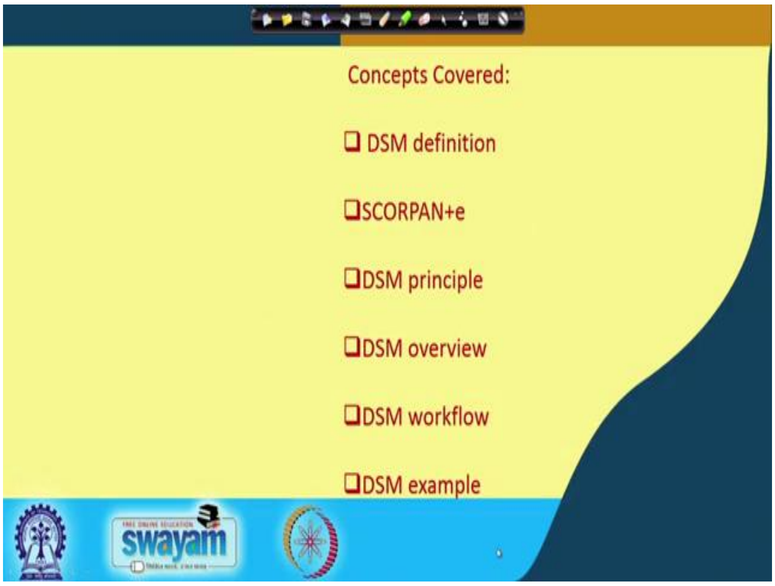
Soil Science and Technology
Prof. Somsubhra Chakraborty
Department of Agricultural and Food Engineering
Indian Institute of Technology, Kharagpur

Lecture – 56
Basic Overview of DSM

Welcome friends, to this week 12, lectures of Soil Science and Technology. This will be our final week of lectures in this course and in this week the coming 5 lectures we will be discussing about several important topics and several current research areas in the field of soil science.

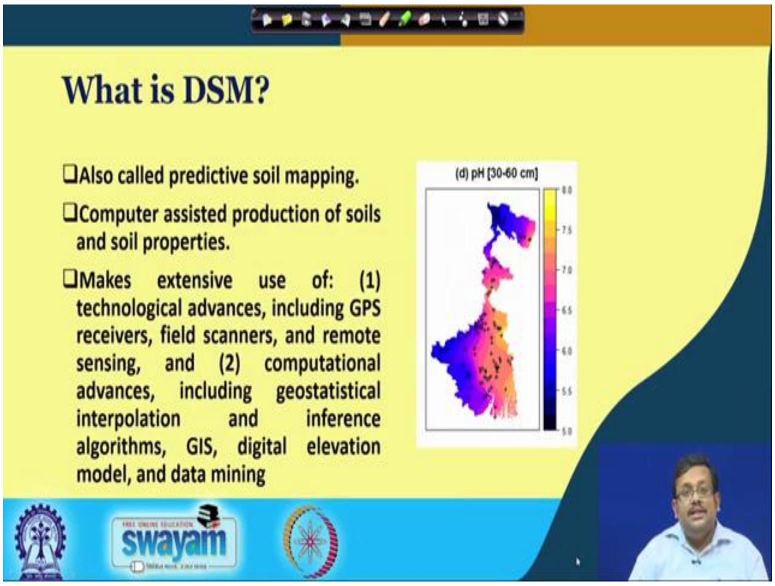
And, so we will be discussing about an important topic in soil science that is called digital soil mapping and there are different aspects. I will talk about different types of modeling which is very much relevant to digital soil mapping as well as in any other modeling of soil properties and we will be talking about different types of continuous models, we will be talking about different types of categorical models and then we will see what is pedo transfer function and then, we will cover what are the uncertainties of digital soil mapping.

(Refer Slide Time: 01:49)



So, in this first lecture we will be talking about this digital soil mapping and we will be basically giving you an overview of digital soil mapping; since it is a very vast subject, it is not possible for us to complete within a very stiff very short time period, but I will try to give you some basic overview of digital soil mapping. So, in this lecture we will be we will be covering these following aspects. First of all we will be giving the DSM, will be, you know, will be having an overview of DSM. And then we will know what is the definition of DSM, then we will know a specific model called SCORPAN plus E model, then we will know about DSM principle, then DSM overview, then what is the work flow in DSM and finally, I will show you some examples of digital soil mapping.

