Geographic Information Systems
Prof. A. K. Saraf
Department of Earth Sciences
Indian Institute of Technology-Roorkee

Lecture-41
How to assess quality of a DEM

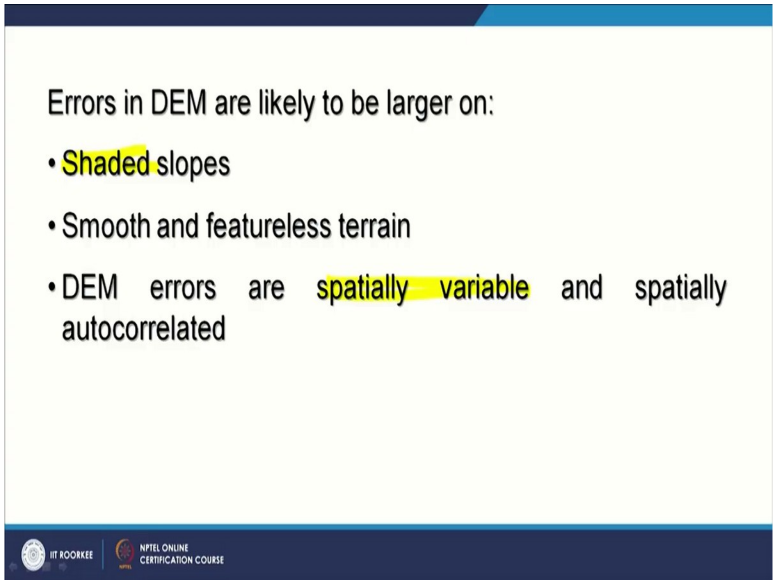
Hello everyone! and welcome to a new discussion of this geographic information systems course. And today, we are going to discuss how to assess quality of DEM. This I have been telling earlier. In earlier discussions that one day, we will discuss about how assessing the quality of DEM. Why it is necessary because nowadays you know, there are various options to generate digital elevation model.

And various digital elevation models are generated using different Remote Sensing techniques or survey toposheets, have been made available on internet. Most of them are free of a moderate resolution DEM's and therefore whenever we employ either an interpolation technique for our own surface or for using already available digital elevation models then we must try to assess the quality of DEM. And you know this should start with fast with the metadata.

If you have downloaded a digital elevation model from internet source then the best thing is first to study the metadata. Also, you should know how it has been generated. Which Remote Sensing or otherwise a technique has been employed. And what are you know, things which are may matter in future while in terms of accuracy. So, in this as you know that whenever we go for these kinds of generations which is surface generation or in form of DEM.

Because it is not possible to visit each and every point on the surface of the earth or part of the earth and therefore some technique has to be employed which also involves your interpolation and then as you know that when we have developed a special database on GIS platform and we start analyzing things.

(Refer Slide Time: 02:49)



Ultimately, we may go up to depending on our requirements special modelling or for prediction and that reliability of that prediction or modelling will definitely will depend on the quality of source data. How good quality the data is? And questions are generally asked that what is the accuracy of your prediction? So, let me give you one example that few years back, I was hearing a presentation by a person on you know a landslide hazard zonation mapping.

And in that landslide hazard zonation mapping, a question was asked to him that what is the accuracy or how would your prediction be? And he had to say that it is 50%. And 50% prediction is virtually, no prediction at all? Because anybody can predict that there may be rain today or may not be rain today. So that does not mean anything in terms of prediction. And whenever we go for special modelling employing digital elevation models and one of the major sources of information is our digital elevation models and therefore, we must assess the quality.

And one inherent error because there can be various types of errors. So first we have to assess the inherent errors and there might be errors during processing. So, processing errors might be in our hands to check and rectify. But if there are errors which are inherent; the organization which has generated the digital elevation model or collected data in the field. And then later on, I have generated digital elevation model using the dataset.

Then I would call as an inherent error or things which are lacking in that one. So, we must know and you can always know if you're downloading that any DEM whether it is SRTM or ASTER DEM, CARTOSAT, if you go through the meta file and also how it has been generated. Some documents must be available then you will get an idea about the inherent errors. So, this is very-2 important part. Before you employ any data, you must know the limitations.

You must do check on the errors also. Now as you know that DEM is a you know, storehouse of information. That means that we can drive various information, various surfaces, various Trends using digital elevation model. And these derivatives are very sensitive to quality of the DEM. So, if the quality of the DEM is poor, you can guess that what is going to be the results when I drive some data. For example, if digital elevation model is having lot of error and if I generate a slope map or aspect map then I cannot say with high level of confidence that my derivative using that particular DEM is very good or accurate.

That is why this inherent error really matters lot and therefore generally you would find that many times people will just download the DEM and will never bothered about inherent errors. And this knowledge or the lack of this knowledge about the errors in DEM still is not discussed much; neither on literature or in practice, people do not really bother but that will make a huge difference in the confidence level in your results.

One definitely spends some time to assess the quality of DEM. How you can assess? There are various ways one can assess like errors in DEM can be very large when you are having shaded slopes. Also, why on shaded slopes because you know even if I am using stereo pairs or I am using InSAR based DEM then in hilly terrain, the shaded sloped where the illumination is not reaching or because of orientation of that particular slope or valley, may hinder data and that ultimately may get reflected in terms of errors in a digital elevation model.

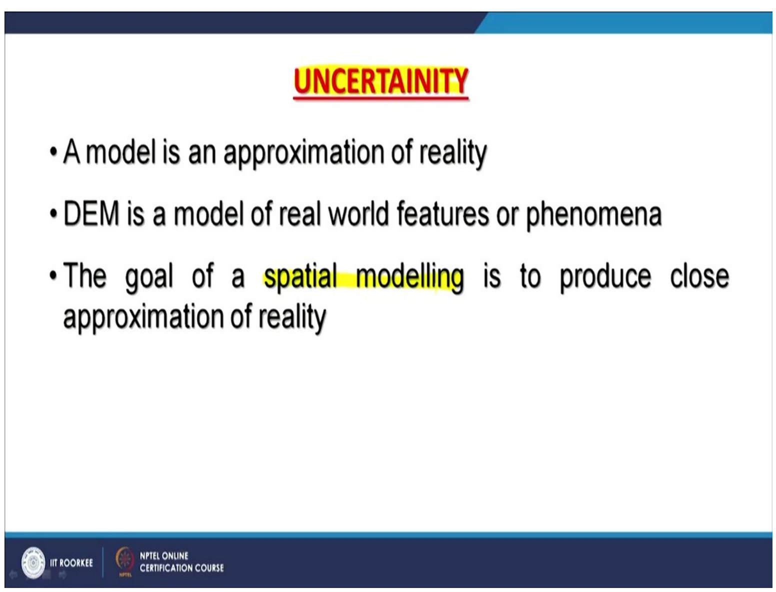
And therefore, one has to be really careful about that. Sometimes when digital elevation models are generated for very flat areas, smooth terrain; feature less terrain, again there might be a lot of errors. For example, if I am working in a desert area and there are no even dunes then it becomes really very difficult and that digital elevation model no matter what ever the technique which has been employed, may also generate lot of errors.

