Remote Sensing: Principles and Applications

Prof. R. Eswar
Department of Civil Engineering and Interdisciplinary Program in Climate Studies
Indian Institute of Technology-Bombay

Lecture-59
RS Data, Data Portals and Processing Tools-Part-1

Hello everyone, welcome back to the next lecture in the course. Till last lecture we have finished discussing about the fundamental principles that a new student is expected to know about remote sensing or which will pay way for further understanding about the subject. We have provided a very broad overview and detailed concepts in some topics which will help them to further dig deeper into the course and understand its beauty and the various ways in which the remote sensing can be used.

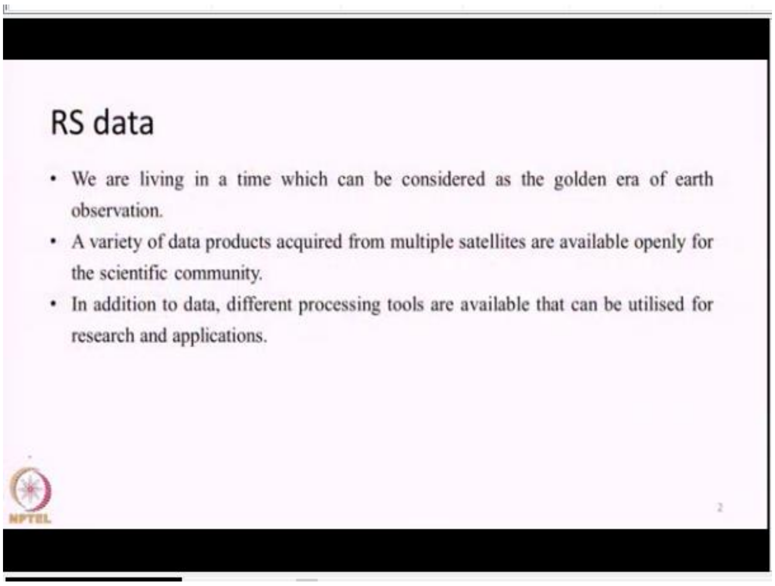
So, we are now reaching towards the ending part of the course where we will briefly get introduced to various data sets and data portals from which we can download the data and some data processing tools. Remote sensing is a combination of both theory and practicals. We need to process the data to get some information. Let us say government agencies want to monitor how much the land cover changed within a city or a state. So, they want to understand how much the landscape is changing; say forest may be diminishing, urban may be expanding, agriculture may be expanding and so on. So, naturally different agencies be it government or private agencies will want to know all these things. This is just one example of application of remote sensing technology. So, there in order to achieve or in order to get this application done we will be in a position to identify some remote sensing based data sets.

All these data should be processed using certain processing tools and using certain algorithms, only after that we will be able to achieve our application. So, essentially a major part of success of remote sensing lies in identifying what data to use and understanding what are the processing we should do for ultimately achieving our objectives.

So, in this lecture I will briefly introduce the various remote sensing data sets that are available, from which data portals we can download them and also certain open source processing tools that are available to us. Some of the commonly used data portals also have publicly available data processing tool which we can use.

First let us discuss about remote sensing data sets. Actually the current time period in which we are living can be considered as the golden era for earth observation.

(Refer Slide Time: 04:24)



Because the amount of information or the amount of data and the variety of data that we get is unimaginable, we are generating petabytes of data every day. Unless we want really specific data like a very high spatial resolution data or for a particular project the data should be operating, unless we are extremely specific about data characteristic, most of the data sets for our applications are available in the open domain and also the processing tools are also available in the open domain, we need not buy expensive commercial tools.