(Refer Slide Time: 02:21)

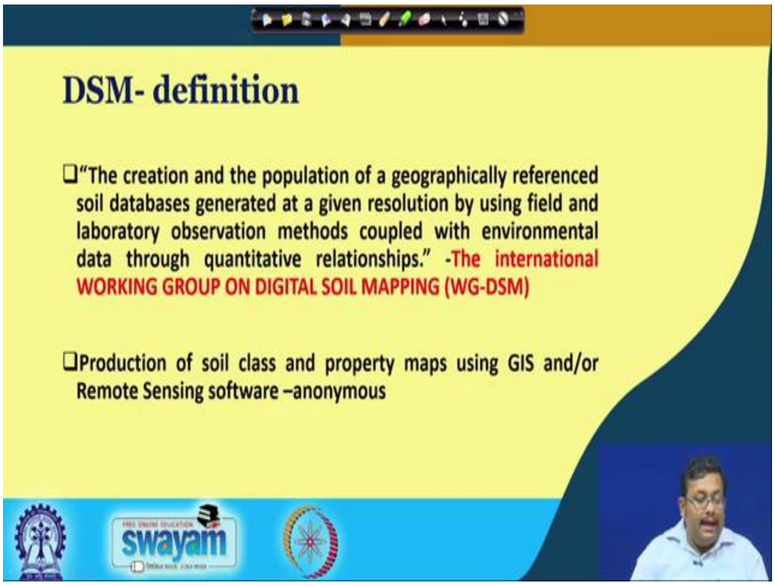


Now, what is digital soil mapping? Well, these digital soil mapping, the another name of digital soil mapping is predictive soil mapping and it is basically a computer assisted production of soil and soil properties. And, you know that mapping of soil is very much important for proper and judicious management of a natural resources. And for that this digital soil mapping, you know, gives us the maximum, you know, flexibility as well as maximum technological advancement. And, using this digital soil mapping, it is possible to properly manage and properly map different, you know, the spatial variability of different soil properties.

So, this digital soil mapping makes extensive use of technological advance including GPS receivers, like field scanners and remote sensing and, you know, computational advances including geostatistical interpolation and inference algorithms and then GIS, digital elevation model and data mining. I have already covered this GIS and geostatistics in my previous lectures. So, those will be basically used in digital soil mapping and we will talk about it.

So, as you can see here in the right picture is basically a picture of soil pH spatial variability map which you produced through digital soil mapping for the state of West Bengal in India. And you can see how these pH values are variable across the state and this map is produced by the use of digital soil mapping techniques.

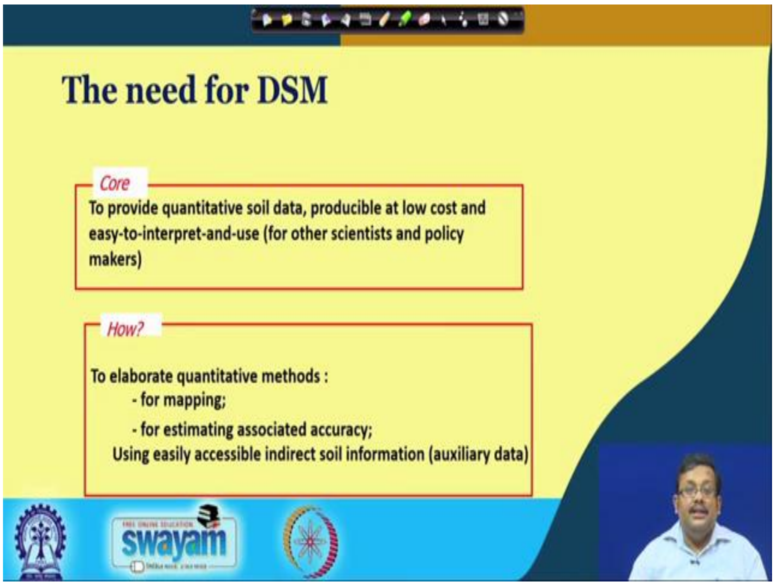
(Refer Slide Time: 04:11)



Now, what is a definition of digital soil mapping when the creation and the population of geologically reference soil database generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationship is the formal definition of DSM. And it is given by "The international WORKING GROUP ON DIGITAL SOIL MAPPING".

And, there is an anonymous definition also which is much simpler than this formal definition and it says that digital soil mapping is basically production of soil class and property maps using GIS and a remote sensing software.