Similarly, when area is fully covered with snow like in Antarctica or in north Polar regions then you may have a lot of problems in a digital elevation model. Now this is another very important point that the errors in digital elevation model are spatially variable. It means that at some location the error may be different, at some location within the same file, within the same surface that might be different because of as I have discussed about these shaded slopes.

In some parts which do not have to shaded slopes, the errors may be less but a in the areas where you are having shaded sloped, error may be large. And errors are spatially variable and spatially autocorrelated also. So, in that way gives a problem about the errors in digital elevation model. And all these errors which we have discussed so far or discussing are inherently errors because we do not have much control except knowing them and if possible, we can avoid a particular digital elevation model if it is having too much error then it is there

But it is very difficult to correct those digital elevation model unless we ourselves generating digital elevation model.

(Refer Slide Time: 09:43)

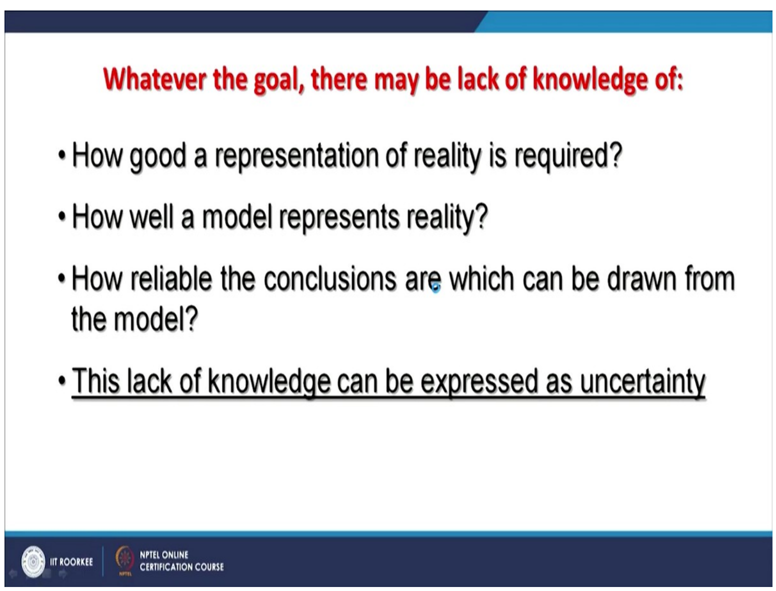


Now while assessing the quality of a digital elevation model, one term we must also discuss here is about the uncertainty. So, basically as you know the DEM is a model and it's an approximation of reality or abstract form of reality. Reality here is the terrain so what we are having? We are having a model of the terrain in abstract form, not in a one-to-one scale and therefore this DEM as a model of the real-world features of phenomena which is what present in abstract form reduced scale.

And as you know that in spatial modelling, what we try? We try to produce whatever results which are approximation or close approximation of reality. And reality of course; the real world you can never generate on the computer or through some modelling. Because modelling itself is a representation. Representation in our case is like in current discussion, representation of the terrain.

So, it is not truly a terrain but it's a representation. Now, whatever is the goal where we are going to use digital elevation model, there may be lack of knowledge about how good a representation of reality is required.

(Refer Slide Time: 11:17)

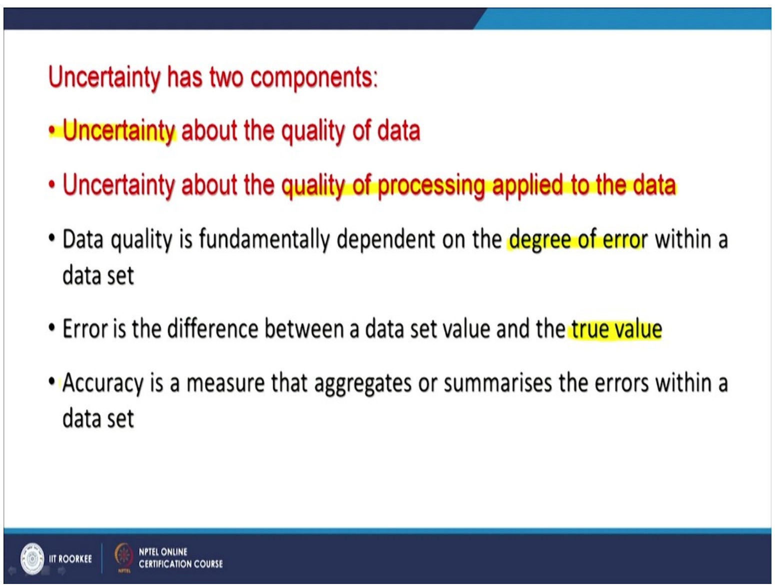


So, if we know what are my targets of project. How accurate the results I have to produce? How reliable I have to produce? There, that can guidance to choose an appropriate digital elevation model and not only that but more accurate digital elevation model. So further is that how well a model represents reality. If it is having errors, definitely it is not going to represent the reality in a good term and how reliable is the conclusions are?

Of course, this reliability or level of confidence will depend on how that model has been generated? What are the errors or extent of errors and whether there is just one error within one file or spatially varying errors are there? If it is there then things may become complicated. So, this lack of knowledge about a model or about a digital elevation model in our case can be expressed as uncertainty.

And this uncertainty has basically two components; one is the uncertainty about the quality of the data which is what we are discussing here. And uncertainty about the quality of processing applied to the data.

