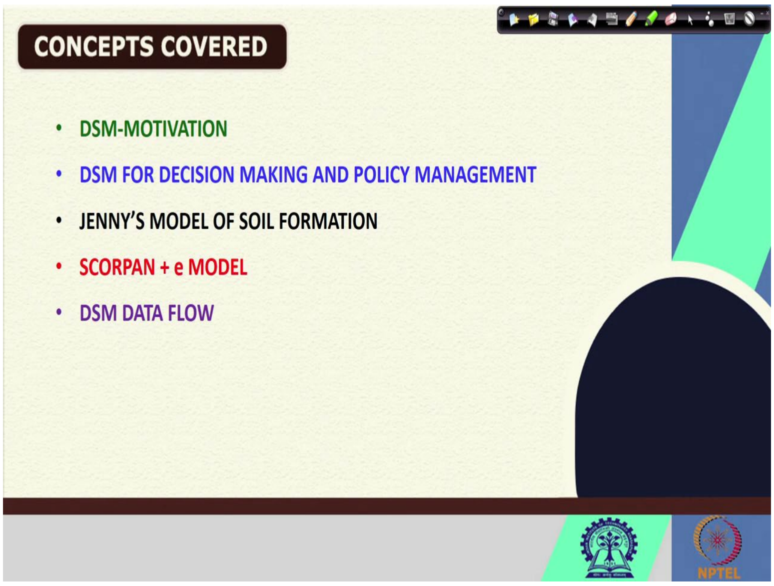
Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur
Lecture 46
Digital Soil Mapping – General Overview

Welcome friends to this NPTEL online certification course of Machine Learning for Soil and Crop Management. And we are in week 10. And in this week we are going to get the overview of a very important and advanced discipline of soil that is called digital soil mapping or in short form me sometime call it DSM.

So, in this week we are will be having a basic overview of DSM, what is DSM? And then what are the different aspects of DSM, we will be also learning about some important concepts of course, those will be basic overview. For example, geographic information system then the data types in geographic information system.

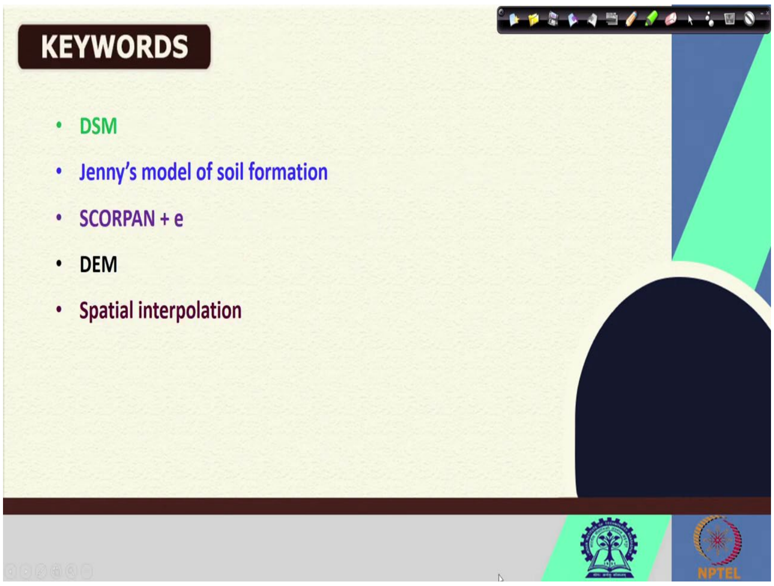
Then we learn about geo statistics, geo statistical interpolations, Variogram, and so many things. And we will also learn how to handle the data large dataset for digital soil mapping with some practical examples from agricultural fields. But before we go for understanding are applying the DSM, and to see the applications of DSM in agricultural field focusing on soil and crop. We will see first, the basic overview of digital soil mapping.

(Refer Slide Time: 2:18)



So, in this first lecture of this week 10 that is lecture number 46. I am going to focus on certain aspects and these aspects are we will be starting discussing about motivation for digital soil mapping, and then how digital soil mapping helps for decision making and policy management. And then we will learn Jenny's model of soil formation, then SCORPAN plus E model, and then we will learn about data flow in digital soil mapping. So, these are the important concepts which we are going to cover in this first lecture of week 10.

(Refer Slide Time: 2:54)



And these are some of the key words for example, DSM Jenny's model of soil formations SCORPAN plus e DEM spatial interpolation. So, these keywords will be also discussed in this first lecture.

(Refer Slide Time: 3:09)

So, if we see the what is the motivation of digital soil mapping. Now, in different type of agricultural operations or agricultural interpretation, we use different types of crop, environmental and socio economic models. For example, when you talk about DSSAT, when you talk about crops seized and other models, we generally use different types of we use different types of models to for crop group simulation, and also for identifying the risk for and also some testing some scenarios comparing some scenarios et cetera.

