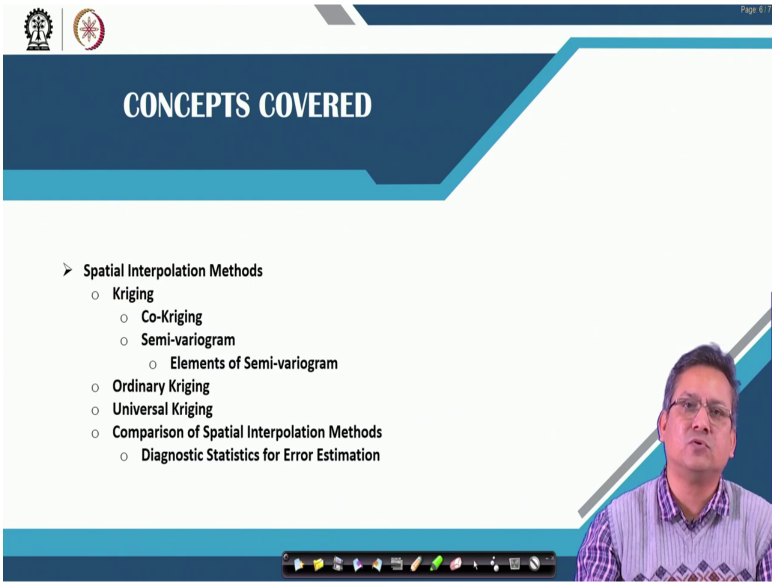
Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture – 13
Spatial Interpolation (Contd.)

Welcome back dear students. So, we are in the module 3 and we are in lecture 13. Today, we are going to talk about Spatial Interpolation. We had talked about different spatial interpolation techniques in the last lecture.

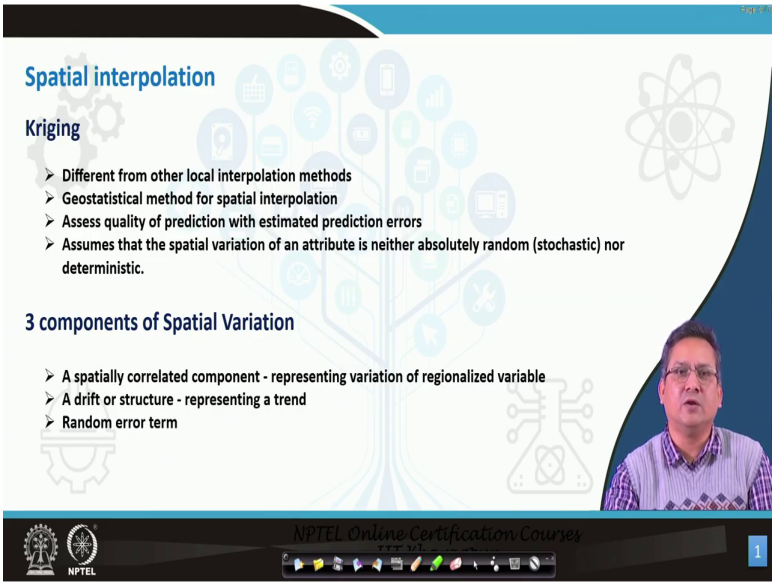
(Refer Slide Time: 00:33)



And we shall continue looking into the different other methods of spatial interpolation as a continuation to the earlier lecture. So, the concepts that we are going to cover today are spatial interpolation methods in which we are going to talk about Kriging.

So, we shall look into the process of Co-Kriging, Semi-variogram, what are the elements of difference semi variogram. Then we shall look into Ordinary Kriging and Universal Kriging. We shall see how the spatial interpolation methods, we have been talking about different spatial interpolation methods. So, how we can compare them; how we can compare the outputs to find out, which is the best spatial interpolation method and which one you should use? So, to do that we run diagnostic statistics for error estimation between the interpolated values and the actual values.