(Refer Slide Time: 12:42)



So, you know this probably to some extent we can control because this is in our hands. I have got a DEM, now I am processing it. I am driving lot of products out of the DEM. So, there I can have a check but if I am using readymade DEM's; available DEM's on internet then I do not have much knowledge about the accuracy or errors of that DEM. So, our efforts should be to learn those inherent errors of any data which I am going to employee for my study.

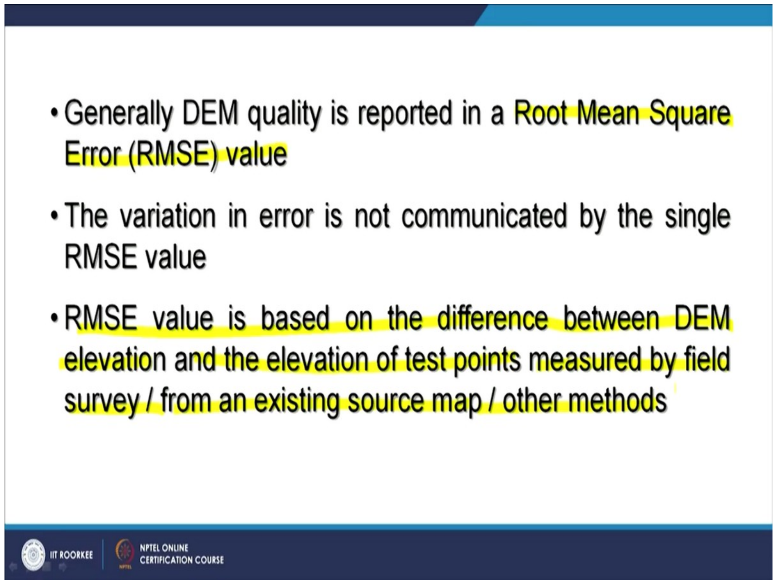
So, the data quality as you know is fundamentally dependent on the degree of error within a data set. If errors are very minimum that means the data in case of a digital elevation model is representing through a good representation of a terrain. But if the degree of error is very high then you can think that what reliability you would have and how big uncertainty would be there. As you know also that error is basically a difference between the data set value and the true value.

Now problem here is in many-2 cases we do not know the true value at each location because a DEM is a continuous surface. That means for each cell, I do not know the true value. If I would have known then definitely, I can create much more highly accurate and high-quality digital elevation model. Since those values were not there for each cell. That is why I have resorted to some Remote Sensing technique or interpolation technique.

So, only we can find the errors where we are having some values; some control point, some spot heights which we considered as reliable height or nowadays the data about few points randomly distributed within my study area can come also through differential GNSS. So that way, I can get some true values and then can compare. Do some statistical analysis and can tell that how good my input DEM is.

And as we also know, this we have discussed that 2 things about accuracy and precision in previous lectures that accuracy is basically a measure that aggregate or summarizes the errors within a dataset. And sometimes we usually used but this because you know, the precision depends on the instrument which I am going to use it but accuracy is basically a statistical term. How close I am to the real value or true value?

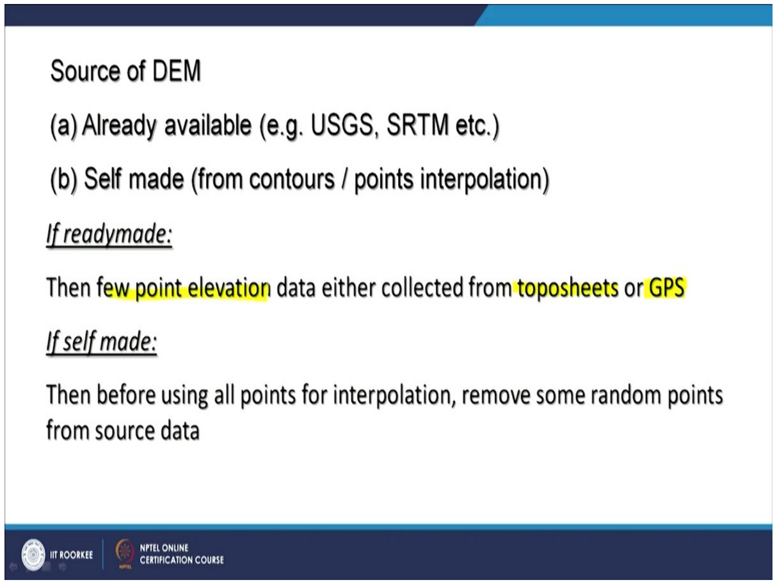
(Refer Slide Time: 15:34)



Now how this error can be represented? So normal way or conventional way of using this root mean square error or RMSE values because there are some ways to check it, how closely that is? So, we get through this RMSE. Now variations in error are not communicated by single RMSE value. So, if you are having for one entire dataset, just one single RMSE value that may not be sufficient in sense because as we have learnt that errors may vary spatially within a data set and that is why, a single RMSE value will give you some idea about the inherent errors in the data set but may not be sufficient.

So RMSE as you know is based on the difference between the DEM elevations and the elevation of test points or true values measured by field survey or from an existing source map or other methods. So, if I am having that data, generally if you started trying to assess the quality of a digital elevation model, I am sure from one way or another, you may get some test points; some true values for your area which you can employ to assess the quality of DEM.

(Refer Slide Time: 17:01)



These are basically RMSE or any other analysis which you would do are basically statistical based analysis. For example, visual examination of the surface model because if surface model or DEM is suffering from voids then visually you can also assess. If it is having seems, visually you can assess very easily but when you involve the statistics then it can be summarizing by key morphometric characteristics of DEM, can provide valuable additional information about quality of DEM.

That means before you employing a digital elevation model for my own work or analysis, I would require to do some analysis like morphometric analysis over a targeted DEM to assess the quality. So, this is another way of assessing the quality. And therefore, we can hypothesize that the distribution and scale of errors with any DEM are at least partly related to morphometric characteristics of the terrain.

Because when you are having a digital elevation model of highly rugged terrain then scenario can be very different. However, if you employ this morphometric analysis, you may get that idea about the error which are inherent in that particular digital elevation model. So, an accuracy surface would give a better representation of the distribution of error within DEM then a single RMSE errors because of our errors are spatially vary.

