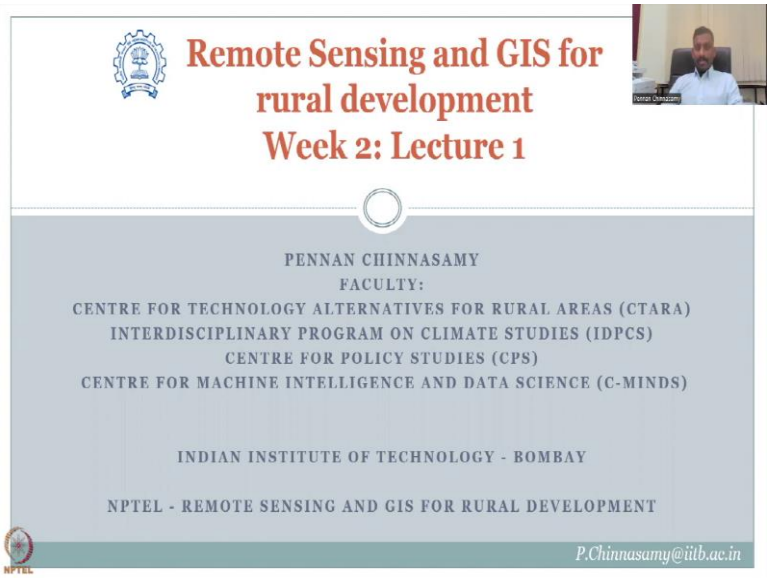
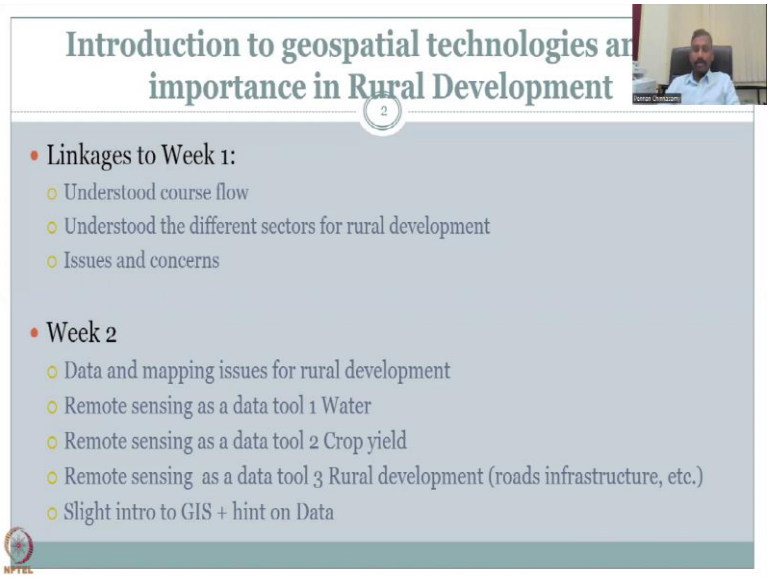
Remote Sensing and GIS for Rural Development
Professor Pennan Chinnasamy
Centre for Technology Alternatives for Rural Areas
Indian Institute of Technology, Bombay
Week - 02
Lecture No - 01
Data and mapping issues for rural regions

(Refer Slide Time: 00:18)



Hello everyone, welcome to NPTEL course on Remote Sensing and GIS for rural development. This is week 2, lecture 1. In the last week, we had some introduction to the course flow and as per the course flow, we will be discussing still the need for Remote Sensing and GIS in rural development in this lecture

(Refer Slide Time: 00:58)



Before that, it will be good to understand, what are the key topics that will be covered this week. We will give some introduction to geospatial technologies and how they are relevant to rural development. But first, why is the question? So, let us link week 1 and week 2 together. In week 1, we understood the course flow as in, what will be covered and in what weeks.

