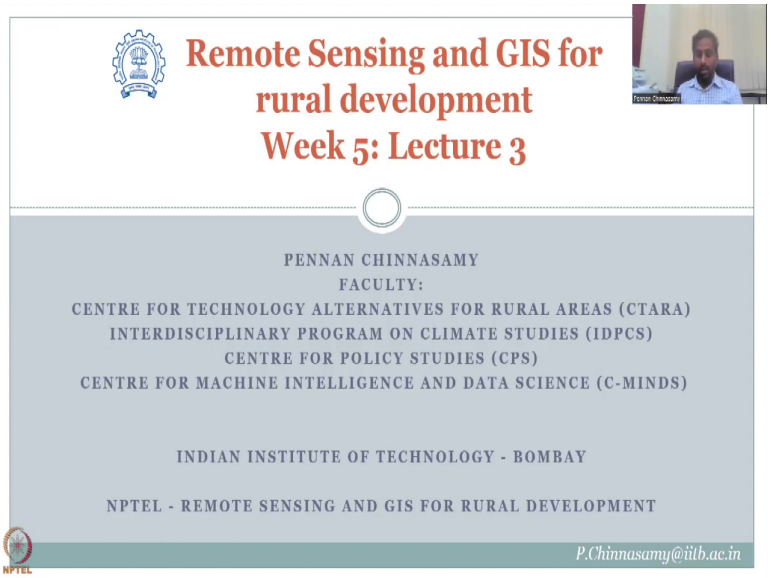
Remote Sensing and GIS for Rural Development
Professor Pennan Chinnasamy
Department of Centre for Technology Alternatives for Rural Areas (CTARA)
Indian Institute of Technology, Bombay
Week-5
Lecture no. 03
Raster data and Vector Data Quality Issues

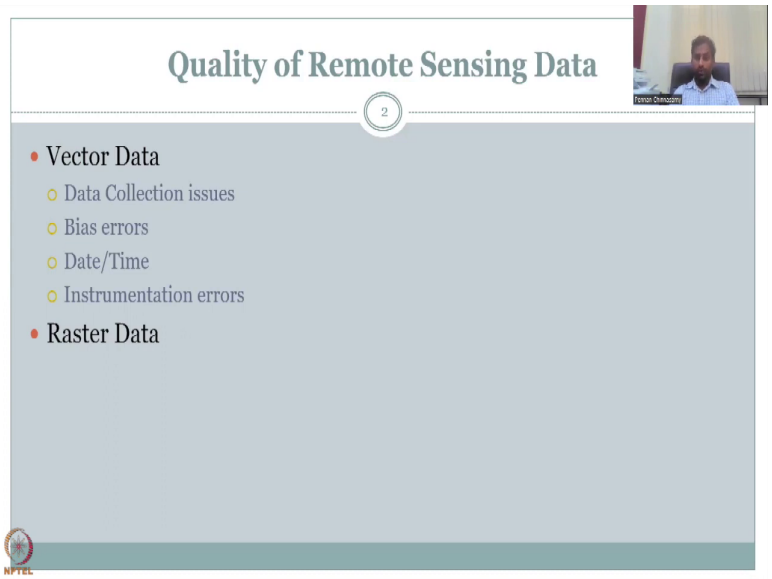
(Refer Slide Time: 00:16)



Hello everyone, welcome to Remote Sensing and GIS for Rural Development NPTEL course, this is week 5, lecture 3. In this week, we have been looking at NPTEL courses, which focus on using raster data for rural development. The definition of rural development has been given earlier. However, in this case, we will be looking at just the data types week 4 we have looked at vector and week 5 we are looking at raster we have defined the raster in the previous lectures in week 5 and showcase some examples of where this data can come handy.

One important thing in week 4 and Week 5 is we need to understand the data before we use it. So, in that note, while we discuss the raster data, let us discuss what issues can come up in remote sensing data plus vector data for GIS while using it for rural development.

(Refer Slide Time: 01:31)



Initially vector data we will discuss, in vector data as we mentioned, it can either be a point for polyline or mine or a polygon. So, mostly these are observation data collected from ground or some data entry on tables and then convert it to geospatial data. So, as the name sounds, observation data and user enter data etcetera. For example, sensors is done through surveys and then convert it to GIS.

So, what happens here is there is lots of data collection issues, I will just label some of them there are detailed issues discussed in your classes which are available in NPTEL, I have given you the links, you could definitely look into what are the other classes that can happen. So, most importantly, we will have data collection issues as bias errors, bias is what, the data collector or the data itself will have an affinity or a bias towards one side.

Let us say I am collecting the incidence of COVID. If too much COVID cases are reported, let us say in a tourist country like Thailand or Singapore then people will be very less coming in, tourism industry will collapse. So, there might be some errors while data collection happens and these we call it as bias error. So, the error is while entering the data, sometimes not the whole picture is captured only some part of the data is collected and that is bias errors.

There is also representation errors wherein the data is collected non representative locations. We will go through this when we do a case study example, but for now understand that both the data types will have errors. We will focus on the raster data with some examples. But it is also necessary to understand the vector data issues. As I said, there is representation errors.

For example, this is your land and it has five parts. If you collect in the centre, maybe it is representative of all the five parts five lines five channels whatever it is, but if you collect only here the data how is this data representative of the whole hand?

So, this is called representative errors or the data is not representative. In a direct rural development problem, I can tell that if a district is consuming a lot of groundwater you should not be measuring the groundwater in a farm well, which is being used you cannot take water from a location which is not being used. Let us say school so that the whole village has a lot of farmland and then there is a school, if you go to the school and take the groundwater level, it would be better compared to the farm because in the farm, they extract the water. So, these errors are called representative errors. Then date and time a lot of errors do not have a date and time or they are outdated.

