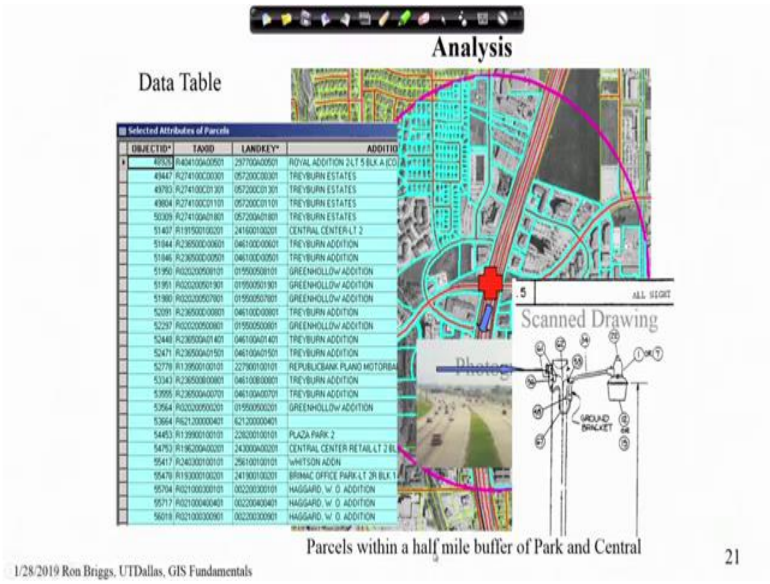
Soil Science and Technology
Prof. Somsubhra Chakraborty
Department of Agricultural and Food Engineering
Indian Institute of Technology, Kharagpur

Lecture - 52
Geostatistics

Welcome friends to this 2nd lecture of week 11 of Soil Science and Technology and in this lecture we will be finishing this GIS and GPS lectures.

(Refer Slide Time: 00:27)



We will be first talking about GPS and then we will be talking about you know a new topic that is the geostatistics and how these geostatistics is useful for spatial interpolation and we will be discussing that also. So, I hope that this lecture will be very much informative to you.

So, in the last lecture we talked about GIS or Geographic Information System and then, we talked about how this geographic information system combines the geographic information or the you know geographic information as well as software and then hardware and the users to produce the required information in a better way. So, if you remember the geographical information system is based on you know overlaying of several layers you know stacking of several layers, and when we stack them together, then they represent any feature of the earth surface and this feature of the earth surface can be represented by two features you know two data features.

One is called vector data ; another is raster data and vector data can be represented by any points lines or polygons whereas, the raster data can be represented by gridded cells.

And raster data generally represents any continuous features like you know like any elevation of a particular side and by using this raster and vector feature, we can get information not only from the spatial point of view, but also we can get the information for different other queries and for getting the answer for different other queries. These individual layers or individual you know vectors or individual raster data always contains some associated attributes and these attributes basically are stored in some attribute table.

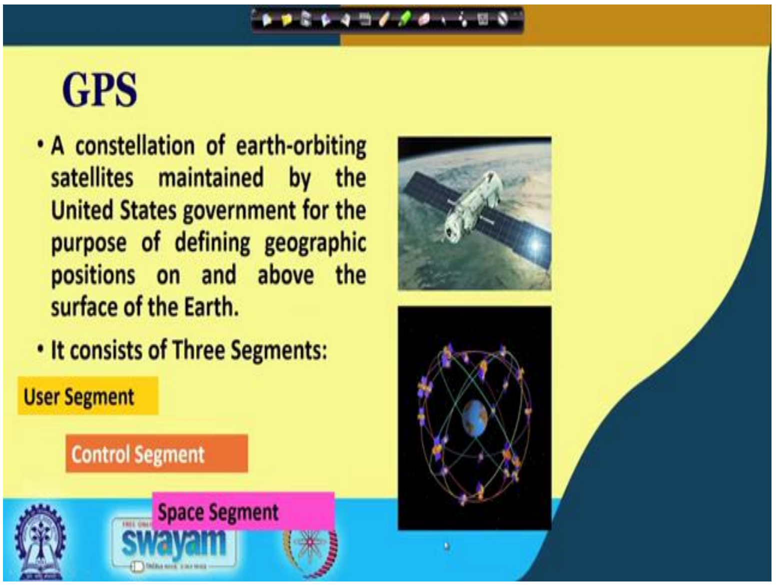
So, as you know this GIS software can simultaneously handle both plotting as well as different types of database management. So, it is basically a specialized type of database management which can show different location giving its proper locational context. So, that is why it is called Geographic Information System.

So, in the last lecture we finished here. I showed you one analysis where we have you know combined three layers where is the base layer was basically digital ortho-photo. You know ortho-photo where we overlaid two different other layers; one is the street layer view another a parcel of layers parcel of lands.

So, these three layers were stacked together and then we try to find some areas which are parcel of land which are located you know within half mile buffer of park and central and remember that we got this area, this circle circular area which basically shows the this circular areas which basically shows the queried or the final output and obviously, the selected attributes for the parcels you can see for all these parcel selected attributes are there some tax id land keys and additional information.