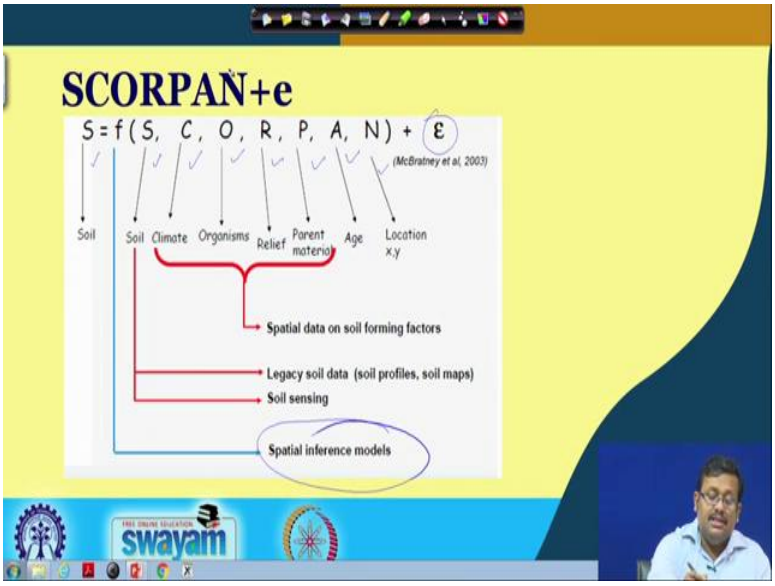
(Refer Slide Time: 04:51)



So, basically the need for DSM is divided into, you know, if you considered, if you are, if you want to know what is the need of DSM; obviously, the core need for DSM is to provide quantitative soil data, producible at a low cost and easy-to-interpret-and-use for other scientists and policymakers. Because policymakers needs this type of data for effective management of natural resources and only digital soil mapping has the capability to produce and arrange and compile and, you know, map the soil resources using this DSM technology.

How? How they can, how we can achieve this core need, you know, to elaborate quantitative methods for mapping and for estimating the associated accuracy and using easily accessible indirect soil information; we call it auxiliary data or covariates and we will discuss about these covariates in auxiliary data in details in the coming slides.

(Refer Slide Time: 06:05)



So, before we go to digital soil mapping let us see what is the base of these digital soil mapping works? Well, this model called SCORPAN plus E is the base of digital soil mapping and it was proposed by Professor Alex McBratney of the University of Sydney. And, this SCORPAN plus E model just like Jenny's soil formation model you remember that the soil is a function of climate, organism, relief, parent material and time. Similarly it this SCORPAN plus E model is basically a modification of this Jenny's factors of soil formation.

And, you can see that these factors of soil formation this SCORPAN plus E model says that soil at any particular place is a function of this SCORPAN. So, S stands for soil, C stands for climate, O stands for organisms, R stands for relief, P stands for parent material, A stands for age and N basically denotes location in x and y coordinates and, obviously, this E is the residual or error which you cannot model.

And, this function is basically the spatial inference model and we will discuss about this spatial inference model. These are some quantitative and qualitative, you know, we can model these this soil formation using some mathematical algorithms or statistical algorithm these are called spatial inference models like multiple linear regression, then random forest regression, then decision tree we will discuss them in details.

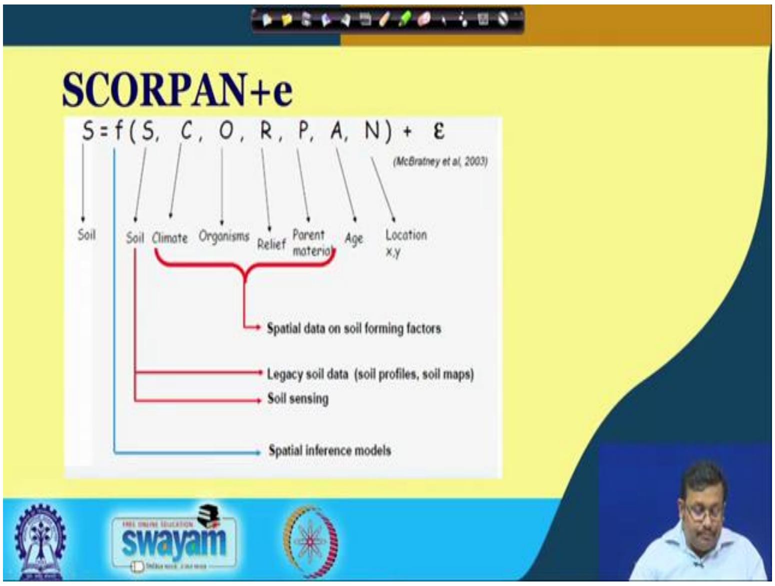
And, so, basically the SCORPAN plus E model says that soil is a, you know, the soil of a particular place can be modeled in a more precise way. Any particular property of a soil can be predicted by combining these factors like soil, climate, organism, relief, parent material then age and location. And, obviously, soil here you can gather the soil information by this legacy soil data which is basically the soil data the legacy soil data is the soil data original soil data which has been analyzed which has been compiled from analyzing the real soil samples.

For example, in India this legacy soil data is maintained by our national agency called National Bureau of Soil Survey and Land Use Planning or NBSS and LUP. And, this NBSS and LUP basically, you know, compile and analyze the real soil samples and, you know, compile their data into digital format and we call it this legacy soil data and the profile wise data is available.

