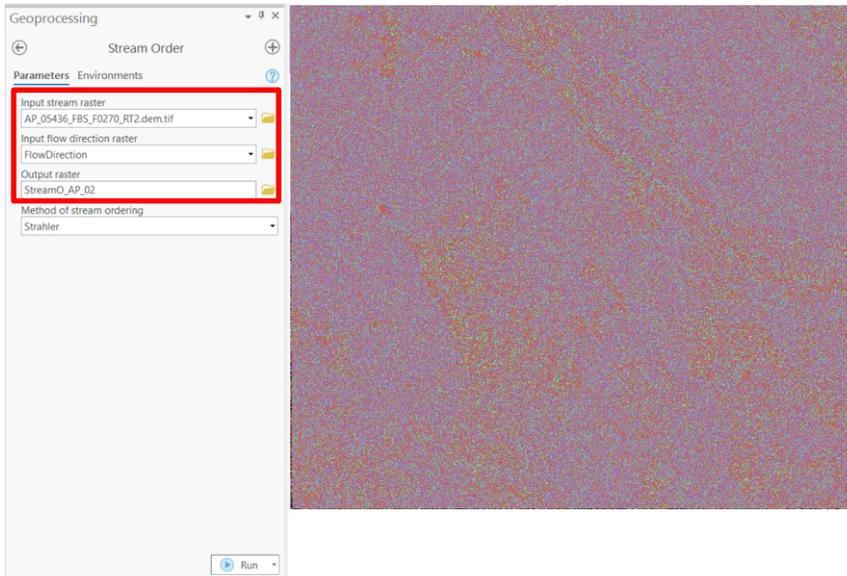
Create Stream Order

- STEP 1: In Catalog Pane, navigate to Toolboxes > System Toolboxes > Spatial Analyst Tools > Hydrology > Stream Order.

- STEP 2: Use a Stream raster as the 'Input stream raster'.

- STEP 3: Verify the path name for the 'Output raster'.

- STEP 4: Input the stream ordering method and Click OK.
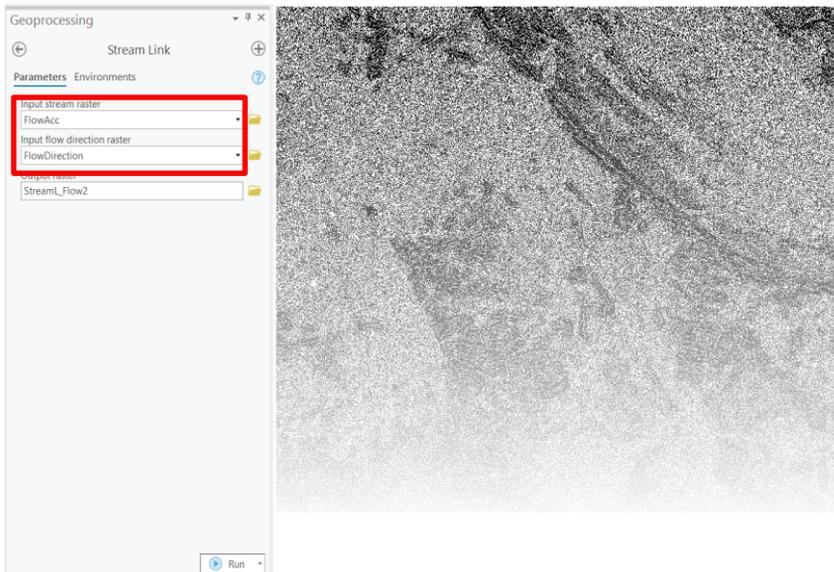
Create Stream Link

- STEP 1: In Catalog Pane, navigate to Toolboxes > System Toolboxes > Spatial Analyst Tools > Hydrology > Stream Link.

- STEP 2: Use a Stream raster as the 'Input stream raster'.

- STEP 3: Verify the path name for the 'Output raster'.

- STEP 4: Input the stream ordering method and Click OK.

18. Digitization

Introduction

Digitizing in GIS is the process of "tracing", in a geographically correct way, information from images/maps. The process of representing an analogue signal or an image by a discrete set of its points is known as Digitizing. This data after conversion is in the binary format, which is directly readable by computer. The data to be converted can be a text, an image, audio or a video. The analogue signals are variable whereas the digital format is the discrete one. These discrete units are called as bits. These bits (8) organized in groups are known as byte. The digital signals are mainly represented in the form of sequence of integers. These integers can be converted back to analogue signal that are approximately similar to the original analogue signals.

Manual digitization: Manual digitizing is done using digitizing tablet for entering maps into GIS. The map to be digitized is affixed to a digitizing table, and a pointing device (called the digitizing cursor or mouse) is used to trace the features of the map. These features can be boundary lines between mapping units, other linear features (rivers, roads, etc.) or point features (sampling points, rainfall stations, etc.). The digitizing table electronically encodes the position of the cursor with the precision of a fraction of a millimeter. The most common digitizing table uses a fine grid of wires, embedded in the table. The vertical wires will record the Y-coordinates, and the horizontal ones, the X-coordinates. Digitizing in GIS is the process of "tracing", in a geographically correct way, information from images/maps. The operator manually traces all the lines from his hardcopy map and creates identical digital map on the computer. It is very time consuming and level of accuracy is also not very good. In this case the map is placed on the digitizer, the corners are selected by a mouse click and the respective coordinates entered through the keyboard. This process is called "'registering a map'". The advantage of this method is that the user is always in control of the whole map. However, the high price of the equipment and the lack of magnification capabilities, e.g. for low-quality maps, are significant disadvantages. Furthermore, the map must not show any signs of lateral stretch to prevent wrong site coordinates.

On-screen or Heads-up digitizing: It is similar to manual digitizing in the way that lines have to be drawn manually but directly on the computer screen. This method is less time consuming and provides increased level of accuracy. In this method, user display an aerial photograph, satellite image, or orthophotograph on-screen as a base map, and then user can draw features, such as roads, buildings, or parcels, on top of it. On-screen digitizing requires a scanned and geocoded raster map (image). All relevant map objects will be vectorized using the mouse. It is not necessary to register such a map as it is geocoded already. The advantage of this method is the support of zooming, hence an improved accuracy can be achieved.

Interactive tracing method: It is improvement over Heads-up digitizing in terms of speed and accuracy.

Automatic Digitizing: It is automated raster to vector conversion using image processing and pattern recognition techniques. In this technique computer traces all the lines, which results in high speed and accuracy along with improved quality of images.

When digitizing user can digitize points, lines, or polygons. By digitizing these features, user make them available for mapping once the tabular data to the attribute table is added. Most of the software packages allow creating point, line, or polygon layer files separately; however, some like TNTMips allows creating all three types of layers in a single file.

While digitizing snapping allows user to create features that connect to each other so that edits are more accurate and have fewer errors. With snapping, pointer will jump, or snap to, edges, vertices, and other geometric elements when it nears them. This enables to position a feature easily in relation to the locations of other features. As the operator moves the pointer around the map, it snaps automatically to points, endpoints, vertices, and edges. All softwares' enable snapping while digitizing. Once created features, users should be able to edit their attribute and shapes, for example, to split a polygon into two polygons, to draw in a new shape for a line or polygon, to shift, add, or remove vertices, to merge lines or polygons, etc. the user should also be able to repair topological and geometrical problems in the layers.

There are a few basic rules regarding digitizing from analogue maps that apply to topological GIS. Obedience to these rules is essential in order to use the topological features of the software.

• Lines must not cross without nodes.

• Lines that use common "node" must hit it exactly. Make use of the snapping function of the digitizing module.

• Common area boundaries must only be digitized once.

• Areas must be closed. Use the snapping function of the digitizing module.

• Store lines and areas in separate maps to avoid problems with crossing lines. It is also easier to assign a certain type (line or area boundary) to a line.