Therefore, single RMSE error is not sufficient. We need a surface which will tell us where the errors are more and where the errors are less. Very quickly, I will go through the sources of DEMs which are available to us and will also see that what are the errors associated with this. As you know, we have already discussed these things earlier so very quickly I will go, that already available DEM's may be USIS DEM which has been generated from toposheets. SRTM which is based on InSAR Remote Sensing Technology or self made DEM's also.

So, from contour or points interpolation that we do. So, if we are employing readymade DEM then the best thing which we have developed in few years' time; a best technique rather I would say is to get some elevation points, reference points, test point from some other sources. From some other source that may be independent, what I mean is maybe from spot heights from Survey of India toposheets or may be from differential GNSS or GPS.

Now here I would like to mention that organisations like Survey of India and National Remote Sensing agency; NRSA, they are having a big library of these control point called as ground control points. We discuss this ground control points when we have been discussing geo-referencing. So, like NRSA has developed a library which is having about 30,000 ground control points of every part of India.

And these ground control points (GCPs) are evenly distributed over India. But unfortunately, that dataset or that library is not open to us. But I know that in order to assess the quality of a digital elevation model which has been generated using CARTOSAT and it is available at 30-meter resolution freely for India which you can download from BHUVAN portal. While assessing the quality, they have used dost 30,000 ground control points to assess the quality.

So, in that way, these few test elevation points have been used to assess the quality. How it can be done which we have also developed that if it is self made digital elevation model then things a little easier in sense because in every process, we are involved. So, on errors, we can check at every step. Keep checking the errors. If found, we go back and do the correction and then go forward.

If it is self made DEM then before using all points, whatever the observations you are having for interpolations, remove some random points from the source data. And I will give you an example, like I am having 100 elevation points or spot heights to generate a digital elevation model to interpolation technique. So, what I mean is that I will remove randomly some points from those 100 points and then keep those points as a separate file.

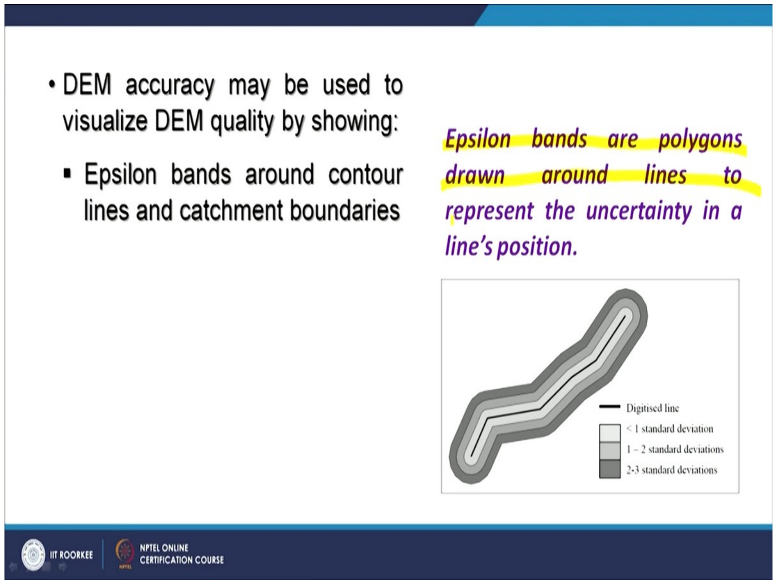
And I will do the interpolation using those 90 points or 90% of points. So let us talk in terms of percentage. So, then those remaining 10% points will be kept as without disturbing them and collected randomly which can be done very easily on a GIS platform then I will use those 10% point to check the accuracy of DEM because those points have never been used while interpolation was being done.

Then I can employ different interpolation techniques and variants of each interpolation techniques. So that means first, I have to do if there are 100% points. I will take randomly select 10% points from that and with 90% points, I will do interpolations and generate surfaces using different interpolation techniques, may be even 10 different surfaces employing different techniques.

And then I will use 10% point to see that which is showing the maximum variation, which surface is showing maximum variation with those 10% points. Because those 10% points have not been used in my Interpolation. Now whichever the surface which qualifies for the closest or you know, with minimum variations with test points and the interpolated surface point then that means now I know that which interpolation techniques have to employ.

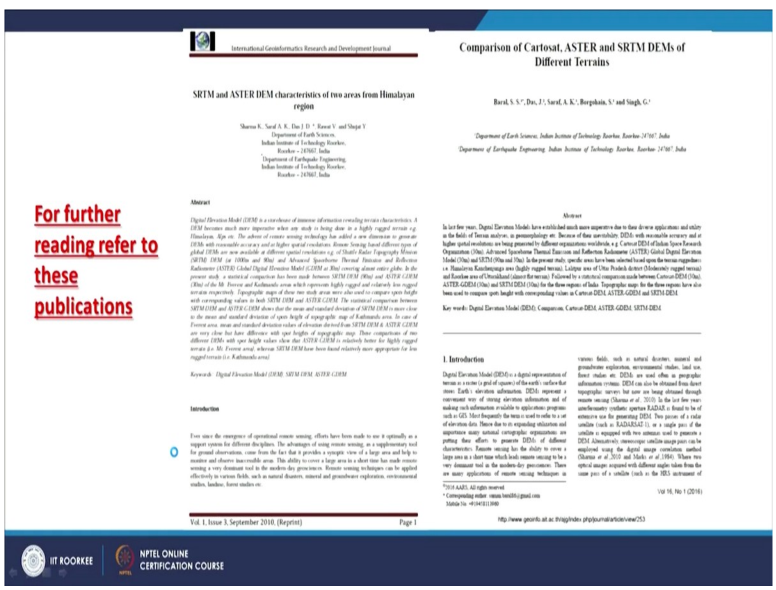
And then what I will do? I will use all 100% points and that particular interpolation technique to generate a digital elevation model. So, I would do know that digital elevation model is better more accurate than any other interpolation technique.

(Refer Slide Time: 24:44)



So, this answer will come from the error which I just discuss about this. Now those who are interested to learn about how to assess the quality of digital elevation model in much more detail.