So, this soil data can also be, we can get by soil sensing methods I have already discussed different soil sensing method. In our last lectures, we talked about diffuse reflectance spectroscopy, we talked about this portable X-ray fluorescence and remember that this soil sensing technologies can be extensively used for gathering different soil information. Then, climate, organism, relief and parent material these are the spatial data on soil forming factors and you can gather this information as auxiliary information or covariates and obviously, these age and location data is also available through different means.

So, this is in nutshell the SCORPAN model, but why this SCORPAN model is important? The SCORPAN model is important from the point of view or from the from the consideration that any soil data can be modeled through quantitative or expressed through some quantitative relationship. We are basically using this SCORPAN plus E model to give some quantitative relationship to a particular property of a soil at a particular space at a particular space. So, here this SCORPAN plus E model is the basis for every digital soil mapping products.

(Refer Slide Time: 10:51)

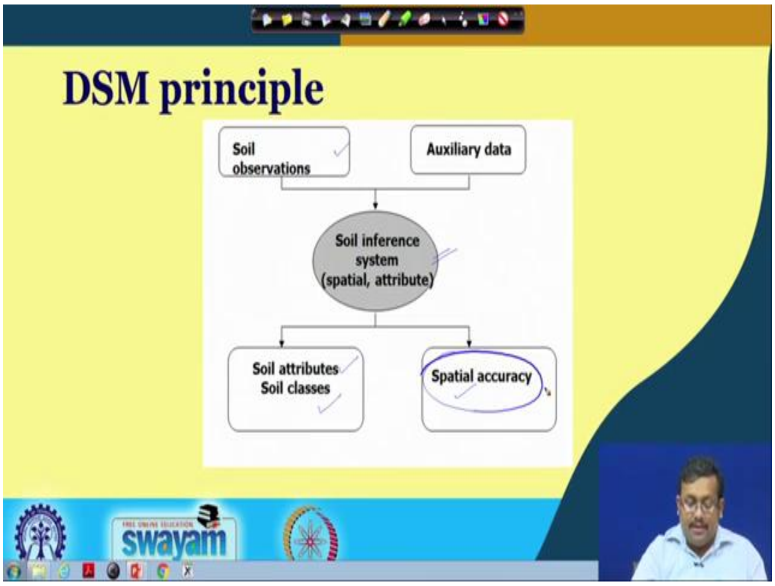


So, if you see why this SCORPAN plus E model is important, well, we already know that this soil information can be gathered through legacy soil data or we can gather this information by using different types of field scanners or proximal sensors or soil sensors.

And, this climatic data you can get from different types of climate model outputs which are available, and organism data you can get it from remote sensing of vegetation and land use, relief data you can gather from digital terrain models, parent material data can be gathered from digital geological maps or digitized geological maps, and age data is also available and location data you can get from GPS receivers. So, basically maps of distance from landscape features. So, you can see all these required data or predictor in this SCORPAN plus E model can be compiled in a digitized format.

So, that is why this mapping technique is called digital soil mapping. Again, all the required information or the predictor variables required for predicting a particular soil property can be gathered in digitized form and that is why this type of technology of mapping is called digital soil mapping.

(Refer Slide Time: 12:22)



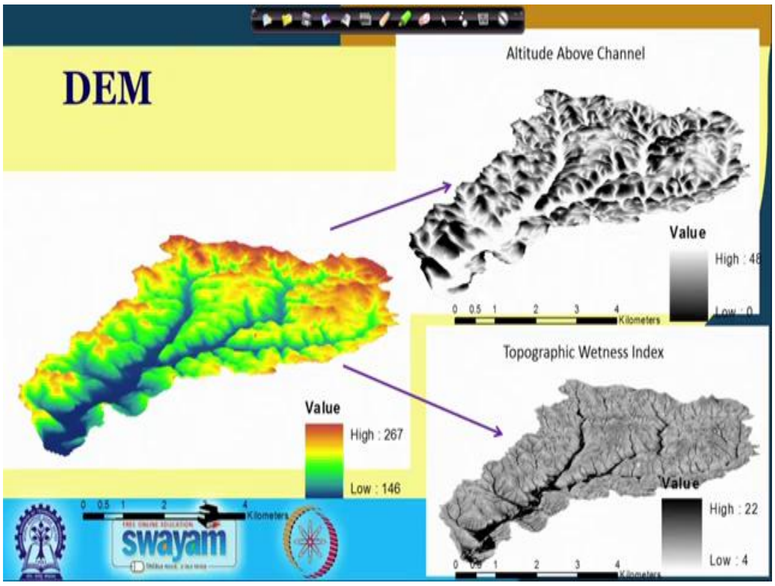
So, let us see the principle of digital soil mapping. Now, the principle of digital soil mapping says that obviously, the starting point is soil observation or in our case it is the legacy soil data. And, obviously, so, this legacy soil data is gathered by collecting the real soil samples and this is the most important part because instead of the soil observation digital soil mapping is not possible.