6.1 Digitization in ArcGIS
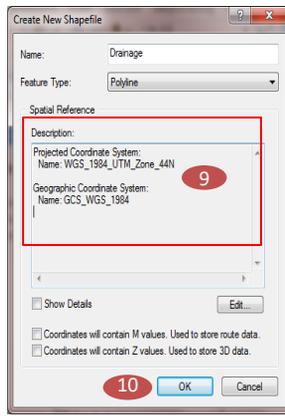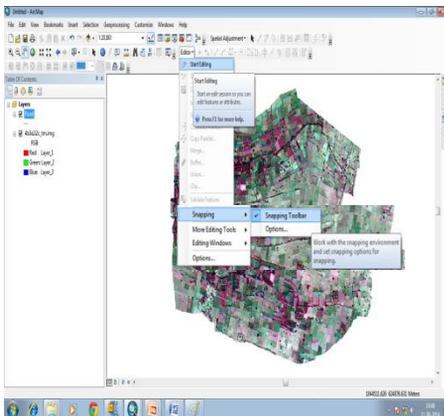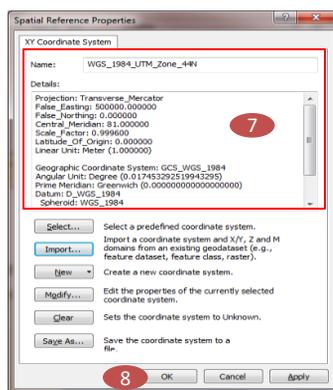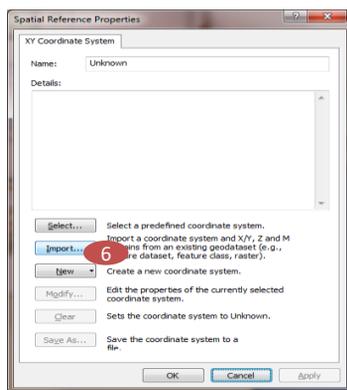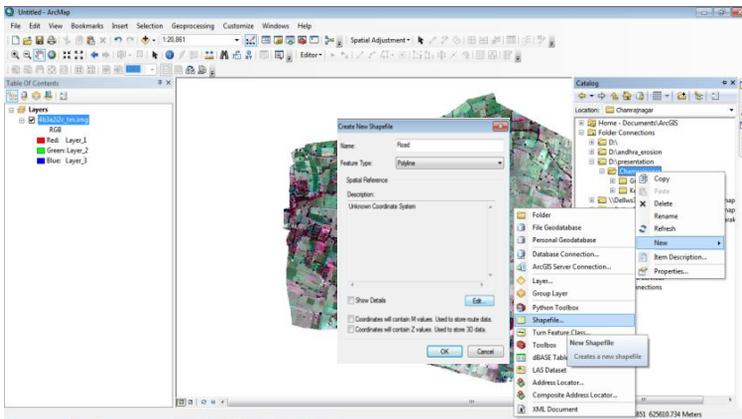
STEP 1: Add the base map (scanned soil map/ satellite image/ toposheet)

STEP 2: In **Arc Catalogue** right click on the folder to which user want to create the shape file and go to **New → Shapefile**

STEP 3: In the **Create new Shapefile wizard** define the **Name** of the layer

STEP 4: Select **Feature Type** as per requirement (**Polyline/ Polygon/ Point)**

STEP 5: Click on **Edit**

STEP 6: In **Spatial Reference Properties** window click on **Import** and browse to the base map

STEP 7: Check the **XY Coordinate System Name and Details**

STEP 8: Click **OK**

STEP 9: Check the Projection details

STEP 10: Click **OK**

STEP11: Click on **Editor** and select **Snapping → Snapping Toolbar**

STEP 12: Click on **Editor** and select **Start Editing**

STEP 13: Start all snapping tools by clicking on them

STEP 14: In **Create Feature** select the layer

STEP 15: Under **Construction Tools** select **Line**

STEP 16: Select **Straight Segment Tool** from **Editor** and start digitizing

STEP 17: After finishing digitization, click on **Editor** and select **Save Edit** and **Stop Edit**
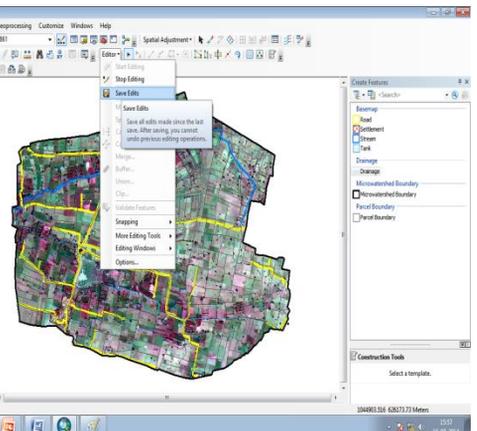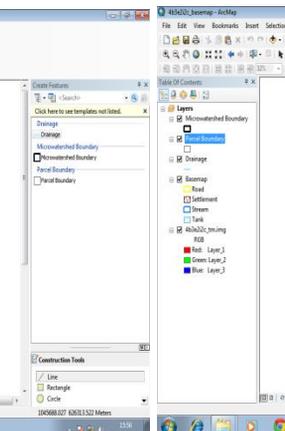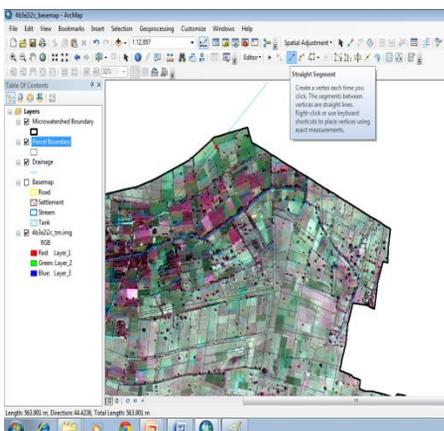
19. Thematic Mapping

Introduction

A thematic map is a map that emphasizes a particular theme or special topic such as the soil depth in an area. They are different from general reference maps because they do not just show natural features like rivers, cities, political subdivisions and highways. Instead, if these items are on a thematic map, they are simply used as reference points to enhance one's understanding of the map's theme and purpose. The goal of a thematic map is to draw the reader's attention to the significance in the distribution of one or few geographic phenomena. The map maker must therefore direct the reader's attention using carefully considered symbology and labelling. This task becomes substantially more difficult if more than one theme of information is included on a thematic map, e.g., soil depth map. The emphasis of thematic maps is on the geographic pattern of the feature attributes. One challenge in making thematic maps is to figure out which features to include as the minimal required locational reference information. In contrast, for general reference maps, the challenge is to figure out which classes of features are of greatest interest and use to a wide range of users. In thematic mapping symbology is a powerful tool for displaying the spatial data in GIS. It allow user to create simple or complex maps that are cartographically pleasing.

Types of thematic maps

Although cartographers can use these datasets in many different ways to create thematic maps, there are five thematic mapping techniques that are used most often.

The first and most commonly used of these is the choropleth map. This is a map that portrays quantitative data as a color and can show density, percent, average value or quantity of an event within a geographic area. Sequential colors on these maps represent increasing or decreasing positive or negative data values. Normally, each color also represents a range of values.

Proportional or graduated symbols are the next type of map and represent data associated with point locations such as cities. Data is displayed on these maps with proportionally sized symbols to show differences in occurrences. Circles are most often used with these maps but squares and other geometric shapes are suitable as well. The most common way to size these symbols is to make their areas proportional to the values to be depicted with mapping or drawing software.

Another thematic map is the isarithmic or contour map and it uses isolines to depict continuous values like precipitation levels. These maps can also display three-dimensional values like elevation on topographic maps. Generally data for isarithmic maps is gathered via measureable points (e.g. - weather stations) or is collected by area (e.g. - tons of corn per acre by county). Isarithmic maps also follow the basic rule that there is a high and low side in relation to the isoline. For example in elevation if the isoline is 150 m then one side must be higher than 150 m and one side must be lower.

A dot map is another type of thematic map and uses dots to show the presence of a theme and display a spatial pattern. On these maps, a dot can represent one unit or several, depending on what is being depicted with the map.

Finally, dasymetric mapping is the last type of thematic map. This map is a complex variation of the choropleth map and works by using statistics and extra information to combine areas with similar values instead of using the administrative boundaries common in a simple choropleth map.

Reclassification

Reclassification operations can also relate to location, as well as purely thematic. One such characteristic is position. An overlay category represented by a single "point" location, for example, might be reclassified according to its latitude and longitude. Similarly, a line segment or area feature could be reassigned values indicating its center or general orientation.

Symbology

Features (Single Symbol), this is the simplest method of symbolizing the data. It is used when there is no need to distinguish between categories within an attribute field. For example, soils (polygons), railroads (lines), and dam locations (points) are typically displayed using a single symbol. This means that every feature in that dataset uses the same symbol.

Categories

• *Unique Values:* This method of symbolizing is appropriate when user would like to group features based on an attribute in the table.

• *Quantities Graduated Colors and Symbols:* If user have quantities in attribute table and would like to display these quantities in relationship to specific features, user can use graduated colors, graduated symbols, proportional symbols, or dot density.

• *Multiple Attributes Quantity By Category:* Sometimes it may be necessary to symbolize on two or more fields, with each field symbolized differently.