So, these attribute table is further helping us to identify these areas, then photographic images of these areas and some scan drawing. So, I hope that now it is clear to you how GIS can be effectively used for mapping a specific you know specific event or specific area.

(Refer Slide Time: 04:13)



And let us see what is GPS. GPS basically stands for Global Positioning System and this Global Positioning System basically utilizes a constellation of earth orbiting satellites maintained by the United States government for the purpose of defining geographic positions on and above the surface of the earth.

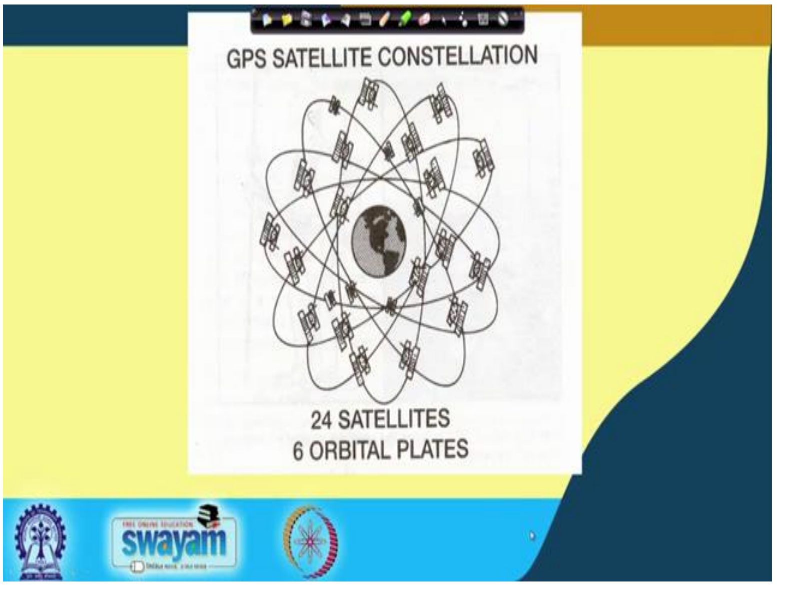
So, you can see here GPS is basically global positioning system short form of global positioning system and this GPS is basically helpful for defining the geographic position in terms of both coordinates you know x and y coordinates or latitude and longitude along with the elevation some time. So, basically we help it helps us to attain or to gather the info locational information from a particular space, a particular place.

Now the GPS remember that this GPS is an important data input tool for GIS because GPS helps in identifying the areas with proper geographical context and then, it you know populate that data in the GIS attribute table with giving some proper locational context and obviously you know these GIS or these GPS is based on this constellation of 24 and sometime you know 24 plus satellites and it consists of basically three segments. One is called the user segment, then control segment and then space segment.

(Refer Slide Time: 05:48)

So, the satellites which are more than 24 in number are called GPS NAVSTAR Satellite and NAVSTAR is a basically short form of Navigation by Satellite Timing and Ranging and using this satellite you know all the GPS applications are being maintained and operated.

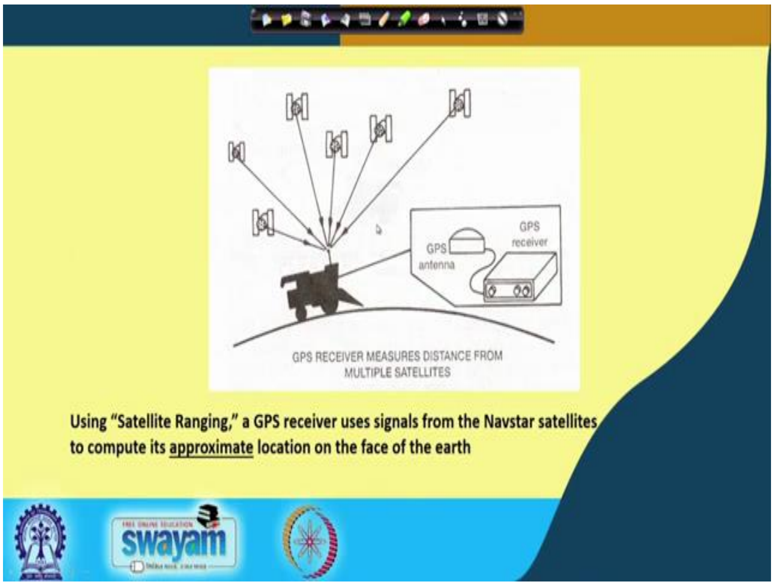
(Refer Slide Time: 06:07)



So, you see here these are the GPS satellite constellations, which shows basically 24 satellites which are moving across 6 orbital plates.

