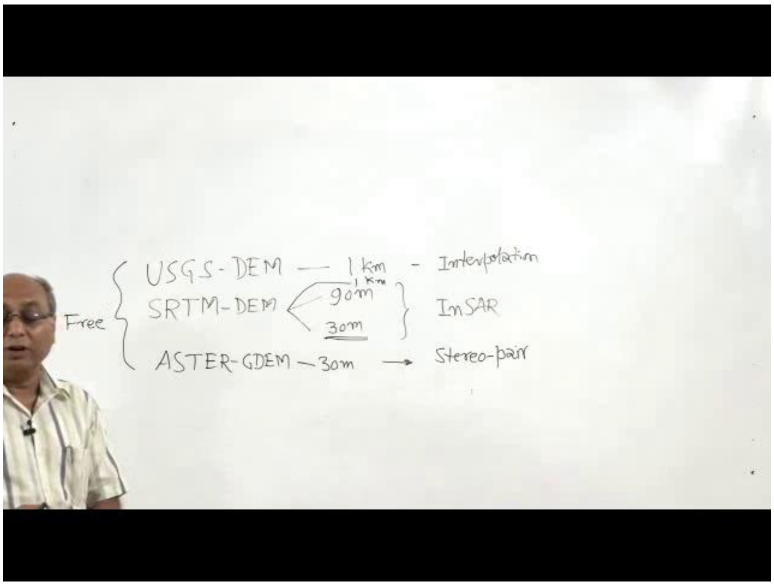
Introduction to Geographic Information Systems
Prof. Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture – 14
Quality Assessment of freely available DEMs

Hello everyone, and this particular lecture we are going to discuss and how access the quality of a DEM and especially quality of DEMs which are freely available.

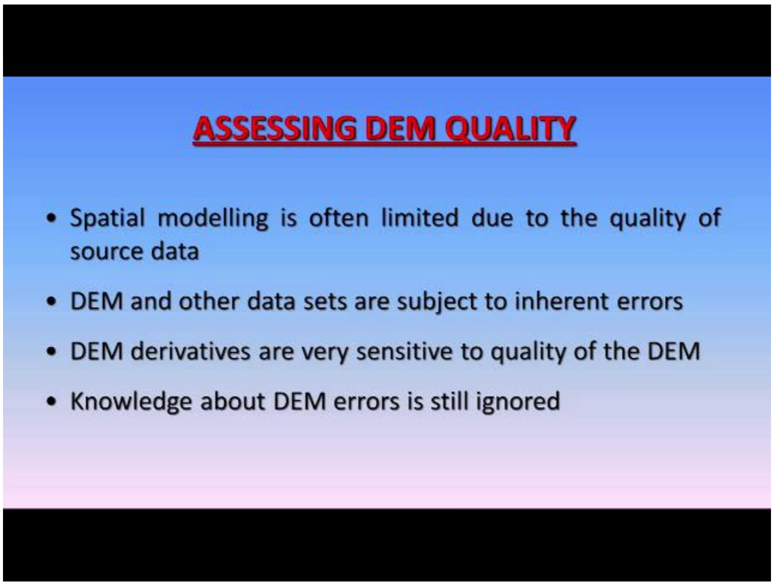
(Refer Slide Time: 00:46).



As in previous lecture, we have discussed that there are several freely available DEMs are available on internet the USGS – DEM, SRTM - DEM at different resolutions even including at one more, one kilometer resolution as well and now a days recently it has they have started giving also 30 meter resolution and then earlier this thirty meter ASTER - GDEM that is global digital elevation model for entire globe at 30 meters resolution is available.

So, depending on the process or the inputs of the data the quality of DEMs would be there, but why quality is, important because we are going to use this digital elevation model to drive various outputs and if we know that if the input itself is having errors, or problems then your outputs will get affected their quality will get affect.