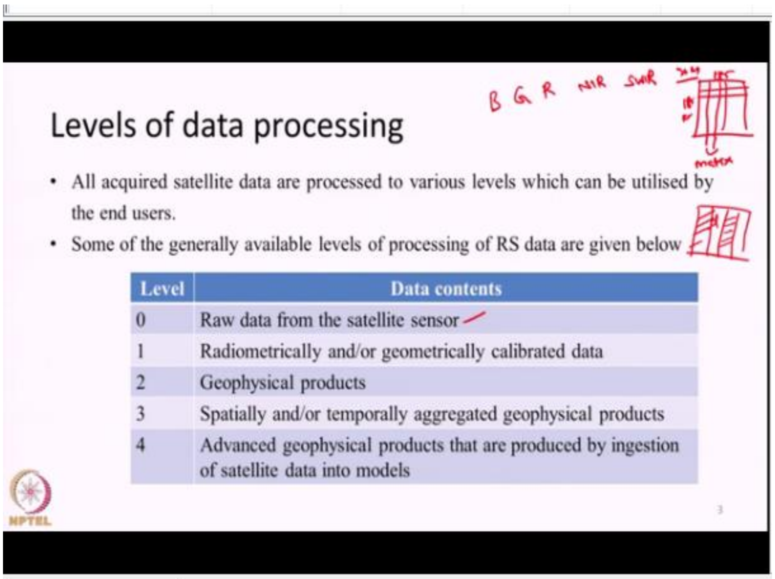
This is the data I need, these are all the processing I should do and this is a tool I should use. So, even for getting this knowledge the amount of information we need is all available publicly in the public domain. So, that is why I said this is like a golden era for remote sensing or earth observation as a whole. So, here what we are going to see is some of the few commonly used remote sensing data sets or publicly available remote sensing data sets which has the potential of being used in large number of application.

First we should know the levels of data processing. Say whenever a satellite or sensor acquires data, the data is not directly transferred to the user, the data from the satellite will be transferred to a ground station and the ground station will receive it. The received information will not only have remote sensing data but all other information about satellite, orbital characteristics, satellites health characteristics, the sensor, how they are performing, etc,. All those important technical and sensitive information will be present. So, naturally the space agencies will collect this data, will process them to various different levels and then it will provide us. And also the data about the earth surface coming in from the satellite may contain a lot of errors; like errors in terms of radiometric quality, errors in terms of geometry quality. So, all these things have to be corrected before being used by the end user.

We have already discussed about calibration and there should always be a correlation or relationship between what a satellite measures and what we intend to measure. Satellite measures radiance, but what we will be getting in the image is DN. So, how to convert it? We should be able to properly quantify what is being measured from the data contained within the image. There should be a relationship, all those calibration may be applied only after receiving the data, they may not be applied from the satellite, a satellite will simply observe the data and may transmit it back to the ground station. The ground station systems will apply this calibration coefficients on those equations that will convert the data to more meaningful form.

So, the data directly coming from a satellite will be of little use to the end users. It has to undergo several levels of processing and normally each space agency which provides data, will do certain level of processing to the data before it is being delivered to the user. First we will understand the basic levels of data processing.

(Refer Slide Time: 08:24)



First is level 0 or the raw data from satellite sensor, whatever is coming in from the satellite people call it as level 0, it will not be shared, because as I told you it will contain not only the data about earth. It has to undergo various error corrections, calibration and so on. And it may also contain lot of other information about satellites, spacecraft itself, many different things. So, normally level 0 data will not be shared with users, it is the raw data received by the satellite.

Then the useful part of the data from a user's perspective will be taken, it will be undergoing some sort of radiometric calibration, geometry calibration, error correction and so on. Then we will have level 1 data. So, a level 1 data is basically the first level of useful information that a user can get. We will get what is known as digital numbers within the image. So, it will be a two dimensional matrix, each band is one matrix.

For example let us say our system has 5 bands, blue, green, red, NIR and SWIR. So, let us say it is covering an area of about 185 kilometers by 185 kilometers, one image at a resolution of 30 meters. So, this is an example for Landsat system. You will have multiple 30 meter pixels and each band will have one such kind of a matrix. So, you can think in terms of a 2D matrix arrange in terms of rows and columns, different, different latitudes, different, different longitudes, along the columns. So, each band will produce one 2D matrix array and will have 5 matrix. So, this particular information is the remote sensing data and each pixel will contain the DN number.

So, the output in level 1 will be this DN, you will have 5 different images, each image will be containing large number of pixels, each one having a DN value, from this DN value we will be able to calculate the radiance recorded in the sensor, because calibration is done. So, the first level of useful product that we may get is level 1. So, within this level 1 there can be several levels, level 1A, 1B, 1C and all which varies with different, different agencies. There are no hard and fast rules, how many different sub levels one can go? Certain agencies say 1A, 1B, 1C certain agency may stop only at 1A, 1B and so on. And each level or each sub level may indicate a different level of processing. Level 1A may be just a calibrated data. Level 1B can be calibrated and radiometrically corrected data. Level 1C may be calibrated radiometrically as well as geometrically corrected data and so on.

