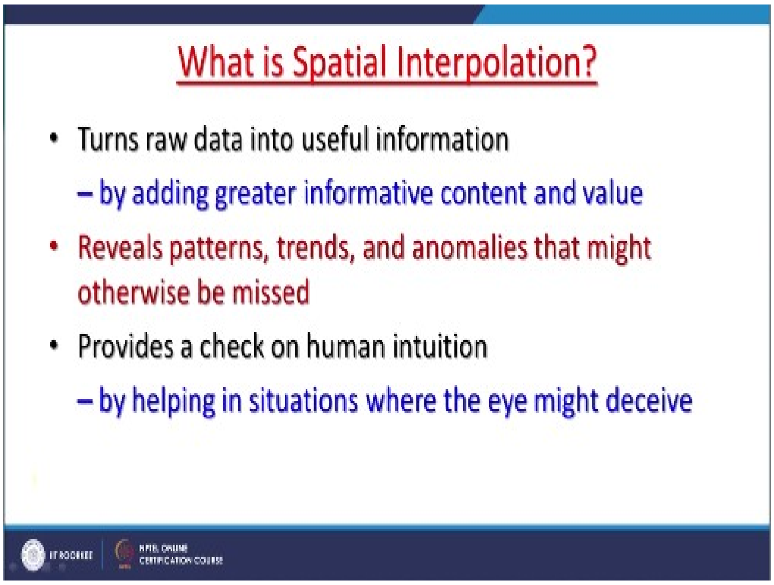
**Geographic Information Systems**
**Prof. A. K. Saraf**
**Department of Earth Science**
**Indian Institute of Technology – Roorkee**

**Lecture – 18**
**Spatial Interpolation Techniques - 01**

Hello everyone! and welcome to new discussion. Today, we are going to have a discussion on spatial interpolation techniques and it is very important to create continuous data from vector data, especially the point data. Even from polylines also, we can create surface or a continuous data. This discussion, I am going to have in 2 parts. So, we will have details in the later part with some examples.

We think about the spatial interpolation. First of all, why we have to do it that we will also discuss.

**(Refer Slide Time: 01:08)**




Before that, let us discuss what is basically spatial interpolation? It is a mathematical technique by which we try to predict something based on our input data. And because we do not have a complete coverage at every desired location where we want the value, say elevation value and therefore we go for interpolation. So, what spatial interpolation does? Basically, it turns raw data into useful information.

So, if you are having simply point data or a spot height, you want to create a surface then this is the best technique which can be used. Earlier when we did not have mathematical you know, models or softwares to do all these interpolations, the surveyors especially for topographic surveying, they used to collect point heights from the field and in the labs, they used to create contour lines.

So, contour lines are also interpolated polylines from input raw data or input point data. But here, what we are doing? We are directly converting into a surface. So, from discrete data which is a point data, we go for the continuous data as a surface or a digital elevation model or any surface. Instead of elevation value, we can have said pH value, we can have some other consultation value and so on.

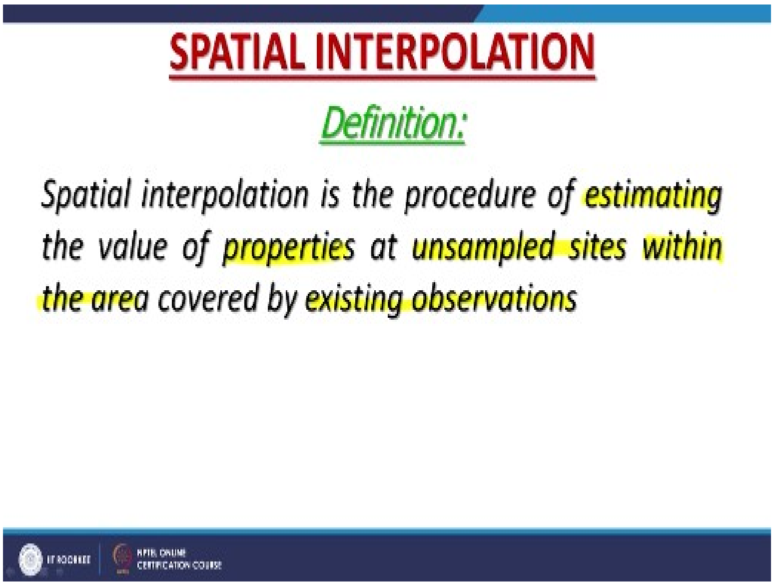
So, by doing this interpolation, we turn data into the useful information and by adding greater informative content and value which we will see, how it becomes very useful through some examples. Once we do the interpolations then we can review a lot of patterns; what are the low areas; what are the high area otherwise from input point data, it is not very easy.

But with this interpolated surface, we get lot of information about patterns, trend, anomalies. There might be lows or highs or anomaly in any data set which might be initially in the point data. After converting through spatial interpolation into surface, we can detect anomalies. We can mask other areas; we can use the anomalous area only. So, lot of usefulness of spatial interpolation is.

Now, it also provides a check on human intuition. Because for every location, we do not have the information or data. So, we try to you know, estimate data by looking whatever the data in between. But if we apply some mathematical model through spatial interpolations then this will have a check on our intuition and will create a surface, hopefully that would be more reliable than just predicting with our eyes only.

So, this helps in that situation where our eyes might deceive. And what exactly definition let us see that one.

**(Refer Slide Time: 04:46)**




This is the more or less the definition you will find in literature that the spatial interpolation is the procedure by estimating. This is the important term here because it is not measuring anything. It is estimating based on input data for unsampled sites and also within the area which we would like to cover by existing observations.

Now here, if we go beyond within the areas which existing; suppose I am having an input layer like point and it has got an extent but, when I do the interpolation between 2 input points, what I do? I exactly do the interpolation but when I do not have any you know, information beyond the existing observation extends then that goes as extrapolation. So, interpolation compared to extrapolation is more reliable.

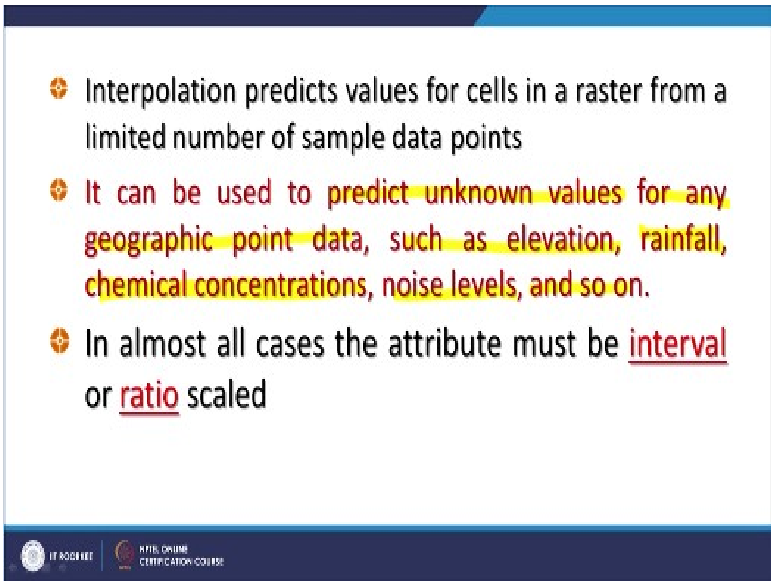
But sometimes we have to do it otherwise we do not have any other information. So, if we can have some information based on some mathematical relationships or estimated value that may be very useful. And basically, this spatial interpolation; the basis of this is on the belief that something might be there like that only. Elevation value at an unknown or unsampled site might be like this.