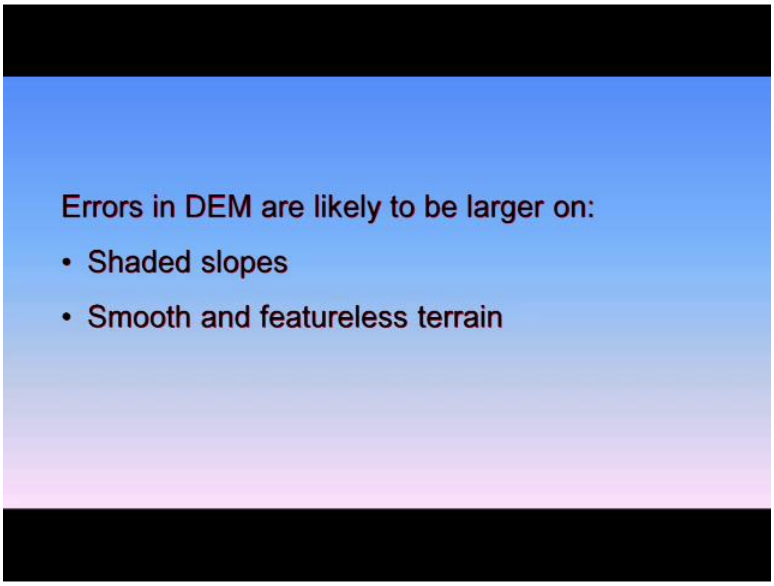
(Refer Slide Time: 01:51)



Therefore, it is very necessary to under extend the limitations of each digital elevation models the errors associated with different type of digital elevation models which we are going to use in our g I s analysis. So, this is one part which is least discussed or least in and cooperated in literate as well the quality and the errors component which we will be emphasizing in this particular error because, we must know the inherent errors which are present in DEMs and others data sets which are going to inside our GIS data base which we will later use for different types of analysis and because there are different DEM derivatives tens of derivatives can be drive from DEM which are very sensitive to quality of the DEM for example, a slope map or a aspect map gradient map through surface hydrologic modeling one might be going for water shed boundary may be the extreme network extreme ordering and sediment power index there are many, many outputs which can come directly from DEMs these derivatives.

So, if the quality of DEM itself is poor then definitely this will affect the our output and knowledge about the assignments that a the errors which are associated with different DEMs even your own interpolated DEM then this is really very much ignored not much discussed, but we must, must understand the inherent errors in the DEM and wherever it is possible try to minimize those errors because otherwise errors will propagate in GIS.

(Refer Slide Time: 03:46)

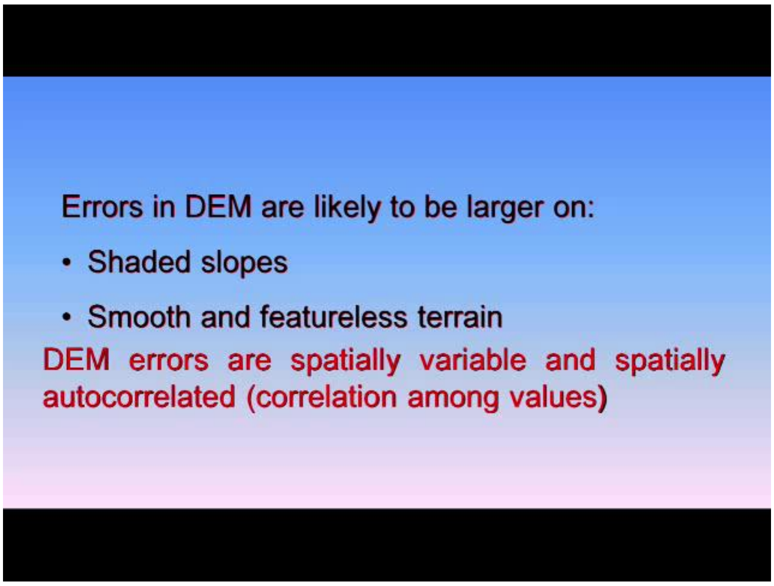


Now, which are the errors in DEM are likely to be larger in which areas especially in the shaded slopes or a smooth and featureless terrain now through extreme terrain conditions can bring errors one is like hilly terrain like Himalaya because technique is a looking in of another direction and if there is a severely. So, it is only looking one side slope not the other side of the slope and therefore, it will have problems in a hilly terrain. Where as a raster DEM which is using the stereo pairs the because in hilly terrain lot of areas in hilly terrain may go in the shadow and then these shaded areas will have problems, each type of a digital elevation model if a, if we consider only say freely available digital elevation models they are suffering for one type of problems or another.

Similarly, if it is a smooth and featureless terrain like a dessert area then you do not have any much features or DEMs very prominent one and then, then it is very difficult to access even the quality of DEM and therefore, there as more likely to have more errors in a very smooth and featureless terrain like desserts or terrain like Himalaya, where it is. So, much raggedness is there that a and a different techniques are having different problems, but the another the requirement of a high resolution digital elevation models for these two extreme areas especially of hilly terrain of Himalaya is much more because there are lot of many things which are been which are done in Himalaya n terrain and you are having contour intervals which are quite large. So, you cannot rely all the time on the because, if you put a say in 50,000 scale map even 100 meter contour interval these two contours are coming very close to each other.