So, these are issues that should be properly managed before converting to vectors. You find these errors a lot there is multiple literature on data errors for each aspect, I am speaking about rainfall, groundwater, surface water, agricultural pesticide, availability, and then data on storage structures etcetera. The final part is instrumentation errors, the instrument that you hold to measure all so has errors, let us say you are calculating the crop yield in a field. How do you calculate crop yield, you harvest you put it in a sack and then you weigh the sack 1 ton per hectare, 2 tonnes per hectare, etcetera.

So, now, if your balance the weighing machine itself has errors, then it is wrong, correct and this lot of people have experienced personally in a gym you will see a different weight, at home you will see a different weight, on the railway station where you can put a coin I used to do that in Chennai you have a coin machine you put and you get a weight. So, all these are different weights. How is that possible?

If the weight is more or less equal, for example, 500 grams change that means maybe you drank too much water you had lunch or big lunch biryani or something and so, you can compensate for that excess but if the difference is like 4 or 5 kilos from morning to afternoon or within 2 days that cannot be explained. So, how is that explained by deeming them as instrumentation errors, that is why you see very very accurate weighing scales in a jewellery shop. And in hospitals, where they for the babies and all they have a very sophisticated weighing scale, all these data are taken and converted to vector these are observation data.

So, we will go through with again on this slide, when we discuss the case studies for vectors, I will pinpoint how the errors can creep in. And it is our duty to represent correct data while doing rural development work. Now, we come to raster, raster is very, very complicated in terms of errors. However, some of these errors are taken care of, let us say for example, bias errors. Bias error is not there because it is unbiased data.

For example, the satellite can take images, it need not take a different image for China and it is the same image, whereas when you ask people to collect data, let us say you want to study the Ganges River, which gives water to almost 1 billion of the population. across countries, Tibet, Japan, China and Nepal and India and also Bangladesh.

What happens here is you will have to look at it in a different angle. Because you cannot collect data in India and apply it to China, China has to give the data for the Ganges water how it is moving and Nepal. So, this is where bias comes they will give different data quality and based on their benefits etcetera. Whereas if you take a satellite image, it is the same. So now you could see how some errors are removed just by using satellite data. So, there is no bias errors. There is no data time issues because while the image is captured automatically, the data has a time and date stamp. You will see this when we download data because every data that we download at the end of the name will have the data name, the version of the data and the date and time of that image which has been populated.

Instrumentation errors could be there for example, a satellite is launched in 2022 it starts to collect data 2023 And there is a lifetime maybe 5 years, 10 years, then after that, that satellite is let go into space, it is not collecting data anymore. These are instrumentation errors, but you have time. Whereas, if you go to, for example, as I said, a hospital to collect hospital record data for children, height, weight, etcetera. there is no replacement of the balance, it is just there.

Only if it breaks totally they have replaced there is no every five years you throw it and then bring a new one in. Whereas this one satellites have a lifetime. But still there could be some errors, these are mechanical errors, we call drift errors, any moving part will have some drift after some time it has to be calibrated again. And if you do not calibrate the data, you will get very very bearing information which is not correct. So, by using raster data, already lot of issues have been taken care of. Now, we will look at raster data issues, other issues which are very important, and we should know.

(Refer Slide Time: 11:27)

One is quality of the image here in photographs, satellite image, etcetera. Seasonal considerations have to be given. So if you are using drones and unmanned aerial vehicles or small planes to take data, you have to make sure that you take it in a correct time and season. For example, you have here 1 and 2.

1 is a leaf off season and a leaf on season. You could clearly see that the leaf off season has issues in terms of the elevations, the elevations can change rapidly. So these elevation changes can be addressed. If we take correct seasonal measurements, these are two same location, same area, but time is different. In one, you could see a lot of leafs, and then in number, the second one you do not see leafs because it has fallen down. So you will have to make sure the date and time you select is correct.

(Refer Slide Time: 12:38)

These are reflection issues. What you could see here is a farm pond, which I took from the plane you could see that so anyone who is going on a plane, ask for a window seat and you can just take an image from your phone, it is as good as an aerial image, you are not going to use it for data analysis. Because commercial plane taking from a window is different. And data plane that goes exactly and takes perpendicularly is different.

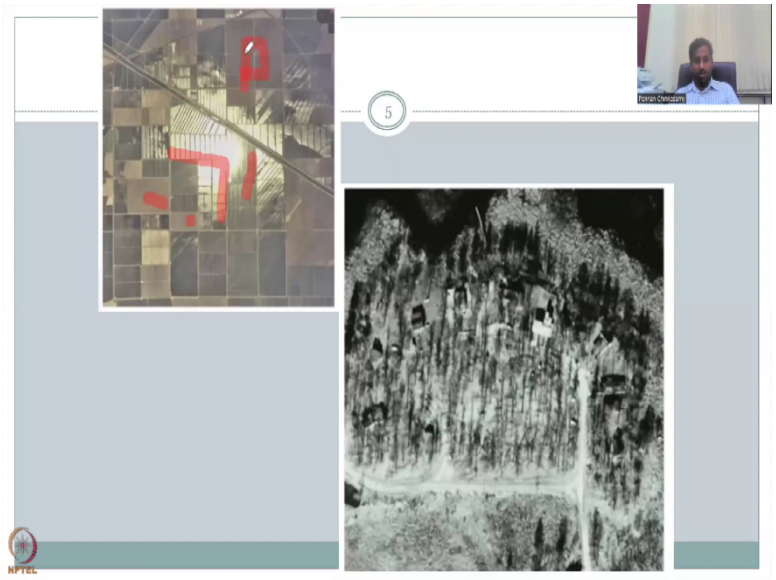
So I am just showing this for quality perspective, you could see that the first one there is some fog some on the screen on the screen there is some fog or condensation happening and that can affect the image as much as you clean it, you still will get the fog. So, those who take a flight can see that in the window pane, you keep on cleaning, but still the fog will come up. So you will it is like a car out. So, you can see it in the car like you always have to keep the ventilation heating on. So, these errors are bad the data is not usable. Because for example you want to see what is the centre part and the centre part is totally fogged. So this data is unusable for that region. There are filters that can be used in remote sensing data. However, these are mathematical filters.