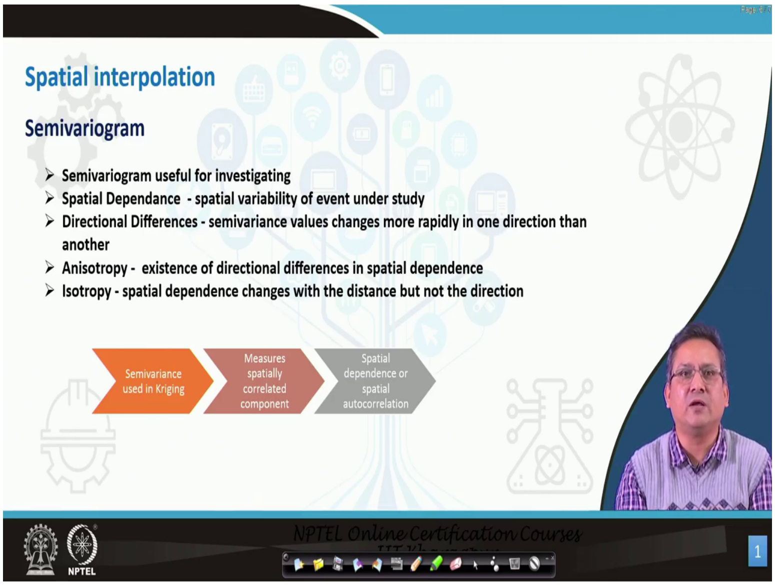
(Refer Slide Time: 01:37)



So, we have diagnostic statistics for error estimation. Now, kriging; this method of spatial interpolation is different from other local interpolation methods. It is a geo statistical method and it assesses the quality of prediction with estimated prediction errors. Now, the assumption of kriging is that the spatial variation of attribute is neither randomly distributed, it is neither stochastic neither it is deterministic. So, I mean in a way kriging is different from the other extrapolation method, interpolation methods that we had discussed earlier.

Now, there are three components of spatial variation. The first one is the specially correlated component which represents variation of regionalized variables. Then, there is the ah next component which is which represents that trend and is known as a drift or a structure. And finally, there could be a term which is error term, it will pop up when you have error between the interpolated surface and the known values. So, if there is a error then you will have a error term.

(Refer Slide Time: 03:17)



So, we can find out the error in terms of interpolation by looking into this error term, random error term. Now, what is the semi variogram? So, it is very useful for investigating the spatial dependence and this spatial dependence that we are talking about is a measure of the spatial variability of any event. I mean if we are measuring the temperature, I mean the spatial variability of the temperature.

So, I mean whether there is a dependence of distance to the I mean and the impact of different points on the you are calculating the values interpolating the values in unknown location based on the known value points. So, I mean we try to work out the spatial dependence using this semi variogram approach.

In this approach we can also identify the difference in the directions. I mean how the values change across different directions can be found out using a semi variogram approach. So, say suppose if you have an urban area and it has been witnessing changes. So, you can draw transects or of the in along different I mean cardinal directions and you can find out that directional differences in terms of urbanization.

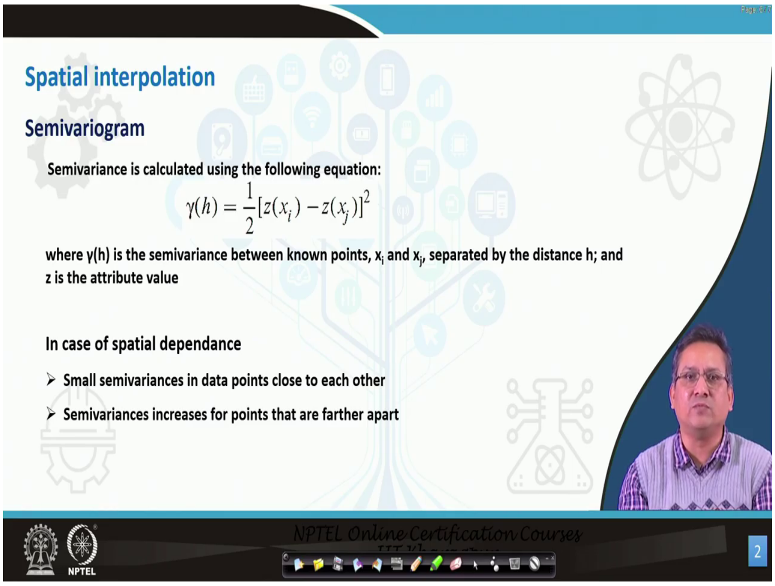
So, another way of doing it is using a semi variogram approach. So, in this case, there could be either spatial dependence or there could be directional dependence. So, I mean the directional dependence would give you the semi variance values and how they change ah in the different directions.