There are plenty of different sub levels that exist, calibration is providing a relationship between what is stored in the image and what actually the satellite measured. But what is this geometry correction, radiometric correction? Say radiometric correction is to remove certain unwanted artifacts within the image. Let us talk in terms of a push broom sensor, there will be

100s of detector elements oriented in the across track direction and as the satellite moves it will collect information across the swath. So, there are many different detectors. Each detectors may have a different response, say one unit of energy's radiance is falling on the detector, the output voltage produced from one detector will not exactly be the same as the output produced from the next detector there can be minor variations.

Finally after arrange everything, calibration is done to produce an image with stripping effect. Stripping effect means the image may not look uniform, within the two dimensional image, this block may appear bright, this block may appear dark, again this block may appear bright and so on. This is because of the difference in the way that the data is recorded by the detectors. Stripping effect is just one example of radiometric error, or there can be one full line of data missing, like one detector might have failed. Or one particular column in the image may be missing. So, this is called line drop out or column dropout. These are some examples of radiometric errors. There are various algorithms to remove it and finally we get radiometrically corrected data.

Geometrically corrected data means say satellite will just take a snapshot, the snapshot has to be properly referenced to a ground point. This point in the image corresponds to a certain point in the ground, unless that relationship is established properly, an image is of very less use, it cannot be properly termed as a remote sensing image. There should be a proper relationship between each point in the image and the corresponding ground coordinate. Sometimes certain points can be mis-oriented, there can be geometric distortions happening like a circular feature may be appearing ellipse, all these things can happen. So, people use ground based information called as ground control points or some other ground based information to correct this geometric artifacts. So, all these things will be done one after the other and finally we may get radiometrically corrected and geometrically corrected data.

All these things may come under the same term level 1, level 1A, 1B, 1C and each agency can have different sub levels of processing. Then comes level 2 data. So, level 2 data is actual geophysical product. Say level 1 just tells us DN, with DN we have to convert it into meaningful quality, say if we are talking about visible and NIR bands, the intended output for us is surface reflectance, we will be interested in calculating the surface reflectance. So, we have already discussed in detail about the steps.

We have to take this DN value, convert the DN to radiance, do atmospheric correction on the radiance, we have discussed about dark object subtraction method briefly to do this. So, there are plenty of other complex methods too. Basically we take the radiance, do atmospheric correction over the radiance and calculate surface reflectance. So, all these processing might have been done by the space agency themselves. So, level 2 basically talks about geophysical product. If it is in visible or NIR or SWIR domain we will get surface reflectance. If it is thermal infrared we may get the land surface temperature and emissivity. If it is passive microwave then level 1 data may contain brightness temperature, because essentially what the passive microwave sensor measures is brightness temperature. So all these are level 1.

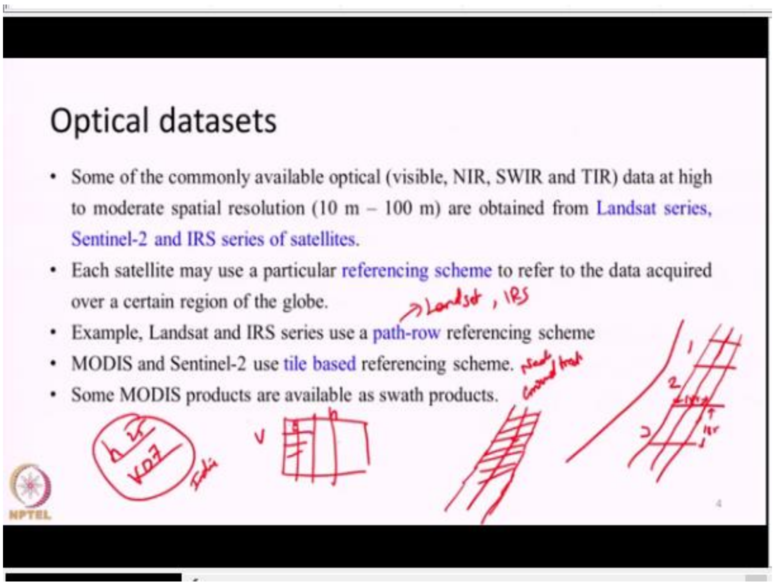
Level 2 maybe soil moisture product or a vegetation optical depth product or sea ice thickness product and so on. So, level 2 products are essentially some sort of geophysical quantities after certain level of processing to the level 1 data at the same resolution or same data characteristics as that of level 1. Let us say Landsat level 1 data is 180 kilometer by 180 kilometer swath image. So, level 2 is same characteristic data, but instead of DN it will contain reflectance value or instead of DN it may contain a temperature value and so on.