So, once we collect this point data from different places we will analyze the soil samples and then we will gather and compile those soil, you know, the soil property values and that is called the legacy soil data. Now, once we get the legacy soil data then we will use the auxiliary data because we think that these auxiliary data also represent different factors of soil formation and all these auxiliary variables are also helpful for depicting the different soil forming factors and different processes which occur for that particular soil formation.

So, this auxiliary data can be any data like relief data, it can be bioclimatic variables like rainfall and then, you know, thousands and thousands of different types of data is available. For example, if you want to add the spectral reflectance vary, you know, reflectance values which you gathered from diffuse reflectance spectrometer that can be also used as auxiliary data. So, apart from this real soil data we will add this auxiliary data for as predictors in this spatial inference models. So, this spatial inference model is known as the soil inference system and this soil inference system ultimately produces soil attributes and soil classes and simultaneously it also produces the spatial accuracy.

So, the beauty of DSM is not only it gives the estimate of soil attributes from this soil inference system or this particular, you know, spatial prediction model, it also produces soil classes and it side by side it produces this is the most important spatial accuracy. A sense of accuracy you will get side by side from any DSM product; so this is an overview of DSM principle.

(Refer Slide Time: 14:57)

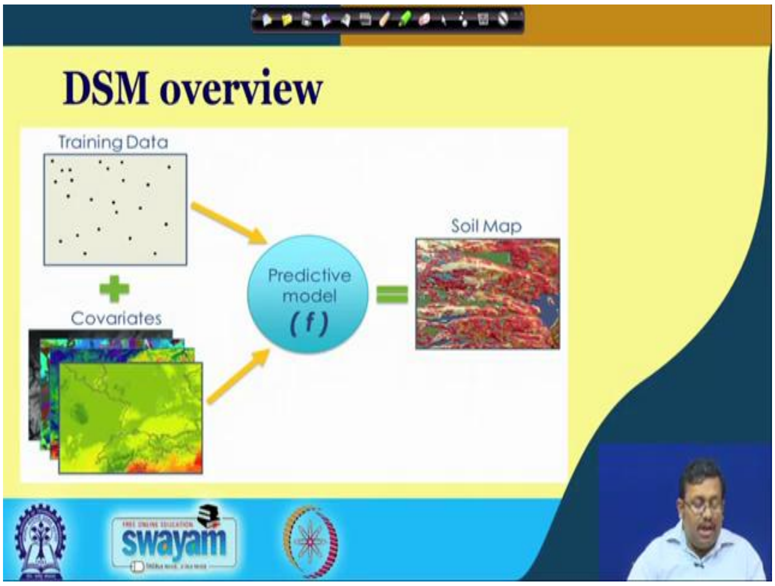


Let us go ahead and see what is digital elevation model. Now, digital elevation model is basically a, you know, it basically the 3D representation of any soil surface and it is basically available in raster format and in the raster format it basically shows thousands and, you know, any digital elevation model for a particular area is present, you know, contains thousands and thousands of grids and each grid contains some values which may represent the elevation of that particular point.

So, you can see this is a digital elevation model and this digital elevation model is available in different resolution. And obviously, if we use the better resolution, we can differentiate and we can identify very minute features in that surface or many minute changes in the elevation in our surface. However, when we use this high resolution digital elevation model, it requires more storage space as well as it requires more computing power.

So, depending upon your computing power and depending upon your storage space, you can select the resolution of digital elevation model, but nevertheless digital elevation model is a very very important, you know you know, source of auxiliary variables which you use in digital soil mapping. Because from this digital elevation model you can generate several, you know, several auxiliary variables or index also, other variables and index like you can see here altitude above channel and the topographic wetness index and all these things.