But, all of these type of models require soil parameters as input to estimate and forecast changes in our future life conditions. So, to help simulating different outcomes from this model, we require soil parameters as an essential input. Now, of course, there are available soil data in scattered in different places, in different organization they are archived but there are some issues.

What are those issues? First of all, these data is missing at the appropriate scale, the scale which is required for inputting this data in this crop model is sometimes missing and most of the data are not digitized and then the these data whatever is there are not well explained for reliable interpretation, we cannot make any interpretation with the available data.

So, to deal with this type of problem, we require a advanced methodology for analyzing, storing and make some, making some query with the digitized soil information. And that is why digital soil mapping comes into action.

(Refer Slide Time: 5:18)

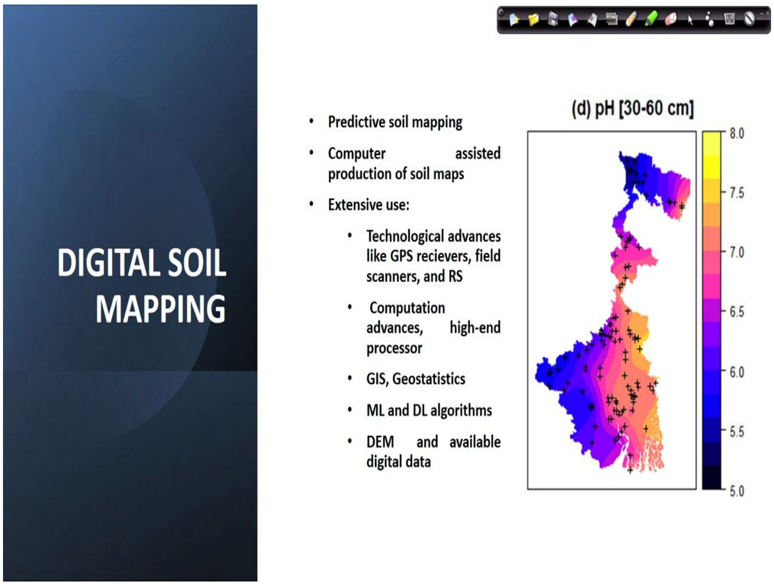
Now, what is a data requirement? We declare the data, specifically soil data, which is easy to interpret and use for future support for future to support this decision making and modeling, then we record the data both at national as well as global scale. And then there is a pressing need for advanced data collection and spatial interpolation methods.

Now, this is very important spatial interpolation method that means, the data not only should have some values, but at the same time it should have some locational context also. Now, all these data, spatial data is required to create a digital soil database for doing some mapping for extracting some informed important information and for spatial interpolation.

Now, what is spatial interpolation? We will learn in details in the lecture of geo statistics, but at the same time I would like to tell you the spatial interpolation is basically the is basically predicting data. For example, in this case soil, if we have a certain amount of sample from a given area, if we use those handful of samples to predict the any property of the soil at the unsample location that will be called a spatial interpolation.

So, digital soil mapping deals with the spatial interpolation of the soil data using several types of techniques like geo statistical technique, we are going to discuss this in coming lectures. But the spatial interpolation is very much needed for creating a digital soil database with locational context.

(Refer Slide Time: 7:27)

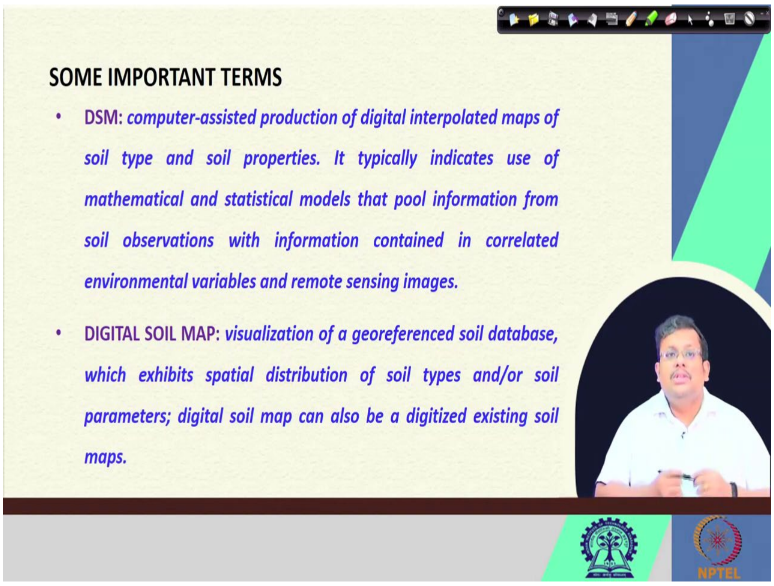


Now, if we see the digital soil mapping, what is the definition of the digital soil mapping? Digital soil mapping is basically also known as predictive soil mapping or computer assisted production of soil maps. As you can see right side of the slide, it is showing the pH of one of the important states of India that is West Bengal. And here we are producing the digital soil map of soil pH for 30 to 60 centimeter depth.