So, in specific directions, it may change very rapidly; in some directions, it may not change altogether. So, we have two terms which is which are anisotropy and isotropy. So, the anisotropy, it measures the existence of directional difference in terms of spatial dependence and isotropy would measure the spatial dependence changes with distance, but not with direction.

So, in case of isotropy, your distribution is not direction dependence. It dependent, but it is dependent on the I mean distance from the given point. So, that is why we turn terminators isotropy. It is a function of the distance, but not a function of the direction.

Now, this semi variance is used this term semi variance or this function semi variance is used while we are doing the kriging and it is a measure of the correlated component, this semi variance, it measures the correlated component and it would also give you an idea regarding the spatial distance or spatial variability of an event or we can also call it as a spatial autocorrelation function.

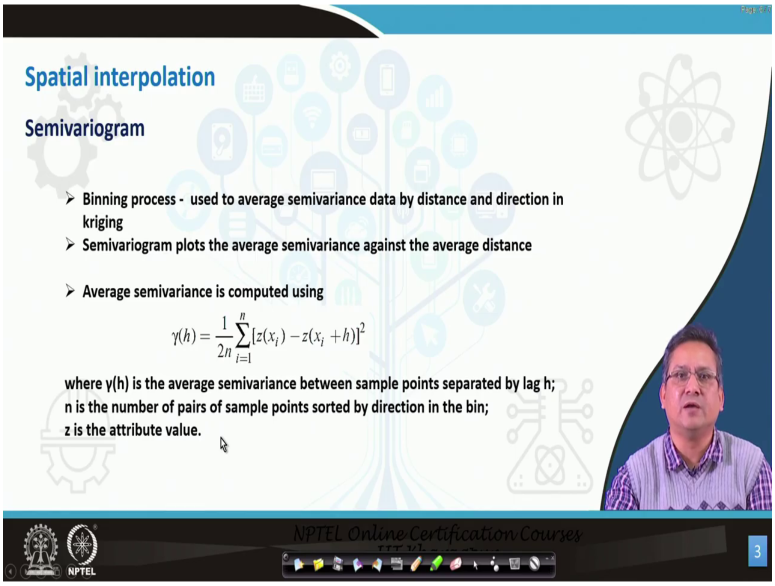
(Refer Slide Time: 06:39)



Now, the semi variance is calculated while we are doing the kriging, the semi variance is calculated using this particular equation; wherein, it is the mean value of z x i minus z x j ah squared, where in your x i and x j are the I mean known points, values of the known points, z values of the known points and y h gives you the semi variance and these two known points are at a distance h. So, I mean y as a gamma as a function of h and your z is the attribute value at x of point x i and x j.

So, I mean when you have spatial dependence, there would be small semi variance for data points that are close together. Now, when the semi I mean this semi variance in case of spatial dependence, the semi variance that you are calculating, it would in start increasing for points which are further away. So, for points which are close together very close to the point where you are trying to interpolate the value, the semi variance would be small; but as you move further away from this point, the semi variance would start increasing.

(Refer Slide Time: 08:11)

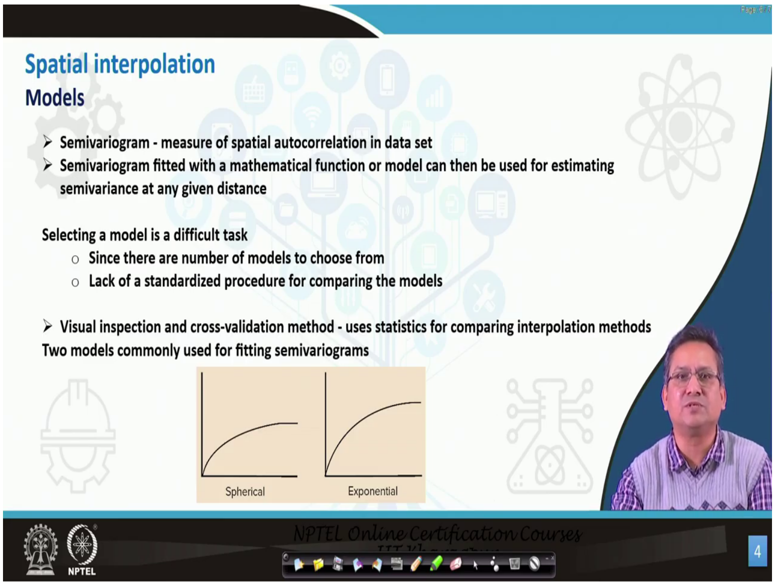


Now, there is a binding process which is used to average the semi variance data as a function of distance and direction and what we do is we create the plot of average semi variance against the average distance. I mean I mean we can we are I mean calculating the semi variogram. So, we can create plots of the average values of semi variance against the distance function average distance.