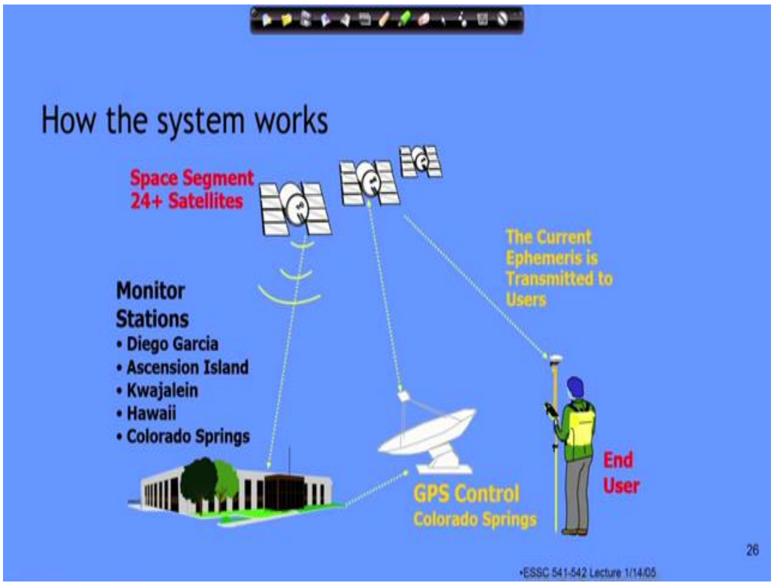
So, these satellites 24 some now a days this number has been increased. So, 24 or more satellites are you know are basically orbiting the earth in 6 orbital plates and they gather the locational information.

(Refer Slide Time: 06:38)



And remember that using satellite ranging a GPS receiver use a signal from the NAVSTAR satellites to compute its approximate location on the face of the earth. So, you know for forgetting an approximate location we have to use a satellite ranging a GPS you know, so that a GPS receiver use a signal from the NAVSTAR satellite to compute its approximate location on the face of the earth. So, we can see here this is a GPS antenna is mounted here whereas, the GPS receiver and this GPS receiver is basically collecting the information from different you know different NAVSTAR satellites.

(Refer Slide Time: 07:24)



So, let us see how this GPS system basically works. So, the space segment or the first segment is basically consists of 24 more satellites and these satellites basically you know monitored by 5 monitoring stations like you know and which are located at Diego Garcia, then Ascension Island and then Kwajalein and then Hawaii and Colorado Springs and they have GPS control system at this you know at this monitoring stations and as a result of this monitoring station, these basically you know satellites basically calculates with their positions and finally it will be transmitted to the end users.

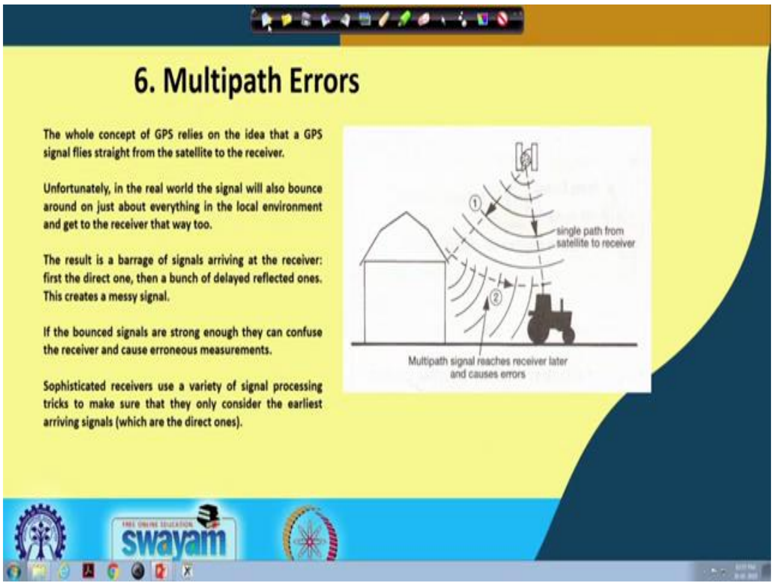
So, the current ephemeris is basically transmitted to the end users which basically uses the GPS coordinates or coordinate they at his particular point using some GPS receiver.

So, again the space segment contains of you know a constellation of satellite which is basically monitored by monitoring stations and this monitoring stations using the using this monitoring station they a particular end user can measure its locational context by you know by measuring the GPS signals at its point by using some GPS receivers.

(Refer Slide Time: 08:53)

So, obviously there are some you know although the GPS is very handy tool, however there is some possibilities of errors all the time. These errors are basically we call it you know there are five types of error. Satellite Clock errors are there, Receiver Clock bias is there, then Atmospheric Dispersion is there, Satellitic Orbital errors are there and GPS Receiver Quality is also there.