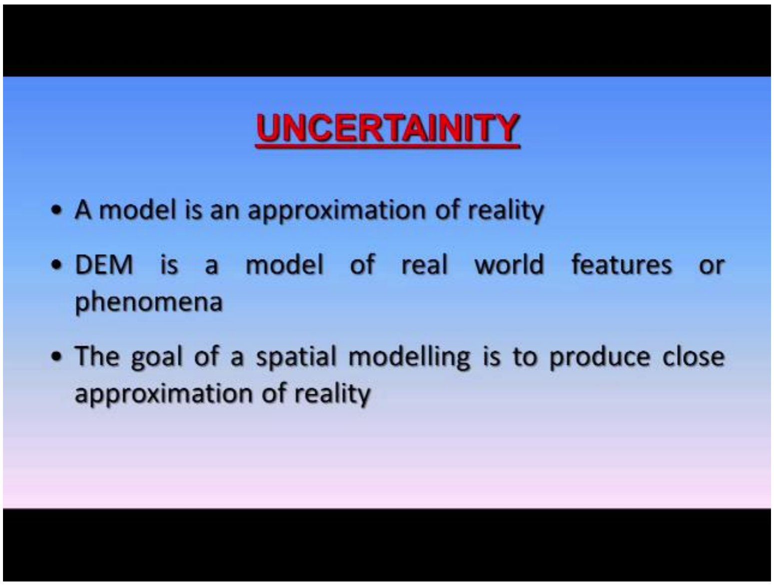
So, you cannot go for even then, you go for larger scale maps; that means, you require more input and it hilly terrain like Himalaya, it becomes very difficult, but the requirements for a high resolution digital elevation model and that too accurate digital elevation models are much, much more difficult for Himalaya n terrain and different techniques are suffering the digital elevation models, which are derived from different techniques are suffering from one type of error or another. So, this assessment must be there before we employ such DEMs and DEMs errors are especially variable because it depending on the terrain they will vary and also it especially auto correlated; that means, the correlation among values is also there.

(Refer Slide Time: 06:51)



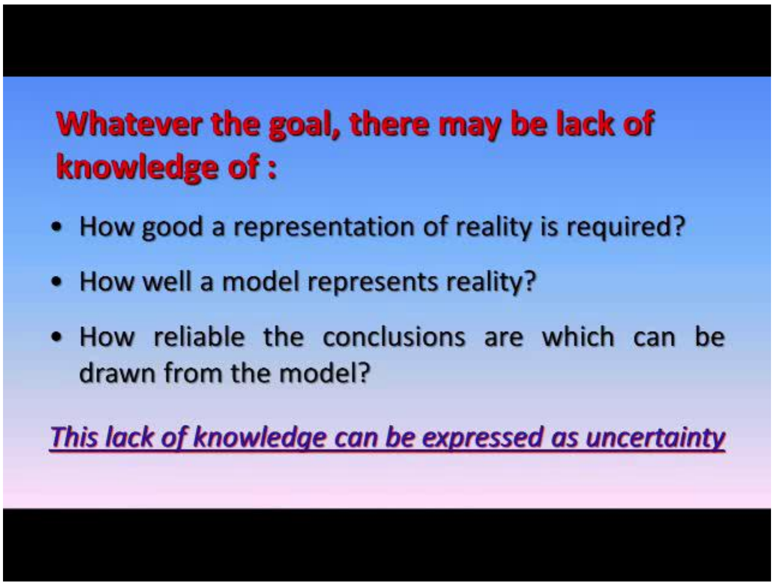
So, sometimes you get a very smooth surfaces sometimes you want to represent a very rugged terrain and then, you are having problem, another issue is about the uncertainty as you know the digital elevation model is a approximation of reality, and the DEM is a model of the real world of features or phenomena which is a terrain.

(Refer Slide Time: 07:25)



And the here the goal of a spatial modeling is to produce close approximation of reality, because it is a representation it is based on the Shannon theory that within that cell all elevations are having the one value. So, this is basically sort of sampling for assumptions is there, and whatever the goal, whatever the purpose there may be there are lack of knowledge about how good representation of reality is required.