If one does not believe in that then only solution is left is to go and measure. But it is not possible to go on every site and measure the say, elevation. So, we have to rely on many times on interpolations. However, there are ways by which we can check how good our interpolation is. So that we will be also discussing you know, in second lecture on this spatial interpolation that how to check these things or maybe little later also.

And much more detailed discussion about how to find out which spatial interpolation is good for my sample data. So those kind of detailed discuss, we will have later. For time being, as you can see that spatial interpolation is the procedure and this is mathematical approach, mathematical procedure of estimating the values of properties. Now, these properties can be elevation, can be pH value, can be concentration of pollutants or any other thing.

And we are estimating for unsampled sites of course within the input data extract. Now, interpolation, what basically it does?

**(Refer Slide Time: 07:57)**




It predicts values for cells in raster from a limited number of sample data points that means that when we are having the input data; which is the point data, which is a vector entity, which is a discrete entity and when we create a surface using spatial interpolation, we are creating a raster. So, in that way, it also converts your vector to raster and this raster surface is basically a predictive surface.

And it will predict values whichever the value; whether elevation or other concentration value for each cell in the raster therefore it is a continuous. Whereas point between 2 points, you do not have any observation. You do not have any sampling. So, you are lacking that information but in case of raster; for every cell, you are having some information, maybe about elevation or any other concentration.

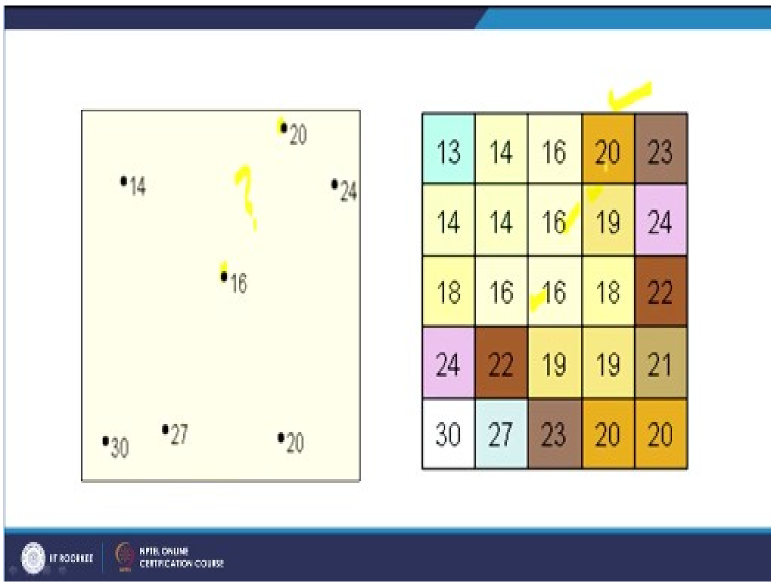
And because of this kind of property of spatial interpolation, basically we predict unknown values for any geographic point data such as elevation, rainfall, chemical concentration, noise label and so on. So, many-2 surfaces though, sometimes loosely we call as a digital elevation model but instead of elevation value, we can have rainfall or we can have some chemical concentration present in the soil or present in the water or maybe water levels, maybe water tables, maybe noise levels and so on.

But in loose terms sometimes, the software will also be called as a digital elevation model but in fact, if value is not elevation for the cell value, then it may be a surface basically. So, in almost all cases, attribute here generally is interval or ratio scale. If you recall that discussion based on the types of attributes, we had various nominal attributes and interval attribute, ratio attributes and cyclic attributes.

So, among them that means that all types of attributes, interpolation cannot be done. For example, on nominal; nominal is just a simple name or a category, now it may not have any sequence or any starting point or any other thing. So therefore, nominal attributes if they have been defined in our database as nominal then those fields cannot be used; those attributes cannot be used for prediction.

Basically, what do you need is numerical value. That numerical value maybe interval value, maybe ratio attribute or maybe a whole number or maybe a real number or floating-point number. So, that is very important. Now, let us take a very simple example.

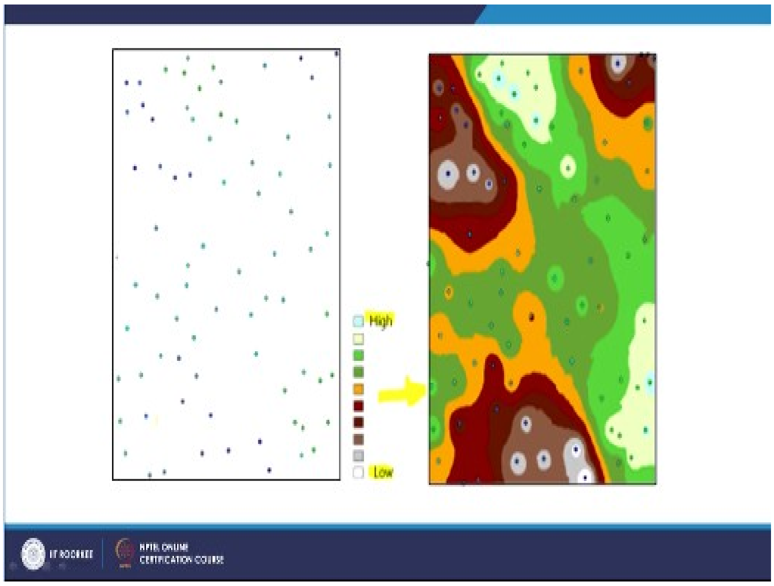
**(Refer Slide Time: 10:58)**




Here you are seeing several points like 7 input points are there. They are distributed in a space within the given extent randomly. Now using these points and implying spatial interpolation, we can create a surface; for each cell now, there is a prediction. For each cell, there is a prediction of, suppose these are elevation values then for each cell, there is a prediction for elevation.

So, instead of like here between this one and this one, we do not have any information in between but if I take these here then between cell 20 or cell which is having 20 and cell which is having value 16, I am having you know, these 2 cells in between. So therefore, our raster or interpolated surface is a continuous surface rather than a discrete or a point or a vector entity.

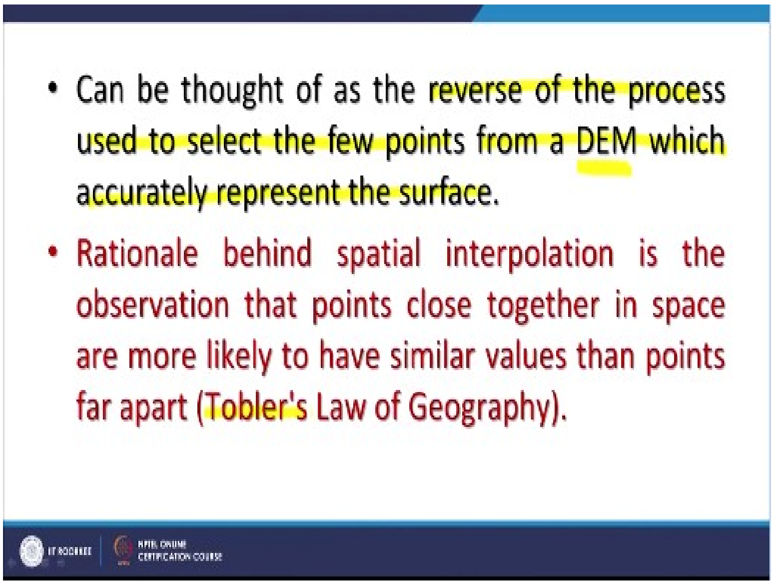
**(Refer Slide Time: 12:03)**




This is another real example that here, we are having the very low values, high values and this is for interpolated surface. Input data is like this. Also, these colors are for this point data as well. And then you can basically create a surface which is a continuous and here what it has been done additionally that point layer has also been superimposed on this one. So that we can correlate things very easily.