14.1 Thematic mapping in ArcGIS
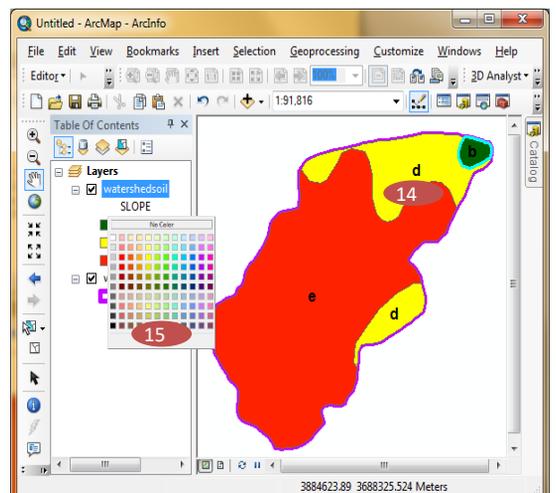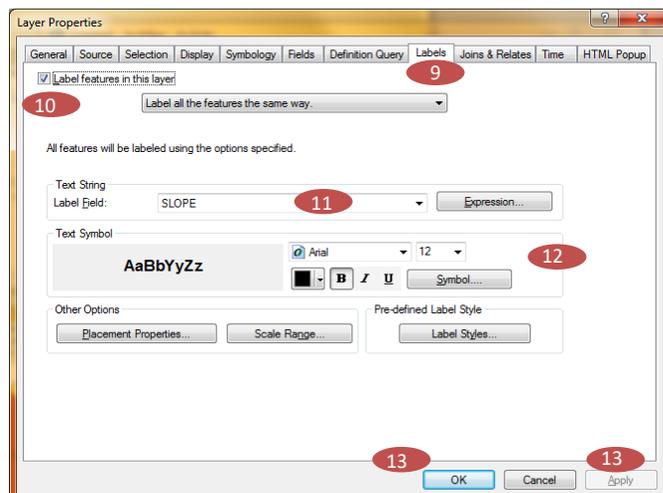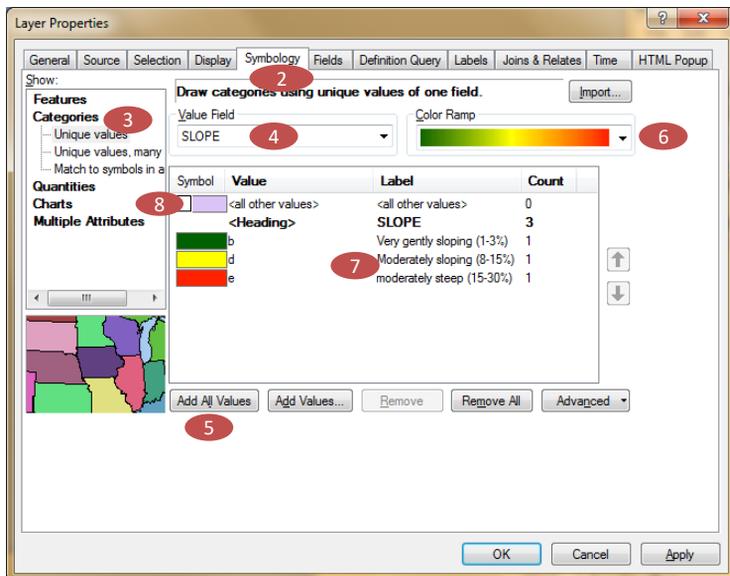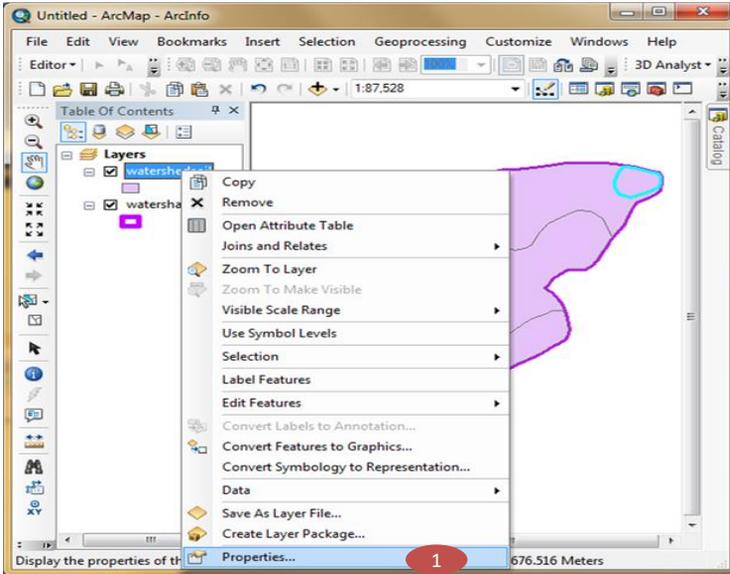
STEP 1: Right Click on the layer you want to map and click on **Properties**

STEP 2: In the **Layer properties table**, go to **Symbology Tab**

STEP 3: In the **Show list,** go to **Categories → Unique values**

STEP 4: In the **View Field,** select the theme to be mapped, e.g., Slope in this case

STEP 5: Click on **Add all Values**

STEP 6: In the **Color Ramp select** the suitable color ramp for the classes of the themes

STEP 7: Rename the **Labels if needed**

STEP 8: Uncheck the **All other value** button

STEP 9: In the **Layer Properties table**, go to **Label Tab** to display the lables

STEP 10: Check the button **Label Features in this layer**

STEP 11: Select **Label Field** *i.e.*, Slope

STEP 12: Under **Text Symbol** define **Font Type** and **Size**

STEP 13: *Click* **Apply** and **OK**

STEP 14: Look for **Label Names**, **Color Boxes** and **Features Labeled**

STEP 15: User can change the color of classes by right clicking on the particular **color** to select the appropriate color

20. Map Composition

Cartographic principles in map composition in GIS

Traditionally, maps have been created to serve two main functions. The first function has been to store information. Creating a map has been a way to record information for future reference. The second function has been to provide a picture to relay spatial information to a user. There are many types of maps, each with general and possibly specific requirements. However, in GIS user is having the flexibility in map composition using various map composition tools in GIS. These guidelines have been organized into seven areas that one can use as a checklist for creating or improving maps.

- **Purpose:** Typically, a map does not have more than one purpose. Trying to communicate too much in one map—having more than one purpose for the map—tends to blur the message and confuse the map reader. Using two or more maps, each focused on a single message, is always a better strategy.

- **Audience:** If map is designing for a few readers or for a large audience of hundreds or millions of people? accordingly , maps to be generated to understand map's message clearly.

- **Size and scale**: The physical size of a map relative to the geographic extent shown on the map will dictate the scale of the map and determine how user will represent the actual size and number of features shown on the map. Data is often collected at a particular scale. However, if user is not displaying the data at that scale, be sure user data "fits".

- **Media**: Media also plays an important role, because a map printed on newspaper will not show fine details clearly, whereas one printed on high-quality paper will. In addition, the details on a digital map could vary depending on the viewing program.

- **Focus:** Refers to where the designer wants the map reader to first focus. Typically, cool colors (blues, greens, and light gray) are used for background information and warm colors (red, yellow, black) are used to capture the reader's attention.

- **Integrity:** User may want to cross-validate some of your information, such as the names or spelling of some features. If the data was produced by another organization, it is customary to give that organization credit on the map.

- **Balance:** How does map look on the page or screen?. Are the parts of the map properly aligned?. The body of the map should be the dominant element. Try to avoid large open spaces. Be flexible in where user place elements (that is, not all titles need to go at the top).

- **Completeness:** A map generally should contain some basic elements, such as a title, legend, scalebar, and North arrow; however, there are exceptions. For example, if a graticule exists, it is not necessary to place a North arrow. Basically, place all the information , which readers need to fully understand the map.

Map elements

- **Map body:** The primary mapped area. One can display more than one image in primary mapped area within the document. For example, to portray change by showing several images with differing but related information, such as population maps of various years. Map may also contain a locator map, an inset map or an index map. All are used to assist in communicating your information to others.

- **Title:** Used to tell the reader what the map represents. This is often placed on a map layout as text.

- **Legend:** Lists the symbology used within the map and what it represents.

- **Scale:** Provides readers with the information they need to determine distance. A map scale is a ratio, where one unit on the map represents some multiple of that value in the real world.

- **Projection:** A mathematical formula that transforms feature locations from the earth's curved surface to a map's flat surface. Projections can cause distortions in distance, area, shape, and direction; all projections have some distortion. Therefore, the projection type is often placed on the map to help readers determine the accuracy of the measurement information they get from the map.

- **Direction:** Is shown using a North arrow. A map may show true north and magnetic north. This element is inserted in the map layout view.

- **Data source:** The bibliographic information for the data used to develop the map.