(Refer Slide Time: 08:01)

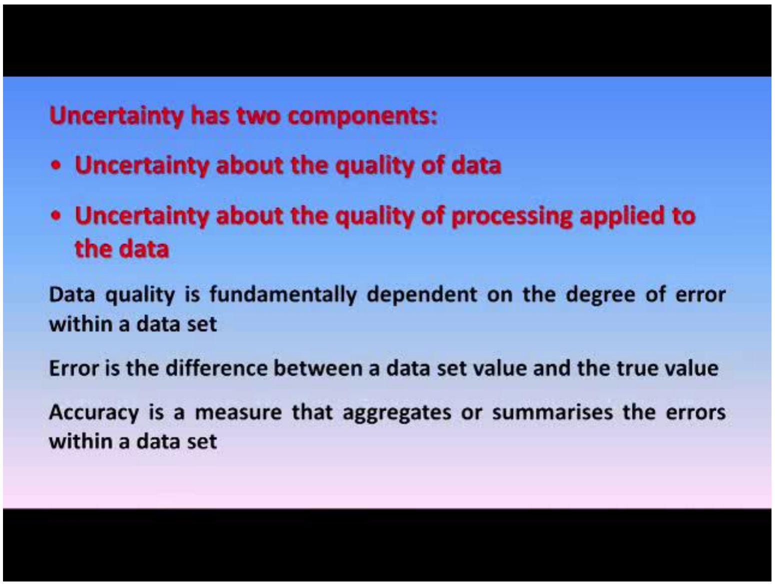


Because we do not know for what if we are very much sure that for what purpose we are going to use then we must know the errors associated with our input data.

So, this say whether our input data is really representing the reality in a in an accurate way or true manner that has to understand. And how well a model represents the reality which particular digital elevation model derived from variety of sources will fit to our requirements and how reliable the conclusions are which can be drawn from the model. So, when you drive the outputs a way different derivatives from a digital elevation how much relive you can have with them, like if you drive use a digital elevation model and drive a slope map.

So, how accurate how reliable that slope map again this reliability or this confidence will combines you are assured that your digital elevation model is really truly representing the real terrain conditions. Then only you can rely otherwise you will have little less level of confidence. So, this lack of knowledge which we can represent or express as uncertainty with the digital elevation model and this uncertainty has got two components; one is the uncertainty about the quality of data and that uncertainty about the quality of processing applied to the data.

(Refer Slide Time: 09:27)



So, because if we talk say about the based digital elevation model which been derived from this SRTM then this is (Refer Time: 09:41) based on one particular remote sensing technique which is a radar remote sensing or active remote sensing. So, there are issues and associated problems for different terrain with the radar data. So, that there will come under the uncertainty of quality of data then, this data when it was acquired by this mission then this data has been processed and then digital elevation models we have created. So, how accurate that process was. So, that uncertainty will come in the quality of processing applied to the data because each processing will might or might bring certain errors.

So, those errors should also be kept in mind, and basically those will produces all these digital elevation models generally they declare that what kind of errors their data is suffering from, but they do not declare a specifically for different terrains. So, if I am going to use for Indo-Gangetic plain I may be sure that yes whatever the data I am using is quite accurate see example related to SRTM, but when I go for hilly terrain like Himalaya then I know that because of radar techniques this is having some limitations some problems and why do so, which might be there.

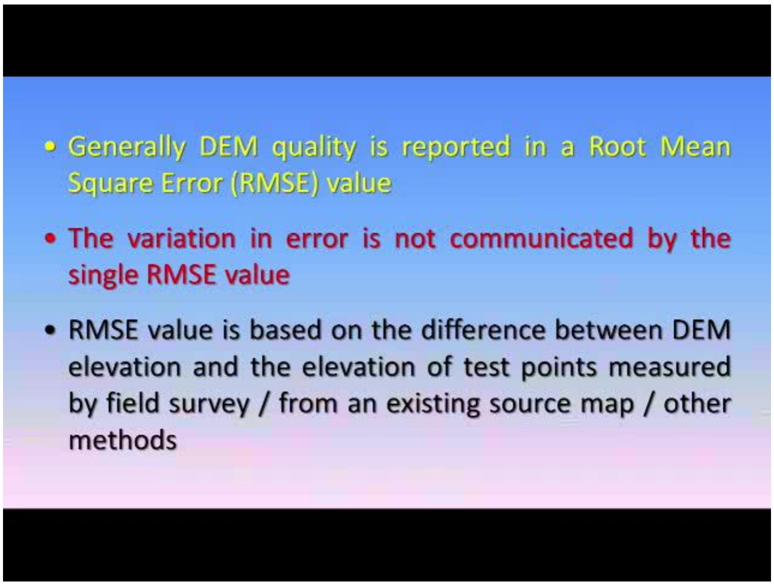
So, depending on the terrain our application or utilization of different DEMs should be there, and this is our own assessment through experience and some analysis that aster in GDEM is quite good for Himalaya n conditions like Himalaya n terrain or terrain which is highly rugged. Whereas SRTM is quite good or quite accurate for the plane area like Indo-Gangetic plain, all DEMs are not suitable for all kind of terrain conditions. So, this thing has to be remembered.