But in computer, both will be in the separate layers because one is vector and one is raster. So, they will decide as a separate layer in the system.

**(Refer Slide Time: 12:51)**

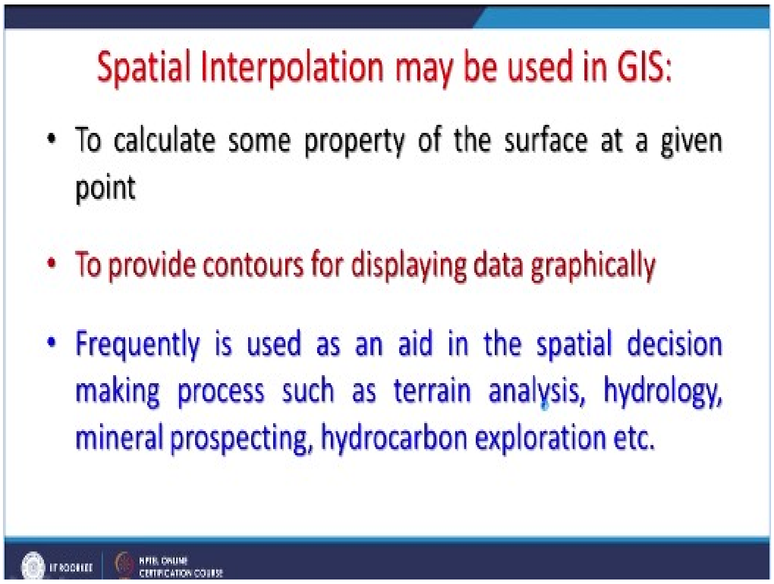



Now, this spatial interpolation can be thought of a reverse process used to select few points from DEM which accurately represent the surface. And DEM as I have already said, in loose term we are using a digital elevation model but that elevation value can be other value. And the rationale behind this interpolation is that observation that points close in a space together or more likely to have similar values.

You recall that earlier also, we have used this Tobler's Law of Geography which considers that whichever is the closest will have the maximum influence in the prediction values. So, this is how that the points close together in a space are more likely to have similar values than points far apart. And this was proposed by Tobler's Law of Geography which is definitely very much implemented into GIS or exist this concept.

Now, spatial interpolation may be used in GIS to calculate some property of the surface at a given point. Now that property may be elevation, concentration, rainfall or whatever.

**(Refer Slide Time: 14:16)**

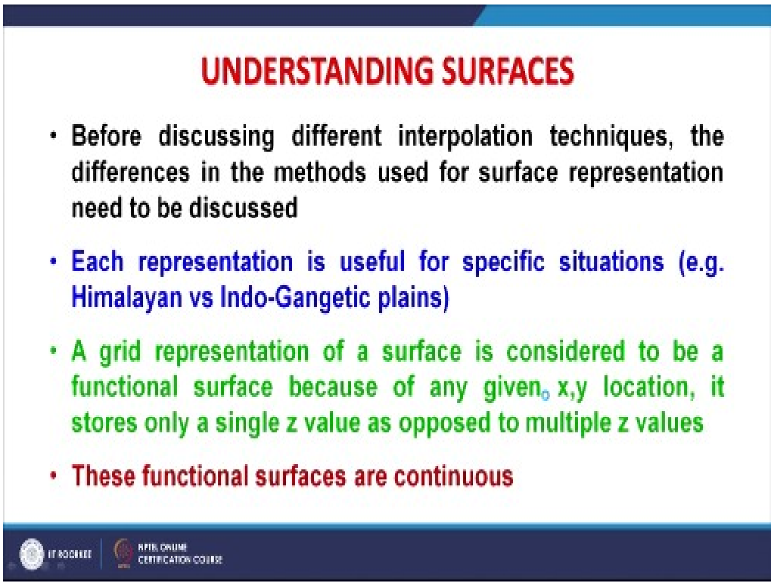



And then to provide contours for displaying data graphical. Many people are very comfortable with contours. Though, they may not be comfortable with point data but they are very comfortable with contours. Contours means isolines. Now contours are also interpolated from point data. So, once we have created a surface which is really end product, we can do a reverse process.

By using that surface, we can also draw contours as per our desired interval. So that is another advantage of having surface from point data that we can draw contours at whatever the interval we want. And because basically, what we are trying to extract from this spatial data or raster surface or interpolate surface about the trends, anomalies and patterns.

And these are used like for example in terrain analysis, maybe in hydrology, mineral prospecting, in hydrocarbon exploration and many-2 fields, surface interpolation can be used and the only requirement is your input data. So, this is an aid in a spatial decision-making process and which may be of various types.

**(Refer Slide Time: 15:43)**




Now, basically what surfaces? So before, we go further in detailing that different interpolation techniques, there are differences in the methods which are used to surface representation so which we would like to discuss first. That each surface representation is useful for a particular or spatial situation. For example, if I am working in a Himalayan condition then the terrain condition is completely different.

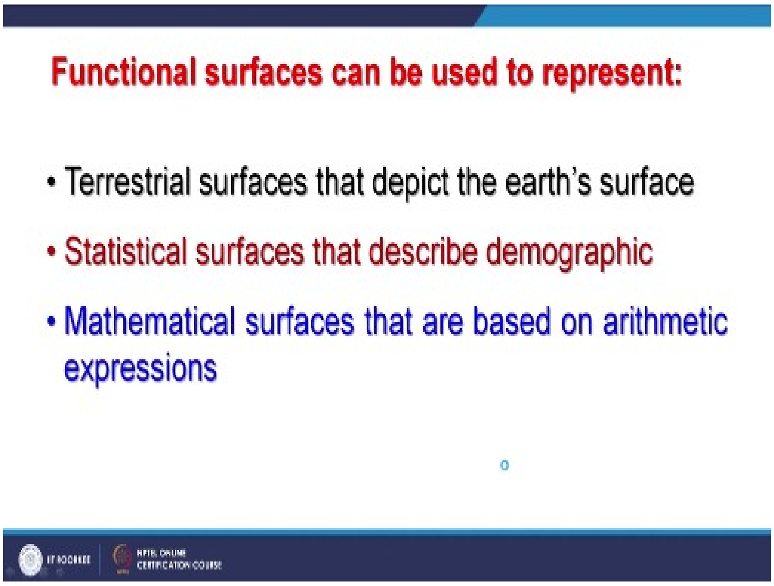
And therefore, my spatial interpolation techniques should also be different compared to if I am working or somebody else is working into Indo-Gangetic plain which is a flat area. So, further if I am having point data for part of Himalayan terrain and I want to create a surface then I cannot use the same interpolation technique as for Indo-Gangetic plain if they are also, I am having input point data.

That means depending on the situation; depending on the terrain conditions, depending on the surface conditions or geologic condition if prior information is available then we can choose most appropriate surface generation technique or spatial interpolation technique. And as you know that this is a raster so, it is a grid representation of a surface which is considered as a functional surface so that because of this, we get not only the x, y location but we also get the z value or maybe height value, maybe concentration value.

And as you know that in case of raster, there is only single attribute that means against each cell, I can have only one attribute and same time we have also discussed that against vector entity like point, polygon or polyline, we can have theoretically n number of attributes. So, limitation here is that against each cell in a grid, we can have maximum only one value but we can create many such layers.