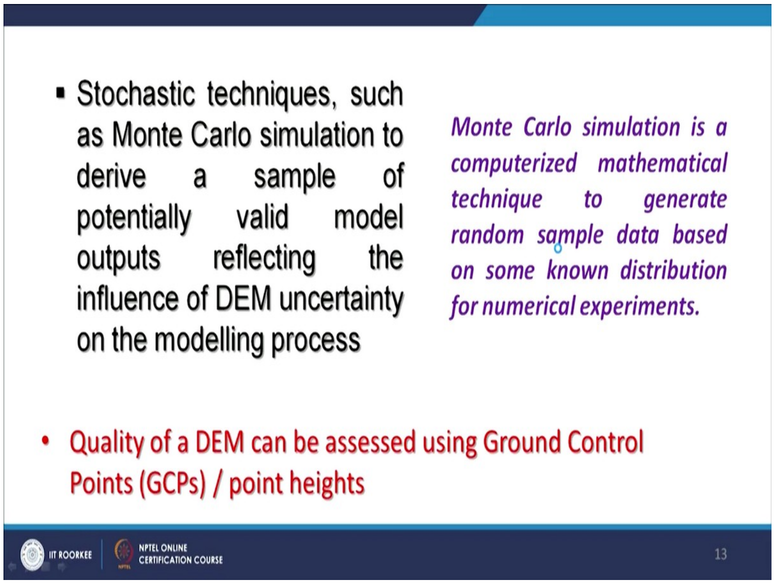
(Refer Slide Time: 25:03)



Then we are having these two publications in which we have extensively describe that how to assess the quality of a DEM. Here is the SRTM DEM characteristics for 2 Himalayan regions. So, one is for very rugged region, one is for less rugged region. And again, there is a comparison of CARTOSAT DEM and other things. So, there also be as we have assessed the quality of a digital elevation model.

Visually, you can do it. I have already told if there are voids, you can check very easily. If there are seams, you can check very easily. So, there are various ways. Now another way is Epsilon bands around control lines and catchment boundaries, like generation of these Epsilon bands are polygon basically or buffer drawn around lines to represent the uncertainty in line position.

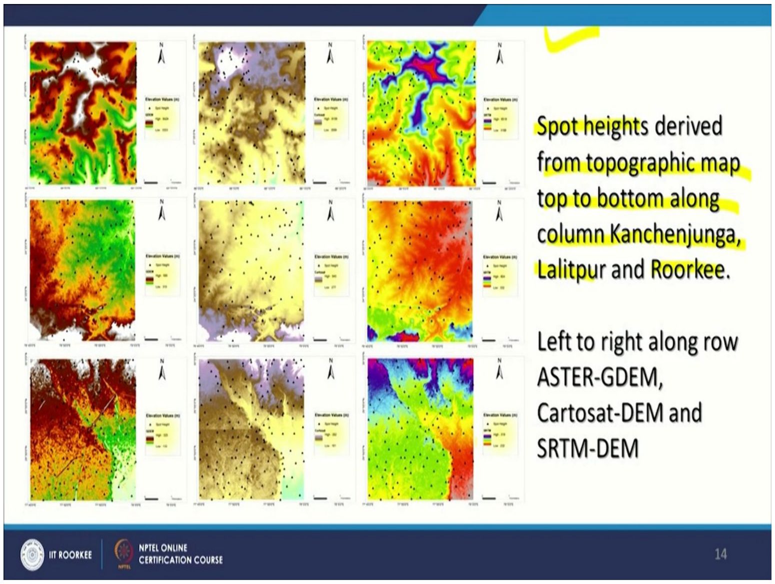
(Refer Slide Time: 26:00)



And there might be some other techniques like Monte Carlo based on stochastic technique which basically simulation to drive a sample of potentially valid model outputs reflecting the influence of uncertainty on the modelling process. So, this Monte Carlo Simulation is a basically computerized mathematical techniques, very common to generate random samples data based on some unknown distribution for numerical experiments.

Now random generators are there or Random selectors are there in standard GIS software like in ArcGIS which you can employ and can instruct the software that I want to select 10% point. It will select. Now you extract those 10% points, keep as a separate file later on to check and then finally with 100% point, you can generate a surface. So, quality of a DEM can be assessed using ground control points or point height. This is what we have discussed.

(Refer Slide Time: 27:12)



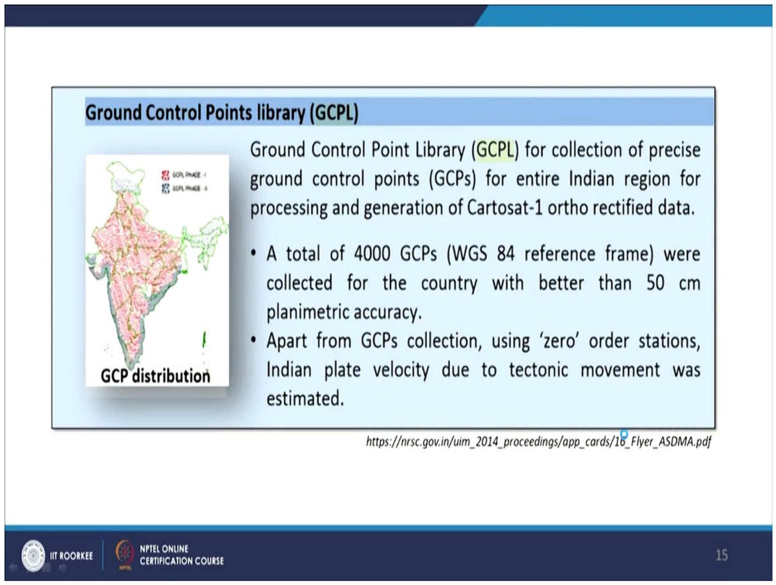
Further in this one, there are various areas which we covered in this particular example. And digital elevation models, we are checked using ground control points or spot Heights. Digitized or brought from these survey toposheets and assuming that they are providing you know ground verified elevation values. So, once we employed for three different types of highly rugged Terrain.

And then you know moderately rugged terrain and on plain area, you can definitely find out that which digital elevation model is good. Though you do not have to do extensive experiments. What you can do in order to optimise this, just use different digital elevation model of your area of interest and employ these ground heights, spot height or if you are having access to that GCP library and then check that which elevation model is closer or having less variations against those tests' points.

And whichever the digital elevation models which is showing the minimum variation is more reliable in that sense. So, this spot heights which are available on survey toposheets which we have derived top to bottom along with so Kanchenjunga area; the top one then this is a less undulated area; the middle part of Bundelkhand region. And the lower one is flat part of Indo-Gangetic plain, that is around Roorkee.