So, this for creating this map you can see that it is high resolution and at the same time for DSM it uses technological advances like GPS receivers, field scanners, and remote sensing platform, it also require computational advances like and also high end processor, it requires GIS geo statistics, it requires machine learning and deep learning algorithms. And also it also required the digital elevation model or DM and available digital data and also some auxiliary data.

So, this map is an outcome of different advance methodologies as well as sensors like GPS, then field scanners, then remote sensing, then computers with high end processors, GIS, geo statistics, machine learning and then DM and other available digital auxiliary data sources. So, all these play in a coherent manner to develop these high resolution maps at any given depth up to 200 centimeter or 2 meter from the soil surface. So, we can develop this map at 6 different depths using this digital soil mapping technology.

(Refer Slide Time: 9:27)



Now, some important terms we have to learn. Now, DSM is basically if you want to define what is DSM, so DSM is basically a computer assisted production of digital interpolated maps of soil type and soil properties. It typically indicates the use of mathematical and statistical models that pool or combines the information from soil observation with information contain in correlated environmental variables which we also call auxiliary covariates or auxiliary variables.

So, using the original soil data which have been gathered in the standard laboratory along it this DSM but also utilizes the auxiliary soil data which are digitally available with a with a machine learning or deep learning algorithm. So, this is in a nutshell, this is DSM and what is digital soil map digital file map is basically visualization of a geo reference soil database which exhibits spatial distribution of soil types and or soil parameters and digital soil map can be a digitize existing soil maps also. So, you understand what is a digital scale mapping and what are the essential components of digital soil map.

(Refer Slide Time: 11:03)

So, what is the use of DSM why we go for DSM. Now, the first one important in use of DSM is to provide the quantitative soil data producible at low cost and easy to interpret and use, it is not possible for us for going for high resolution sampling all over the study area, because there are budget construction because there are limited resources.

So, when we have this type of budget construction limited resources and the area is vast then we require some simple and I mean low cost and easy to interpret interpolation technique for producing a map and that will be that is basically the digital soil mapping. And also it helps us for better mapping of soil properties at different depths and helps in policy decision.

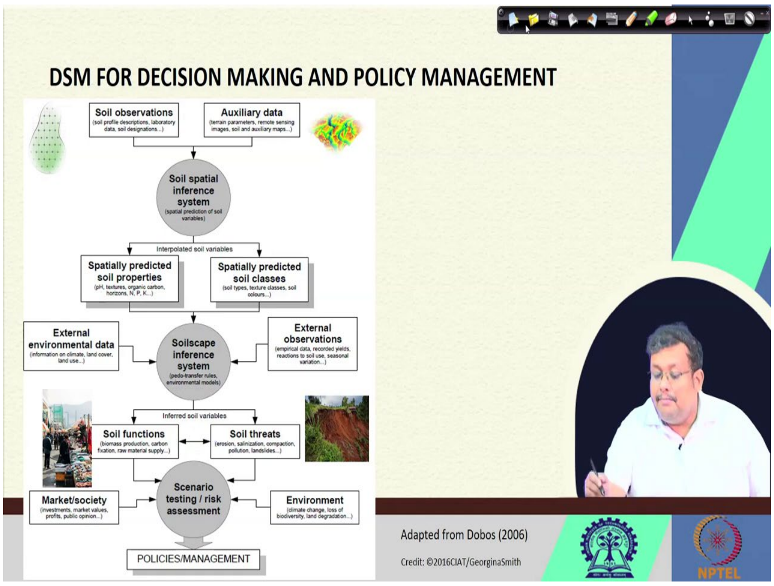
It helps in environmental management by interpolated maps and identifying the pollution hotspots I will show you how to do that. And then digital file maps also illustrate the spatial distribution of soil classes or properties and can document the uncertainty of the soil prediction.

This is very important not only it can produce the interpolated maps of soil properties but at the same time it gives the uncertainty prediction of that soil property also this is the so, the mapper will have an idea about the area of the produce digital soil map also. So, that we give him enough confidence for developing for utilizing or for interpreting that soil map.

So, digital soil mapping can be used to create initial soil survey data, refine or update existing soil surveys generate specific soil information and also assess the risk and finally, it can facilitates the rapid inventory reinventory and project mismanagement of lands in a changing environment.

So, you can see that digital soil mapping has a wide scale application for producing different management decision policy decision agricultural planning and also it can help you to get the easily interpretable digitize form of maps of soil properties.

(Refer Slide Time: 13:31)



Now, how this digital soil mapping helps in policy may for decision making and policy management, these this flowchart gives you a better idea. So, we start with the digital solving as I told you that we declare the original soil observation the original file observations may be soil profile description in the field also laboratory data soil designation. So, all these are known as the legacy soil data.