How we can access this thing that for that you need to have some independent collection of some say point data which you can use as a control points or controlled ground controlled points and then compare what the value in a cell of a DEM of different DEMs is being carried here. So, whichever the DEM which is having the close representation of your controlled or these ground controlled points you can say that that DEM is representing truly my terrain that particular terrain then a different DEM and therefore, the data quality is fundamentally dependent on the degree of errors within a data set this is very important here about the data quality errors is a difference a data set value and the true value. Because when, the data is being acquired to remote sensing techniques or some other techniques we do not have all the time the true values and the since DEM is a continuous surface we do not have the true values to check continuously corresponding continuous values and therefore, accessing the errors becomes very difficult the true values be might be having only few point values only again those point values we can compare with the value which has which are present in a continuous fashion in the digital elevation model.

So, even I am having the ground controlled points or some independently collected elevation values and then only for those points I am having comparison otherwise for remaining area I do not have any comparison in this point. So, this point becomes very important especially the terrain which is highly rugged accuracy thumb is also is there is a measure of that a measure that aggregates or summarizes the errors within a data set. So, how much errors it is being carried suppose in case of digital elevation model a elevation suppose to we say hundred meters it is coming say in your digital elevation model it is coming hundred 10 meter.

Now, so that means, 10 meter plus value are this coming or there might be case where suppose to be 100 meter, but it is coming 90 meter. So, we can say this plus minus 10 meter.

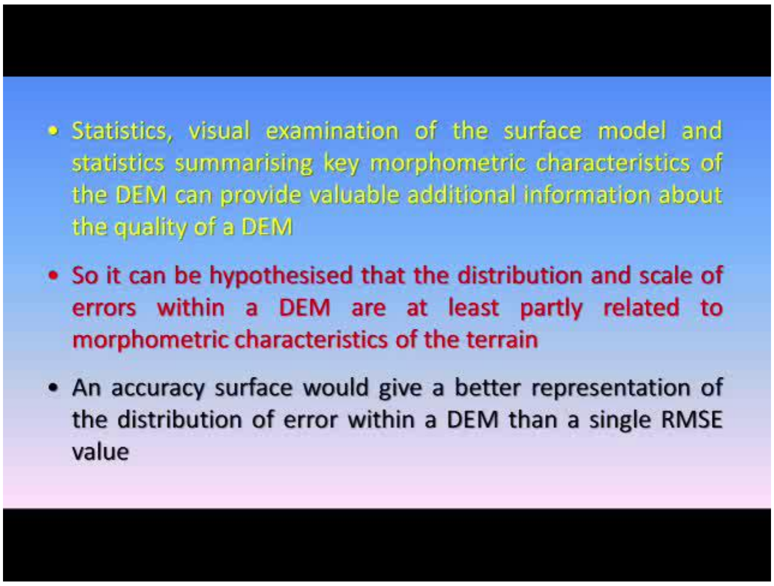
(Refer Slide Time: 14:28)



So, this information must be available before we employ that DEM and derive certain derivatives or outputs from that generally d e m quality is reported in a root mean square or RMSE value by the developers of those DEMs or if you are using interpolation techniques that time also you can record the errors and the variation in error is not ah communicated by the single RMSE value. Now this RMSE is based on the differences between DEM elevation and the elevation of the test points or ground controlled points measured by field survey from an existing source map and other methods.

Now, when you go for a very accurate assessment then one question can also we raised that how accurate these test points are because after all you would be using certain instrument say for example, you are using GPS or differential GPS technique now differential GPS techniques will also have their own errors. So, your test data or which you are considering as a standard data is also suffering from some errors. So, these things have to be kept in mind about these products which we are going to use in our g I s analysis.