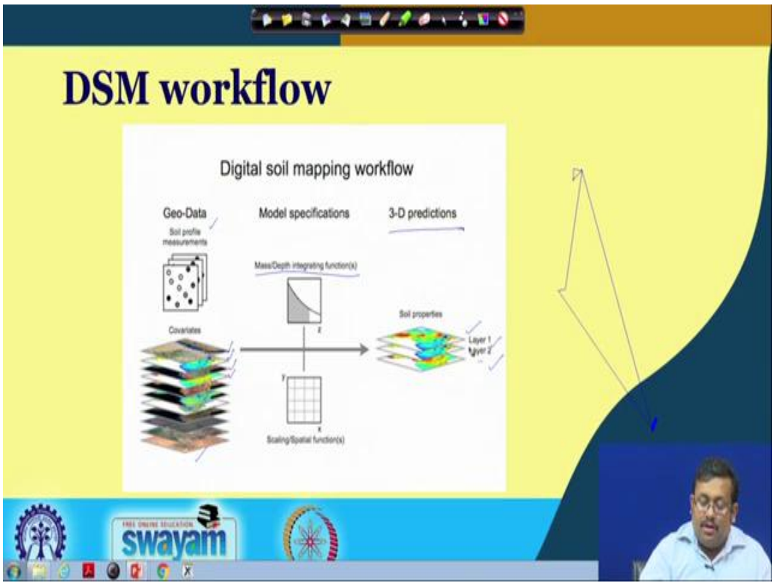
(Refer Slide Time: 16:47)



So, so, if you see the overview of digital soil mapping obviously, if the first starting point is our training data, as you can see here these are some point observation and we compile the data in the form of legacy soil data. So, this is our starting point and then we add different layers and covariates. Different layers and covariates can be generated from digital elevation model or we can gather some more other information like soil spectral signature, soil elemental values so on so forth.

And, we will use these covariates plus training data as predictors in this predictive model and ultimately we will produce the soil map. And, this soil map will be giving us higher accuracy as compared to if we use only the legacy soil data. So, this is the overview of digital soil mapping.

(Refer Slide Time: 17:48)



So, what is the workflow of digital soil mapping? This question always comes to our mind. Now, digital soil mapping workflow says that this workflow digital soil mapping says that again the geo data is the starting point we call it soil profile measurement or legacy soil data; so it is our starting point.

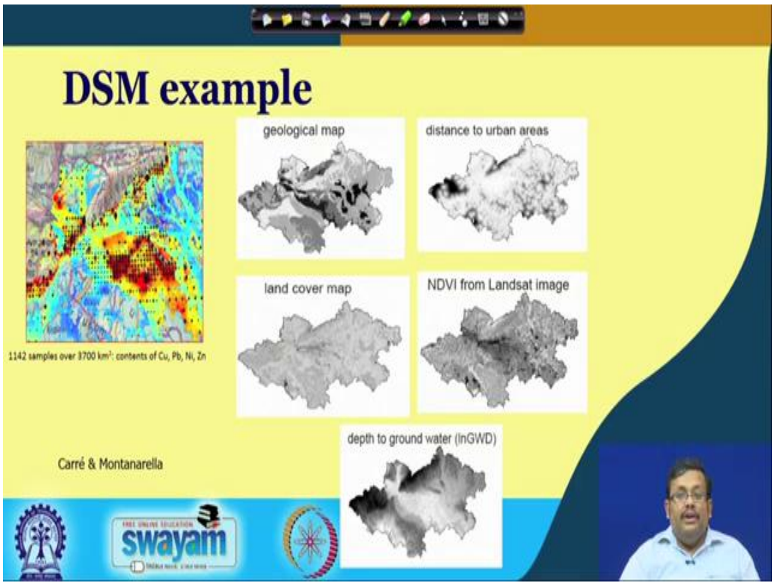
Similarly, ok; so once we get these, you know, once we get this geo data or legacy soil data, the next step is to harmonize these data across different profiles of, across different horizons we know that soil is divided into OABC and these are the master horizons. And these OABC horizon these, you know, these legacy soil data needs to be harmonized across this horizon using some mass or depth integrated function, we will discuss that later.

So, once we harmonize this data into this, you know, depth into these individual layers, the next step is to gather several covariates. Now, these covariates can be gathered from digital elevation model, from other sources. And then from you know spectral reflectance values, elemental values and each of them will be considered as an individual layer.

So, basically we will stack them together to produce a stacked raster file obviously, all these covariates are available in rasterized format and then we will stack them together and these covariates will be used along with this geo data. So, ultimately we will be producing these soil properties for individual layer as you can see here this is layer one layer two and so on so forth.

So, remember that these digital soil mapping not only it is not a two it is not only produce two-dimensional products because it gives 3D prediction across different layers; so this is very very important.