(Refer Slide Time: 09:19)

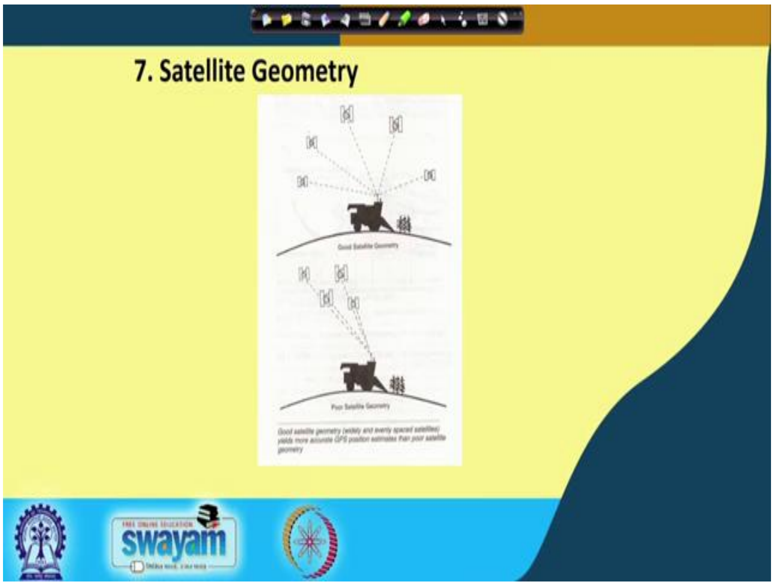


So, multipath error is another type of error. Remember that the whole concept of GPS relies on the idea that the GPS signal files you know signal files straight from this GPS signal file straight from the satellite to the receiver. So, from this you know again whatever we have seen in the last slide that this GPS signal file straight from the satellite to the receiver, however unfortunately in the real world this signal will bounce around just about anything in the local environment and get to the receiver that way too.

So, you can see this is one way to which GPS signal is expected to reach from satellite to the receiver, however in reality it may get reflected by different other features or other you know other features over the earth surface and as a result of that there are several bouncing backs. So, the result is a barrage of signals arriving at the receiver first the direct one and then a bunch of delayed reflected one.

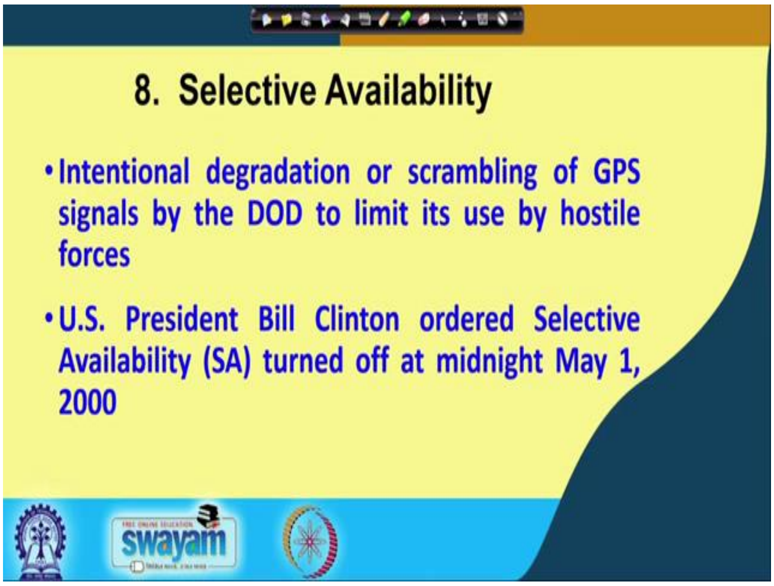
So, this create a messy signal. So, if the bounce signals are strong enough, they can confuse the receivers and cause erroneous measurements. So, sophisticated receivers use a variety of signal processing tricks to make sure that, they only consider the earliest arriving signals which is the direct signals most of the time. So, that is why if this is the multipath error and this multipath error can be removed by using some advanced GPS receivers.

(Refer Slide Time: 11:00)



Another source of error is Satellite Geometry. So, you can see here good satellite geometry where all the satellites are widely and evenly spaced you know yields more accurate GPS position you know estimates, then poor satellite geometry where all the satellites are cluster at a particular side. So, this is how this satellite geometry you know controls the source of error.

(Refer Slide Time: 11:28)



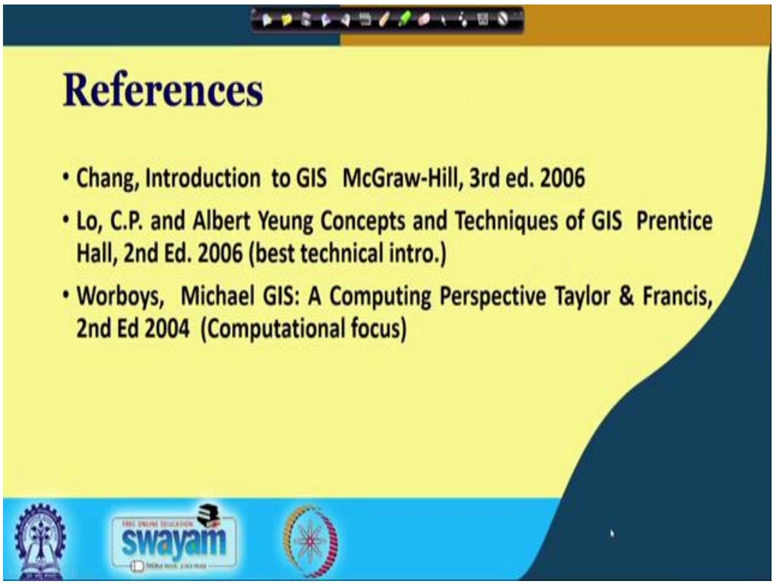
Another important term is Selective Availability. Now selective availability is basically the intentional degradation or scrambling of GPS signal by the department of defence or United States you know Military to limit this use by hostile forces.