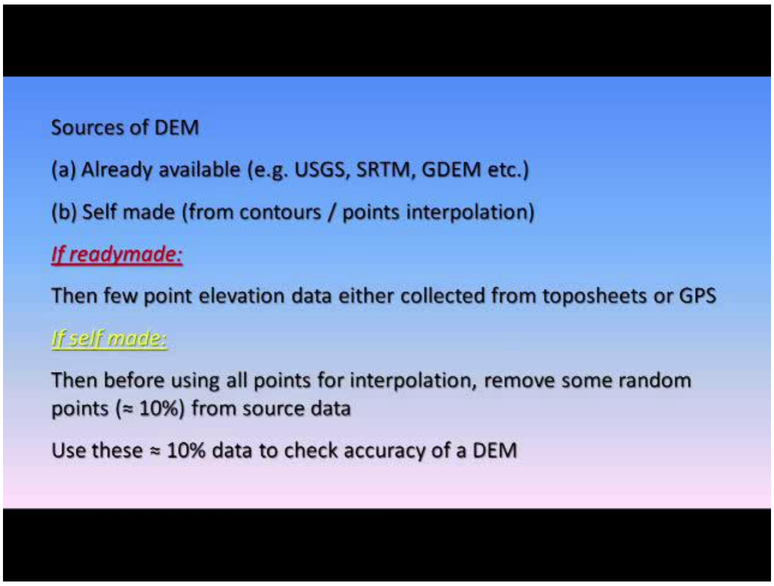
(Refer Slide Time: 15:50)



now there are various ways statistics visual examination of the surface model and statistics summarizing key morph metric characteristics of the DEM, can provide valuable additional information about the quality of a DEM and. So, it is hypothesized that the distribution and the scale of errors within a DEM are at least partly related to morphometric characteristics of the terrain. Because if it is a elevation model the presenting the elevation or terrain conditions then one we are in sometime we are going to use for morphometric analysis.

So, if we can perform before that then we can also use this tool to access the errors in our input digital elevation model and accuracy surface would give a better representation of the distribution of errors within a DEM than a single RMSE value. So, this if this can be derived then this surface and even you can you know, exactly in which part of your input DEM it is having more errors than compare to other errors rather than just having a single value through this root mean square estimation now a sources of DEM which we have already discussed initially, but I am just going quickly through this one.

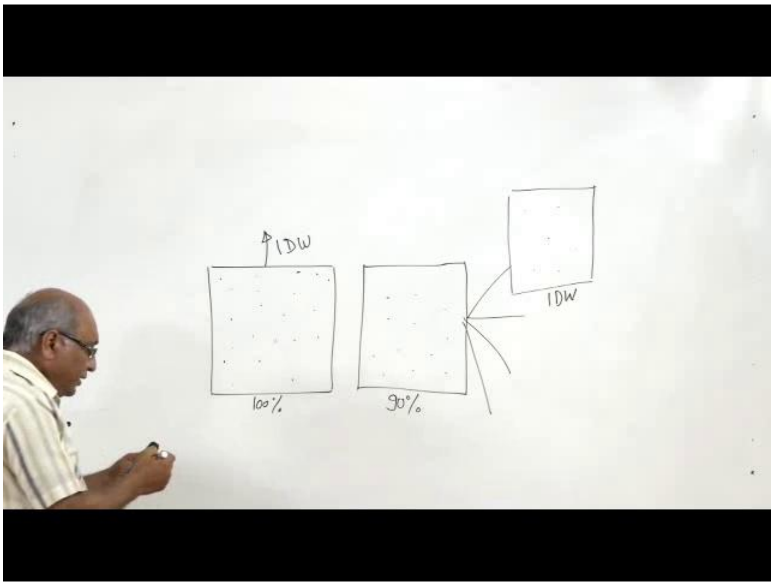
(Refer Slide Time: 17:12)



Now if a readymade DEMs are their then, how you can check that if readymade DEMs are you are using in g I s then if you are having some test points available through your Indo-Gangetic you can compare those test points with an digital elevation corresponding cell values of your digital elevation model and find out standard deviation and other things and then you can access that which is closest and that you can say that this is more true representation of the terrain, but it is only against certain points or these test points or control points can also come through GPS there are libraries of GPS ground control points are available which can also be implied to access the quality of DEM how accurately it is representing a particular part of the earth.