15.1 Map composition in ArcGIS
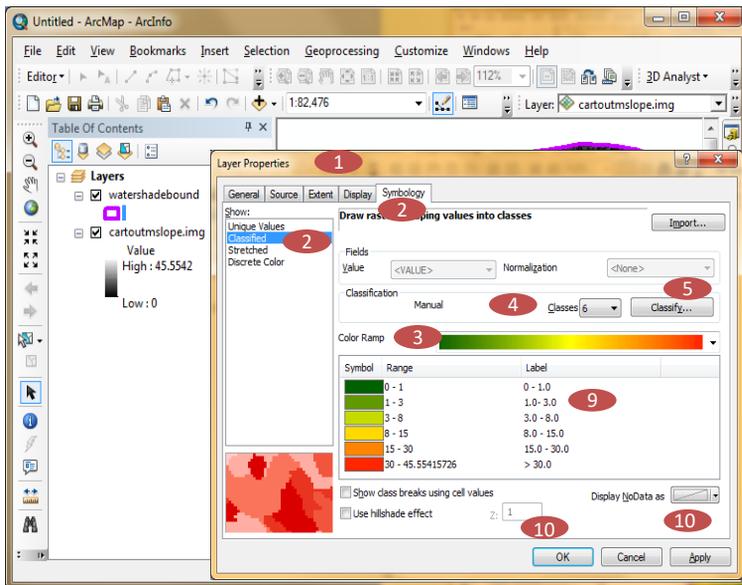
STEP 1: Right click on the layer to be mapped and select **Properties** to open the **Layer Properties** window

STEP 2: Go to **Symbology** Tab and select **Classified** in the show list

STEP 3: Select appropriate color ramp

STEP 4 : Select number of classes in the class drop down

STEP 5: Click on **Classify** tab to open **Classification** window
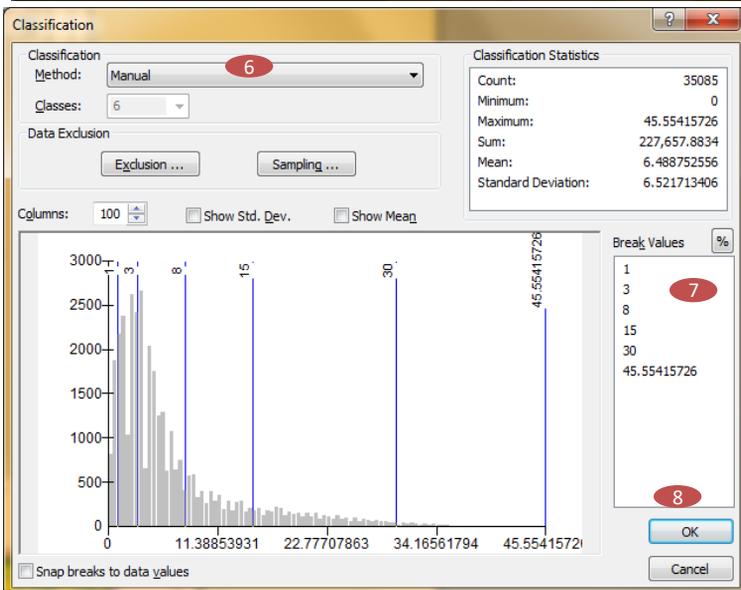
STEP 6: Select Classification Method as **Manual**

STEP 7: Modify break values as shown

STEP 8: Click **OK**

STEP 9: In the **Layer Properties** window recode the **Labels** as shown

STEP 10: Click **Apply** and **OK**
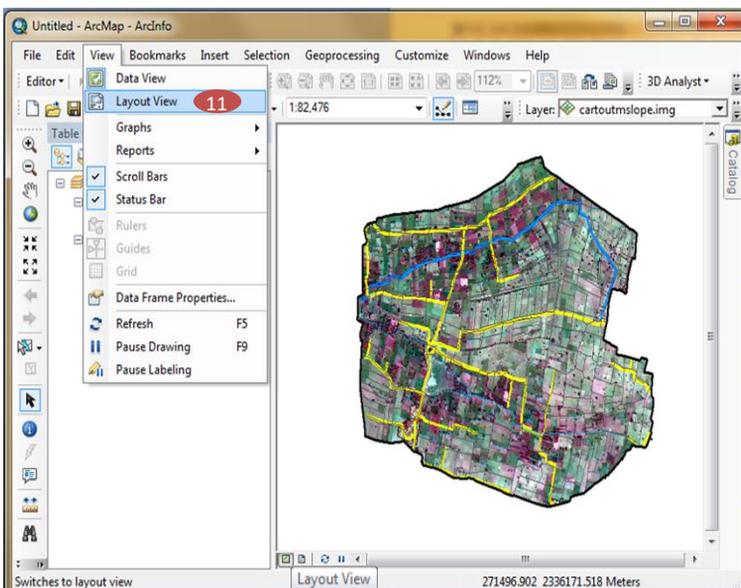
STEP 11: Go to the **View Tab** on the Menu Bar and select **Layout View**
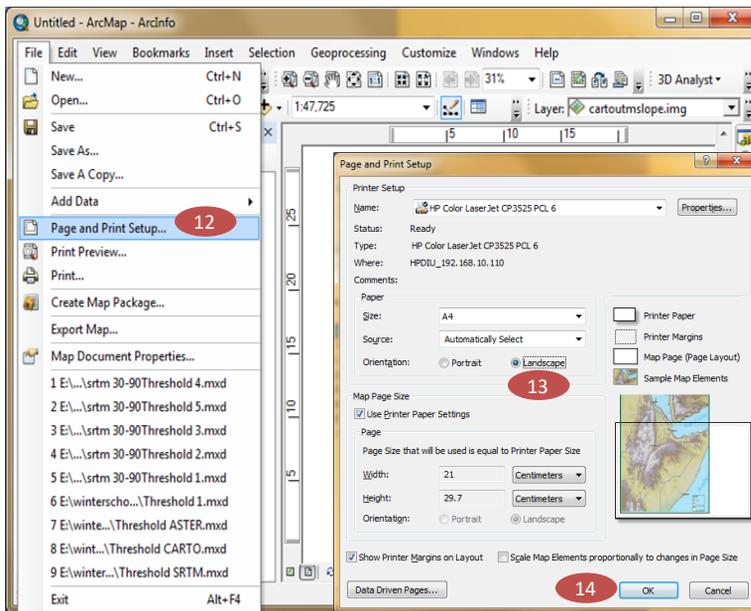
STEP 12: Go to the **File** Tab on the Menu Bar and select **Page and Print Setup** to define the output page setup

STEP 13: Select **Orientation** as **portrait/ landscape**

STEP 14: Click **OK**
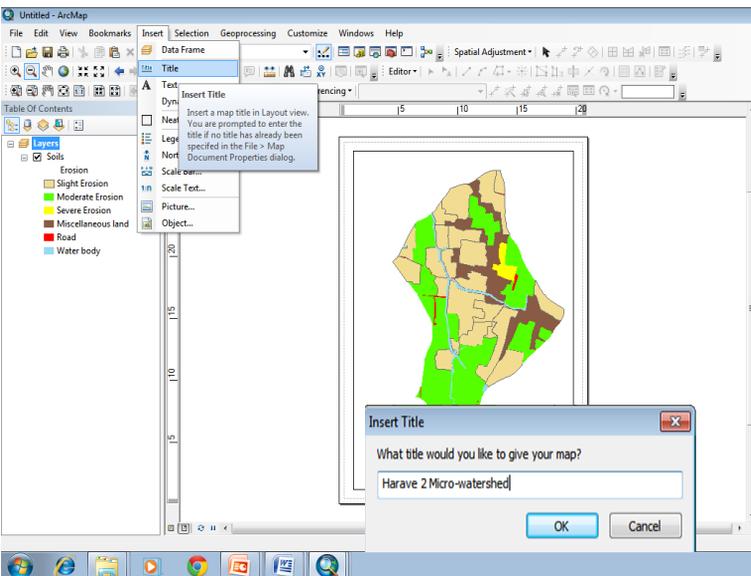
Fix the image in position with zoom out, zoom in, and pan tools

STEP 15: Go to **Insert → Title**

STEP 16: Write the **Title** of the map

STEP 17: Click **OK**
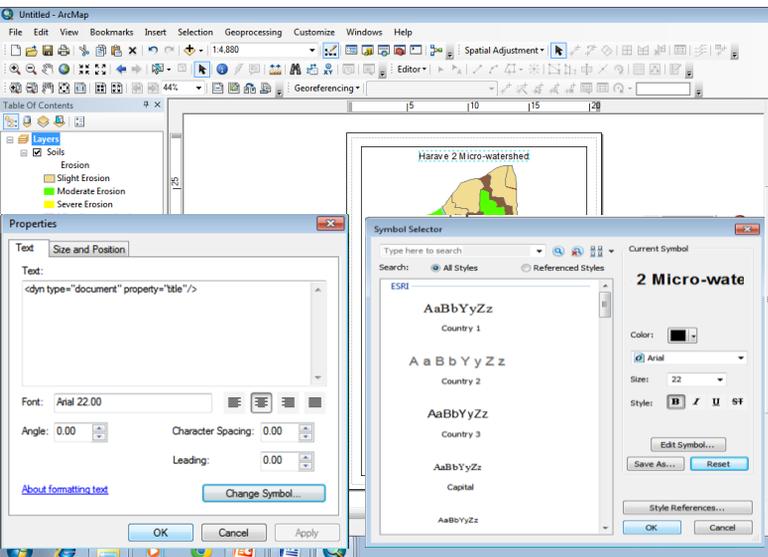
STEP 18: Fix the **Title** in appropriate position by dragging it and double click on it to open the **Properties** Window