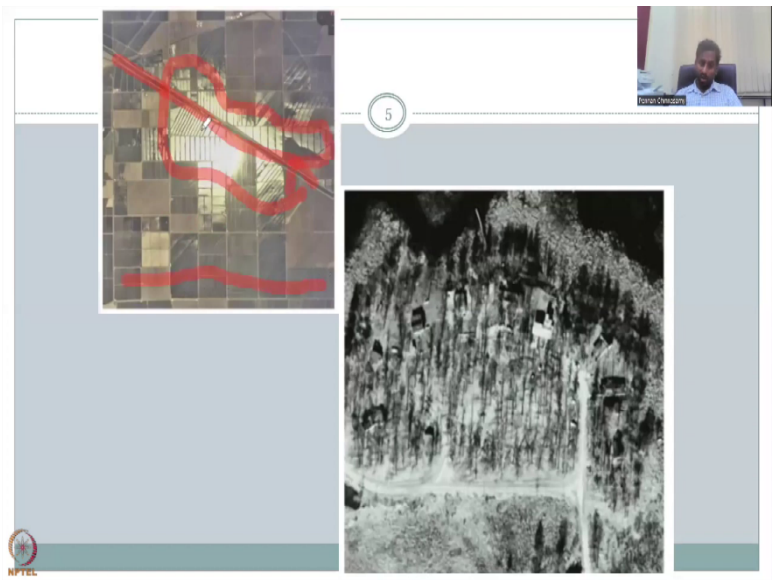
These are not exact filters because it will just remove but what is underlying it, we do not know. The second one here is reflection from the body itself. So here most of you would not guess it in the first sense it is a farm pond, which is just full of water. And this water, part of it is reflecting because there is a wind blowing and the wind creates ripples. So if you go to a beach, on a beach, you could see that when the wind comes the front part of the water is white in colour, but the water is blue or white in the centre. So, if you take water, it is it is

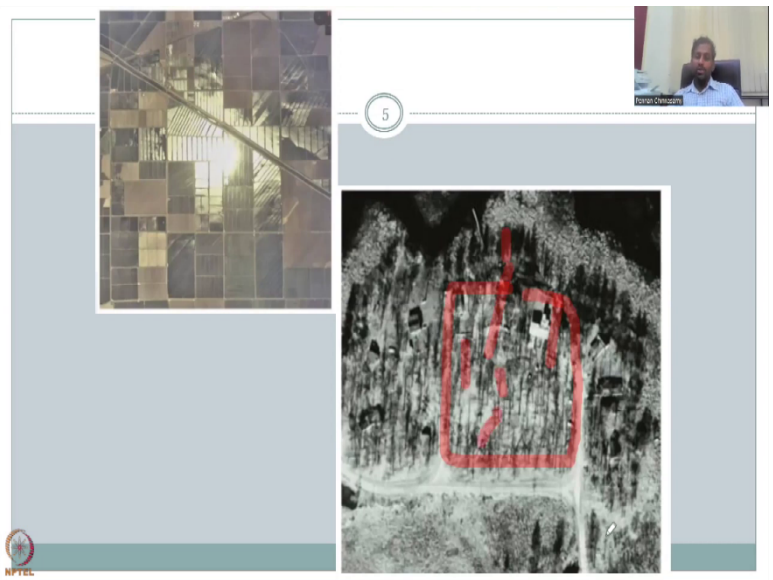
blue in the centre and then when it comes and then forms a ripples it forms up and creates a white colour for visually I am saying. So, visually these errors can happen.

However, you should be careful by just looking at how do you differentiate it, we can differentiate by saying there is a boundary. So, this boundary is constant, can you see and so, this is what you need for a water body you are mapping the water body, government of India's very important missions of Gati Shakti and mapping water bodies will use these kinds of techniques where they say okay there is a boundary and then there is a water body because some part of it is water this should also be water is not concrete because it is not built, if you take another image the best is to take another image.

(Refer Slide Time: 15:51)







Then we have other types of errors, which are the top is reflection error, again this is pure reflection, the previous one was there was a wind and the ripples happened the reflection was white, here you could see it looks like a glass pane glass window, but it is not it is not actually water blocks. So, you see the canal area and then water is being applied to the fields and only the fields which are full of water is reflecting sunlight.

So, the sun is on the top and so, it is reflecting the light and it looks blurry. So, here if you want to take this image you could have taken early in the morning or late in the afternoon when the sun is not perpendicularly up and shining on the land. And there are filters for it again as I said filters help only for a particular wavelength and it becomes expensive, but this image is still very valuable, because you could see like what percentage of land is water.

So, in rural development, one important aspect is availability of water for all. So, you could see that these lands have water, where does this land this land, these lands do not have water. Or maybe the crop they are using is different. So they do not have to water it for this particular period, the water sharing is important and you could see that along the channel, the plots along the borders get water whereas the ones away from the canal area does not get water.