For example any hostile forces can use this GPS signal to target something or some you know target some you know building or some features. So, that is why there is was an intentional degradation or scrambling of GPS signal by the department of defence. However US President this selective availability was removed by US President Bill Clinton in the midnight of May 1, 2000. So, right now there is no selective availability.

(Refer Slide Time: 12:18)

And common uses of GPS you know this you know there is a wide ranges of users for GPS starting from land, sea and air navigation and tracking and then surveying and mapping obviously, military applications and recreational use. So, all these can be done from your GPS.

(Refer Slide Time: 12:38)



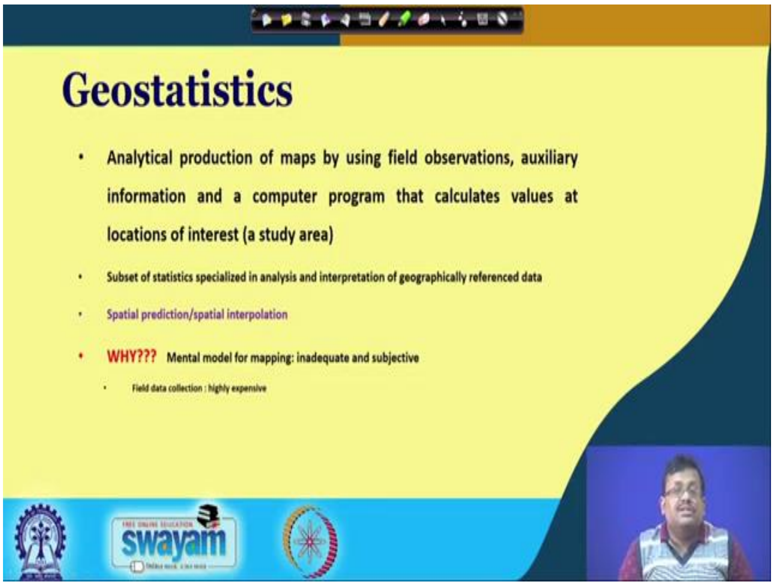
Now the references are obviously these are three references. So, you can follow this introduction to GPS by Chang which is the 3rd edition and then the then Lo, C. P. Albert

Yeung concepts you know basically they wrote these concepts and techniques of GIS and it is published by Prentice Hall.

So, it give you basically gives you best technical introduction for GPS and if you want to focus on computational aspects obviously this computational perspective GIS a computational perspective by this Worboys and Michael I am sorry it is been a single author. So, Michael Worboys you know he has authored this book GIS a computing perspective Taylor and you know by Taylor and Francis. So, you can follow that which also gives better computational focus.

So, I hope that guys you have got that some basic knowledge about GIS and GPS and I hope these are helpful for you, will be helpful for you and you should go ahead and read some more literature regarding this GIS and GPS applications.

(Refer Slide Time: 14:02)



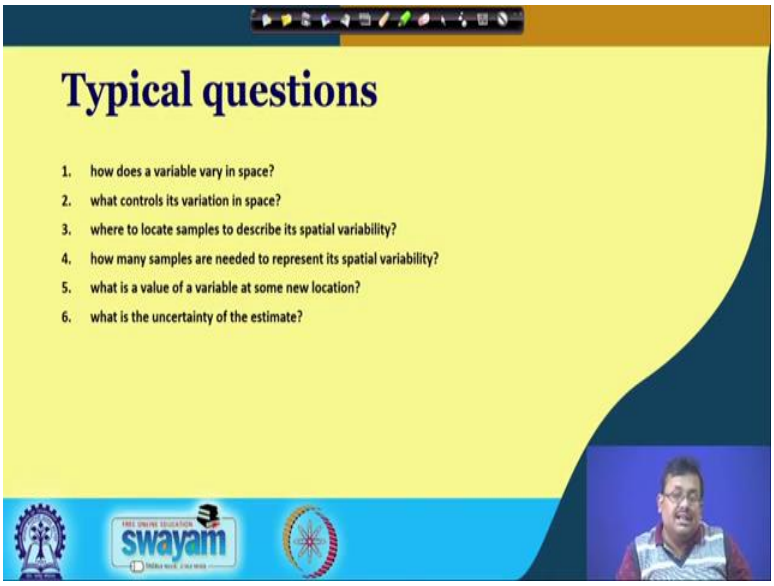
So, let me move ahead and go to a new topic that is Geostatistics and Geostatistics is basically a branch of statistics which shows the analytical production of maps by using field observations, auxiliary information and a computer program that calculates values at location of interest or a study area.

So, again this geostatistic is basically an analytical production of maps by using field observation, auxiliary information and computer programs and the other name of

Geostatistics is Interpolation Techniques. So, sometime although geostatitics is not solely based on interpolation, but also they can be used interchangeably most of the time.