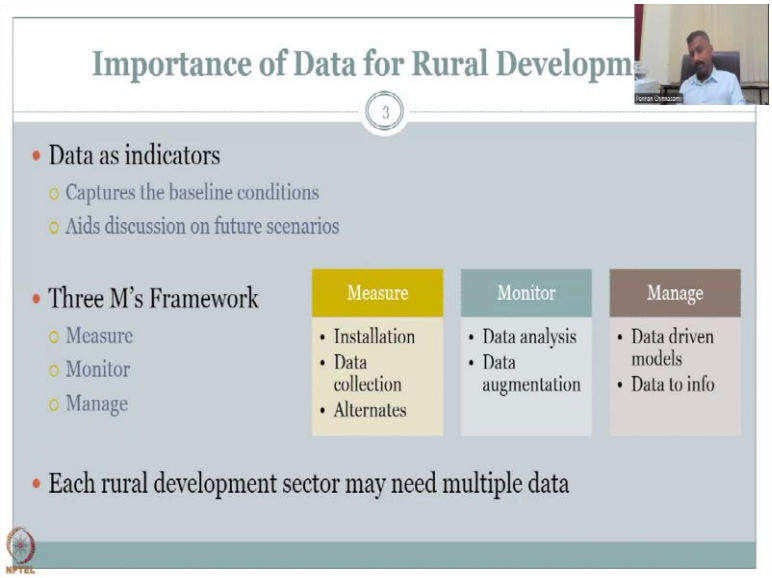
We also looked at the possibilities of using other lecture series especially from NPTEL course archives into your current course. We also understood the different sectors for rural development. And under the Ministry of Rural Development, we looked at the different departments. We also looked at similarities for rural development or overlaps or you could say commonalities between different ministries.

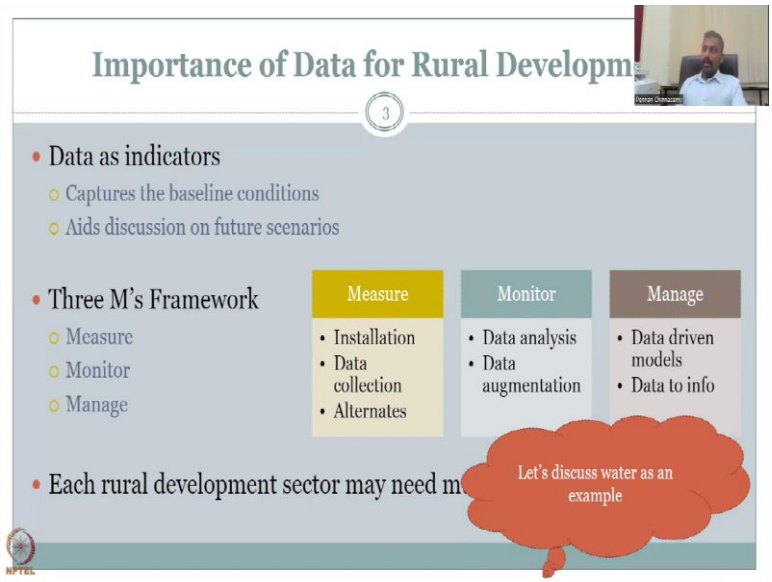
Let us say environment, food, farmer welfare, agriculture, all these are interment. There are separate ministries, but can overlap with Rural Development Ministry. We also looked at issues and concerns and how we will to address, the major issue is there is a lot of schemes a lot of positive work that can be done.

However, mapping the needy with the schemes is tricky. For example, we need to map the water to the water demand scenarios. If we do not know the water demand scenarios, it is hard to map the water either we will supply surplus to a less needed water place or we would be supplying less, or somewhere where water need is predominately high.

So, let us see how we are going to do about it in week 2. In week 2, we will still look at rural development from the angle of remote sensing and data development. We will discuss data and mapping issues for rural development. Basically, looking at where the data exists and how there are mapping issues in the current scenario. We will also look into how this can be mapped to rural development. As an example, we will consider topics on water, in one lecture. We will consider topics on crop yield, in one lecture and infrastructures. If you can recollect, these are the same schemes that we discussed earlier. And then I would also like to close this week by giving some intro on GIS and what are the data that is needed for GIS assessments.

(Refer Slide Time: 05:03)





Let us look at the importance of data for rural development. As you would know, there is tremendous need for all development in India and data can support decision makings in rural development. The first thing is data as indicators. So, the data you capture will assess the baseline condition. What do you mean by baseline condition? Baseline is the stage at which you start or you keep as an average.

So, you want to build on top of it. So, only when you know what is the current status you can build on top of it, and that is where data need is very very high, the data is converted to information and information can be converted to indicators, and these indicators capture the baseline condition.