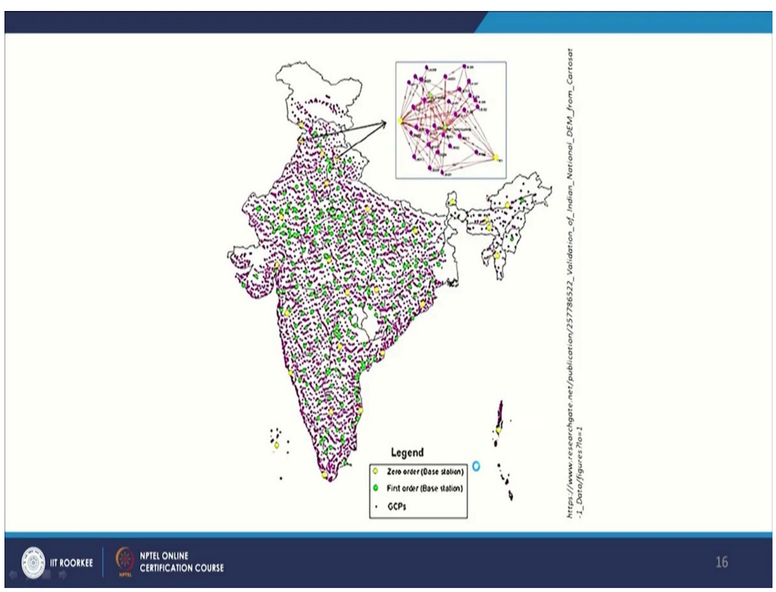
And this is what we found that there are three digital elevation models we employed. So, ASTER DEM was on the left side and then CARTOSAT DEM in the middle and then SRTM; all are having 30-meter spatially resolution.

(Refer Slide Time: 29:24)



Now, this is the library which I was mentioning about the ground control points library; GCP. This is not available to public but some details are available about this library.

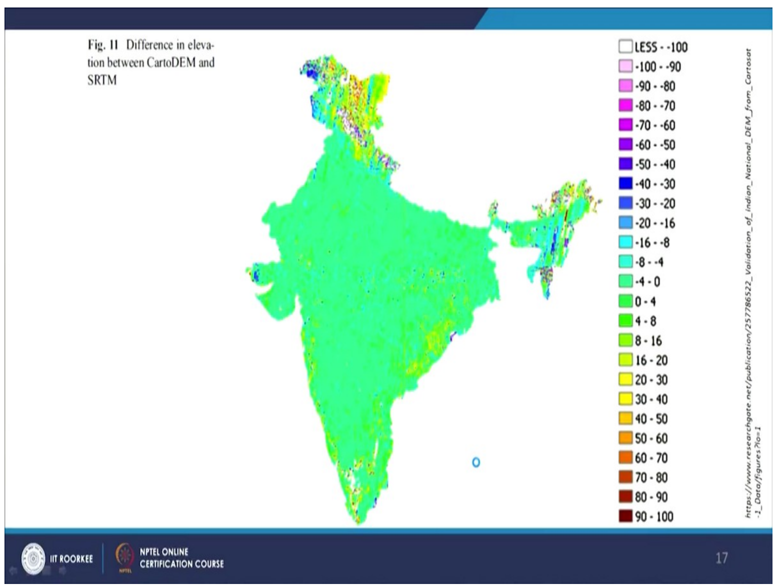
(Refer Slide Time: 29:49)



However, as I have already mentioned that these 30,000-ground control point which are spread all over the India except for some parts of J and K, have been used to validate or to find out the errors in CARTOSAT DEM which has been generated by the same organization NRSA. Of course, it would have been much better if some independent people would have checked rather than the same organization which generated digital elevation model and ground control points are checking.

But none the less, we believe that they have done the justice to the checking and with whatever the errors were there, they must have got rid of those to the maximum extent and that is why you go for Bhuvan portal, you would find that there are different versions. So first they had version 1 then version 2 and currently having version 3. So, assuming that version 3 is more accurate than previously version 1.

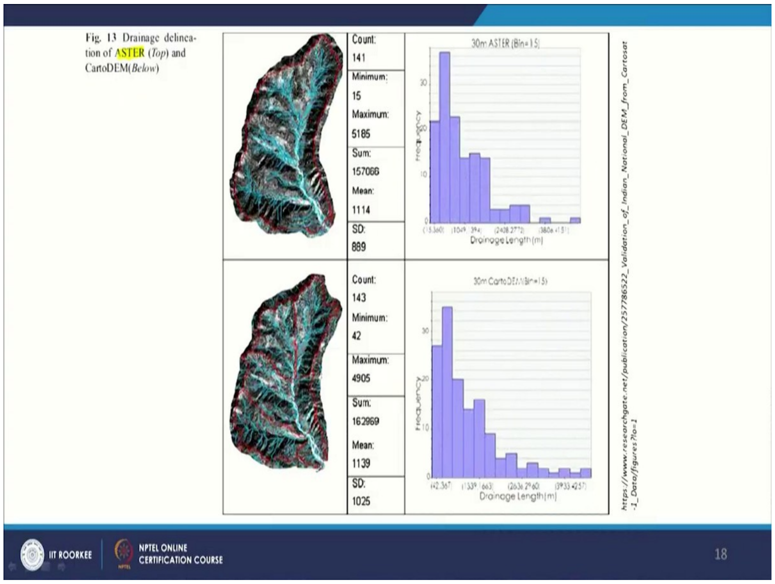
(Refer Slide Time: 31:00)



Likewise, one can you know really find out that the errors. This is what they said when they have used the CARTOSAT in DEM and see the difference in elevation, if I talked about the CARTOSAT DEM generated by NRSA. So obviously in hilly regions, the error is much larger. But in flat areas, as you can see that most of these colours are falling in this category and so wherever undulations are there, remember one point which I discussed about shadow regions.

So, in hilly regions, you might have a problem. So same in the Northeast India or in the Western Himalayan region including Uttarakhand, the error part is much more same in the Kutch region where you are also having some undulated terrain. Also, on the Eastern Ghats and Western Ghats region also and some in the central India, Narmada shown areas. So, those areas had the large errors compared to flat areas in that particular model CARTOSAT DEM and validated or checked using those GCP's.