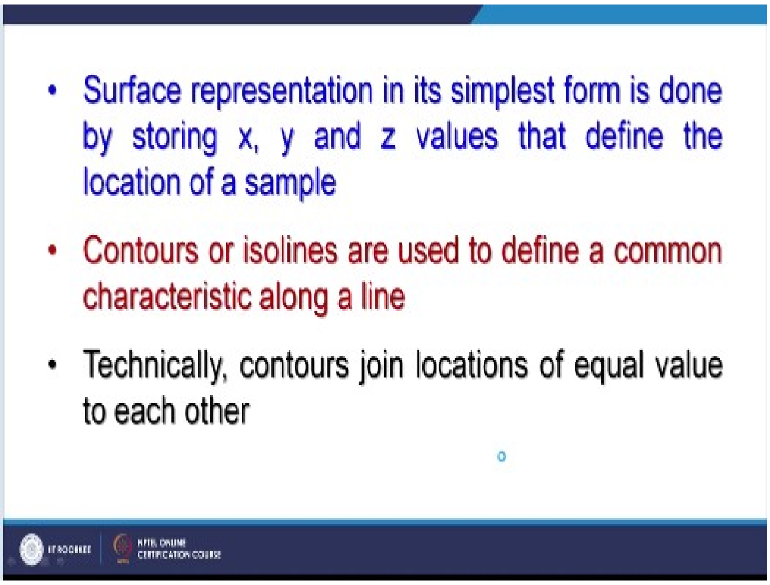
So, one layer is representing elevation. One layer is representing rainfall. One layer is representing concentration of certain elements in the soil. You can create n number of layers and by that you can represent different surfaces having different z values. So, this way also. And as you know that these functional surfaces are continuous surface.

**(Refer Slide Time: 18:26)**




Now functional surfaces can be used to represent terrestrial surfaces because in topography, we use lot about identifying or even extracting watershed boundary or drainage network or stream ordering. So, for terrestrial applications, this digital elevation models have got really a lot of use. And of course, the statistical surveys that describes basically demographic. Mathematical surfaces that are based on arithmetic operations.

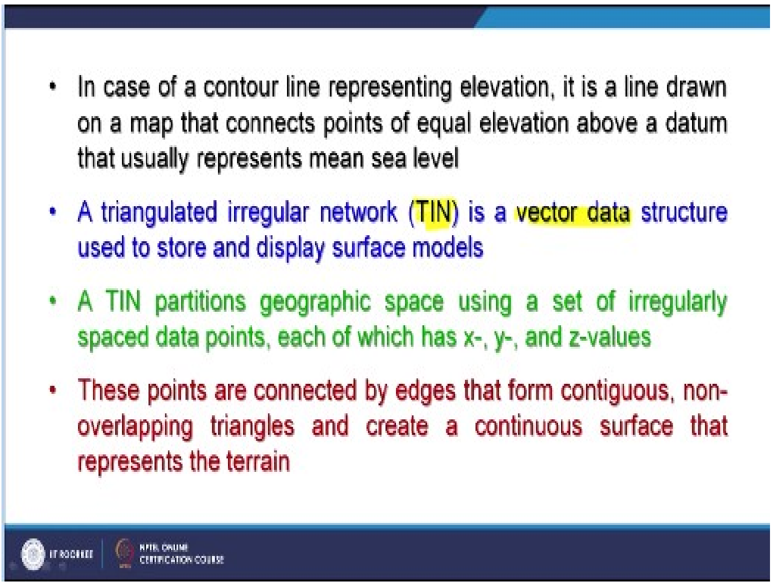
**(Refer Slide Time: 19:02)**




And as you know that raster is quite simple compared to the vector. So, the surface representation having any attribute value is the simplest form which is done by storing x y and z which for every location then you are having a value or property. Further as we have discussed that you can go back one step backward and drive the contour or isolines.

And as you know that these are isoheights and they join you know, equal values to each other if it is a point data.

**(Refer Slide Time: 19:56)**



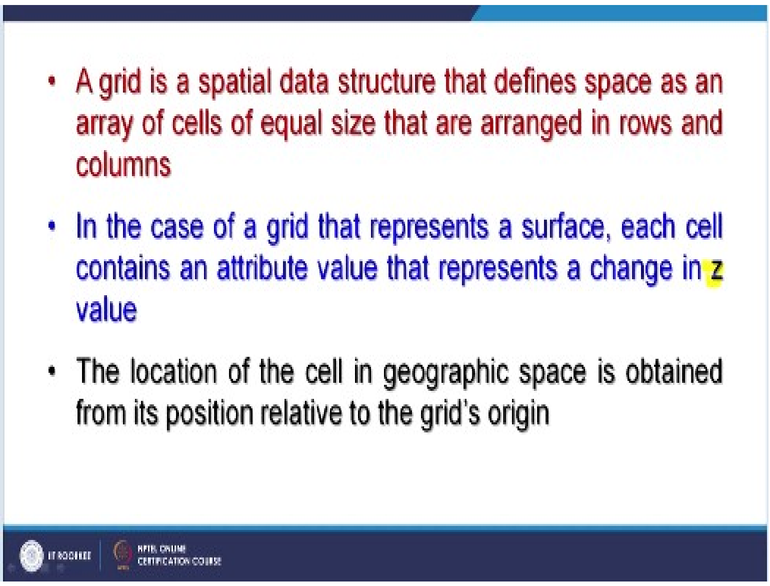

And in case of contour representing elevation which is generally typical digital elevation model, a line drawn on a map that connects points of equal elevation above a datum that is usually represent above mean sea level. But when we represent above mean sea level is only true in case of digital elevation model. But if I am creating a surface for rainfall or for some other concentration of elements in water or soil then there is above mean sea level issue will not come.

Now there is another way which we have also discussed to create surface and that is called TIN; triangulated irregular network. Though it is not really a vector data structure but, in some way, it is quite close. But we considered earlier in our discussion as it completely separate data model. So, we had 3 spatial data models; one is a vector, raster and TIN. So, TIN in many ways is not really close to vector neither raster but the best part is, it represents the surface.

So that is why it is very advantages from representing point of view but while discussing TIN, we also discussed that up to certain extent, TIN may be very good to represent the ruggedness into topography because it is adaptive to changes. Later on, in analysis when we are using other surfaces; which are grid surface, raster surface then using TIN along with that may become very difficult.

So, later on we might have to convert TIN into raster. So, that is a separate issue for discussion.

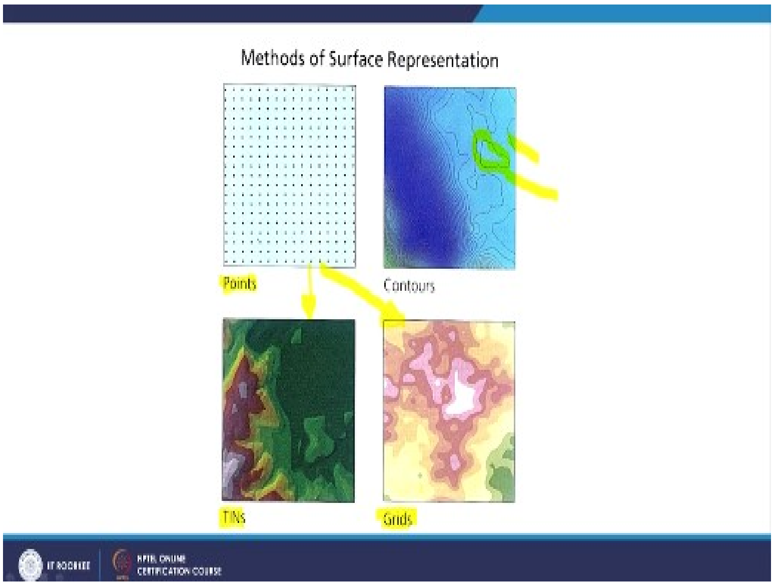
**(Refer Slide Time: 21:48)**




Now, as you know grid is a spatial data structures which defines the space in array cells which we know already and this is a basically surface and the single z value, that is here. And the location of the cell in geographic space is obtained by its position relative to grid's origin.