So, once you know the baseline condition, then you can build scenarios that say I know how much population it is. So, I would know how much to produce food for attained food security. If I do not know how much population it is, and what is this consumption rate, then how will we invest in food security another quick example is, children and milk supply, if we know how many new-borns and mothers that need milk is there, then we can set up cooperatives and bring in stocked milk, this is very important because the shelf life of milk is very less you cannot keep it for weeks you do not, freeze it and keep it for weeks, you normally buy in a packet, use it within a day or two and then buy another packet. So, this is where knowing the baseline condition is important. And data plays a vital role for that. And once you know that you can add on other scenarios future scenarios to the baseline.

So, there is very critical thing is that there is three M's for data in rural development. First thing is measure you need to measure something, so the volume of this bottle needs to be measured so that I can refill, you need to measure the number of babies that are new-born that need milk. So, measuring is the process of collecting data. If you do it only one time, it is not monitoring.

So, monitoring what is the difference biggest differences is? You collect one month the next month you come and then collect again maybe some more new-born kids or there or some people have left the villages to a different village to update you need to monitor. So, once you measure you can monitor in different intervals and these two M's give to the third M which is manage. So now, you can manage what is needed, let us say water for agriculture or as a example I have given milk for the new-borns. So, three M's framework is very very important in this aspect.

Hear data collection comes in the first aspect itself measuring, measuring how much the population is, measuring the volume of the container for filling water etc. And then monitoring is regular measuring, regular data collection. And then how you collect the data and convert it to analysis is also part of your monitoring. Once these two are achieved, you get into managing the resources better.

So, measuring involves installation for data collection. It could be instruments, it could be tools to be a satellites that can be traced and then collect data. So, there is an installation phase. Then there is a data collection phase, where data is collected in regular intervals. And ultimately, data is identified. So, this for example I do not have the historic data. So, how can

we go back and install and collect data we cannot. But we can identify alternatives, proxy data that can be used to measure.

Now, once you have this measurement data at regular intervals, that becomes monitoring. And as I said, data analysis is the process where you convert this data into information, some kind of result that you can use for future steps. Data augmentation or value addition is where you bring other data to your results and add value to the data.

Let us say for example, the rainfall that is recorded by the government in India, they monitor at different points, but what happens between the points. So, here is where you can bring a satellite image and correlate these two values to interpolate the data you get a better data of rainfall across India, not in two points.

For example, there is one in Mumbai there is one in Pune but what is the distance between that and how do you account for variations in typography and climatic factors? So, that is done by data augmentation. Then we have the final step, who is this managing the resource and to manage the resource there are data driven models, we are only talking about data here.

So, it has to be driven by models and the models given output which can be used for management, there are plenty of data driven models and how it works is, by definition of the problem or how the model was made. Finally, your results become an information as I said, you have a decision support system or a model you feed in the data that you collect through the installation and data augmentation etc.

And the data is driven in the models to finally give an output, the output becomes an information for example, you do the rainfall model as I said, you interpolated and then you predict rainfall in different regions. Now, you can send the rainfall data to the farmer through a mobile device as an information or you could send it as a broadcasted message through TV or radio as a warning.

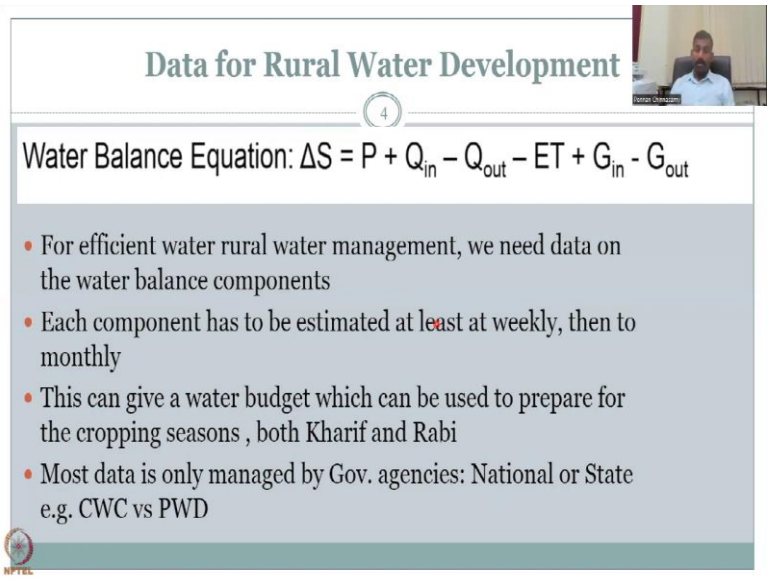
So, these are needed for rural development, for example, if I could save farmers losses by protecting them from agricultural activities during a storm that will lead to rural development because they save the economic resources, the social resources and the labour cost which can they use for other reasons.