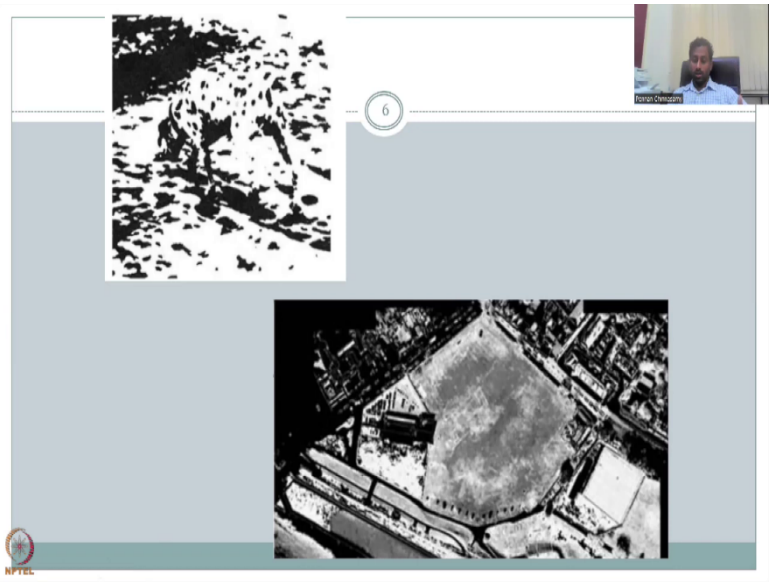
So, this is one of the benefits of raster images, satellite images. However, as I said, there are issues with these kind of reflections. This is a drone image and if the drone is tilting, if the angle is there, then you can see shadows and elongated image. So this is taken at an angle but most importantly, in the first image the sun was on the top the second image the sun was on the side. So, you see a image with reflection I am sorry, with shadows. So this shadow part

reflection is this one, this one is shadow. So, shadow part what you could see is the trees are here, but the shadows are elongated, because see many of these trees have elongated shadows and that is because of the angle of the sun and the time of the image.

So, these can actually be reduced by using correct time images and most importantly flying the drones and UAVs in a particular angle not a time. So make sure it is stable and horizontally flying perpendicular to the land. If it is like this or like this, then issues will happen in the image. And this is purely because you cannot control the wind see the plane is moving as per a particular wind speed but suddenly the wind speed can come up.

And once the wind speed comes up, it will tilt, the plane tilts or it goes like this and this can actually affect the camera power. But nowadays there are a lot of sensors to say that, you do not have the motion. But it is our important aspect to make sure that the data has all these errors removed.

(Refer Slide Time: 19:36)

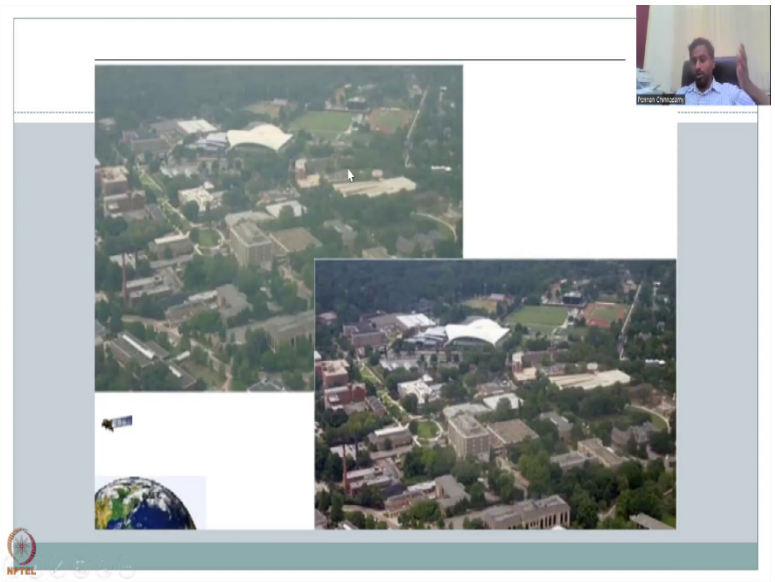


Then we have contrast and brightness issues. The first one you have lot of contrast difference between it is a black and white image, this is just a photograph that you can use. But the same principle applies to drones and UAVs also for capturing raster data. You could see that the contrast issue is there wherein you have a black and white and the white overtaking the black.

So, higher contrast ratios are given. So, there are some corrections you can retrieve the information. The bottom one has a playing field, it also has a contrast issue and a brightness issue on the borders you can see that these parts the image is not as good whereas, the central

region it is good. So, it is very important to make sure that you have taken an image with good contrast and brightness.

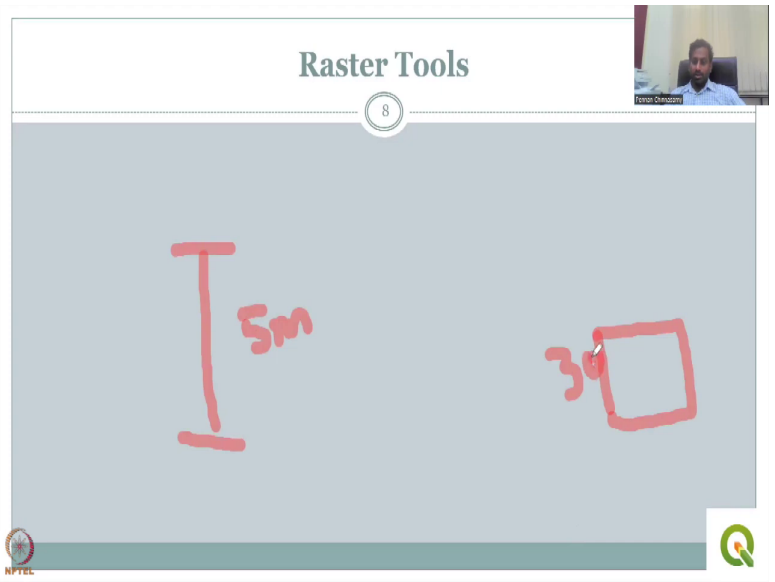
(Refer Slide Time: 20:31)