So, if it is in geographic domain, there will not be an issue. We will get a complete coverage like that.

**(Refer Slide Time: 22:09)**




Now just to compare you know, 3 or 4 ways which we represent the surface. One way is, we are having a systematic points information which is not very easy to get because nobody collects data in systematic manner but anyway theoretically, we are having systematic values against each point. If I consider this is all elevation value so against each point here on the top left, we are having some elevation value.

Using those elevation values, one can create contours and fill in between and you get those things like here, you are seeing. Using the same point data, we can create TIN like here which you are seeing. And again, using the same point data, I can also create grid that is raster. Now as we have just discussed that using grid this grid, we can also create contours that means the contours can be created by 2 methods.

Directly from point to polyline or from surface to polyline; both are possible. Now sometimes we may not be having directly point data for this spatial interpolation. Why I am bringing this discussion here because all these softwares expect that you would provide input data for creation of surface as point data, not polyline data. So that means you have to provide point data and that point data will be then interpolated through different interpolation techniques whichever you choose and a surface will be created.

But if suppose I do not have point data, I am having polyline data or contour data. Can I create a surface from there? Because so far that part we have not discussed that means what I am trying to say, can I create a surface from polyline data using existing softwares? So, in some softwares, what they do in background that first convert polyline into point data.

And then using of course point, you can create surface. But suppose a software which you are using is not capable of doing that then there is a very simple method that you can convert these polylines first into the point. Such tools are always available. And then using point, you can create surface. So, because the software tools you have to use judiciously and very intelligently.

Basic tools are there. Only you have to find the best way how to create a surface. Now though, I have been given examples of various softwares while discussing any such technique or any subject or topic in GIS but the main purpose in this course is to you know teach you or give you the concept of GIS in generic form, not software based. And therefore deliberately, I have been trying to use different softwares to demonstrate certain things.

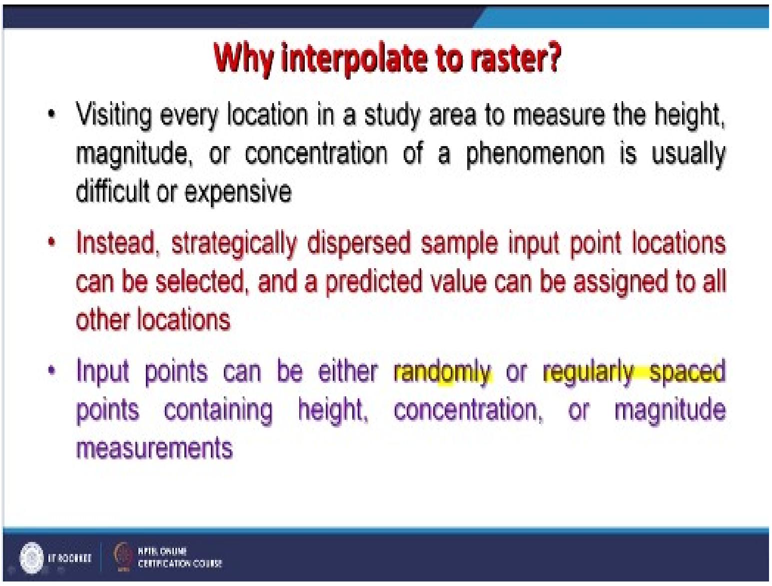
I may resort only to one software. Indirectly that means I am promoting that software which is not really correct. So, what we are discussing is the generic form of GIS and therefore, we must understand the fundamentals about the data itself. So, if we understand like contour lines are interpolated from points that means these have been generated from points and therefore, it becomes much easier to convert polyline to point.

Now, there will be also a question about can I create a points from surface? Of course, you can create very easily like using surface or a grid, we can create polylines. Same way we can create points also. Either systematically for each cell, you can get value in point form or you can have a random distribution of points which can also be generated using a standard GIS softwares.

So, first you can generate say 100 or 500 points within your area and against each point then z value can be picked. So, your point becomes basically a 3D point or point is having x, y and z value. So, reverse process in this is very much possible that means from point to surface and from surface to point, point to polyline and polyline to surface and reverse is also possible. So, all kinds of possibilities are there in surface interpolation.

Now as you know that why we do interpolation because it is not possible to visit each and every site, especially when terrain is very difficult like Himalaya.

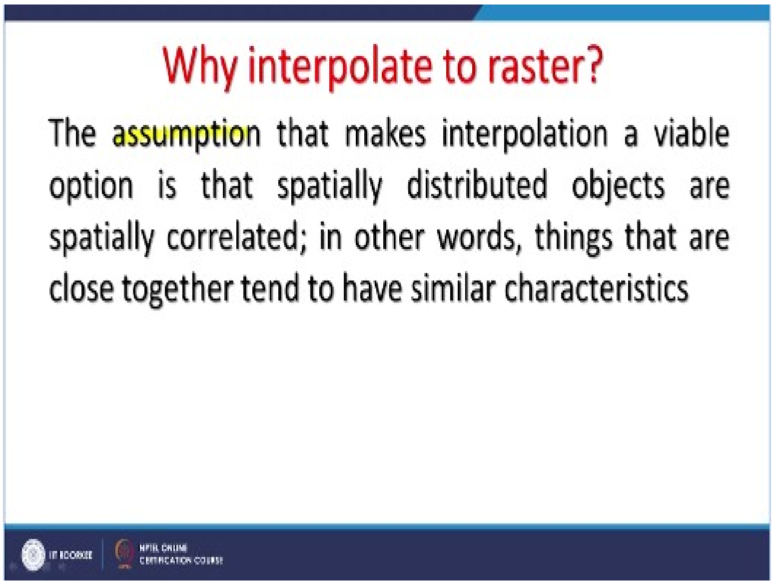
**(Refer Slide Time: 27:39)**




So, wherever it is possible people go and collect the height. This is how our Survey of India to proceed have been created. Or there are many such phenomena where it is not possible to collect information systematically. So, wherever the information is available that is used to create a surface which becomes a systematic information. So, because of this, instead strategically dispersed the sample point locations can be selected and a predicted value can be assigned to all other locations.

Now as I have said that the points may be randomly distributed or regularly spaced. So, in previous example, what we have seen that points were regularly spaced but in many-2 cases and more natural or real projects, we will have the distribution of points in random space. So therefore, still there is no issue; still, we can create surface without any problem.