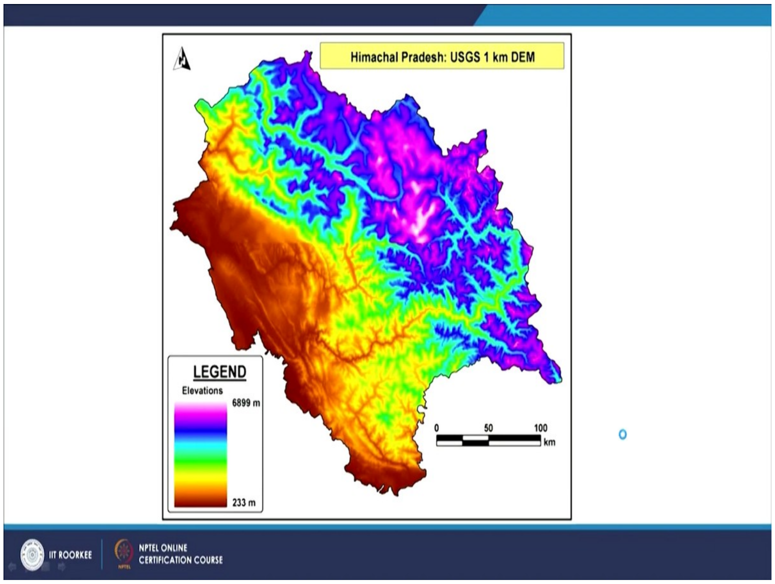
(Refer Slide Time: 32:16)



There are some other checks based on you know using the ASTER DEM and CARTOSAT DEM and driving the drainage network that means through surface hydrologic modelling. And again, you can see that how things are varying and these one like when they derived you know, counts difference are there because sources of 2 digital elevation models are different. Only one thing which is common between CARTOSAT DEM and ASTER DEM is because both are based on stereo pairs and employing photo grammatic technique.

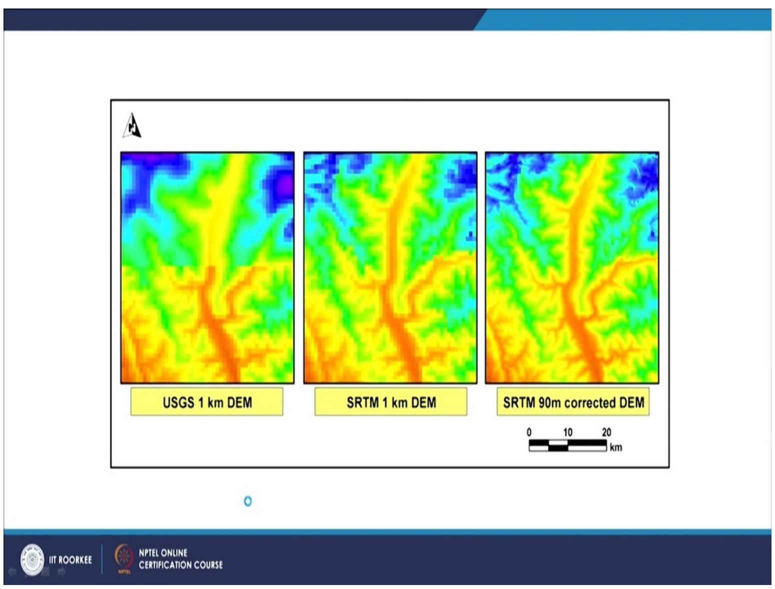
So therefore, this comparison in that sense is more valid because the same technique has been used, though sensors are different but the resolution wise, they are same. So, the variation which you can see in different drainage length is definitely much-2 different in both these models.

(Refer Slide Time: 33:23)



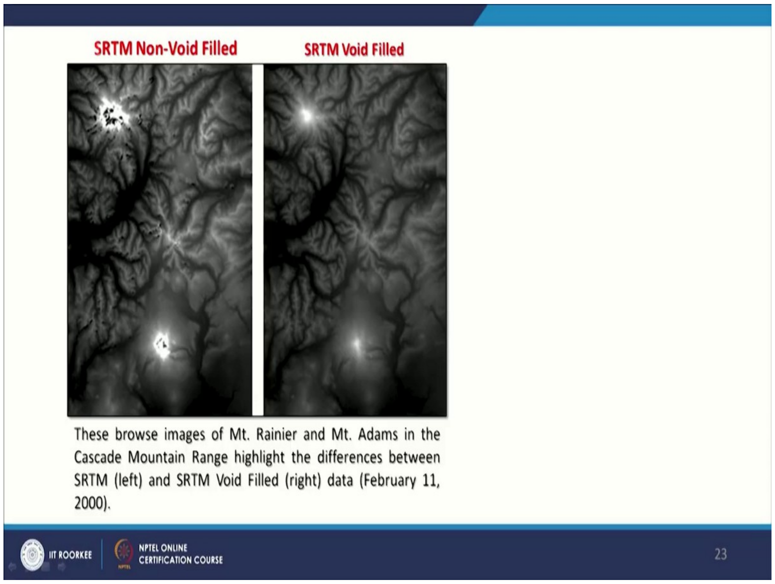
This part, I have already discussed how to assess the visuals so very quickly I will go that this is a digital elevation model generated using a survey toposheets 1-kilometer resolution. This is the digital elevation model at 90-meter resolution. But voids as you can see there is marked as no data value and then you can make void free as well.

(Refer Slide Time: 33:49)



However visually when you meet, you may find the seam or you may find some areas which are not as sharp within a digital elevation model because those have been voids and those have been filled with low special resolution DEM below.

(Refer Slide Time: 34:05)



Similarly, like SRTM void; this is Null void field scenario where you are seeing the voids and in high elevated regions, the voids problem was much there in case of Himalaya. This is also in case of Mount Rainier and mount Adams which are near Washington DC USA. And this is what the void free or void filled digital elevation model is.