STEP 19: Click on **Change Symbol** to open **Symbol Selector** window

STEP 20: Define the **Font Color, Type, Size** and **Style**

STEP 21: Click **OK** on both windows
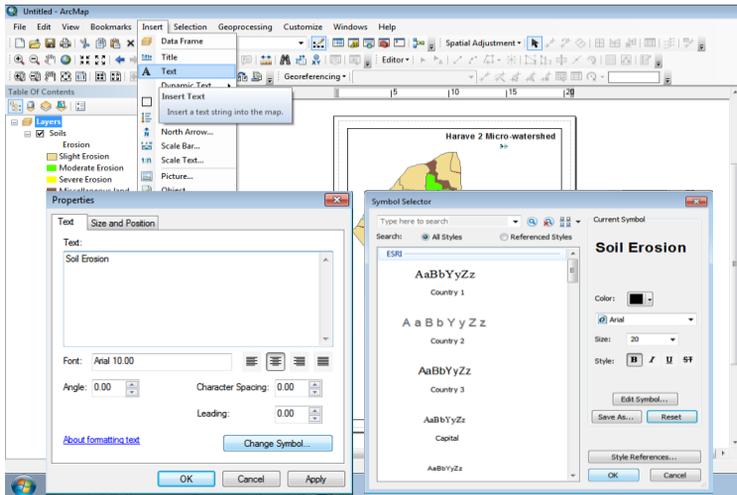
STEP 23: Go to **Insert → Text**

STEP 24: Drag the text to appropriate location and double click on the text to open **Properties window**

STEP 25: Write the **Theme name (Slope in this case)** under the **Text**

STEP 26: Click on **Change Symbol** to open **Symbol Selector** window

STEP 27: Define the **Font Color, Type, Size and style**

STEP 28: Click **OK** on both windows
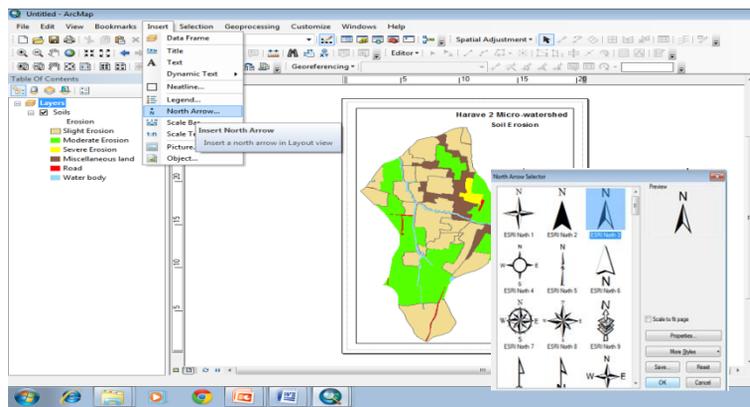
STEP 29: Go to **Insert → North Arrow**

STEP 30: Select the suitable **North Arrow** and Click **OK**
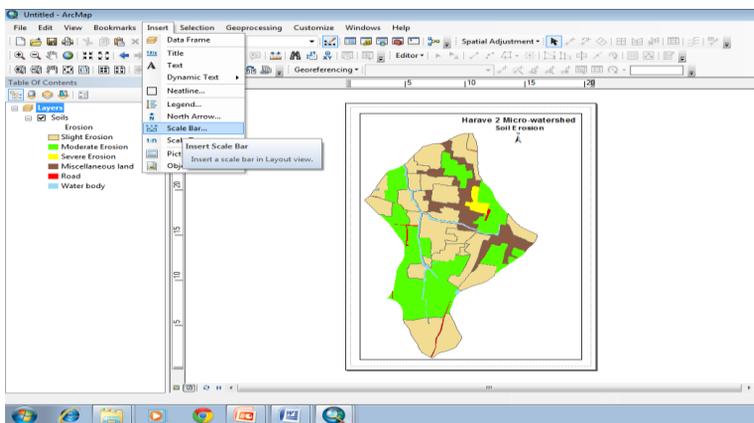
STEP 31: Drag the north arrow to appropriate position

STEP 32: Go to **Insert → Scale Bar**
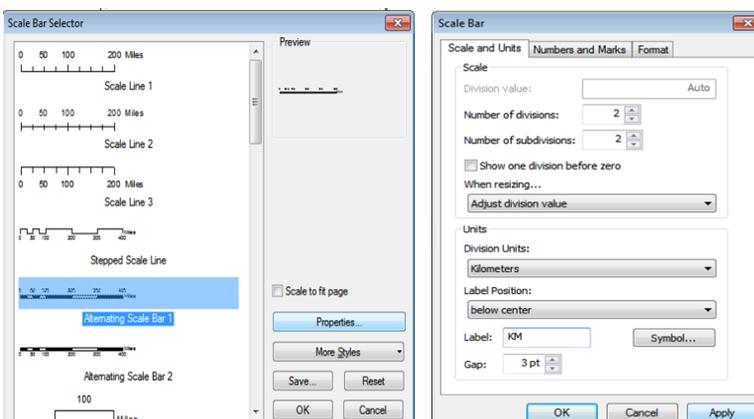
STEP 33: Select a scale bar and click **Properties** to open **Scale Bar** window

STEP 34: Select **Number of Division and Subdivision**

STEP 35: Select suitable **Division Unit** from the drop down menu (**Kilometers in this case)**

STEP 36: Write suitable **Label** (**"km"** in this case**)**

STEP 37: Click **OK**
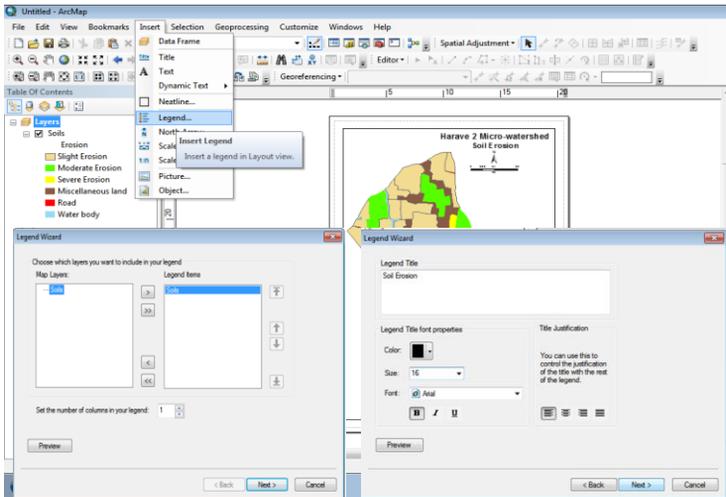
STEP 38: Drag the **Scale Bar** to appropriate position

STEP 39: Go to **Insert → Legend** to open **Legend Wizard**

STEP 40: Select the layer whose legend is to be added and click the forward arrow to add it in **Legend Items**

STEP 41: Click **Next**

STEP 42: Write the **Legend Title**

STEP 43: Define **color, Size, Font, and style**

STEP 44: Click **Next**
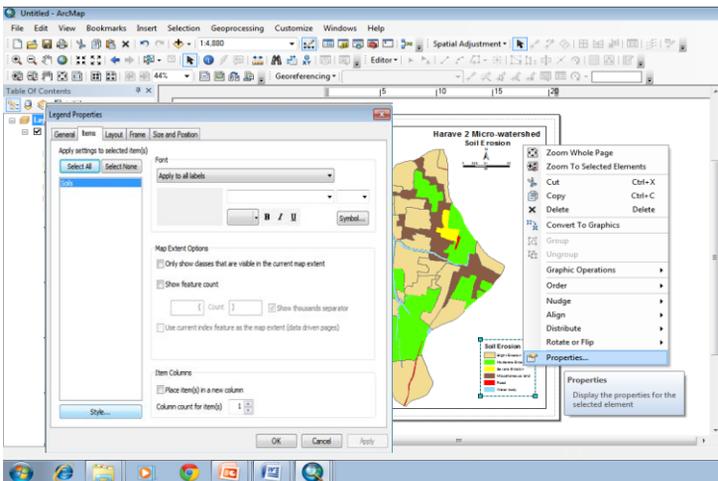
STEP 45: Drag the **Legend** to appropriate position and double click on it to get Legend Properties dialogue box