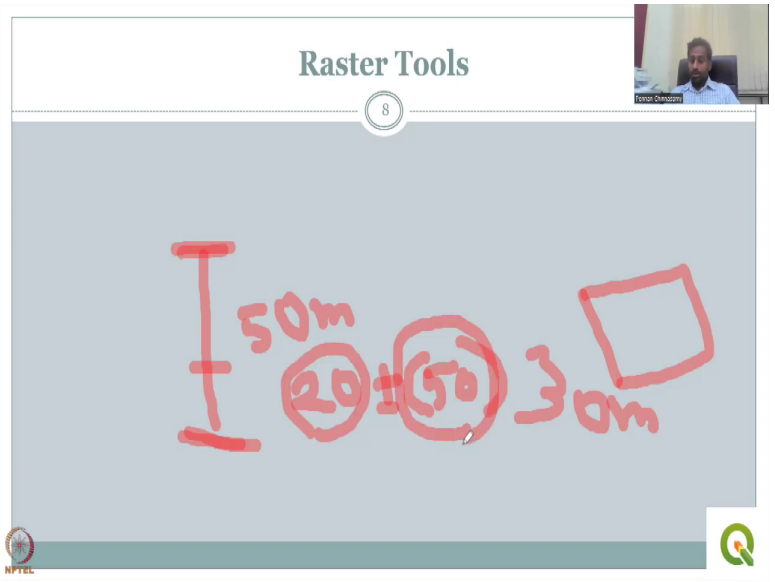


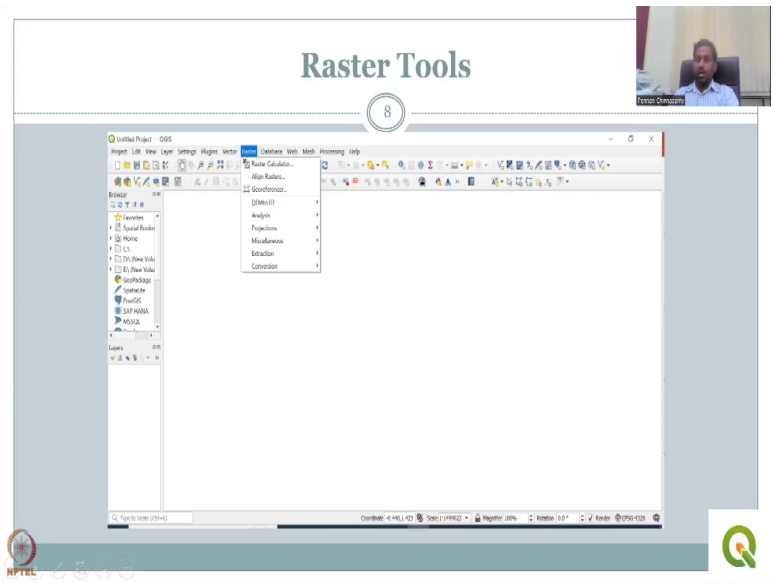
This is the condensation part where you see foggy and misty conditions in one image whereas, the other image is clear, both are at the same location same college my university but what has happened is in one it is clear whereas the other it is foggy and with the foggy you cannot take exact locations and measurements of the objects.

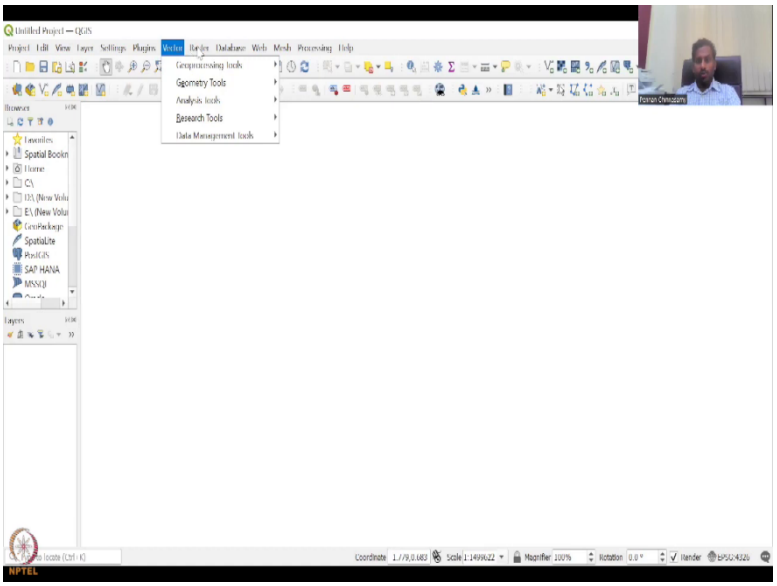
So, planning all these is very important for your future aspects of data collection. Make sure that you have these data checked do not see that why is this colour like this and if it is wrong and you cannot clean it, do not use it.

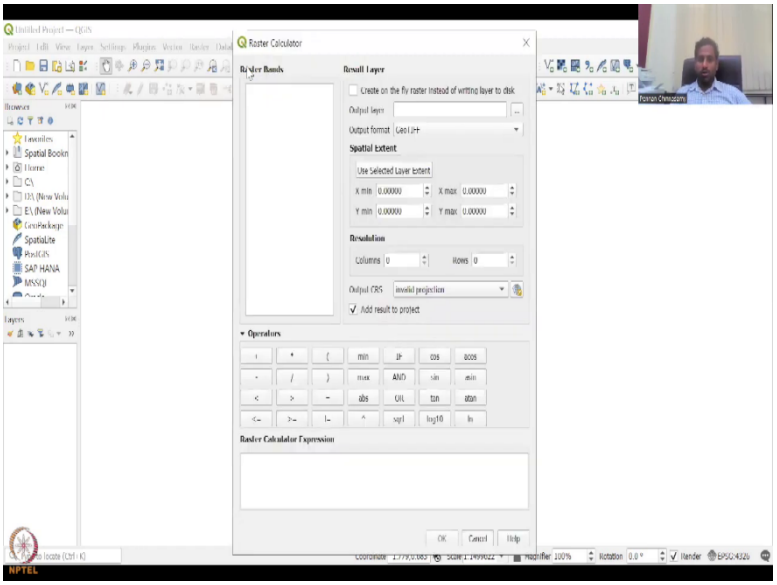
(Refer Slide Time: 21:22)