If it is self made; that means, using interpolation techniques then I can suggest a very quick quite accurate method of accessing which your digital elevation model or which interpolation techniques is suitable to you.

(Refer Slide Time: 18:40)



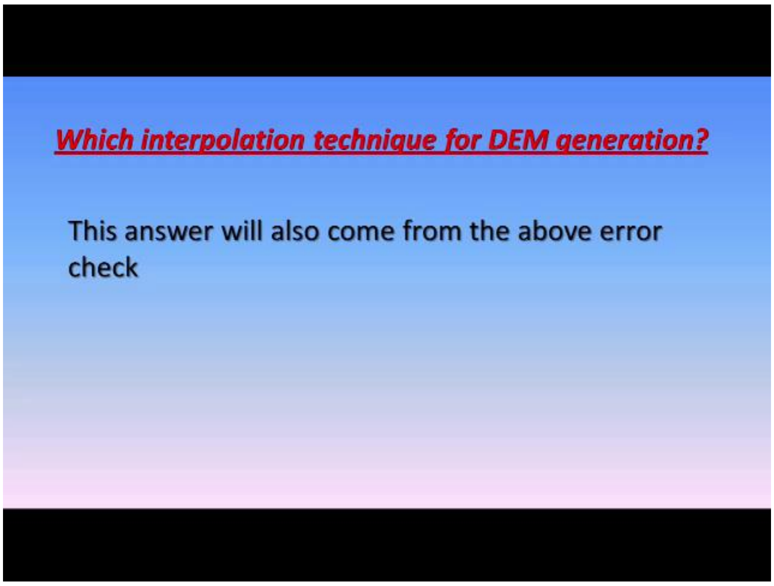
For example, if you are having a hundred points elevation points available to you which are say distributed in a area now before going for interpolation, but you do that you select randomly 10 percent of the points these random selectors are also available with this standard g I s softwares and you remember from original data sets. So, now, say here we had the hundred percent data now, we are having here before interpolation we are having now ninety percent of the data randomly 10 percent data has been selected and removed temporarily from the original data set and do that interpolations implying three four and different interpolation techniques.

So, you will have different surfaces here one two three four surfaces then use those 10 percent which you kept separately those points you use and compare with the your different DEMs which you have generated. So, for example, I am having one DEM which has been generated now those 10 percent points I will use as test points or control points compare the corresponding cell values and whichever will give the closest values I will consider that there is the best digital elevation model be a is suitable for my input data set.

Now, at this is stage I am having a input of only ninety percent points. So, what do you then what I have to do that now I know that say for example, I reach to a conclusion that I d w is more suitable for my data sets. So, then I will go back to hundred percent points and then create a surface using I d w method because I know that I d w is giving the best results then I compare with 10 percent points and those 10 percent points you have selected randomly.

So, there is no biasness and this say this way I can access that which DEM or which interpolation technique is suitable for my set of input points and this is what is mentioned here that, if you are going for self made digital elevation models then before using all points for interpolation remove some random points say 10 percent from the source data. Then use these 10 percent data to check the accuracy of a DEM or different DEMs which you have derived from using different interpolation techniques and once you are assured that which one is giving the closest value use then hundred percent use that particular aj tool or interpolation techniques and create a final surface.

(Refer Slide Time: 21:49)



So, which interpolation technique for DEM generations as I have answered that this answer will come from the above error check then we will go for the this kind of the exercise first a hundred percent 10 percent randomly selected then 90 percent various DEMs using various techniques and then comparing finding the best one and once I have identified the best one I will use 100 percent points and create. So, this will remove my sort of uncertainty about which method is suitable for that particular set of points, now this exercise has to be done with each new set of the point because we do not know what it ah what it is representing you may end up with the same result, but it has to be done with each set of new points.

So, this this will really help us to find out which interpretation technique is suitable for us. So, this way you can have knowledge of the errors in your data in your input data in your digital elevation model and also you can earn when you yourself are preparing then you can find out the best interpolation techniques by doing a sort of exercise and then finally, preparing a digital elevation model I have also mentioned in previous lectures that the user or g I s or the data which is coming into your g I s he must be aware about the inherent errors which are already inside the data because, if you are aware you will you will be using that data set very carefully because when you derive a output using that data sets that output may not be that reliable which you are intending then whole exercise may go in waste. So, it is always to have knowledge of errors and keep errors at the minimum this is very, very important component in g I s.

Thank you very much.