STEP 46: Go to **Item** tab

STEP 47: Click on **Style**

STEP 48: Select the desired style and click on **Properties**

STEP 49: In the **legend Item** dialogue box go to **General** tab

STEP 50: Define **color, Size, Font, and style**

STEP 51: Click **OK** on all four windows

STEP 52: Go to **File** and click on **Export Map** to export the map in **.jpg, .png, .gif, .tif , .bmp, or .pdf** format
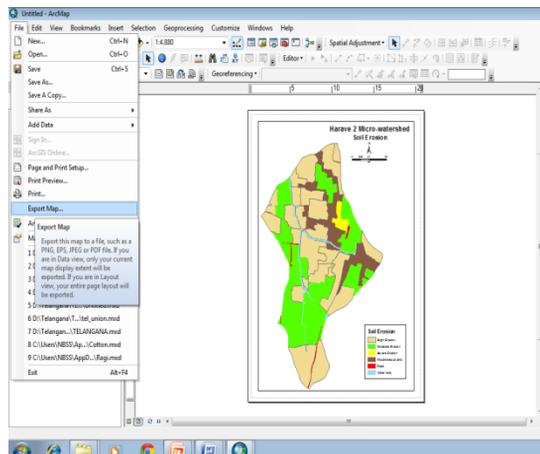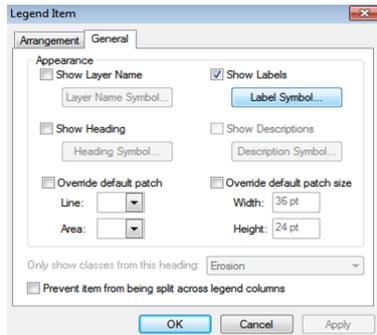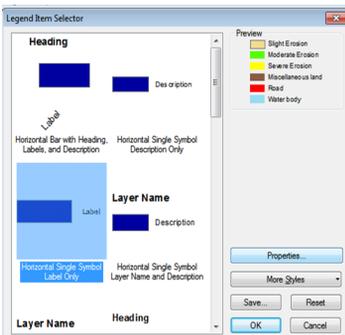
21. Glossary of GIS

Glossary of GIS

Attribute: a characteristic which describes a feature. Attributes can be thought of as questions which are asked about the feature.

Arc: a line connecting a set of points.

Buffer: A zone of a specified distance around coverage features. Both constant and variable width buffers can be generated for a set of coverage features based on each features attribute values.

CAD: Computer Aided Design. An automated system for the design, drafting and display of graphically oriented information.

Cadastre: Public record of the extent, value and ownership of land within a district for purposes of taxation.

Cartesian coordinate system: A two dimensional coordinate system in which x measures horizontal distance and y measures vertical distance. An x,y coordinate defines every point on the plane.

Clip: The spatial extraction of those features from one coverage that reside entirely within the boundary defined by features in another coverage. Clipping works much like a cookie cutter.

COGO: Abbreviation for the term COordinate GeOmetry. Land surveyors use COGO functions to enter survey data, to calculate precise locations and boundaries, to define curves, and so on.

Contour line: An imaginary line joining points of equal elevation.

Control points: A set of points on the ground whose horizontal and vertical location is known. Control points are used as the basis for detailed surveys.

Coordinate: An x,y location in a Cartesian coordinate system or an x,y,z coordinate in a three dimensional system. Coordinates represent locations on the Earth's surface relative to other locations.

Coverage: A digital version of a map forming the basic unit of vector data storage in ARC/INFO. A coverage stores map features as primary features. A coverage usually represents a single theme, or layer, such as soils, roads, or land use.

Coverage units: The units (e.g., feet, meter, inches) of the coordinate system in which a coverage is stored.

Data: raw collection of facts.

Data conversion: The translation of data from one format to another. ARC/INFO supports data conversion from many different geographic data formats in addition to routines for converting paper maps. Those data formats include DLG, TIGER, DXF, and DEM.

Database: A logical collection of interrelated information, managed and stored as a unit. A GIS database includes data about the spatial location and shape of geographic features recorded as points, lines, and polygons as well as their attributes.

Database management system: a set of programs for features drawn and characterized with scale and direction to represent the real world.

Datum: A set of parameters and control points used to accurately define the three dimensional shape of the Earth. The corresponding datum is the basis for a planar coordinate system.

DBMS: Data Base Management System. Software that manages, manipulates and retrieves data in a database.

DGPS: Differential Global Positioning System. A positioning procedure that uses two receivers, a rover at an unknown location and a base station at a known, fixed location. The base station computes corrections based on the differences between its actual and observed ranges to the satellites being tracked.

Digital map library: A series of directories and subdirectories designed to uniformly organize a collection of spatial data. Map libraries organize geographic data spatially as a set of tiles and thematically as a set of layers.

Digital Elevation Model (DEM): Terrain elevation data organized by quadrangle and provided in digital form.

Digital Terrain Model (DTM) : A three-dimensional model of the Earth's surface, provided in digital form.

Digitize: To encode map features as x,y coordinates in digital form. Lines are traced to define their shapes. This can be accomplished either manually or by use of a scanner.

Dissolve: The process of removing boundaries between adjacent polygons that have the same values for a specified attribute.

DLG: Digital Line Graph files from the U.S. Geological Survey.

DXF: Data Exchange Format. A format for storing vector data in ASCII or binary files; used by AutoCad or other CAD software and convertible to ARC/INFO coverages.

Edge matching: An editing procedure to ensure that all features that cross adjacent map sheets have the same edge locations.

Feature attribute table: A table used by ARC/INFO to store attribute information for a specific coverage feature class. Feature attribute tables supported include:

Feature class: The type of feature represented in a coverage. Coverage feature classes include arcs, nodes, label points, polygons, tics, annotation, links, boundaries, routes, and sections.

Geocode: The process of identifying a location by one or more x,y coordinates from another location description such as an address.

Geographic Information System (GIS): Geographic Information System. An organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information.

Global Positioning System (GPS): A satellite based device that records x,y,z coordinates and other data. GPS devices can be taken into the field to record data while driving, hiking, or flying. Ground locations are calculated by signals from satellites orbiting the Earth.

INFO: A tabular DBMS used by ARC/INFO to store and manipulate feature attribute and related tables.

Intersect: The topological integration of two spatial data sets that preserves features that fall within the spatial extent common to both input data sets.

Item: In an attribute table, a field of information commonly displayed as a column. A single attribute from a record in an INFO data file.

latitude-longitude: A spherical reference system used to measure locations on surface. Latitude measures angles in the north south direction and longitude measures angles in the east west direction.

Layer: A logical set of thematic data described and stored in a map library. Layers organize a map library by subject matter, e.g., soils, roads, wells, and extend over the entire geographic area defined by the spatial index of the map library.

Line: a geographic feature defined by connecting at least two points.

line-in-polygon: A spatial operation in which arcs in one coverage are overlaid with polygons in another to determine which arcs, or portions of arcs, are contained within the polygons. Polygon attributes are associated with corresponding arcs in the resulting line coverage.

Logical selection: The process of selecting a subset of features from a coverage using logical selection criteria that operate on the attributes of coverage features (e.g., area greater than 500 m). Only those features whose attributes meet the selection criteria are selected. Also known as feature selection by attribute.

Many-to-one-relate: A relate in which many records in one table are related to a single record in another table. A goal in relational database design is to use one to many relates to reduce data storage and redundancy.

Map extent: The rectangular limits (xmin,ymin,xmax,ymax) of the area of the Earth's surface, which user want to display using ARC/INFO. The geographic extent specified by the minimum bounding rectangle of a study area.

Map projection: A systematic conversion of locations on the Earth's surface from spherical to planar coordinates. Several of the more popular projections are: State Plane Coordinates (SPC) which uses feet for units of measure; Universal Transverse Mercator (UTM) which uses meters for units of measure; and latitude and longitude which uses degrees, minutes, and seconds of arc for units of measure.

Map scale: The extent of reduction needed to display a representation of the Earth's surface on a map. A statement of a measure on the map and the equivalent measure on the Earth's surface, often expressed as a representative fraction of distance, such as 1:10,000 (one unit in the map equals 10,000 units on the ground).

One-to-many-relate: A type of relate connecting a unique value in one file to many records (that have the same value) in another file.

Orthophotography: The process of aerial photographs that have been rectified to produce an accurate image of the Earth by removing tilt and relief displacements which occurred when the photo was taken.

PAT: Point Attribute Table. Polygon Attribute Table. A coverage can have either a point attribute table or a polygon attribute table, but not both.

Photogrammetry: The science of deducing the physical dimension of objects from measurements on photographs.

Planimetric: The horizontal (x,y) locations of non-topographic features, such as rivers, lakes, buildings, roads, etc.

Point-in-polygon: A spatial operation in which points from one coverage are overlaid with a polygonal coverage to determine which points fall within the polygon boundaries. Points assume the attributes of the polygons within which they fall.

Polygon: A multisided figure that represents area on a map. A feature defined by the arcs that make up its boundary. Every polygon contains one label point within its boundary. Polygons have attributes that describe the geographic feature they represent.

Polygon overlay: A process that merges spatially coincident polygons from two coverages, and their attributes, to create a third coverage, that contains new polygons and describes new relationships.

Quadrangle (quad): Typically refers to a map sheet published by the U.S. Geological Survey, a 7.5 minute quadrangle series or the 15 minute quadrangle series. Also known as a topographic or topo map.

Raster: Data displayed as discrete picture elements (pixels).