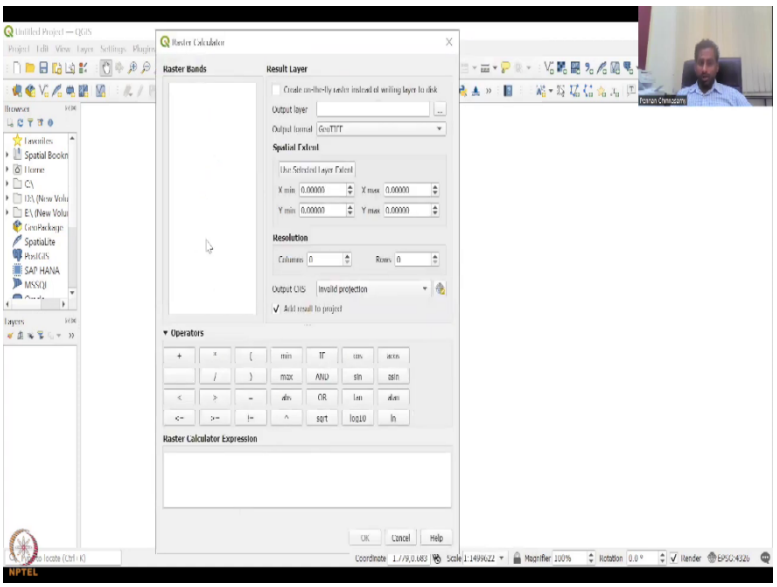


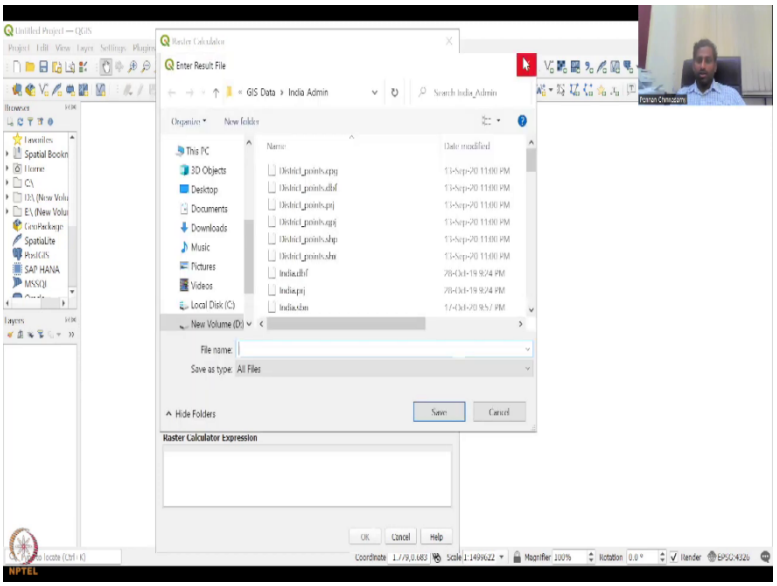


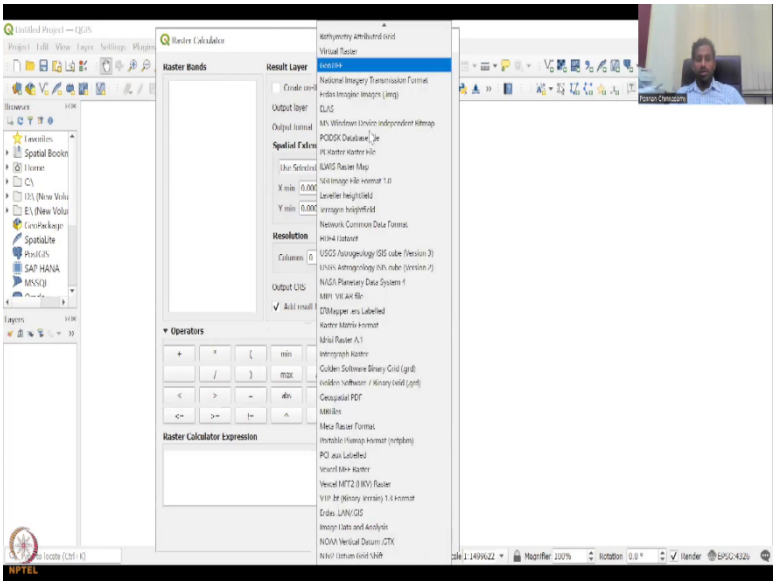

Raster Tools
8
Raster Calculator...
Align Rasters...
Georeferencer...
Analysis
Projections
Miscellaneous
Extraction
Conversion
NPTEL

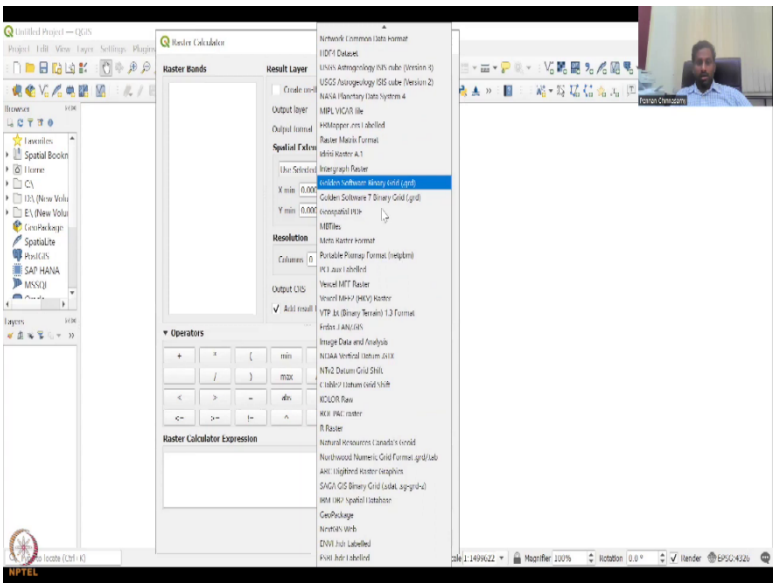

Geoprocessing tools
Geometry Tools
Analysis tools
Research Tools
Data Management tools
SpatiaLite
PostGIS
SAP HANA
MSSQL
NPTEL