So, the average semi variance is calculated using this particular function, wherein you can see that earlier function did not use the function h, the distance between these two values; two values of z that is the known values at x i and x j.

But in this function, we are adding this value of h. So, I mean the y h is the average semi variance between the sample points. So, we if we have n sample points, we can find out the summation of that n points and we divide it by 2 n to get the semi average semi variance and these are separated by the distance function or it is also known as lag, which is I mean defined as h. So, we have the n as the number of pairs of sample points sorted by direction and z is the attribute value.

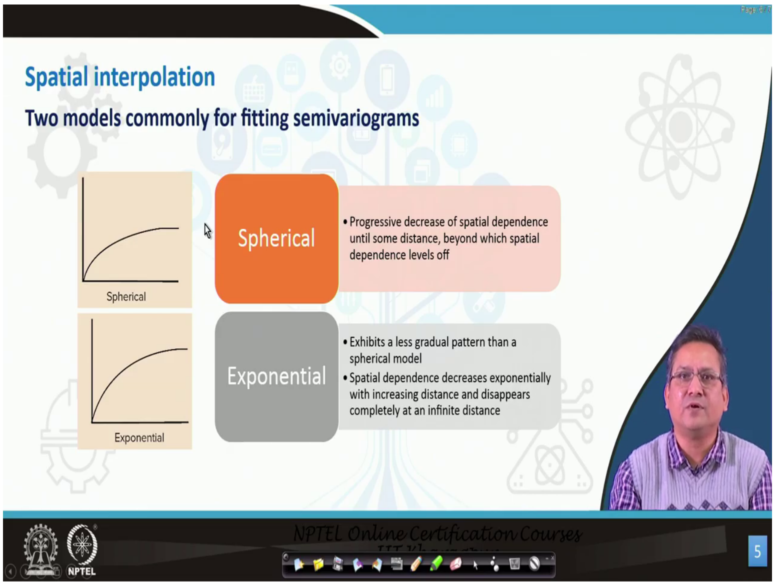
(Refer Slide Time: 09:51)



Now, we have different types of models. As we had said that semi variogram measures the autocorrelation, the spatial autocorrelation in the data set and it is fitted with a mathematical model or function which estimates the semi variance at different distances. But selecting the model is a difficult task as there are lot of models to choose from. I mean if you go into edges there are several models to choose from or if you go into other software's you will see there are different types of models for calculating the semi variogram graph.

So, there is a lack of standardized procedure, when I mean we are trying to select a model to calculate the semi variogram. So, what we do is we do a visual I mean assessment, a visual inspection or we can also do a kind of an validation approach, wherein we can measure the statistics error statistics for comparing the different interpolation methods. So, two commonly used models for fitting semi variograms is the spherical model and the exponential model.

(Refer Slide Time: 11:17)



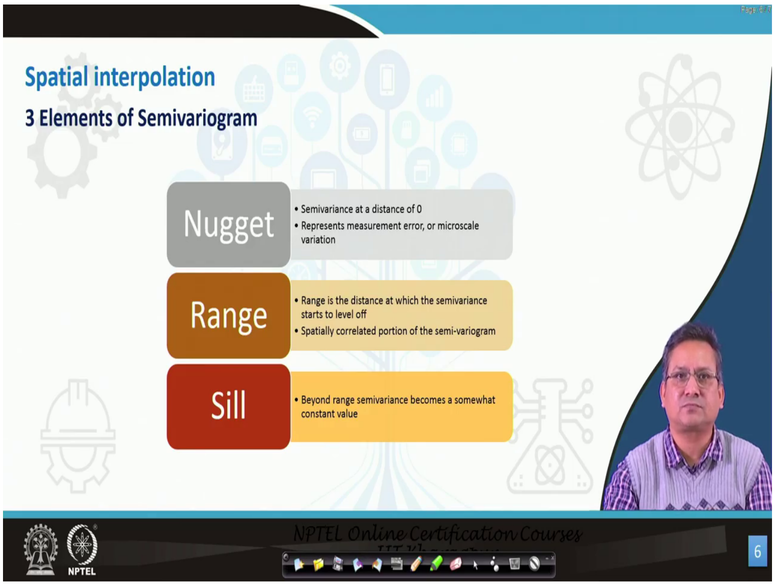
Now, the spherical model, it progressively decrease; decreases I mean there is a decrease of the spatial distance dependence until some distance and beyond which spatial dependence levels are off. So, you can see in the this particular function, after this it tapers off. So, there is no dependence of I mean the I mean spatial dependence of these z values beyond a certain distance, I mean it tapers off.