Level 3 is same geophysical product but aggregated to different spatial or temporal resolution. Say sensors like MODIS or VIIRS may produce more than one image every day over a same location which is a high temporal resolution. But not all the days you will get image over all ground points, certain points may be covered with clouds, over certain points the look angle may be very wide leading to a very large geometric destruction and so on. So, normally MODIS data products will be aggregated, they may provide rather than providing a daily image they will also provide an 8 day composite in which all the observations done during the 8 day will be analyzed and only the best observed variable will be populated within the 8 day. So, this is a level 3 temporally aggregated data. Similarly, it can be spatially aggregated. Say the natural resolution of the satellite may be 500 meters, but some users may be requiring data at say 0.05 degrees climate grid, MODIS provide such a data.

So, spatially and temporarily changed geophysical product is called level 3 and finally comes level 4 where level 4 means it is advanced geophysical product which is not directly observed by the satellite. But by ingesting the satellite data into various models we will get those variables as output. Say for example using the surface reflectance and other ancillary data we can calculate what is known as leaf area index. Leaf area index is one of the important variables related to vegetation growth or you can calculate the biomass or gross primary productivity. So, satellite will not directly observe all these things. Satellite will observe radiance, brightness temperature and so on. We need to do some sort of processing in order to get to level 4. So, level 4 are advanced geophysical products produced by ingestion of satellite data into models. So, for each advanced level the previous level will be the input. To produce level 1 you need level 0 data. To produce level 2 you need level 1 data and so on. So, it comes as like a sequential chain and all satellites or all space agencies may not provide all levels of data, some may provide only level 1 data, some may provide only level 1 and level 2 data, some may provide all 4 levels of data and so on. So, this is the most generic representation of levels of data processing which we can use or which is naturally available to us.

First we will discuss about optical data sets. Optical dataset means visible, NIR and SWIR as well as TIR domain. So, what are the commonly developed data sets available to us? So, we can use data from Landsat series of satellites. That is one of the most widely used optical data sets starting from visible to thermal infrared wavelengths. Then the recently launched sentinel-2 is again a very good source of optical data sets. Sentinel-3 has its ocean colour sensor and sea and land surface temperature and all these things are example for optical sensors which provides us data in visible NIR, SWIR and TIR domains.

(Refer Slide Time: 21:18)



Also we can use data from Indian remote sensing series of satellite. India has a very broad range of satellites providing different data sets which we can use. Based on application we should understand which data to use.

If you have decided then we should properly understand what data we should download? Whether to download level 1, level 2, level 3, we should analyze and understand which is needed to us. After understanding this we should download the data. So, when we need to download the data we should go to the data portal and download it. Each satellite may have its own spatial referencing scheme to refer to one data. Say I want image acquired over Mumbai. So, I can just go to some data portal and search for Mumbai and download the data. When data comes, Mumbai is a region and Mumbai city may be falling within a small portion of one full image or a small portion divided into 2 or 3 images depending on the satellite data collection, how it happens? So whenever an image downloads for me, we will have a referencing scheme, if your location is this then the data about that particular location will be covered in this particular image. Each image or each data is provided with a proper tag for us to identify it. So, that particular geographical tag is what we call referencing scheme. There are plenty of referencing schemes available.

For example path row referencing scheme, a tile based referencing scheme and so on. So, what is a path based referencing scheme? This is widely used for Landsat series of satellites and Indian remote sensing series of satellites. Essentially there can be unique number to each orbit; orbit 1, orbit 2, orbit 3 and so on. So, let us say after 16 days orbit 1 repeat. So, within the 16 days it will start from orbit 1 it may end in say orbit 233, orbit 255 and so on. So, there can be that many unique orbits to cover the entire globe. So, each orbit or each ground track covered by the satellite will have a unique number. That is called different, different path.

So, they will have path information; say 233 unique paths in order to cover the entire globe. Whenever the satellite covers this entire 233 unique paths it would have produced image over the entire globe, maybe once every 16 days or once every 24 days and so on. So, for Landsat series of satellites they will provide, this is the path and within the path it will be continuously producing strip of images. We may not need to download the entire strip along the swath it will be enormous amount of data, they will cut this path into several small, small segments. One full orbit will be divided into small, small rows. This is maybe row 1, row 2, row 3 and all. This may be covering 185 kilometers. So, again the swath may be covering 185 kilometers. So, this particular block is called one image. So, that particular image will download. This is actually fixed for a certain location; this is your path row. So, that particular location will always be covered in that particular path row for Landsat series of satellite.