Raster Calculator
Raster Bands
Result Layer
Create on the fly raster instead of writing layer to disk
Output layer
Output format
Spatial extent
Use Selected Layer Extent
X min
X max
Y min
Y max
Resolution
Columns
Rows
Output CRS
invalid projection
Add result to project
Operators
Raster Calculator Expression
OK
Cancel
Help
NPTEL


Raster Calculator
Raster Bands
Result Layer
Operators
Raster Calculator Expression
OK
Cancel
Help


Raster Calculator
Enter Result File
GIS Data › India Admin
Raster Calculator Expression
OK
Cancel
Help


Raster Calculator
Raster Bands
Result Layer
Operators
Raster Calculator Expression



So, with this I will get into the raster tools, we have covered the basics, most important part of the raster errors and issues most of the raster errors and issues in satellites are taken care of by giving a correction but it is our duty to read through the manual and apply the correction they will not apply the correction in some satellites for example, Grace, Grace has leakage error issues, this is an algorithm based error which has gravity leakages from the borders of land because sea also has gravity.

So, these can be taken care of by applying an algorithm and a multiplying the raster. However, Gray's team will not give you the full you know, corrected version, they will give you the version to be corrected, then the raw data you will have to multiply to get the corrected version. And this is important because some people do not want to use this correction, they want their own correction terms and they can use it for that the raw data is important and that is given.

So, multiple satellites are there. And every satellite has errors, because it is a mechanical machine with some moving parts. And it is an algorithm that uses a data digitalization. So, you can actually get more input from the metadata of the satellite, which we will cover. And these metadata are always attached to the satellites download page, you will have to download and read them before you use the data. Because a lot of times the data might be sounding good.

However, the errors might be too much. Let me show you a small example. That if your error bar is 5 meters in a data set, but you are only measuring 3 meters, your object is only 3

meters, let us say you have a land and 3, on each side you have 30 meters, 30 meters, and but the error is 50 meters. Let us say 50 meters is the error.

However, what the smallest size you want to measure is 30 meters. So how can you measure this with a scale which has 50 meters error? So, this is what you need to apply. The error bar is the difference between the errors the min and the max is 50 meters. So maybe you are doing this this is the measurement, it says 20. But how do you write it 20 plus or minus 50? Is this a good way of using the data? No.

Why? Because your measurement is much smaller than the error. So always look at the error before you download the data. If it is not you cannot use it. Another example I will give is a scale, if a scale you are using to measure this pen. So, the scale is 15 centimetres, but I want to measure this nip, this nip the front part is only millimetres thick or millimetres height. So, can you use a centimetre scale to measure this, you cannot, can you use a meter scale to measure this, you cannot, you need millimetre graduations in your scale. So, that is what you will be careful in writing and measuring. So, you will need a dataset that actually caters to your objective and problem. Moving on, what we will discuss is we will look at the raster tools.

Where do you find the raster tools? I am going to open the QGIS and then show you and then this is how the layout looks like. But we will show you in a minute. And the next class we will discuss on one or two of these raster tools that you will be using not all will be used in a basic scale and also for rural development not all are used in priority only 2 or 3 are very, very important, we will go through them.

The most important case is again, these tools have been extensively explained in a GIS course or a remote sensing course I am taking only one week of the entire course to explain these raster tools, it is our duty if you want to have more information, please get more info. We will be using the learned message to apply to rural development rural development is a core. But how do you use it and neglect these errors we will look at. So, moving on, we will look into the QGIS interface. Let me share the QGIS part. Whereas we have here and in this QGIS template. Once you open it, you open a blank window and you have this toolbars and panels etcetera. In the tool bars, you will be clicking the tools where you have vector initially, in week 5 we will have the raster. So each tool bar has its own help session.

So these are very important to look at the tools and then read them because these tools get updated now and then, for example, let me open one tool. We will go through this tool later. But I am just going to show you how it looks like in the interface. So you will have the input data here. So, once you input the data, which is in the layers that will come here as raster bands, and then you have do you want to create on the fly raster instead of writing layer to a disk.

What does this mean is it is a temporary storage, all the tools will have a temporary storage where it will get deleted after some time because you will be doing these calculations again and again, using the tools. You do not want to keep storing on your folders and then you do not know which one to use. So the best is to create on the fly. And then once you like it, then you can right click on it and save it. We will show you an example of how it is done. And then you have the output layer where do you want to store the file location those kinds of things. You can even add and delete based on it.

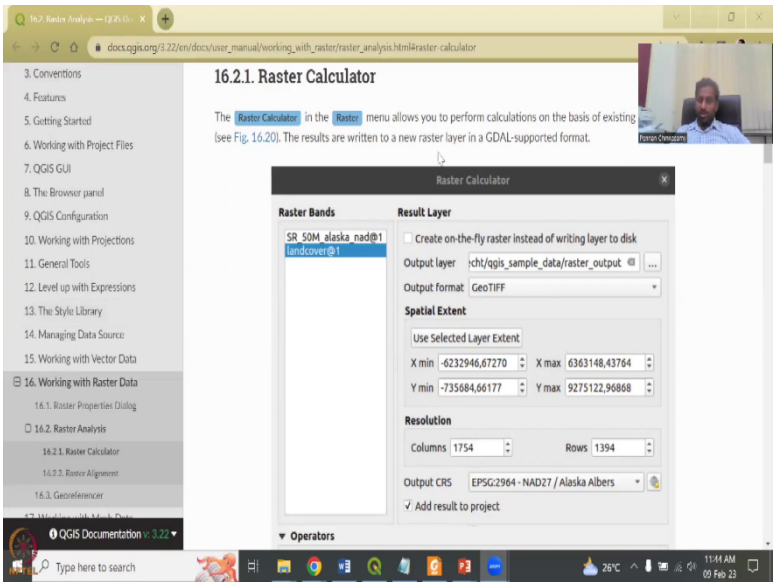
Then you have your geotiff, what type of raster do you want to store. In the class I have only said about geotiff, jpeg, raster, image all those things are raster, but just look at how many different types of raster are there. This is getting updated now and then as I said, while satellites are getting more powerful, sometimes they prefer a particular type of raster extension and that they choose based on their needs.