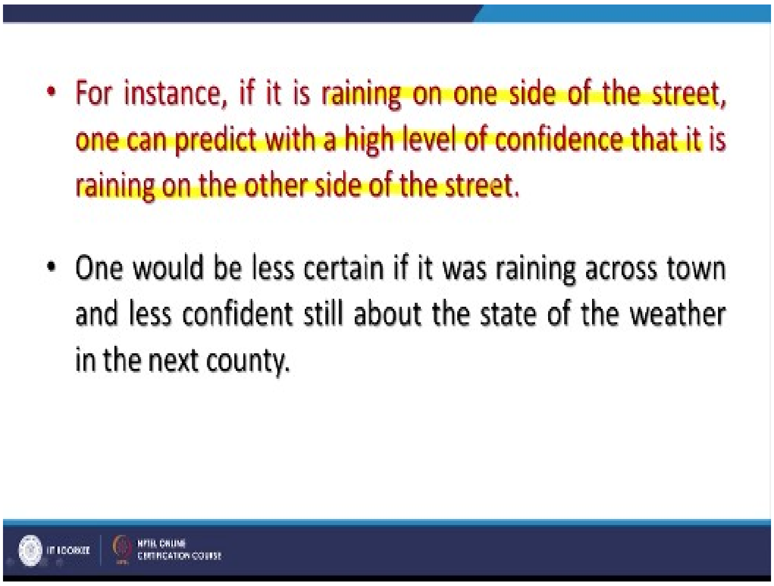
**(Refer Slide Time: 27:39)**




Now why interpolation because as I have said that it is the belief or assumption which makes interpolation a viable option is that a spatially distributed objects in this our case are the points in most of our discussion, are spatially correlated. That means whichever is having closer together, the tabular concept and in other words, the things that are close together tend to have similar characteristics and this is what is exploited in spatial interpolation.

But when we go for extrapolation then beyond a certain extent of the input point data, we do not have any observation and therefore, extrapolation is done. Extrapolation is always less reliable as compared to interpolation. But sometimes we have to resort to that even extrapolation as well. Now there are complications. When we come to the real situation then things start becoming little complicated.

**(Refer Slide Time: 29:52)**

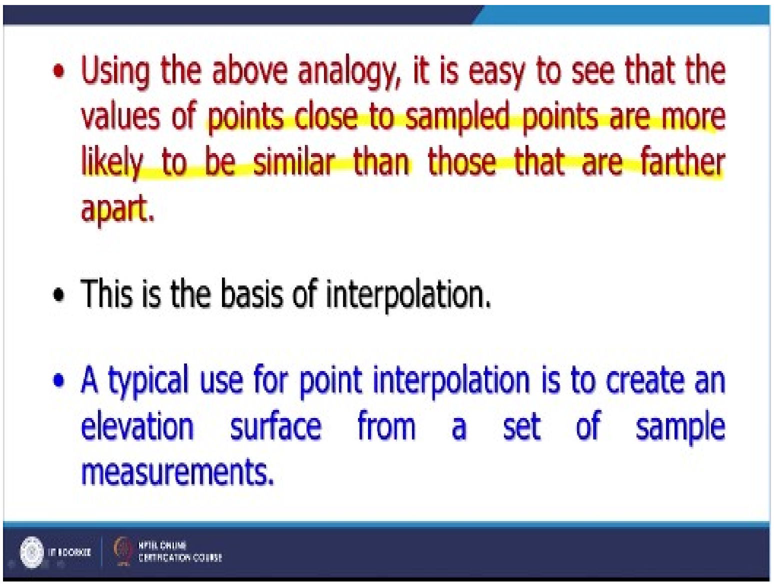


For example, raining on one side of the street; one can very much predicted the other side of the street also will have rain. This is very much possible because it is not much distance between one side to another and with the high level of confidence, one can predict. But it is a little difficult to predict if we predict say, for one town and we cannot say even 10 kilometers beyond whether rainfall or similar weather conditions exist.

For example, suppose rain is occurring in Delhi, it does not mean that rain is also occurring in Gurgaon or in Noida. So even they may have just few kilometers distance but still, it is not possible to predict because the phenomena are like this. So, the important point is coming about here is that what is that phenomena for which we are creating a predicted surface using our interpolation techniques so, it depends on that phenomena

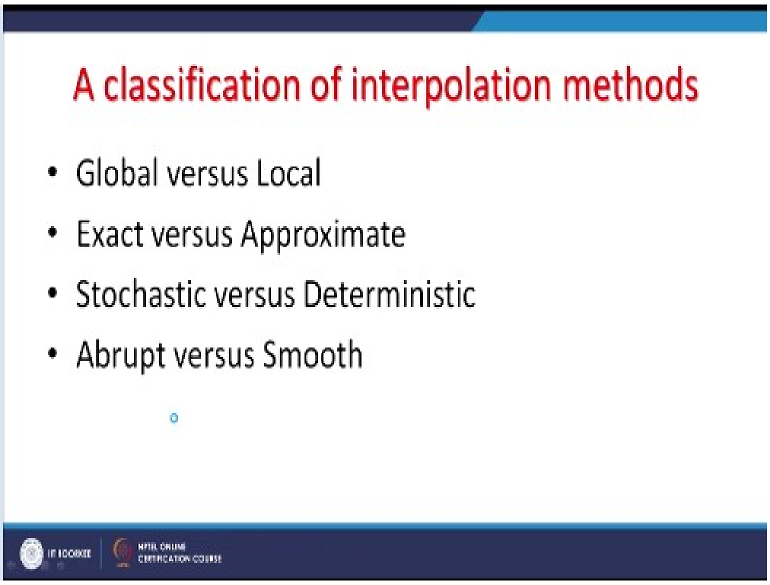
And of course, our accuracy of the surface will also depend on that. So, we have to be very careful. In such case, when our phenomena are so unpredictable like rainfall then we need more observations to have better reliable our interpolated surface but if phenomena are predictable then we can go for you know, simple interpolation techniques and those surfaces which will be created through these techniques may be reliable also.

**(Refer Slide Time: 31:37)**




So that means close sample points are more likely to be similar then those are further apart. Same Tobler concept and this is basically the basis of all interpolation techniques which we will be discussing later. And as you know that typical use of point interpolation is to create an elevation surface. Instead of elevation, any surface basically from set of sample measurements.

**(Refer Slide Time: 32:05)**




Now different types of interpolation techniques exist since these are coming from mathematical domain and mathematicians have developed variety of interpolation techniques and like if I say linear and nonlinear then against linear and nonlinear, there are various variants are also existing and within one interpolation technique for example say Kriging, again variants are there. So those case, we will see later.

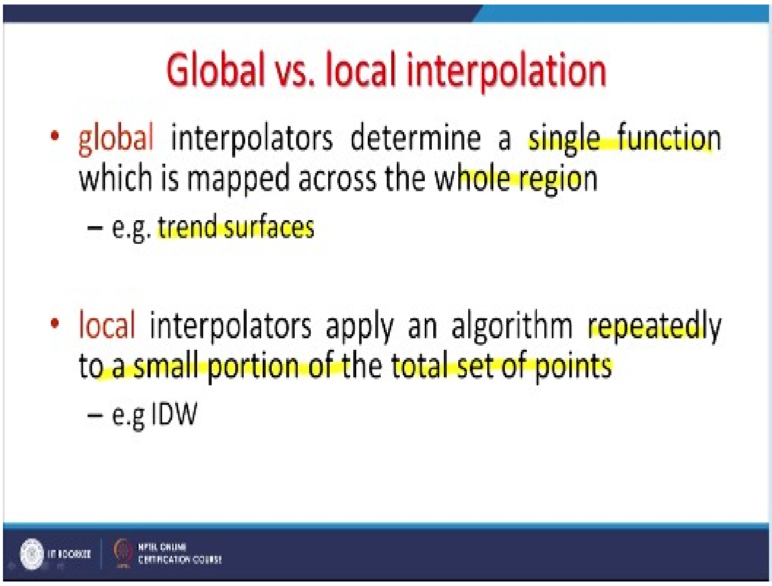
Now, this interpolation techniques can be classified based on like global versus local or maybe exact versus approximate. Of course, interpolate surface is estimated surface; it's an approximate but sometimes we can force interpolation that against my observation, it should keep those values. It should not change those original point values which are getting converted into cell values but the cell values will not change.