So, I mean it levels off after certain distance. Where in, your exponential model, it exhibits less gradual pattern. I mean you can see the change is very steep out here, it changes drastically varying the change is pretty gradual in case of a exponential decay function.

So, this exponential function, in this case the spatial dependence, it decreases exponentially. It decreases exponentially with increase in the distance. So, and when you approach a very large distance, we let us say our distance infinity. So, if we approach a very large distance, then this spatial dependence, it disappears completely ok.

So, we can use to find out the spatial dependence I mean we can use this two methods for fitting the semi variogram either the spherical method or exponential method depending on our data and the requirements.

(Refer Slide Time: 13:05)

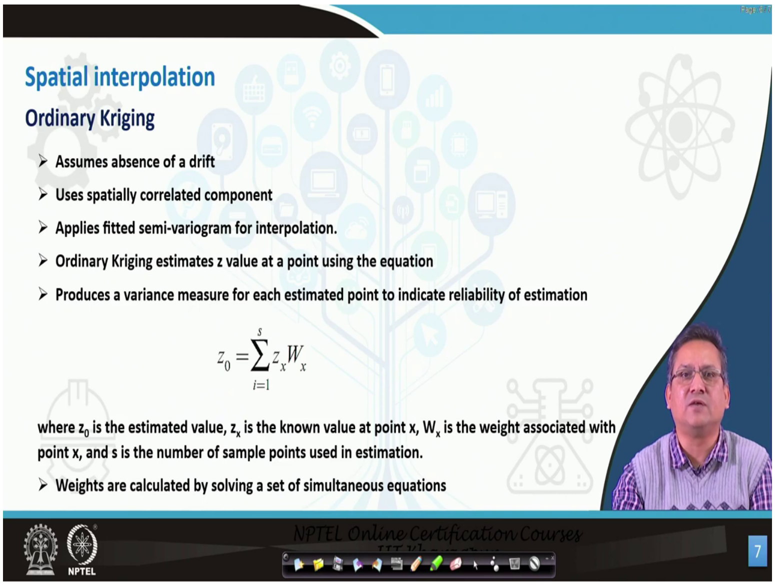


Now, we have three elements of semi variogram. These are Nugget; the second element is Range and the third one is known as Sill. Now, this nugget; it is the semi variance at distance 0, ah at the same place what is the semi variance and it represents it generally would represent a measurement error or it I mean it would also represent the variation at a very small scale at a micro scale, it would represent the variation.

Now, the second one which is the range is the distance at which the semi variance starts to level off. I mean in our earlier slide we had seen the I mean the two different models. So, in that the semi variance starts to level off at a certain distance. So, I mean in the exponential model, we had say seen that it levels of beyond a certain distance. So, it is the range or the distance wherein the semi variance starts to level off. The third one which third element of area semi variogram which is the sill, it is the beyond range semi variance which is constant value.

So, after it tapers off, you will see that nugget is the distance wherein you have some amount of change as a function of distance in terms of z value; then, range basically it would be the point where it would start tapering off and sill is the point where is the range beyond which the semi variance becomes a constant value.