Relate: An operation that establishes a temporary connection between corresponding records in two tables using an item common to both. A relate gives access to additional feature attributes that are not stored in a single table.

Relate key: The common set of columns used to relate two attribute tables.

Remote sensing: Any of the technical disciplines for observing and measuring the Earth from a distance, including satellite imaging, Global Positioning Systems, RADAR, SONAR, aerial photography, etc.

Resolution: Measures the sharpness of an image.

Slope: A measure of change in surface value over distance, expressed in degrees or as a percentage. For example, a rise of 2 meters over distance of 100 meters describes a 2% slope.

Spatial analysis: The process of modeling, examining, and interpreting model results. Spatial analysis is the process of extracting or creating new information about a set of geographic features. Spatial analysis is useful for evaluating suitability and capability, for estimating and predicting, and for interpreting and understanding. In GIS there are four traditional types of spatial analysis: spatial overlay and contiguity analysis, surface analysis, linear analysis, and raster analysis.

Spatial modeling: Analytical procedures applied with GIS. There are three categories of spatial modeling functions that can be applied to geographic data within a GIS: geometric models, such as calculating the distance between features, generating buffers, calculating areas and perimeters, and so on; coincidence modeling, such as polygon overlay; and adjacency modeling such as redistricting and allocation.

SQL: Structured Query Language. A syntax for defining and manipulating data from a relational database. Developed by IBM in the 1970s, it has become an industry standard for query languages in most relational database management systems.

State Plane Coordinates (SPC): A map projection that measures distance in feet. By providing an SPC easting (x) and northing (y), the state name, and the zone number, any location in the United States can be identified by a unique coordinate value.

TIN: Triangulated Irregular Network. A series of triangles constructed using elevation data points taken from coverages. These triangles are used for surface representation and display.

Topography: Shape or configuration of the land surface; represented in map form by contour lines.

Topology: The spatial relationships between connecting or adjacent coverage features.

Transformation: The process that converts coordinates from one coordinate system to another through translation, rotation, and scaling.

Triangulation: A method of surveying in the location of an object may be calculated from the known locations of two other objects. Creating a triangle from the three items, the angles and sides of the triangle can be measured and the location of the unknown object is calculated algebraically.

Vector: A geometric element, stored as a point with x,y coordinates within a computer database

Soil Map Unit (Surface Features) for example - cB1g2R2

Texture denoted as small Alphabet
Slope is denoted as capital Alphabet
Erosion denoted as Numeric
Gravel is denoted as g1 or g2
c is texture class
B is Slope class
1 is erosion class
And **g2** is Gravel
If gravel exists it is indicated g1 or g2.
Other units without any gravel indication we should assume it is
non gravelly
If rocks exists it is indicated **R2, R3** etc.
Similarly Other units without any erosion we should assume it is **Nil**

Surface Soil Texture

a - Sandy	Pink	--------(205,102,153)	
b – Loamy sand		-----------(255,190,190)	
c – Sandy loam		-----------(255,255,190)	
d – Loam	Light Yellow----(230,230,0)		
e - Silt loam		----(255,255,115)	
f - Clay loam		----------(85,255,0)	
g - Silty clay loam	Green	-------(56,168,0)	
h – Sandy clay loam		---------(76,230,0)	
i – Sandy clay	Yellow	------(255,255,0)	
k - Silty clay		----(220,220,52)	
m – Clay	---- Brown	---------(168,135,0)	
Eroded Area/ Miscellaneous	-- Brown	------(205,186,103)	

Soil Slope

A- Nearly level (0-1%)	Green	--- (56,168,0)	
B- Very gently sloping (1-3%)		--- (163,255,115)	
C- Gently sloping (3-5%)	Yellow	-- (255,255,190)	
D- Moderately sloping(5-10%)		--------(255,255,0)	
E- Strongly sloping (10-15%)	Gray	--(178,178,178)	
F- Very Strongly sloping (15-25%)		---(104,104,104)	
G- Moderately Steeply Sloping (25-33%)		---(255,190,232)	
H- Steeply Sloping (33 - 50%)	Red-	(205,137,102)	
I- Very Steeply Sloping (>50%)		------(255,85,0)	

Erosion

e0 - Nil	- 0	Green -- (163,255,115)	
e1 - Slight	- 1	------- (56,168,0)	
e2 - Moderate	- 2 - Yellow	----- (255,255,0)	
e3 - Severe	- 3 - Gray	----- (178,178,178)	
e4 - Very severe	- 4 - Red	--- (205,102,102)	

Soil Gravelliness

g0 - Non gravelly (<15 %)	- Green ------- (85,255,0)	
g1 - Gravelly (15-35 %)	- Yellow ----- (255,255,0)	
g2 - Very gravelly (35-60 %)	- Gray -----(204,204,204)	
g3 - Extremely gravelly (60-80 %)	- Pink -----(205,102,153)	
g4 - Considered as part of the top soil (>80 %)	- Red- (230,76,0)	

Rocks

No to very few rocks (<2%)	- Green ------ (152,230,0)	
Fairly rocky (2-10%)	-R1 -----(115,115,0)	
Rocky (10-25%)	-R2 – Yellow ----(255,255,115)	
Very rocky (25-50 %)	-R3 – Pink ----(255,190,232)	
Extremely rocky (50-90 %)	-R4 – Red ----(205,102,102)	
Rock out crops (>90%)	-R	

Miscellaneous

Rose dust (215,158,158) ……. Mining/Industrial

……. Railway

Note: If profile study is done the soil series is also part of map unit (Soil Phase) [example N1bB3g1R2 – N1 indicate Nagenahalli series 1). For each soil series there will be one depth class (more than one series may have same depth class). In most cases the same is applicable for soil drainage map and soil calcareousness map. Normally series wise depth classes are indicated on the soil map otherwise we have to refer the master profile sheets of each series.

Soil Depth

0 - Extremely shallow (<10 cm) Red/Pink	------ (205,102,102)	
1 – Very shallow (<25 cm)	--- (255,190,232)	
2 - Shallow (25-50 cm) Gray	------ (178,178,178)	
3 - Moderately shallow (50-75 cm) Yellow	---- (255,255,115)	
4 - Moderately deep (75-100 cm)	---- (255,255,190)	
5- Deep (100-150 cm) Shades of Green	-- (163,255,115)	
6 - Very deep (>150 cm)	--- (56,168,0)	

Drainage

0- Extremely poor Brown	----- (205,137,102)	
1- Very poor	----- (137,68,68)	
2- Poor Pink	---- (255,190,232)	
3- Imperfect	---- (255,0,197)	
4- Moderately well ---- Yellow	---- (255,255,115)	
5- Well -------- Green	---- (152,230,0)	
6- Somewhat excessive -- Gray	---(178,178,178)	
7- Excessive -- Red	---(255,127,127)	

Calcareousneous CaCO$_3$ (%)

0- Non calcareousness (Nil) --- Green	-- (163,255,115)	
1- Slight (0-5%) --- Yellow	----- (255,255,115)	
2- Moderate (5-15%) --- Pink	---- (255,190,232)	
3- Strong (>15%) --- Red	------ (255,85,0)	

Preparation of Thematic maps for Surface Texture.

Based on map unit (Soil Phase) decide number of Thematic map to generated like texture, Slope, Erosion, Gravel etc and create two column for each theme for example for texture map create text_code and Text_leg same for other classes also (slope_code & slope_leg, eros_code & eros_leg, grav_code & grav_leg) if profile study is done you can create other maps dept_code & dept_leg, drai_code & drai_leg. For uniformity let us use first 4 letters of the theme nade followed by underscore (_) and code or leg.

Suitabilities

Suitability	Color	Swatch
S1 - Highly Suitable	Green- - (56, 168, 0)	
S2 - Moderately Suitable	Yellow -- (255, 255, 0)	
S3 - Marginally Suitable	Light Pink - (255,190,232)	
N or N1 - Currently Not Suitable	Brown - (205,137,102)	
N2 - Permanently Not Suitable	Gray – (156,156,156)	
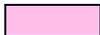
Limitations
g- gravelliness/stoniness
l – topography
r- rooting condition
t- texture
w - drainage
n - nutrient availability
e -erosion
z-excess salt/calcareousness
NOTE: "Remove Duplicate Labels" from Placement Properties of Labels

To avoid the "Others" labels
Expression:

```
Function FindLabel ( [Amla] )
if ( [Amla]<> "Others") then
  FindLabel = [Amla]
end if
End Function
```

```
Function FindLabel ( [CLU_Code] )
if ( [CLU_Code] <> "Waterbody") then
  FindLabel = [CLU_Code]
end if
End Function
```

```
Function FindLabel ( [Jack] )
if ( [Jack] <> "Rockout crop" And [Jack] <> "Others") then
  FindLabel = [Jack]
end if
End Function
```

To avoid the "Soil_No" labels
(eg., if Soil_No= 30 given for "Others" then follow as below)
```
Function FindLabel ( [Soil_No] )
if [Soil_No] <>30 then
  FindLabel = [Soil_No]
end if
End Function
```