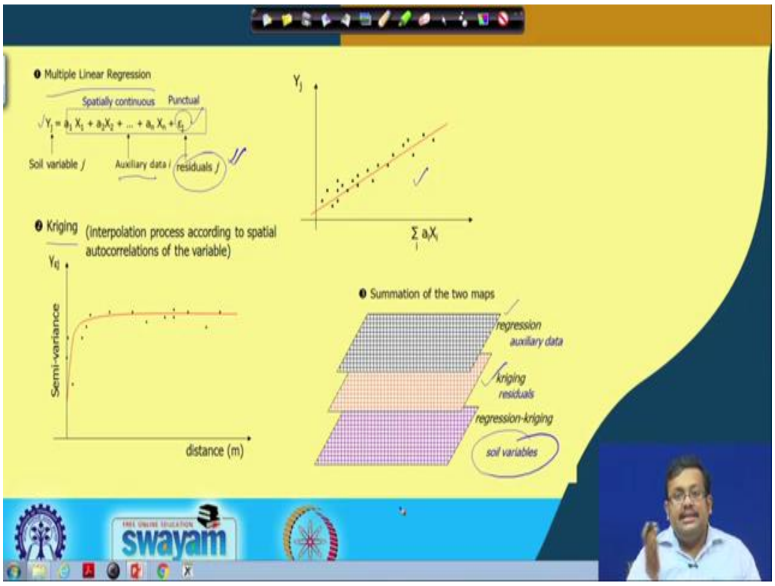
(Refer Slide Time: 20:09)



Just like to give you an example of DSM in Europe and these scientists have collected these 1142 samples over 3700 square kilometer and they measured the copper, lead, nickel and zinc for these 1142 samples and once they collect and gathered those information they combined this legacy soil data with these five, you know, different auxiliary predictors. They downloaded the geological map, they used the distance to urban areas, they use this land cover map, they use this NDVI from landsat image and they also use this depth to groundwater.

Remember this all these things are available and you just have to use geographic information system and remote sensing methods for getting this information. All these information are available and you have to use those techniques together to generate these individual layers of covariates.

(Refer Slide Time: 21:28)



So, these scientists when they have used this legacy soil data of 1142 samples with these five covariates, they use a specialized technique called a regression kriging. Now, we will talk about this regression kriging later on, but just remember that you know they if you know ultimately their aim was to go back.

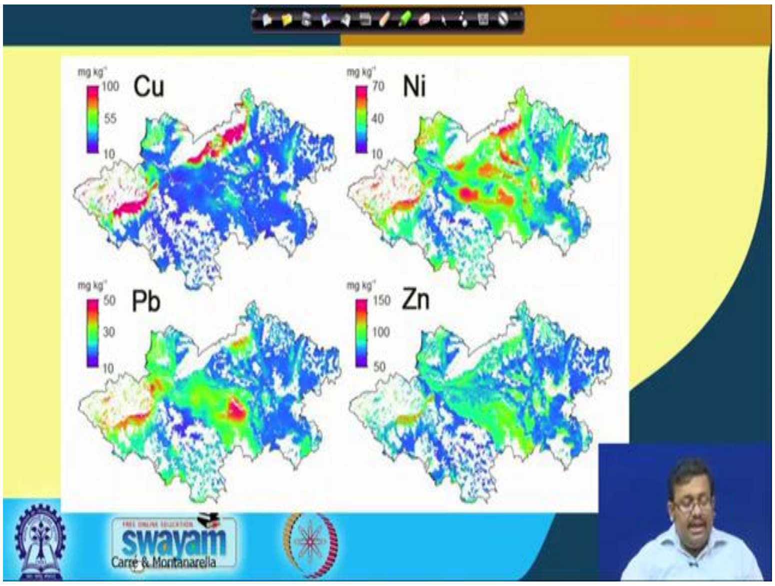
So, ultimately their aim was to produce so have, you know, a particular model and this particular model to predict to predict the particular soil heavy metal for example, copper and you know if this is soil variable j in our case it is a copper then they have used this model thing. So, obviously, these a 1 X 1, a 2 X 2, a n X n these are spatially continuous whereas, this is the residual, you know that anything any predictor model or this is basically a linear predictive model. And in this linear predictive model they have used these auxiliary data as predictors as you can see X 2 up to X n all these are auxiliary data.

So, they have divided the prediction into two major part; one is the multiple linear regression, another is kriging. So, basically first they use this. you know, spatially you know these a 1 X 1, a 2 X 2, a n X n to predict a particular soil property, for example, in our case it is copper and then they use this residuals. These residuals they use separately for kriging and you know kriging is an interpolation technique. So, this kriging was this kriging use this, you know, residual values for interpolation process according to spatial autocorrelation of the variable because I have already told you what is spatial autocorrelation these residuals can show spatial autocorrelation.

So, these residuals basically was used you can see this is a semi variance to model the spatial autocorrelation structure. So, obviously, we will get the prediction from the regression data which is the linear regression and also we will be getting the prediction from kriging of the residuals and regression kriging is basically the combination of these two, you know, these two outcome to finally, give you the prediction of soil variables. Again, for predicting a particular soil variables they have used the covariate data and they have used this in two, you know, in two fraction; one is called the linear regression, another is, you know, kriging of residuals.