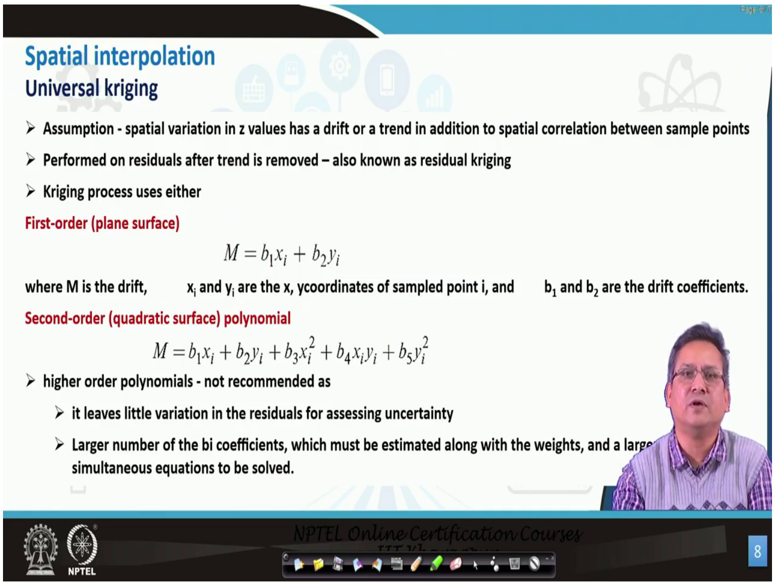
(Refer Slide Time: 15:03)



Now, there are different methods of kriging. The first one is the ordinary kriging. Now, in the ordinary kriging it is based on the assumption there is absence of drift. So, it generally would use the spatially correlated component when you are calculating the ordinary kriging and it applies fitted semi variogram for interpolating the data.

So, it I mean uses a set of equation to calculate the ordinary kriging estimate, the value of z at a given point using this particular equation. It is it produces a variance measure. So, and it also gives a reliability of the estimate. So, I mean we have this function that is z 0 which is summation of the number of sample points from i; starting from i 1 to point the number of samples s and w i W x is the weight associated with the each of these points and z x are the known value say suppose elevation values or temperature values at known points. These are basically the z x values. Now, how do we calculate this weight; this W x needs to be calculated? So, this W x is calculated using a set of simultaneous equation.

(Refer Slide Time: 16:41)



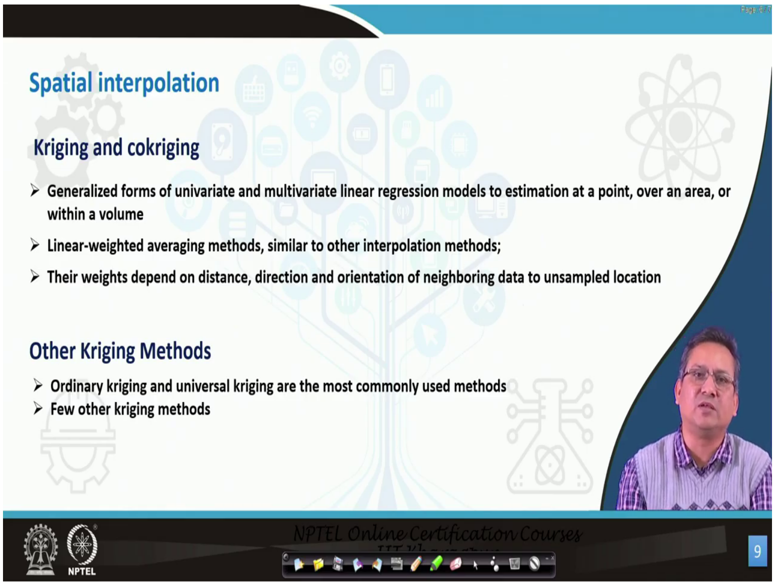
We solve a set of simultaneous equations in I mean this ordinary kriging for calculating the value of z. The next method of kriging is known as the universal kriging. Now, the universal kriging is based on the assumption that your spatial variation in terms of your z values that is your different values that you are projecting, it has a drift. In our ordinary kriging, we assume that there is no drift or there is no trend in the spatial correlation between the sample point. But in universal kriging we assume, I mean it is based on an assumption that there is a drift or a trend in addition to the spatial correlation between the sample points.

So, I mean it is performed on residuals, after we remove the trend and it is thus, it is also known as a residual kriging. So, this kriging the universal kriging uses either a first order equation which is kind of a equation which uses your x i and y i and we need to find out the values of b 1 and b 2 which are the drift coefficients. Your x i and y i are the x and y coordinates of the points that you have sampled and M is the drift.