So, specifically the laboratory data which has been compiled together is known as the legacy soil data. Soil profile data is also part of your legacy soil data. So, soil observation, when we combine the soil observation with auxiliary data, auxiliary data means some correlated variables which may impact, the soil development is known as the auxiliary data.

In case of DSM for example, terrain parameters as you can see here at this digital elevation model, we will be discussing this digital elevation model in details. So, terrain parameters then remote sensing images, soil and auxiliary maps climatic data, all these will be included in the model for producing the high resolution soil maps.

So, when you combine these soil observation with the auxiliary data with the help of a machine learning or digital or deep learning model that is called soil spatial inference system.

Now, using the soil spatial inference system or digital soil mapping model or DSM model, it will basically help in the prediction spatial prediction of different soil properties like pH, texture organic carbon horizon NPK and any other soil properties and also it can also gives you the spatially predicted soil classes.

So, from basic from both prediction point of view and for classification point of view, these digital self-mapping can produce variable outputs. So, it can directly gives you the predicted the a map with the predicted soil properties and also it can also gives you the map of spatially predicted soil classes and this will be important for soil types and identifying the texture classes soil colors and so on.

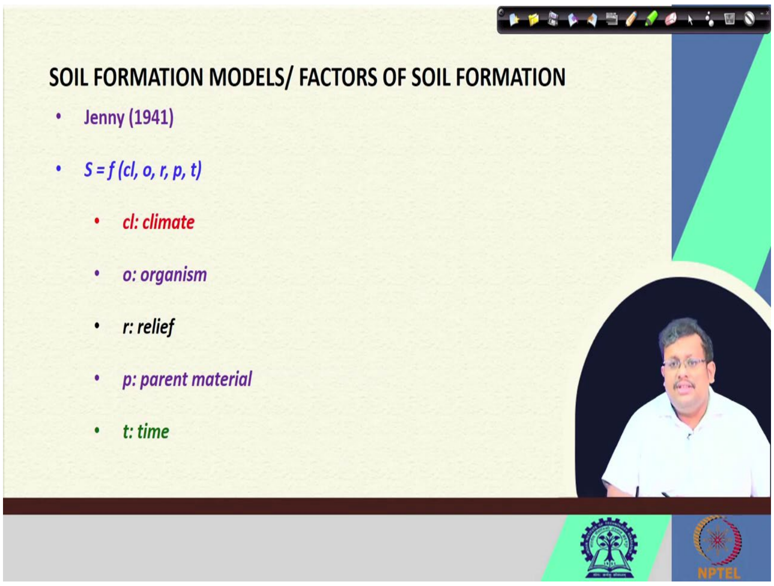
So, you can see that digital soil mapping models can give you both the spatially predicted soil properties as well as the spatially predicted soil classes. So, when we combine these spatially predicted soil product properties and spatially predicted soil classes in the soil scape inference system using different types of pedro transfer rules and environmental models, then, we can do some more important in interpretation.

For example, when we combine these spatially predicted soil properties and spatially predicted soil classes and we combine the information from the external environment like information and climate land cover land use and some external observation like empirical data recorded yields, reaction to soil use seasonal variation, then it can produce different types of outputs like it can infer the soil variables in terms of soil functions and soil threats.

What is the soil function biomass production carbon fixation raw material supply all these could be produced as an output also the soil inference system can produce the erosion and salinization a compaction other soil threats. So, come considering both the soil function and soil threats and also different markets investments, economic variables and also different types of environmental variables we can go for scenario testing or risk assessment and finally, based on that, we can go for the policy management.

So, this is how step by step the digital soil mapping helps us for decision making and policy management because, when we have soil function soil traits, environmental data as well as market or societal data, we can go for the policy management, better policy management using the DSM map. So, this is how step by step DSM helps us for better policy management.

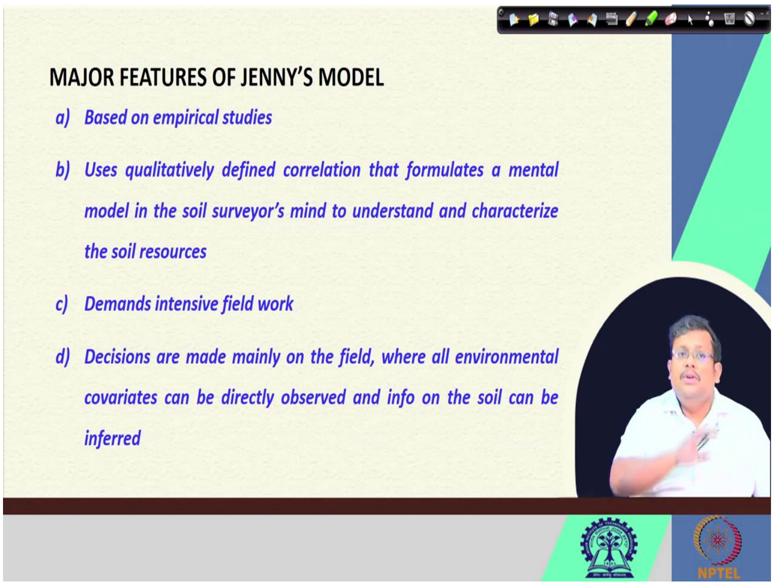
(Refer Slide Time: 18:19)



Now, so, what is the inception of digital soil mapping concept. So, that there are different models which are been which have been given by different scientists to describe the formation of soil cover up the most important soil model soil formation model is Jenny's model of soil formation given in 1941.