Moving on each rural development sector may need multiple data. This we have already seen. There are multiple sectors water, soil infrastructure, education, so each would require its own data, for example, health, the new bonds in rural villages, mostly some of them are managed

malnutrition. And how do you assess their malnutrition is by assessing their body weight to the height and body mass. That is not appropriate. To be used for rainfall, for example.

So, each role government sector may need multiple data not one or two data. Sometimes the data overlaps like water availability or storage for food etc. As an example, to drive this point very clearly of importance of data for rural development that is discussed, as I said, one of the most important sector which is water, as an example.

(Refer Slide Time: 15:34)



So, what you see here is we are going to discuss data for rural water development by using a mass balance approach, a mass balance is a water budget. And this water budget gives the net water availability or net stored water that can be useful future developments. For example, here is your del S, which is your net storage or the remaining water in the storage, it could be a check dam, it could be soil storage, it could be groundwater storage, anything.

We will call it for now, del S is equal to precipitation, which is your rainfall, snow, mealt anything that is input to the system precipitation plus Qin, which is a discharge that comes into your watershed. So, for this equation, the watershed an area has been demarcated and that area inside that all parameters are assessed.

So, you have rainfall coming to that area, which is called P and Qin is the water that comes in to the surface, it could be a pipe connection, it could be a river channel flowing into the piece of land that we are assessing. Q out is how much water leaves the system. So, as water comes, water gets used and then some water can leave the system that is minus you can see the negative sign Q out. And then evapotranspiration is the water used by trees, plants, when

they transpire and evaporation is mostly from the open surface and water opens up surface, sunlight acts when atmospheric conditions and then water vapor is evaporated.

So, there is evaporation and transpiration which is a combined effect of your land, water and plants that transpiring So, it is a loss to the system. So, you see a negative sign here. Then you have plus Gin which is groundwater in. So, because your land can take groundwater from using pumps etc you have a positive sign, so that can be used to the field for watering the plants and then we have groundwater out, groundwater out is the groundwater that you take out and it is lost from the system.

So, it is not stored because groundwater in is the recharge water it can be coming from another land parcel or a watershed can come into your watershed balance. So, we have the positives and negatives, now already you would have known that wow, we need to know the rainfall, the discharge, river discharge coming in, the evaporation the transpiration, the groundwater in, groundwater out, those are all data points that you need.

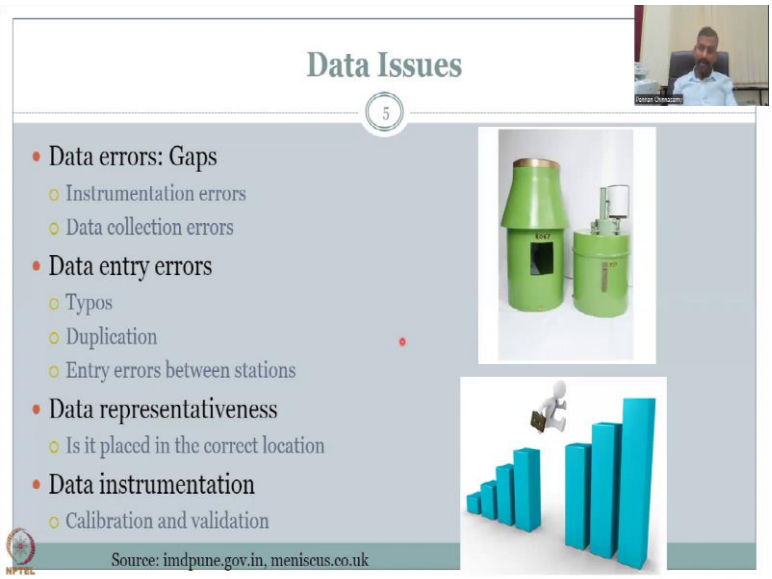
For efficient water management in the rural regions, we need data on the water balance components. And this happens by looking at each and every component individually and as together in a water balance equation and there is a temporal issue. So, there is a spatial saying that every block has to be measured as I gave you the example of a rain gauge sitting in Mumbai and then Pune, so in between maybe you are interpolating.