So, we can calculate the drift using this first order equation or we can also have a second order surface, polynomial surface. So, I mean in this case your higher order polynomials would have more number of coefficients. So, the observed data points has to be more because the unknown coefficients are more. So, I mean this is generally not recommended, as it leaves little variation in terms of the residuals when it comes to assessing the uncertainty.

Now, there are large number of this coefficients, I mean in case of a higher order polynomial. So, it has to be assessed along with the weight values. So, you would need to have lot of simultaneous equations to solve it. So, it becomes computationally very intense.

(Refer Slide Time: 19:17)

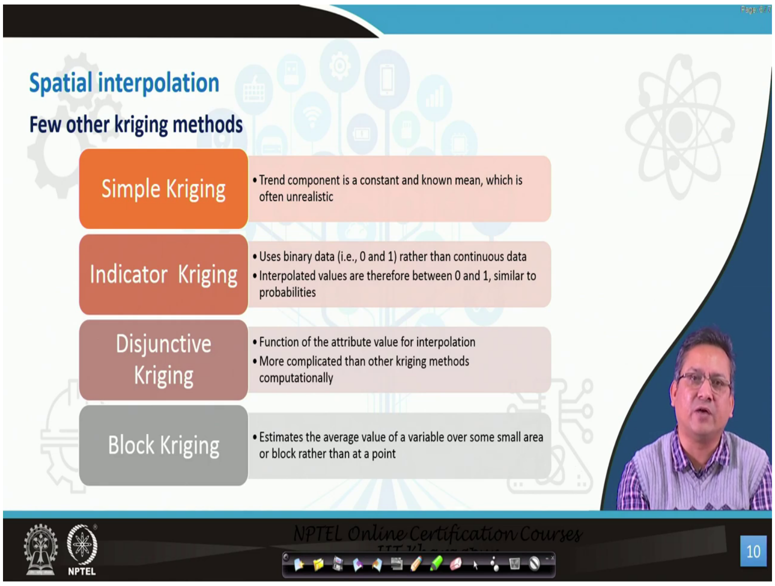


So, that is why we do not generally use the higher order higher order polynomial equations. Now, kriging and co kriging, now when we are extrapolating or interpolating these data values? We are talking about a special dependence or a correlation. So, ah I mean the kriging and co kriging, these are the generalized forms of multivariate or univariate regression models. So, what we are doing is we are estimating a point, the value at a given point, the z value at a given point ah I mean in an area given area or within a given volume.

So, these are linear weighted average methods both this kriging and the co-kriging processes and we had seen the other methods and these methods are similar in terms of outcome or in terms of the outputs that we get out of these processes, they are similar to the other interpolation methods that we had already seen. So, their weights, it depends on the distance and the direction and the orientation of the neighbouring data points and the data values of the to the un sampled location. I mean the data point that is un sampled.

So, these weights would be dependent on the distance as well as the direction or orientation of the neighbouring data points. Now, there are other kriging methods as well. We had talked about the ordinary kriging and the use universal kriging which are the most commonly used methods. So, there are also other kriging methods which are used ah.

(Refer Slide Time: 21:09)



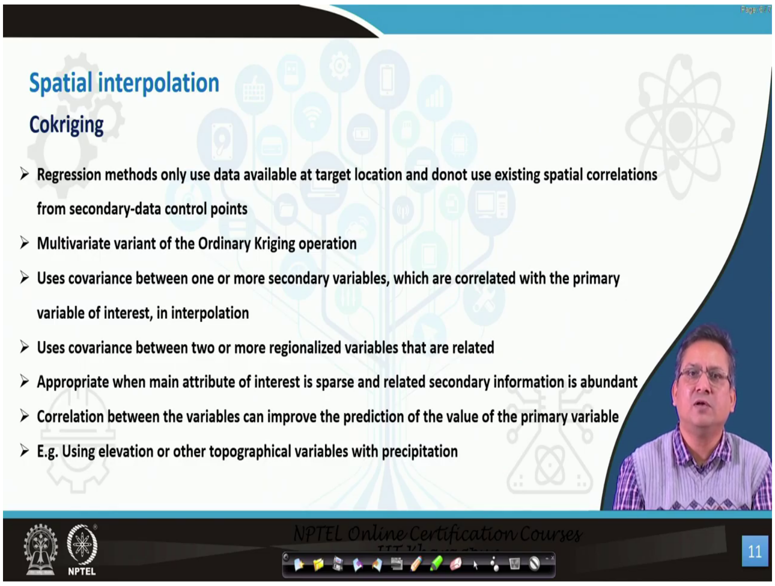
So, these are the simple kriging which gives you the trend component, which is I mean in the simple kriging the trend component is constant and it is it has known mean which is not always I mean realistic. I mean it would be not the idealistic or realistic scenario. There is another method which is known as indicator kriging which uses a binary data; data which are 0’s and 1 instead of using a continuous data.