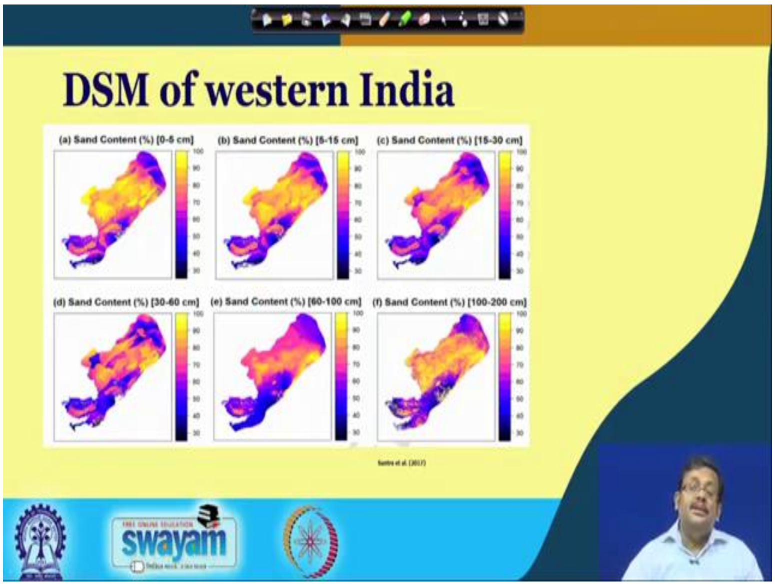
So, in the linear regression they use the covariates data to produce the, you know, regression based outcome. And then use this residual to produce the kriging interpolation and outcome of the kriging interpolation and outcome of the regression they combine together to produce the regression kriging. It is the most advanced method of predicting a soil properties through DSM and they have used this and ultimately they produce this digital soil map of copper, nickel, lead and zinc.

(Refer Slide Time: 24:43)



So, this is the example of actual digital soil mapping.

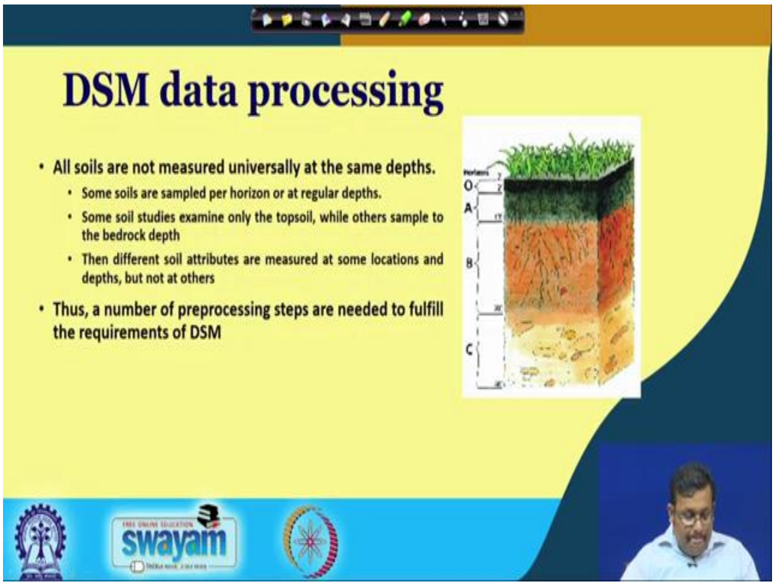
(Refer Slide Time: 24:55)



And, if we consider the Indian digital soil mapping product; obviously, Santra et al in 2017, they have produced the digital soil map of western India where they have used this that they have produced the spatial variability maps of sand content for different depths, as you can see here from 0 to 5 centimeter, then 5 to 15 centimeter, 15 to 30 centimeter, 30 to 60 centimeter, 60 to 100 centimeter and 100 to 200 centimeter. So, as you can see here, these are the DSM products and beauty of it these DSM products and maps are available for each of the standardized horizons standardized depths.

And, these depths, why these are standardized depth because they are prescribed by the global working group of DSM we call, you know, by according to the global soil map dot net project we will discuss this later on. So, these are some prescribed soil depth and you can see using the DSM technique, we can produce the spatial variability map of any soil properties at these different depths. So, it is basically a 3D representation and remember that these spatial variability maps or DSM products are highly accurate because it considers not only the soil legacy data, but also it considers the other covariates from different sources.

(Refer Slide Time: 26:29)



So, let us wrap up here and let us, you know, start from here in our next lecture and then we will discuss the different types of data processing for digital soil mapping and we will discuss other aspects of digital soil mapping.

Thank you and let us next in the meet in the next lecture. Bye.