(Refer Slide Time: 34:37)

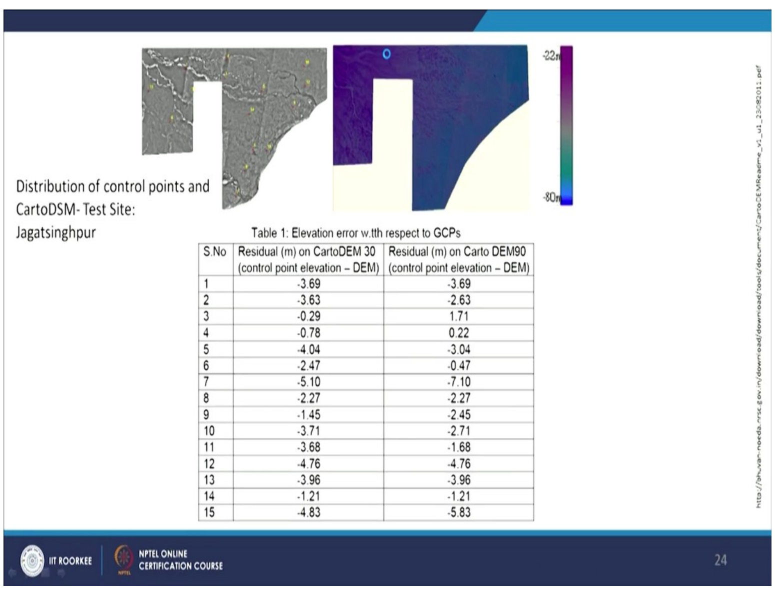
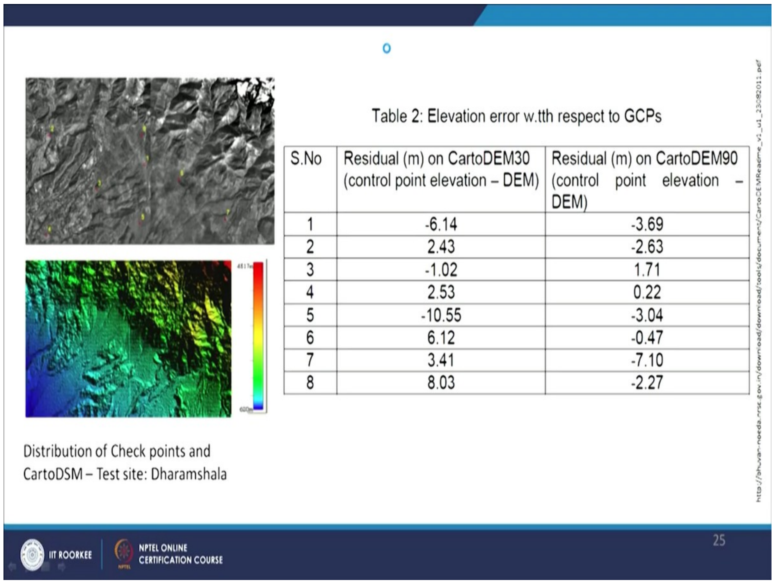


Table 1: Elevation error w.tth respect to GCPs

S.No	Residual (m) on CartoDEM 30 (control point elevation – DEM)	Residual (m) on Carto DEM90 (control point elevation – DEM)
1	-3.69	-3.69
2	-3.63	-2.63
3	-0.29	1.71
4	-0.78	0.22
5	-4.04	-3.04
6	-2.47	-0.47
7	-5.10	-7.10
8	-2.27	-2.27
9	-1.45	-2.45
10	-3.71	-2.71
11	-3.68	-1.68
12	-4.76	-4.76
13	-3.96	-3.96
14	-1.21	-1.21
15	-4.83	-5.83

Similarly, lot of such work is being carried out to assess the quality of different elevation models employing almost the same technique that you have to have something true value or you have to assume some true values which are independently collected and then you compare with the values which have come through a digital elevation model against each cell. So, against those corresponding tests point, you get that value and you asses that one.

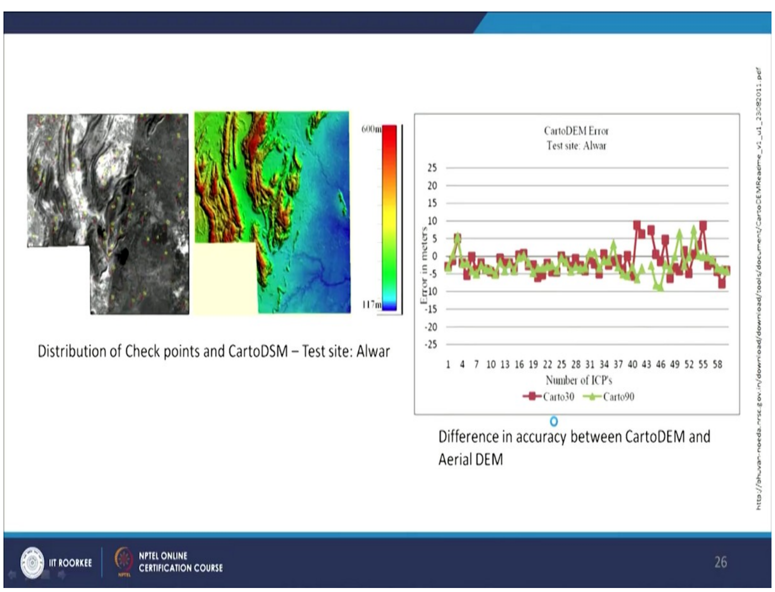
(Refer Slide Time: 35:14)



Similarly, here for SAR interferometry again, the GCP's are used and different residuals have been assessed. So, this is CARTOSAT DEM, this is again CARTOSAT DEM at 13-meter, this is 90-meter. So, what you can find is that at 13 meter in different location, the error is much bigger than in 90-meter. Similarly, also at another location so generally as you go higher and higher in spatial resolution, the error may become larger.

It depends on the terrain basically and the technique which has been employed to generate a digital elevation model.

(Refer Slide Time: 35:51)



Same thing here so CARTOSAT analysis has been done. And as you can see this is not highly rugged like Himalaya but this is Alwar region in Rajasthan. A visible image is there and then you are having a digital elevation model and so checkpoints which are shown here is yellow colours or your test points and when you go through, this is how the number of GCP's or ICP's are there and you can keep checking the errors in meters.

So various ways, one can represent how errors are varying with digital elevation model. And of course, RMSA will give you overall picture but not the spatially varying errors present in a digital elevation model. So finally, we come to end of this discussion. The important point which we have to carry that is we have to differentiate between inherent errors and also errors generated through our processing.

We may not have much control over inherited errors but the knowledge is very much required because our reliability of result or level of confidence will definitely depend on inherent error. Second type of error: the processing error which is in our hands and at every stage, we can keep checking, keep finding the errors if at all and if they are there, correct immediately then move forward. So, with this I end this discussion. Thank you very much.