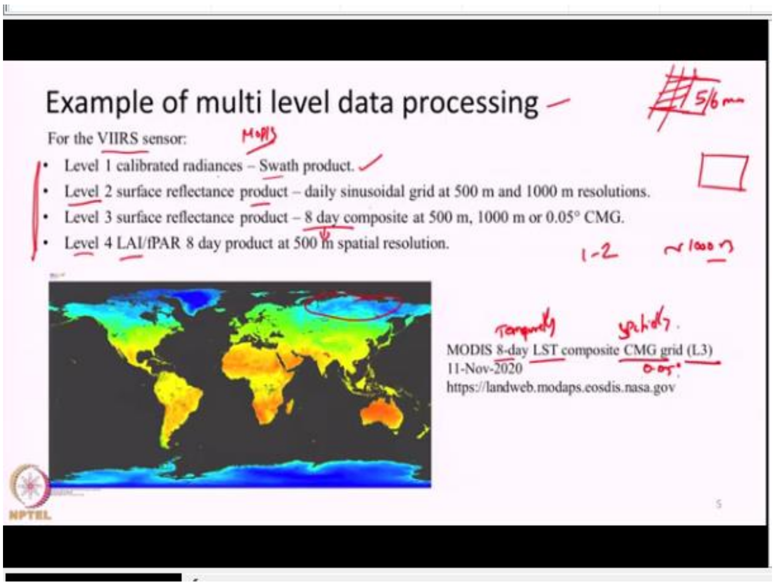
If you talk in terms of MODIS, MODIS will not provide you images in terms of this path row, because the orbits will differ. They have divided the earth into fixed tiles and each location will be located within a particular tile say they will give hv(horizontal, vertical), say h25, v 07 it is a two dimensional coordinate system, it may be located over India, we can identify h 25 v 07 is like MODIS coordinate system over southern part of India. So, as MODIS sensor has a very wide scanning angle same point on the globe can be covered from different, different orbits. So, from whatever orbit the ground point is covered the data will be processed and that will be populated onto this grid. So, we will not be knowing from which orbit it was imaged or from which scan angle it was imaged. That information will be there but naturally when we look at the image we will not know, we have to dig deeper into the metadata.

So, the swath information will be removed, everything will be converted into this kind of grid. Each satellite system may follow their own data referencing scheme and with this data referencing scheme we may be able to download the data over the part of the globe where we need it. So, this varies with satellite. So, whenever we download satellite from Landsat series of satellite or IRS series of satellite we may get image with path row system. So, when we search for data we should know these things to certain extent, we can always search with the ground location we need. Say if I need data over Mumbai I can just go search for Mumbai and whatever data available over Mumbai will be downloaded, but knowing this sort of referencing scheme available for each satellite may actually help us.

So, just one example of this multi-level data processing, say there is a sensor called VIIRS visible infrared imaging radiometric suite which provides data in various levels. Actually if this VIIRS sensor and MODIS sensor are all combined together gives us 100s of different data products. Say the level 1 data can be a calibrated radiance. They would not provide DN values, they will provide you radiance which is a swath product, that will be available in terms of swath. This is the nadir track, this is the orbit coverage, you will get it. So, every 5 minute of along track information or 6 minute of along track information will be cut, it will be given to you as level 1 data.

Level 2 product may be a surface reflectance product in a sinusoidal grid. Say h 25 v 07 like that. That is one. So, level 3 may be an 8 day composite surface reflectance product. So, level 4 may be LAI product which ingested this surface reflectance data and other information. So, this is just one example for data from an optical remote sensing sensor, like visible, NIR, SWIR and TIR bandwidths.

(Refer Slide Time: 28:46)



And what I have shown here is an example for MODIS 8 day LST composite in a climate modelling grid (CMG). So, this is both temporally as well as spatially aggregated. So, this is a level 3 data. MODIS observes the earth every 1 to 2 days at thermal band. The spatial resolution is roughly 1000 meters. So, this is aggregated to 0.05 degree grids and average to 8 days. So, the blue portions are cold areas, temperature is less, yellow portions are slightly warmer, red portions are actually hot desert areas and so on. So, this is an example for level 3 data.

So, just as the overview of this lecture, we are discussing about the various remote sensing data products that are available and we discussed about different levels of data processing and we just started discussing about the optical datasets. We will discuss further in the upcoming lectures.

Thank you very much.