So, each component as they are both said equation has to be estimated at least at weekly then to monthly. So at least weekly then you convert it to monthly and then to annual decades etc. So, this can give a water budget which can be used to prepare for the cropping season both Kharif and Rabi for those who are new to these terms. Kharif means monsoon crops, which are rain fed. So, rain water is used for growing their Kharif crops, mostly your rice, paddy, wheat, etc come under this.

Your Rabi crops are the crops that are grown in the non-monsoon time, by using water resources that are stored for example, you have captured the water and put it in a dam you have captured the water and put it in the groundwater you can use it during the Rabi season. So, what it explains is? This water budget can give an idea about how much water is coming in and used and how much water remains for the Kharif and Rabi season. So, you can plan better for a better crop yield and crop production, which leads to rural development.

However, mostly data is managed by government agencies national or state example CWC and PWD for the water boards. So, CWC is Central Water Commission, PWD is the Public Waters Department, for example Tamil Nadu and in Tamil Nadu, there are two agencies that are monitoring these data. They may talk to each other, but they may also not talk to each other, which means sharing of data. So, there is a lot of duplicates in data there is also issues and data that we will see.

(Refer Slide Time: 21:50)



So, what are the issues in data? The top one image the green image is a rain gauge, an instrument used to capture rainfall and convert it into the water balance we saw in the earlier slide. It is like… what are the issues in this, capturing data there are instrumentation errors, instrument always have some kind of friction or errors, they may malfunction, you have data wrong data coming in.

So, always they may have errors, it is not a perfect system. And data collection errors is how you collect the data. For example, you go to the field and you write it wrongly or while writing you mistakenly do a typo. So, the data entry errors that I am talking about are typos, even one word or the decimal point you figure out that you kept in a different location, then you change the order of magnitude of the sample.

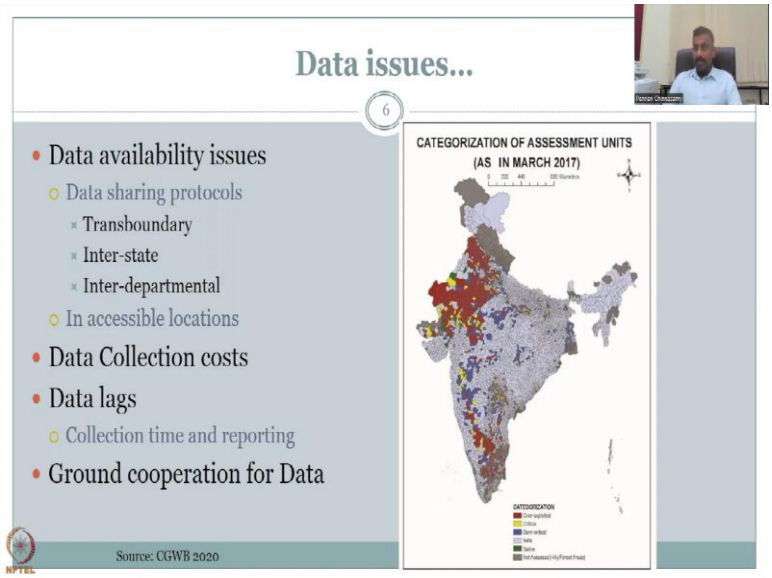
Duplication sometimes the same data is duplicated an entry errors between stations there is a lot of entry errors between the stations some people do not take data. So, they just manipulate the data. So, this is not good enough for rural development because all this data has to be accurate more importantly and they have to consider the instrumentation errors and data collection errors.

Data representativeness is also important. For example, if I am going to monitor rainfall for farmers, I should be putting a station in the farming location, but most of the time the station is placed in the airports because of land undisturbed location those are fine, but sometimes your rainfall occurs kilometres away from the airport. So how is that going to be correct? Is the question.