So, subset of statistics basically which specialized in analysis and interpretation of geographically referenced data and you know it helps in spatial prediction or spatial interpolation. Why? Because whenever we do mapping with our mental modelling, this is basically inadequate and subjective and field data collection on the way on the other hand is highly expensive. So, these are some obstacles that is why we go for this geostatistical interpolation or geostatistical prediction.

(Refer Slide Time: 15:20)

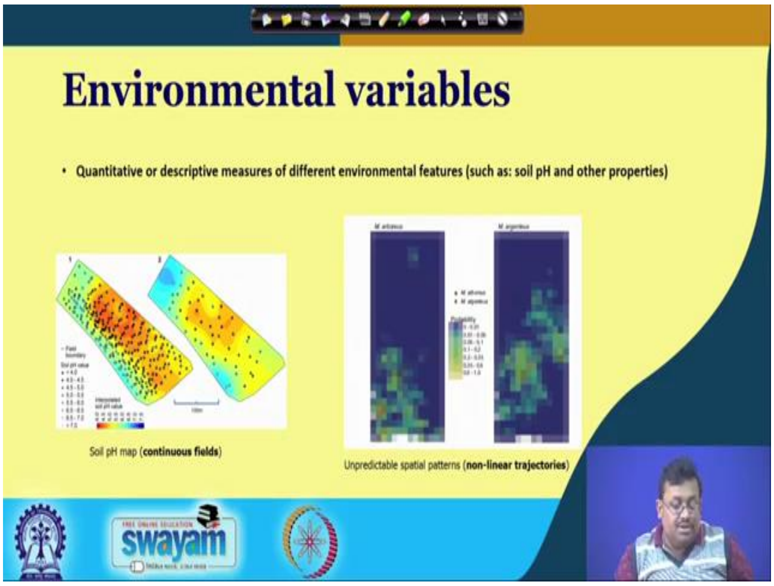


So, these are the typical questions which we generally ask for getting the answer from geostatistical analysis. First of all how does a variable vary in space? Secondly, what controls its variation in space? Thirdly where to locate samples to describe its spatial variability? Fourth is how many samples are needed to represent its spatial variability and finally, what is the value of a variable at some new location and finally, what is the uncertainty of the estimate? So, not only it gives you the not only it gives you map and estimation of that particular area, but also it estimates the uncertainty.

(Refer Slide Time: 16:07)

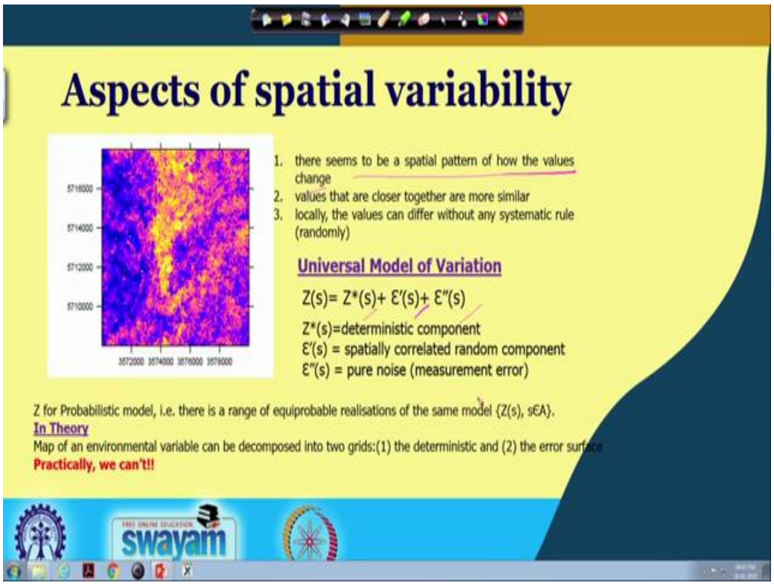
So, what are the contemporary applications of geostatistics ? Obviously to analyze the point data and combination with various GIS layers to explore the spatial variation in remote sensing data, secondly to identify noise in the image and for their filtering, then third is to improve the generation of digital elevation model and for their simulations and fourth one is to optimize the spatial sampling selection and or spatial resolution for image data and selection of support size for you know ground data.

(Refer Slide Time: 16:47)



So, remember that you know this spatial interpolation can only be done in case of their in case it is a continuous data just like the soil pH map which is the negative logarithm of hydrogen and concentration. So, in this case this ph value is called that you know is termed as an Environmental Variate or Environmental Variable and it is basically represented by this soil pH map whereas, there are other types of data for example movement of a hard of a ship along the you know, which basically shows the non-linear trajectories as well as the growth of the plant which basically does not particular type of plant species. So, these are basically non-linear trajectories and this non-linear trajectories are you know create several problems for geostatistical interpolations.