So, it says that soil is a function of cl, o, r, p, t that means, it is a function of climate cl stands for climate O stands for organism R stands for relief or unevenness of the soil surface and then p is the parent material and t is the time. So, all these factors influence the soil formation at a given place. So, this is the Jenny's model of soil formation this is the book the old classical model of soil formation.

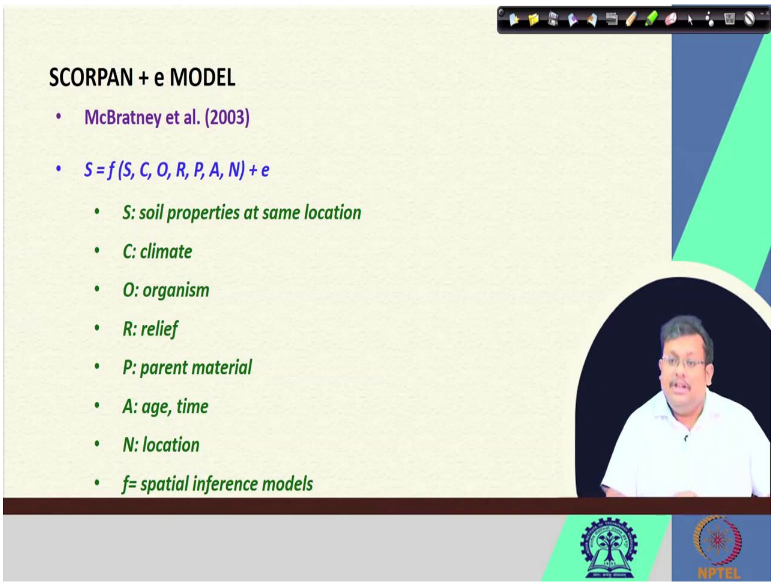
(Refer Slide Time: 19:23)



And these major features of the Jenny’s model is it is based on empirical studies and it uses qualitatively defined correlation that formulates a mental model in the soil surveyor’s mind to understand the characteristics of the soil resources. And so, basically it is mostly depends on the mental model in the soil surveyor’s mind which he did use it when going for the soil observation and seeing the address related conditions and variables.

So, it requires intensive field words and decisions are made mainly in the field on the field, where all environmental covariates can be directly observed and information on the soil can be inferred. So, here in this model the soil surveyor’s hardware goes to the field and then does an observed the soil profile and then observes the environmental covariates which are there in close vicinity of that soil profile and then they can infer the soil properties. So, this is how this Jenny's model basically works.

(Refer Slide Time: 20:45)

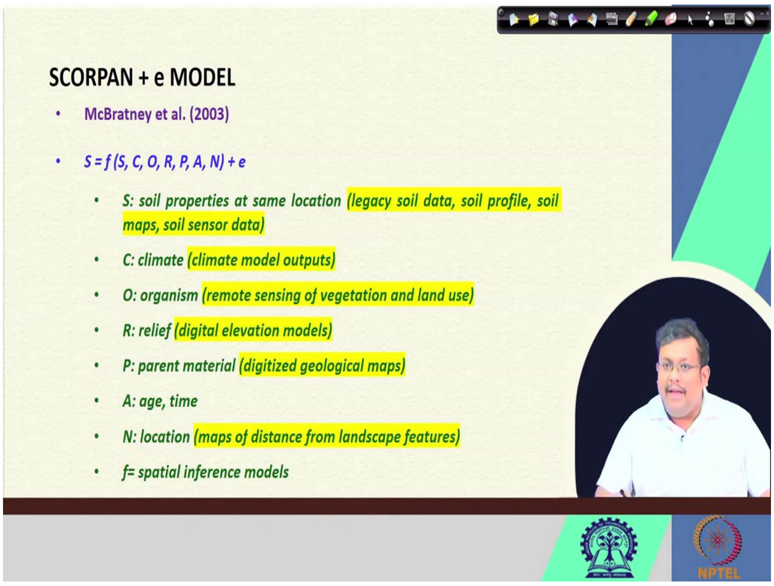


Now, in 2003 McBratney et al they made a paradigm change in the soil formation theory by giving this equation that is called SCORPAN plus e model. So, soil properties at any at any given location is a function of soil property at the same location climate, then organism relief parent material A stands for age or time and n for n stands for the location. So, all these are important for predicting the soil properties at a given space.