So, there are issues with data representativeness, where you put the data, monitoring instrument is it representing your goal all these things, so example, that is why you collect temperature in the year on the forehead now using the gun or in the armpit. Because those are better than putting in the mouth, they have found the thermometer. You do not put it on the cheek, the thermometer on the cheek or on the hand like this.

So, there are specific locations, so these are more representative of the body temperature. So, they asked you to put in the armpit or in under the tongue or nowadays the gun under forehead their temperature mapping. Data instrumentation also lacks calibration and validation, for example, it may be recording correctly as per the instrument setup, but it has to be calibrated to the current situation, and validated. So, all these are needed for better data.

(Refer Slide Time: 25:20)



There are more data issues. So, data the previous one was on the data per se. Now, I am going to talk about procuring the data. So, always there are data availability issues, which means data is not shared freely. There are transboundary issues between countries, between states and inter department, as I clearly said, maybe the data they do not share, maybe they share.

So, the government is also focusing on multiple schemes to bring all the data together on one platform. And this enables everyone to use it for their own goal for rural development. Data availability in inaccessible locations is also a concern, because how do you go collect data from the Himalayan regions, for example, often going and collecting data it is tedious time consuming. So, there are inaccessibility issues, and maybe sometimes they do not collect there. Because getting the data is also going to be difficult.

Data collection cost, as I said, there is a lot of manpower that is needed. And time that is needed to go collect the data. There is a data lag. By the time you go collect the data, bring it to the station process it, and then write a report. It is already two years now. So, for example, the census data, we know that it is once every 10 years. But does that mean on every 10th year they will go and collect the census data and now, they start slowly and in parcels. And then they combine all the data within the 10-year frame and release the data.

So mostly, what you are going to see is? A data that has been worked over 2, 3 years, that comes as census data. So, there is a difference between collection time and reporting, 3 years, as I said, and ground cooperation of data, for example, you are a farmer. In a rural setting, you will not readily share the data because you think that maybe you are going to, the government is going to put taxes on water use.

So, all these data issues are there. However, with all these data, we do get some kind of policy documents to the best of their knowledge on data aspects and where the blocks are red colour, which means the critical, etc. So, these kinds of assessments still continue to come. Without fully addressing the data issues that we have discussed about.

(Refer Slide Time: 28:00)

And the last one is data insufficiency, which means there is not enough data. So, with all the other issues that I have said, if you do not have enough data, then a clear picture cannot be drawn. So, as the court says that is yet insufficient data for meaningful answer, you will need to collect as much data and it is dangerous to reason from insufficient data.

That is why, it is always good to have multiple data for a particular indicator, combine it and then bring out a indicator based on multiple sources or multiple data. For example, this is what we had told in the project for the World Bank. Where we looked at climate change science, knowledge and impacts on water resources across South Asia. The recommendation given worse, the data is insufficient. So, please invest more in collecting data at least now, so that we can understand the climate change better.

(Refer Slide Time: 29:03)

So, challenges and issues or lack of good quality, quality observation data, understanding the current and future scenarios is always going to be challenging. If there is less data, than the physical processes or the drivers that accelerate these changes are very less to be understood. There is lack of cooperation between the agencies that collect data. So, there are agency disputes in data transboundary, interstate data etc. Lack of capacity building. For example, the Ganges basin plans across four countries and a lot of states.

So, all the countries have to tie up and share the data for example, part of small part of China, Nepal, India, Bangladesh, all are dependent on the Ganges. If only one country says I would not give the data in this Ganges basin then the other countries cannot make a good picture or often… So, all have to be brought up to capacity and cooperate to share data.

So, lack of capacities there, where warning systems models, resilient, prototypes are trying to be put, but it does take time. So, as you see the recommendation given is to map flood and drought prone areas to assess and predict water resources in basins, we need to improve capacity of analysis remote sensing products.

So, the hint here is, the satellite data and remote sensing data can aid in bridging these gaps, creating trans boundary data and a same bias for the data, which means the data is of the same quality across the zone, rather than some for example, Nepal is monitoring water at every week and we monitor it at every day. So, there is a big difference of data quality. So, all this had to be addressed in the current scenarios. So, now we have looked at how rural development is important and how data per se is very important for rural development. With this, I conclude today's lecture. I will see you in the next class. Thank you.