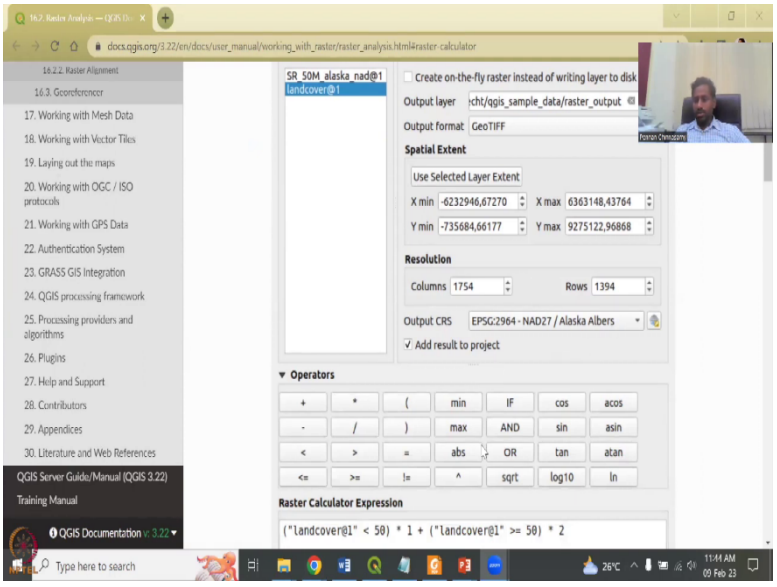
There is a NASA planetary data system, USGS data system etcetera. The most common ones are geotiff, raster HDF 4d data set, NASA planetary data set and yeah, image, etcetera. So you can you can pick and choose which output format you want. Geotiff is the default. Sometimes if you do not know just keep it to the default. It is always easy to interchange but if you keep it in the default it is good for use and then If you want to use an extent for the layer x y, z, the boundaries, you can use it.

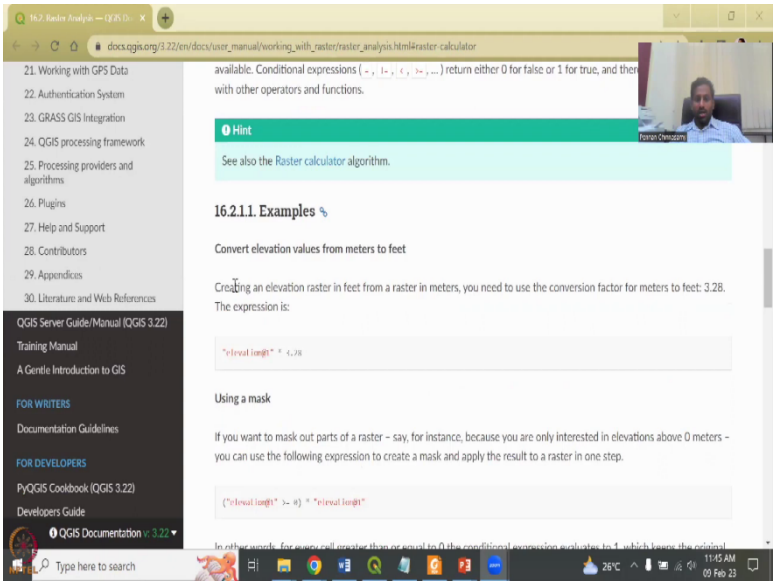
But again, default is fine resolution, what type of resolutions you want to get. And the coordinate system, the coordinate system was given here in this map, if you want to change it to a different coordinate system, you can apply that on here. See this, this coordinate system is different from this coordinate system. And you want to add yourself to the project. Which is this your project or you want it differently. Some tool boxes will have operators like this and an expression window. When you type an expression. It is something like programming you write a code, but understand here that you will not be writing the full code and running.

Because that is why you have a GUI the graphical user interface, you will only be using a code syntax, and all tools, as I said, have a help window, you could click the help window like this, I am going to click it will open a window page, which will open right now. And that is being stored. Whatever website you are using it will open up.

(Refer Slide Time: 31:21)







So let me share that window which has just opened and here you go So, you have the raster calculator, what did I switch there, in the raster calculator click help and this manual in QGIS opens up. So, this warrants that we need to look into this, but you can see here that the raster calculator is being explained the figure and how it is. So, as I said two data have been put in the output layer, they want it to be write it on a disk.

So they have clicked output folder and then put it in GeoTIFF is being used the layer extent has been used columns resolutions, Alaska example has been taken and more importantly operations. So, if you look at these operations, what is the operation saying that land cover one has been taken and only the ones lesser than 50 is multiplied by 1 plus land cover one greater than equal to 50 is multiplied by 2. So, those cells which have land cover less than 50 are multiplied by 1 and land cover above 50 or above an equal to 50 multiplied by 2. So now you have a updated or upscaled raster and that raster is stored here the names is hidden, but you can store it in a different name and then it writes it to the layer.

So all these are there, these are examples of the expression we will go through this in the thing, but more importantly, this is how you will learn by yourself by using the help window. So, the objective of this exercise was to show that each tool has a help window and you will be using the help window to maneuverer between your work because that is very important to have in terms of, in terms of using a tool.

So with this, I would like to conclude today's session or and we have looked into multiple formats of data, raster and vector and in particular, we have looked into a tool box and where to find help to meet and understand the tool. Why this is important is for you to learn as the

tools get updated frequently and you should not be left back because you did not know how to use the help box. With this I conclude today's session. I will see you in the next session. Thank you.