So, here F stands for the spatial inference model, which you have discussed and of course, there will be some amount of error. So, this model is known as SCORPAN plus e model. Now, while we discuss this SCORPAN plus E model this is important to understand that the major difference one of the major difference between Jenny’s model and this model is here, we are taking the locational context for calculating the soil properties.

So, that is why this spatial context what I told you that DSM involves the spatial data with the locational context. So, this is important inclusion in case of SCORPAN plus e model.

(Refer Slide Time: 22:07)

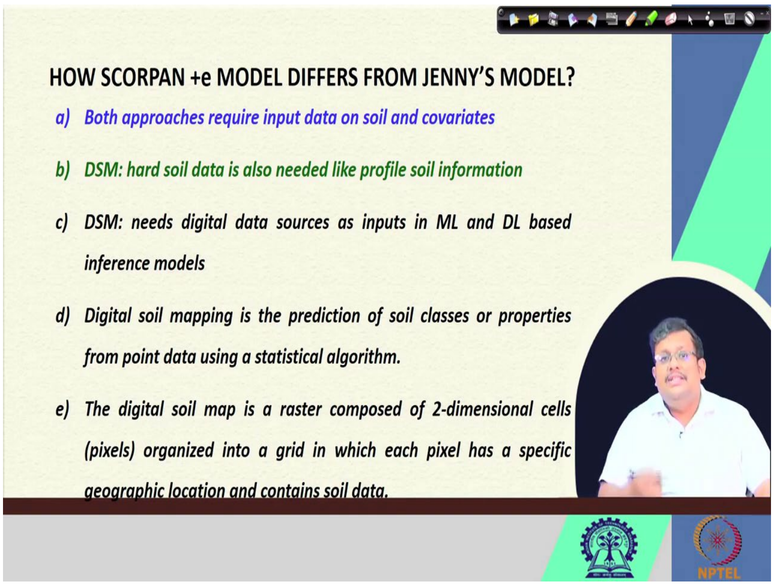


Now, why it is called digital soil mapping that is the question and you see that most of these parameters nowadays are available from in some in some digitized form for example, the soil properties at some location can be gathered as legacy soil data you can convert this legacy soil data into digitized form and then file profiles soil maps and also soil sensor data we have already discussed several soil sensors data proximal sensors data with hyperspectral remote sensing data.

So, all these sensor data are digitized are available in the digitized form. So, you can use this the S variable in digitized form then c variable you can get the climatic model out outputs. So, that is also digitally available or stands for the organism. So, you can get the organism data from remote sensing of vegetation and land use are you can get digital elevation models p is the digitized geological maps and in maps of distance from the landscape features.

So, use you can use all these data in the digitized form and that is why the map which you are going to create is known as digital soil map. So, this is the difference between the Jenny's model and SCORPAN plus e model.

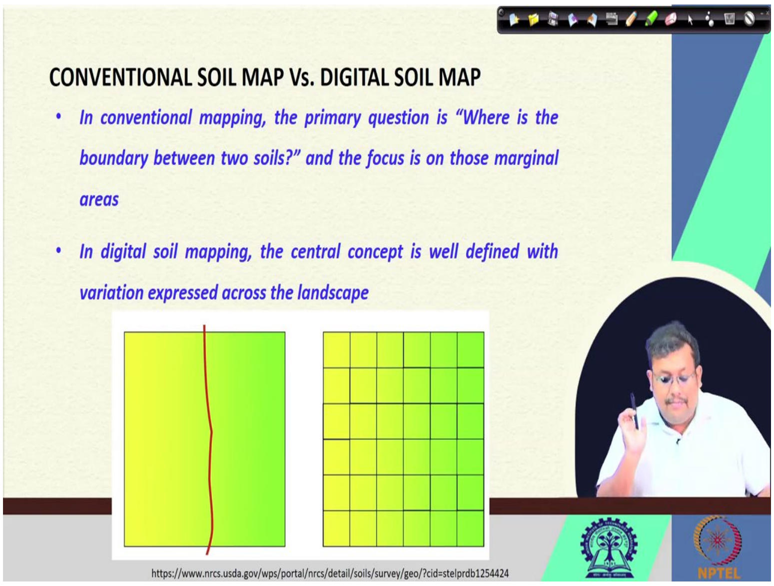
(Refer Slide Time: 23:29)



So, the next question comes how this SCORPAN plus e model differs from Jenny's model. So, both approaches require input data on soil and covariates and DSM it is hard soil data is also needed like profile soil information. And in case of it also required digital data sources as inputs in machine learning and deep learning based spatial inference model.