FERTILITIES - LEGEND CLASSES

	LOW	MEDIUM	HIGH
	Orange-(230,152,0) ▮	Yellow -(255,255,0) ▮	Lt.Green - (76,230,0) ▮
Nitrogen (N)	< 280	280 – 560	> 560
Phosphorus (P₂O₅)	< 23	23 – 57	> 57
Potassium (K₂O)	<145	145 – 337	> 337

NOTE: If we receive data from soil lab as P and K then we have to multiply with the following figures

P to P_2O_5 factor is **2.29**

K to K_2O factor is **1.205**

	LOW	MEDIUM	HIGH
Sulphur (S)	<10	10 – 20	> 20
Boron (B)	< 0.5	0.5 – 1.0	> 1.0
Organic Carbon (OC)	< 0.5	0.5 – 0.75	> 0.75

Iron (Fe)	< 4.5 ppm – Deficient	> 4.5 ppm – Sufficient
Manganese (Mn)	< 1.0 ppm – Deficient	> 1.0 ppm – Sufficient
Cupper (Cu)	< 0.2 ppm – Deficient	> 0.2 ppm – Sufficient
Zinc (Zn) 1	< 0.6 ppm – Deficient	> 0.6 ppm – Sufficient

For **Deficient** - Lt.Brown (215,194,158) ▮ &   **Sufficient** - Lt.Green (85,255,0) ▮

Zinc (Zn) 2 ... Low (< 0.5 ppm), Marginal (0.50 – 0.75 ppm), Adequate (0.75 – 1.50 ppm), High (> 1.50 ppm)

Red--(255,85,0) ▮   Lt.Pink - (255,190,190) ▮   Yellow -- (255,255,0) ▮   Green- - (56,168,0) ▮

SOIL REACTION (pH) CLASSES

Class	Range	Color
Ultra acid (pH <3.5) ⟹	<3.5	Lt.Pink - (255,190,190)
Extremely acid (pH 3.5 – 4.5) ⟹	3.5 – 4.5	Pink---(255,190,232)
Very strongly acid (pH 4.5 – 5.0) ⟹	4.5 – 5.0	Pink--(205,102,153)
Strongly acid (pH 5.0 – 5.5) ⟹	5.0 – 5.5	Red-(255, 85, 0)
Moderately acid (pH 5.5 – 6.0) ⟹	5.5 – 6.0	Lt.Red --(255,127,127)
Slightly acid (pH 6.0 – 6.5) ⟹	6.0 – 6.5	Lt.Green--(85,255,0)
Neutral (pH 6.5 – 7.3) ⟹	6.5 – 7.3	Green- - (56,168,0)
Slightly alkaline (pH 7.3 – 7.8) ⟹	7.3 – 7.8	Lt.Yellow – (255,255,190)
Moderately alkaline (pH 7.8 – 8.4) ⟹	7.8 – 8.4	Yellow -- -- (255,255,0)
Strongly alkaline (pH 8.4 – 9.0) ⟹	8.4 – 9.0	Brown---(205,137,102)
Very strongly alkaline (pH > 9.0) ⟹	> 9.0	Lt. Brown (215,194,158)

Calcium (Ca) (a) ≤1.5
≤300 ppm — **Deficient** - Lt.Brown (215,194,158)

(b) >1.5
>300 ppm — **Sufficient** - Lt.Green (85,255,0)

Magnesium (Mg) (a) ≤1.0
≤120 ppm — **Deficient** - Lt.Brown (215,194,158)

(b) >1.0
>120 ppm — **Sufficient** - Lt.Green (85,255,0)

*Lt.=Light

Electrical Conductivity (EC)

<2 dsm $^{-1}$ Non saline

2 – 4 dsm $^{-1}$ Low

4 – 8 dsm $^{-1}$ Medium

8 – 12 dsm $^{-1}$ High

12 – 16 dsm $^{-1}$ Very high

>16 dsm $^{-1}$ Extremely high

Rock outcrops

AWC(Available Water Capacity)

Very low (<50 mm/m)   -   Sahara Sand- (255,235,190)

Low (51-100 mm/m) –   Medium Sand - (245,202,122)

Medium (101-150 mm/m) – Yucca Yellow - (255,255,190)
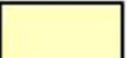
High (151-200 mm/m)   Leaf Green -- (56,168,0)

Very high (>200 mm/m) – Medium apple - (85,255,0)

*Do not forget to add the "mm/m" unit and the brackets "()"

Land Capability (LCC)

I – Light Green

II - Yellow

III - Red

IV – Cadet Blue

V – Dark green

VI - Orange

VII - Brown

VIII - Purple

Key: I - Very good cultivable lands

II - Good cultivable lands

III - Moderately good cultivable lands

IV - Fairly good cultivable lands

V - Very well suitable lands for grazing

VI - Well suitable lands for grazing and forestry

VII - Fairly well suitable lands for grazing and forestry

VIII – Suitable lands for wildlife and recreation

Limitations:

e - erosion/run off limitation

s - Soil limitation (soil depth, gravelliness, light/heavy texture, and

salinity/alkalinity)

w - Excess of water (wetness high water table, drainage) limitation.

c - Climate limitation

Stoniness

Code	Area covered	Class
St1	0.01 to 0.1%	Strong
St2	0.1 to 3%	Very strong
St3	3 to 15%	Extremely strong
St4	15 to 50%	Rubbly
St5	50 to 90%	Very rubbly
St	>90%	Stone

SWC

TCB – Trench cum bunding   -   Orange- (255,211,127)

GB – Graded bunding -   Brown - (205,170,102)

Field bunds – Strengthening of existing bunds - (208,227,61)

Terracing -   Yellow -- -- (255,255,0)

CB - Cresent bunds   -   Pink - (255,115,223)

Note: if Slope (A) is there then **TCB** will be "TCB/Strengthening" and **GB** wil be "GB/Strengthening"

Maintaining "Legends" in Attribute Tables

Phosphorus (P_2O_5)
Low (< 23 kg/ha)
Medium (23 – 57 kg/ha)
High (> 57 kg/ha)

Potassium (K_2O)
Low (<145 kg/ha)
Medium (145 – 337 kg/ha)
High (> 337 kg/ha)

Organic Carbon (OC)
Low (< 0.5 %)
Medium (0.5 – 0.75 %)
High (> 0.75 %)

Electrical Conductivity (EC)
Non saline (<2 dsm)
Low (2 – 4 dsm)
Medium (4 – 8 dsm)
High (8 – 12 dsm)
Very high (12 – 16 dsm)
Extremely high (> 16 dsm)
Note: while generating map you have to insert text in super script as "-1"

SOIL REACTION (pH) CLASSES
Ultra acid (pH <3.5)
Extremely acid (pH 3.5 – 4.5)
Very strongly acid (pH 4.5 – 5.0)
Strongly acid (pH 5.0 – 5.5)
Moderately acid (pH 5.5 – 6.0)
Slightly acid (pH 6.0 – 6.5)
Neutral   (pH 6.5 – 7.3)
Slightly alkaline (pH 7.3 – 7.8)
Moderately alkaline (pH 7.8 – 8.4)
Strongly alkaline (pH 8.4 – 9.0)
Very strongly alkaline (pH > 9.0)

Maintaining "Legends" in Attribute Tables

Sulphur (S)
Low (<10 ppm)
Medium (10 – 20 ppm)
High (> 20 ppm)

Boron (B)
Low (< 0.5 ppm)
Medium (0.5 – 1.0 ppm)
High (> 1.0 ppm)

Iron (Fe)
Deficient (< 4.5 ppm)
Sufficient (>4.5 ppm)

Manganese (Mn)
Deficient (< 1.0 ppm)
Sufficient (> 1.0 ppm)

Cupper (Cu)
Deficient (< 0.2 ppm)
Sufficient (> 0.2 ppm)

Zinc (Zn)
Deficient (< 0.6 ppm)
Sufficient (> 0.6 ppm)

23. References

Burrough, P.A. (1986). Principles of Geographic Information Systems, 1st ed. Oxford University Press, New York 336 pp.

Carter, J.R. (1988). Digital representation of topographic surfaces, *Photogrammetric Engineering & Remote Sensing*, 54(11):1577-1580.

Cooke, R. U. and Doornkamp, J.C., (1990). *Geomorphology in Environmental Management,* Oxford Press.

Drăguţ, L. and Blaschke, T. (2006). Automated classification of landform elements using object-based image analysis. *Geomorphology,* 81, pp. 330-344.

ESRI, (2010). What's new in ArcGIS 10, *ArcGIS Manual*, Environmental System Research Institute (ESRI), Redlands, USA, p 177.

Fix, R. and Burt, T. (1995). Global Positioning System: an effective way to map a small catchments, *Earth Surface Processes and Landforms, 20, 817-827.*

Fundamentals of Remote Sensing by Canada Centre for Remote Sensing.