So that is why exact versus approximate. And the third type of classification methods of this interpolation methods can be stochastic versus deterministic. Now here, one more type can be abrupt versus smooth. Now a natural condition sometimes because of some natural phenomena, there might be sudden changes in say, in elevation if it is a escarpment like it is very common in the Himalayan terrain.

Or in case of if I am creating a interpolate surface for groundwater conditions then there might be sudden change in lithology, there might be dike or a reef or a fault and that will be considered under the abrupt. So, all surfaces which we are going to create using different interpolation techniques are not going to be this smooth. There might be some conditions were abrupt.

And nowadays, these softwares have been developed to the extent that we can even incorporate this abrupt that means we can bring this extra information as these are called barriers and can create more realistic interpolated surface.

**(Refer Slide Time: 34:37)**

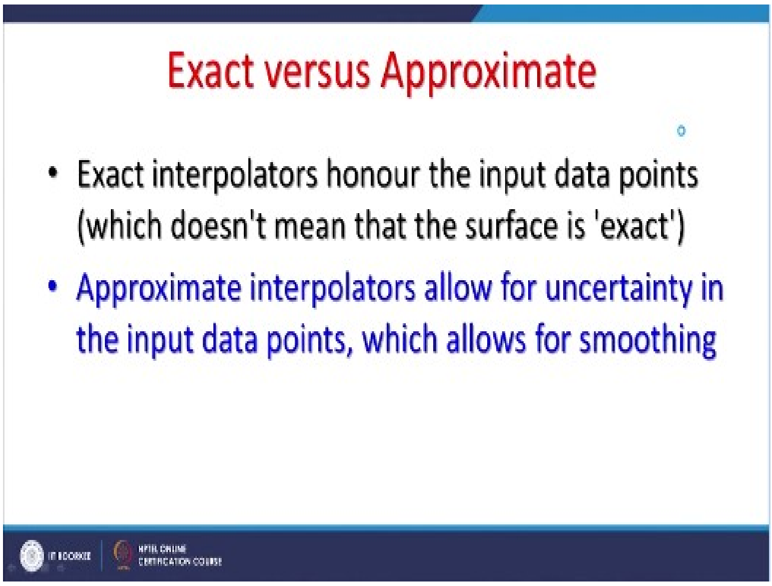



So, let us take first global versus local interpolation. So, global interpolators basically determine a single function which is mapped across the whole region. And best example is trend surfaces which is one type of interpolation technique. Whereas in case of local that algorithm is applied repeatedly to a small portion, not for the whole region and set of points; only selected points it is done.

And the best example is IDW which stands for inverse distance weightage interpolation method, very popular interpolation technique which we use. Because all phenomena may not have that kind of global characteristics and therefore, these trends surfaces may not represent or may not be the close to the reality. Whereas local interpolators or local interpolation techniques may be very close to the reality and therefore, we may resort to a local interpolator rather than global interpolators.

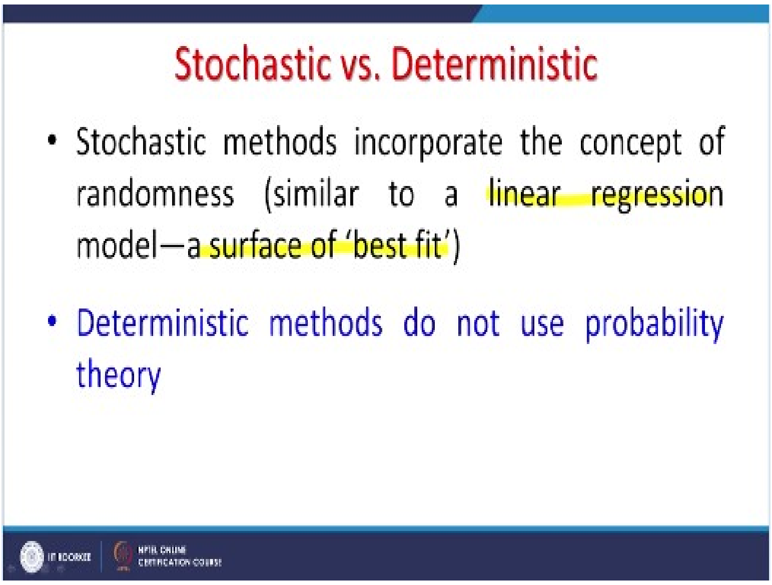
Similarly, exact interpolators versus approximate interpolators; so in exact interpolators as I also mentioned few minutes back that it owners the input data points that means it keeps the exact value against that cell where that point is falling.

**(Refer Slide Time: 36:04)**



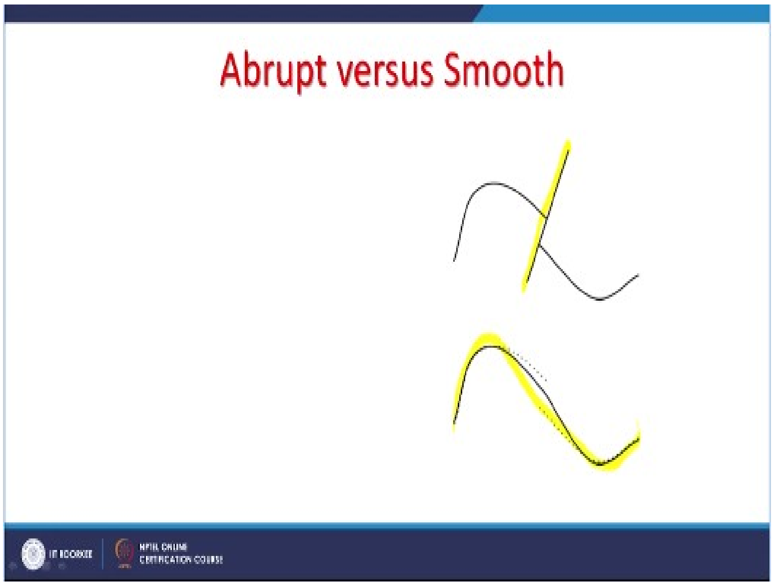

So, it does not mean that basically entire surface is exit but only against that point it is exist. And we compare exact versus approximate then approximate interpolators allow for uncertainty in the input data points because such situations may come where uncertainty may be very high and which basically allows for a smoothing; smooth in surface.

**(Refer Slide Time: 36:38)**




Now third type of classification of interpolation technique can be done based on stochastic versus deterministic. Now stochastic basically, which incorporates the concept of randomness and this is similar to a linear regression model or we call as a best fit model. Compared to this, we are having another you know versus stochastic, there is another interpolation technique deterministic method; do not use probability theory so very far from probability.

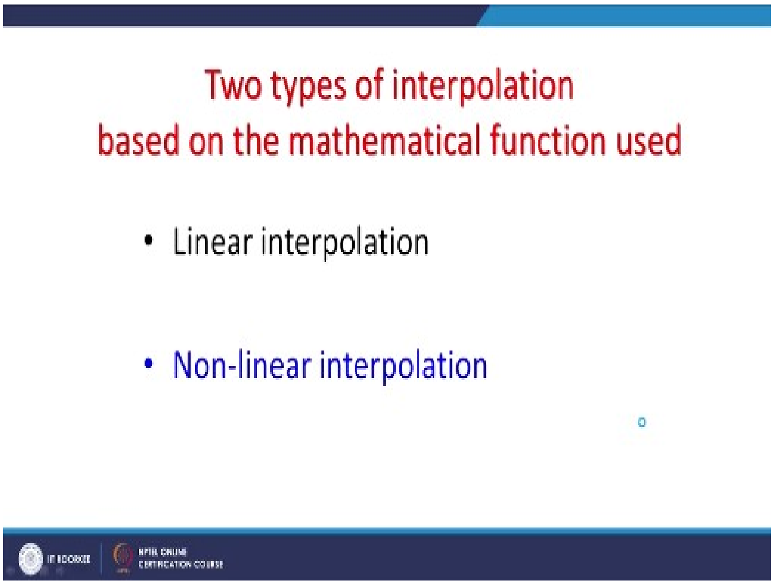
**(Refer Slide Time: 37:18)**




Now, one more we will discuss then we will see that which one to adopt for which condition. So abrupt versus smooth we look, this is the example that if there is a barrier in natural conditions, there may be a stream, there may be a canal and I am working from groundwater point of view then this is very important part. There may be a dike, geological dike or maybe quartz reef or maybe a geological fault.