(Refer Slide Time: 17:52)



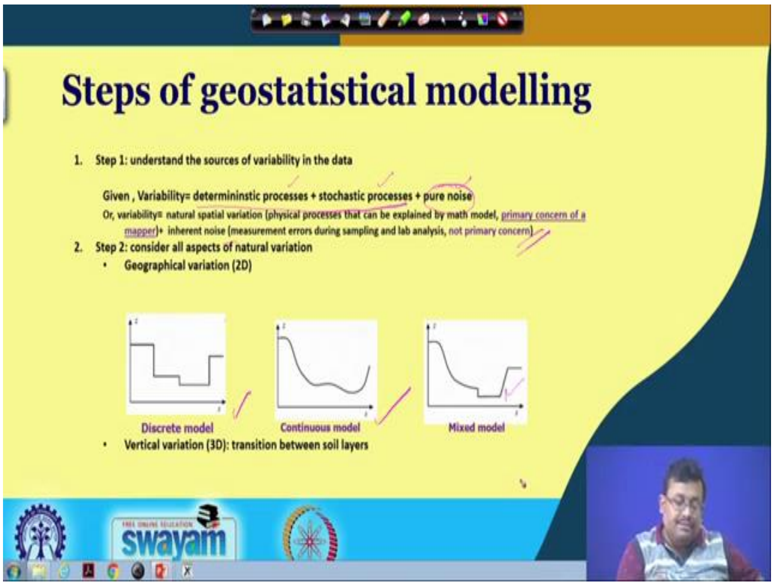
So, let us see what are the aspects of spatial variability. So, if you see this is a study area and this study area you can see here there are three you know types of shades. One is yellow, another is red, another is green, another is blue. So, you can see these three inferences we can draw from this graph first of all there seems to be a spatial pattern of how these values are changes. So, you can see the values basically changes using some spatial pattern for example, values that are close together are most similar. So, these values which are close together are more similar to those you know as compared to the values which are you know which are far apart and secondly, locally thirdly locally the values can differ without any systematic rule or randomly.

So, you can see here there are some you know in this region the values can differ without any systematic rule. So, any spatial distribution of an environmental variable can be represented by universal model of variation and this universal model of variation takes this form where you know Z S can be you know value of a particular variable at a particular location where it can be decomposed into you know into three parts. One is called the deterministic component and then, the spatially correlated stochastic component or random component and the third one is pure noise which is basically measurement error.

So, again this universal model of variation assumes that any you know the value of a particular soil property at a particular space is basically a combination of you know a deterministic component and then, a stochastic or spatially correlated random component and finally a pure noise. So, Z is used for probabilistic model that is there is a range of equiprobable realization of the same model, which is represented by this and in theory map of environmental variable can be decomposed into two grids. For example, the deterministic grid and the error surface practically it cannot be achieved.

So, I mean, so that is why there is a need for geostatistical analysis. Again the universal model of variation can be represented by a combination of deterministic component, then stochastic component and then you know pure noise. Obviously, this deterministic component can be modelled through simple models; however this computation of the stochastic component is kind of difficult. So, we will be seeing how we can you know how we can model this deterministic component and spatially correlated random component using geostatistical means.

(Refer Slide Time: 21:19)



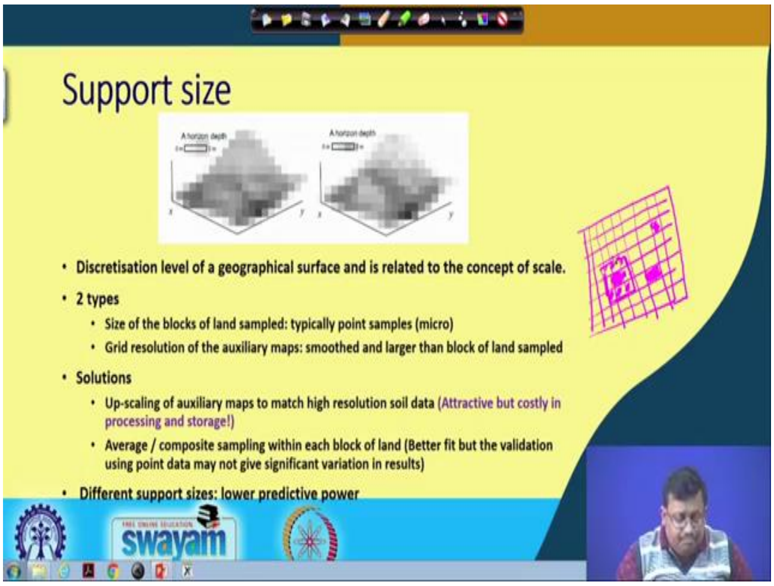
So, what are the steps for Geostatistical Modelling? There are two steps for Geostatistical Modelling the first step says understanding the sources of variability in the data. So, we have already seen that a variability is basically combination of deterministic process as well as stochastic process and pure noise. So, if we combine this deterministic process and stochastic process, this is what we are interested to capture through geostatistical modelling because we cannot capture this pure noise. So, the variability can be represented as a combination of natural spatial variation which is physical which is, which can be represented by physical processes that can be explained by math model that is a primary concern of the mapper.