Digital soil mapping is the prediction of soil classes or properties from point data using a statistical algorithm. And the digital soil mapping is a raster composed of two dimensional cells or pixel, what is pixel organized into grids in which each pixel has a specific geographic location and contains some soil data. So, these are some of the important features and that is how they differ from the typical older Jenny's soil formation model.

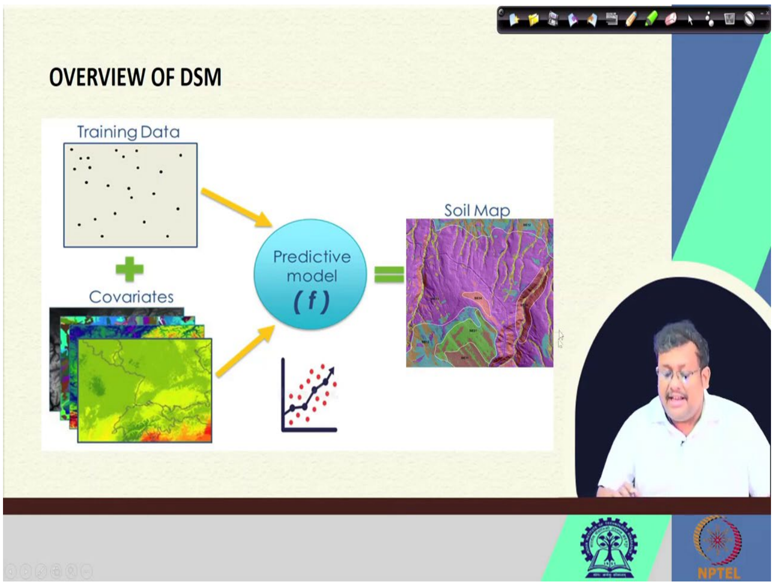
(Refer Slide Time: 24:25)



Now, what is the difference between the conventional map and digital soil map? Now, in case of conventional mapping the primary question is where is the boundary between two soil as you can see here in this figure so, where is the boundary between the soil so and the focus here is on those marginal areas.

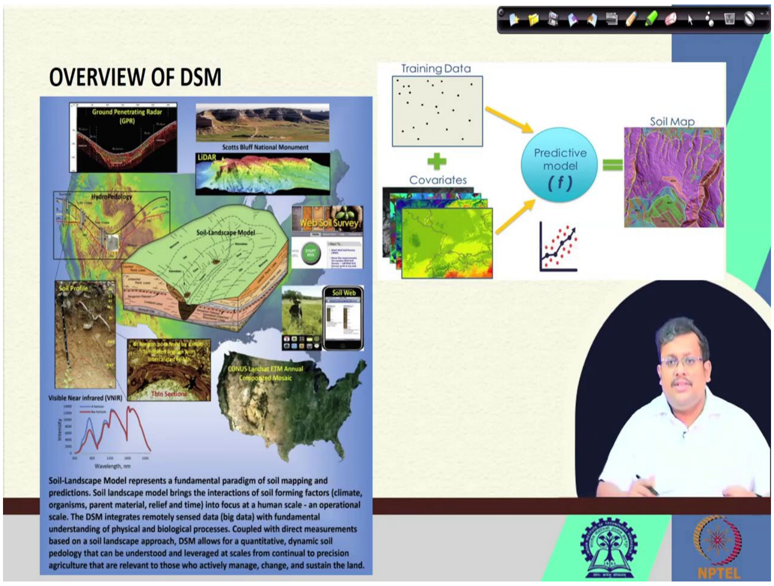
However, in digital soil mapping, the central concept is well defined with variation expressed across the landscape. So, this is the digital soil map. So, the central concept is well defined within variation expressed across the landscape. But here in the conventional mapping, we always try to find the boundary between the soils.

(Refer Slide Time: 25:10)



Now, so, what is the overview of the DSM as you can see, we start with the legacy data or training data, then we add the covariate data from different sources climatic data, terrain data, remote sensing data and also crop model data for example and then we combine them together in the predictive model finally get the soil map. So, this is the overview of digital soil mapping.

(Refer Slide Time: 25:38)