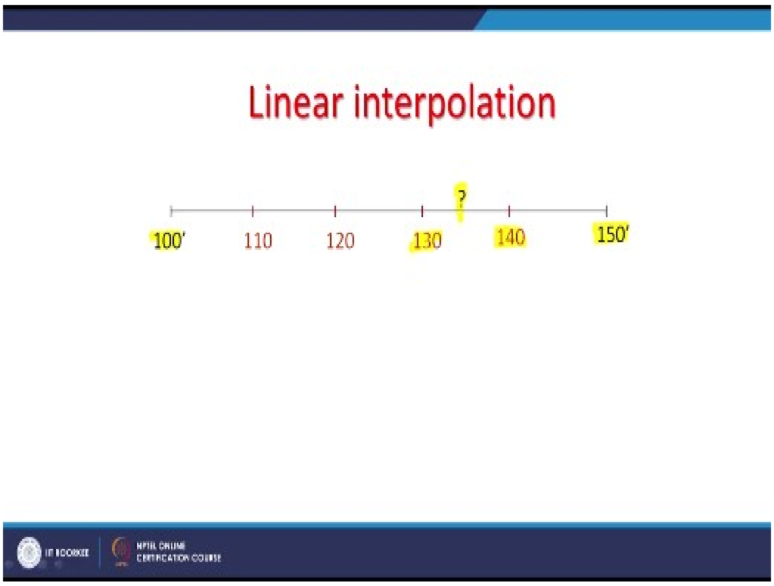
And therefore, these abrupt changes will occur in my surface instead of a smooth surface as you are seeing because all-natural phenomena do not have their representation in a smooth fashion; there might be abrupt, so abrupt interpolators allow for barriers which we can use in interpolations whereas smooth interpolators produce a smooth surface. Now based on mathematical functions, we can again classify these interpolation techniques into 2 major categories.

**(Refer Slide Time: 38:30)**




One is linear interpolation and of course, another one is nonlinear interpolation. Now in literature, you may find people have explained these interpolation techniques in a various manner or they have classified in various manner but most comfortable I find is the linear versus nonlinear.

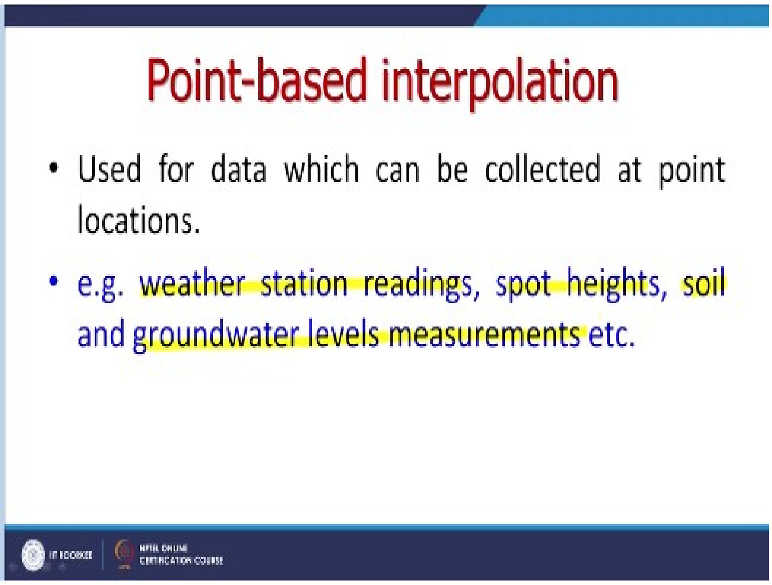
**(Refer Slide Time: 38:52)**




Because we understand these 2 concepts; linear and nonlinear so it is very easy to classify our interpolation techniques in these 2 categories. If I take the linear example and a very simple example, I am taking that I am having say 100 feet here, 150 feet here and there are markings. So, exactly between 130 and 140 feet if I want to know the value is this point then it is very easy to predict because based on this linear distribution of values, I can predict that it should be 135.

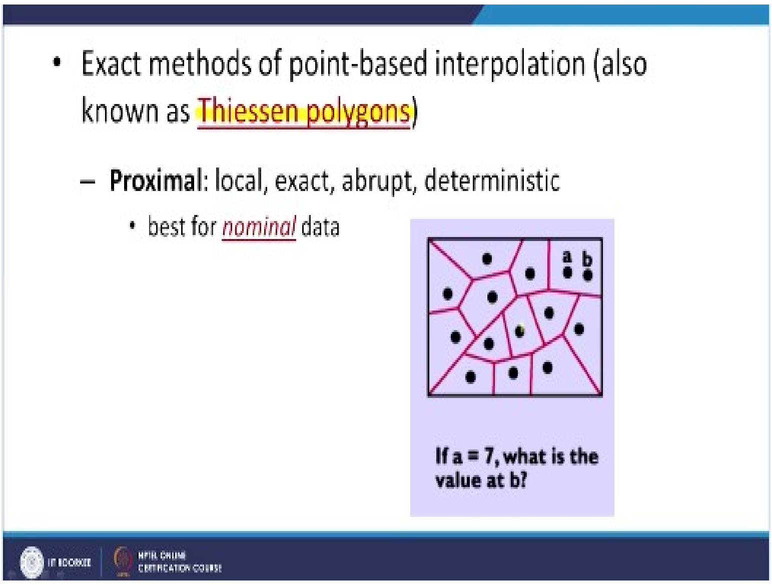
But only if things are having linear relationship. Of course, your interpolate surface is a 2D surface but here we are for simplicity, we are discussing line only. But the same concept exists in 2D also. So that means the prediction in case of linear interpolation become much easier. So, this surface changes that is in linear fashion and can use simple mathematical functions to generate a linear surface?

**(Refer Slide Time: 40:17)**




Now this point based, of course, generally the input is always point data so that is why we are discussing now point based interpolation which are used data, which can be collected at point locations. And for example, may be for rainfall or temperature and any other things, the data can come from weather stations, spot heights, sail properties, maybe certain elements or maybe concentration certain elements or pH value or also in groundwater. So, there are various places from where that data can come.

**(Refer Slide Time: 40:54)**




Now when I go for exact method the most popular when we did not have the GIS software, it was very common in geo-hydrologist or hydrologist to go for Thiessen polygons and Thiessen polygon that means here these are the observations which are shown as black dots. And so, keeping the same distance all around, a polygon used to be drawn like that. So, basically this point is representing the centroid of each point.

And this kind of interpolation techniques; Thiessen by the mathematicians so it is called Thiessen polygon. But it is a proximal that means it is a local, exact, abrupt and deterministic. Abrupt why? Because here you are having very sharp boundary against each polygon. And if you are having nominal data, generally if nominal data in form of some numeric value, then you can use Thiessen polygon. On other datasets also but nowadays it is not common at all.

After this, in the second part we will be going to various other you know interpolation techniques and their variants as well. So, for the time being, Thank you very much.