And secondly the inherent noise which is basically measurement error during the sampling and lab analysis which is not the primary concern of a mapper and step 2 is basically to consider all aspects of natural variation. So, you can see the geographical variation in 2D and vertically variation in 3D. So, here it shows the three types of geographical variation in two dimension.

So, X is basically is the distance. So, you can see the value when the value is changing discretely along with the increasing distance, then it is we call it discrete model. When the when the value changes continuously along with the increasing distance, we call it continuous model and the mix model basically shows the combination of both discrete model and continuous model and obviously, there is a another type of variation that is called vertical variation and vertical variation can be seen in terms of transition between soil layers.

If you remember the soil basically consists of several layers. So, soil profile so this soil profile you know consists of several horizons. So, the transition between in horizons each of the horizon can be described in terms of vertical variation.

(Refer Slide Time: 23:28)



Support size- Support size basically shows the discretization level of a geographic surface and it is related to the concept of the scale and it is of basically two type. For example, the size of the block of the sample which is typically point sample or micro and the second one is grid resolution of the auxiliary maps that is smooth and larger than block of the land sample. So, let me draw you and give you some better example.

So, suppose this is our area of interest and in the area of interest we have a auxiliary data. For example, that is a raster file which is arranged in grids. So, we see these are the you know we can we can de we can produce or we can represent this raster in terms of grids just like this. So, once we produce these grids or grids of rasters let us see that one sample was collected from this side, another sample was collected from this side, another sample was collected from this side.

So, or in other way this is the point from which one sample was collected, this is the point from which another sample was collected; this is the point from which another sample was collected. So, for auxiliary map for example so, we can so first of all we can consider this as a point data just like we have seen in our GIS lecture that how this point data can be represented by a single grid cell.

So, if we have a auxiliary map say we have a digital elevation data or DEM file and from there we have we see that the size of the grid is somewhat bigger than the size of the size of the individual point. So, you can see this is the support size difference. So, the first we collected this is the individual cell represents the support size of the point data and this is now is the support size of the digital elevation model. So, obviously this is quiet larger than that of the block of land sampled.

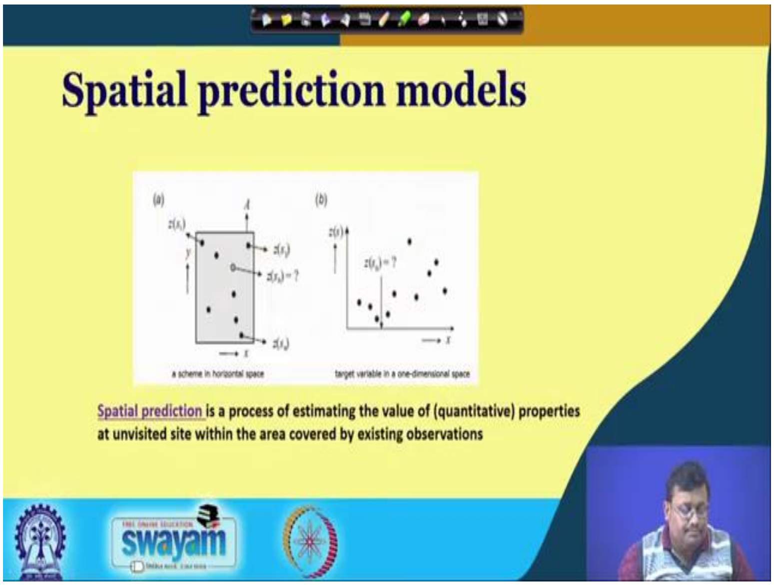
So, that creates the problem because you know discretization of the support size or when there is a difference in the support size, we cannot interpolate the data effectively. So, there are two types of solution. First of all we can upscale the auxiliary maps to match the high resolution soil data. So, in other words we can decrease the grid size or support size of this elevation data to match with this single grid cell. So, that is good, but the problem is that requires huge amount of calculation and also that is costly and it requires huge amount of storage because we are decreasing the resolution, we are sorry we are increasing the resolution by decreasing the grid size and that will create huge volume of data.

So, that is hence one solution. Another solution is to average or composite sampling within each block of land. So, we will take the sample from this area, all of these area and then, we will average them together to match the you know we will average them together to get a composite value and this composite value will match with the support size of the auxiliary map, however the problem is you know although it is a better fit, but the validation using point data may not give significant may not get significant variation in results.

So, there are some problems. So, different support size obviously lower predictive power. So, you will see while we discuss the digital soil mapping in week 12, we will see that how we can you know homogenize the support size, so that we can better interpolate any property using geostatistics. So, remember then remember that again for you know for better management of geostatistical analysis or for better geostatistical interpolation, all the auxiliary map support size has to be homogenized with the support size of the point data or the you know point data set.

So, this is what is support size you know looks like.

(Refer Slide Time: 28:10)



And let us stop here and we will start from this slide in the next lecture and we will be talking about Spatial Prediction Models and we will be talking about different types of other interpolation techniques also and we will be starting you know some advanced sensors also in the next lecture. So, I hope that you have got that some knowledge about this Geostatistics. Let us meet in the next lecture to discuss it further.

Thank you.