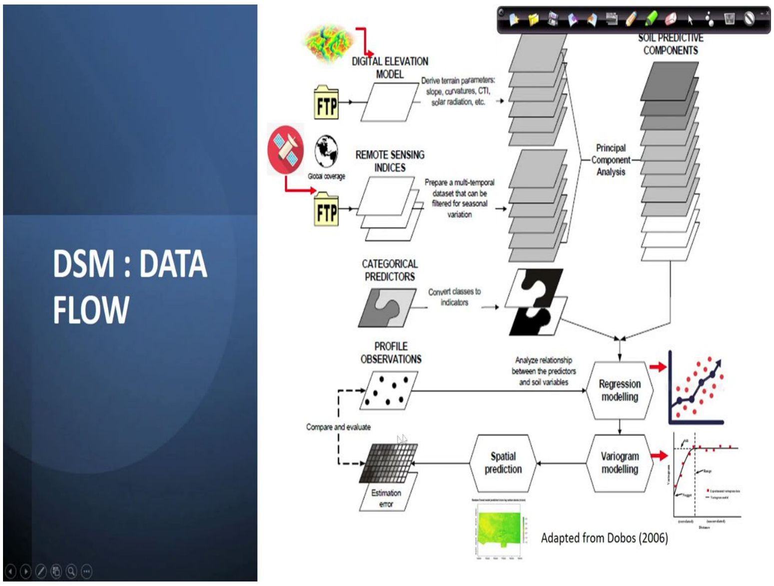


Also overview of digital soil mapping says that these the soil landscape model which represents a fundamental paradigm of soil mapping and prediction these being say interaction of soil farming factors they are add interaction of different soil forming factors like climate, organisms, parent material, relief, time so add into focus into the human scale and operational scale.

So, the DSM basically it integrates the remotely sensed data which is also known as big data. So, big data you have already we have already discussed what is big data. So, these DSM can handle the big data from gathered from different remote sensing platform and then coupled with the direct measurement based on the soil profile.

Or laboratory measured soil data along with some other point spectrometer data or other easily available digitized co variable covariate data in they combine them and they produce the both qualitative and quantitative maps of soil properties using some machine learning and deep learning models. So, this is how this the advent of digital soil mapping has enhanced the soil landscape model outcome for better interpretation of soil processes.

(Refer Slide Time: 27:13)



So, if we see the finally, the what is the data flow in the digital soil map you can see that here we can gather the data from digital elevation model we can download this digital elevation model from different sources and then we can use some software to derive the terrain parameters like slope curvature CTI solar radiation et cetera.

And also we can gather the remote sensing indices, we can prepare a multi temporal data set that can be filtered for seasonal variation, we can get some categorical predictors like soil, types and then we can convert those classes into some indicators. So, we can combine these auxiliary co variable data and then we can do some principal component analysis to select the important principal components.

And then we can combine them with these categorical predictors to in the regression model. And in the regression model, we use all these information categorical predictors, remote sensing indices, digital elevation model, derived terrain parameters to predict the soil properties of some profile or soil observations.

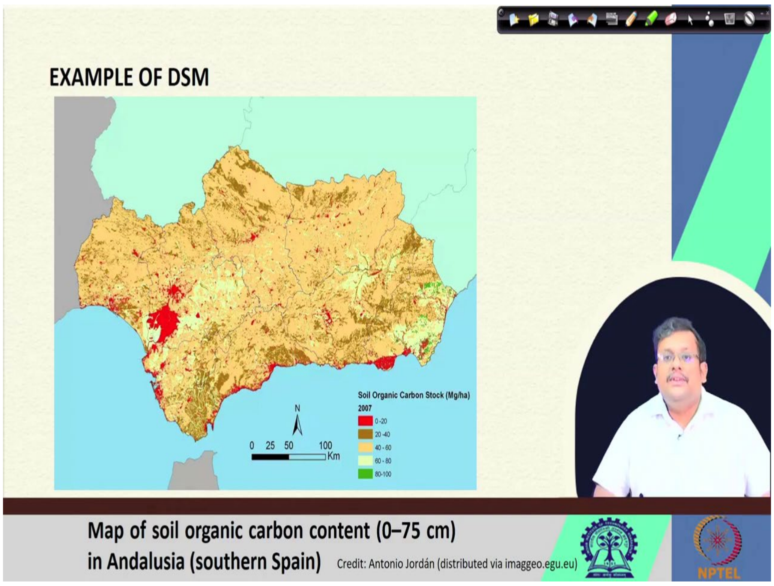
So, these are soil observation, we take the soil observation we take we measured their soil properties using reference methods. So, those reference values will be predicted by using these parameters, terrain parameters, and then remote sensing data categorical predictors using some regression model.

And then, when we have these regression outcome, we get these Variogram modeling, we will be discussing what is Variogram and then we will be producing the spatial prediction.

And finally, from both the regression model output and Variogram model output will be our interpolation output, we will combine them together to produce the final digital soil map.

So, this is how these digital soil mapping workflow operates. And so, this is how we utilize the DSM to get the 5 days and workflow to get the final the output that is digital soil map.

(Refer Slide Time: 29:34)



So, this is an example of DSM that is, this is a map of soil organic carbon content of 0 to 75 centimeter in Andalusia in southern Spain. This is the organic carbon digital soil map of an region. So, this is this has used the traditional methods of DSM and use the spatial inference model to create the soil organic carbon map at a given depth of 0 to 75 centimeters.

(Refer Slide Time: 30:05)

So, guys, I hope that you have learned something new. We have got the basic overview of DSM. This is the reference which I have used in this lecture. So, thank you for joining and let us go from here in our next class in the next lecture. And we will see some more application of DSM. And also we will see what is GIS and why we required GIS. Thank you.