So, in case of binary data fields, when you have I mean data values which are binary in nature, then we can use indicator kriging. So, it interpolates the value and I mean between zeros and ones. So, it is like a probability distribution surface.

I mean the surface is like I mean probability surface ah, the outcome of these values are in the range between zeros and ones for indicator kriging. So, we have the next method kriging method which is known as disjunctive kriging. So, disjunctive kriging is a function of the attribute function for interpolation. So, it is computationally intense and it is a very complicated method of kriging which is that is that is the reason it is not so popular and not often used.

And the finally, we have the last method which is known as block kriging. So, it estimates the average value of the variable over some small area or a small block. So, than a point than at a given point. So, instead of estimating the value for a point that we have been doing in the other methods, in this case we have a small area for which the point data values are estimated. So, that is why it is known as block kriging.

(Refer Slide Time: 23:23)



Now, we have been talking about a term which is called co-kriging. So, I mean in the earlier methods, wherein we are using the regression methods or models to I mean interpolate the data ah, it uses only data available at target location and it does not use data or the it does not use secondary data points. I mean data control points when we are trying to do a spatial correlation analysis.

So, sometimes it is more I mean it gives you better results if along with the primary data set you also have a secondary data set. Suppose, you are you want to interpolate due into the spatial interpolation of say rainfall. So, if you are in a hilly area, if you also have the digital elevation model or the height information as secondary data control points that will give you a better result.

So, we use this process of co-kriging in that case to do the interpolation. I mean which is a limitation that we had talked about of the regression models or regression methods, wherein which uses only the data available at the target location. So, the spatial correlation with secondary data points are not explored in the regression based model. So, it is a multivariate I mean type of operation, I mean we had seen the ordinary kriging which is a univariate type of a model. So, your co-kriging basically is a multivariate type of a data model, wherein you use multiple sets of input controlled data points.

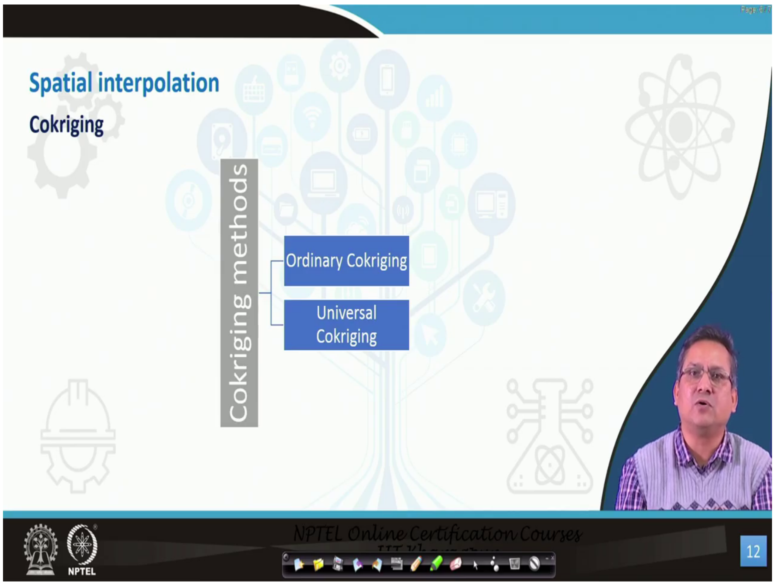
So, it uses the covariance between either you can have one secondary variable or there could be more than one secondary variables. There could be multiple secondary variables. So, they are correlated with the primary variable that you are I mean interpolating. So, these I mean region, I regionalized variables are I mean we try to work out the covariance between these regionalized variables.

So, I mean this could be 2 or even more number of your variables that we are talking about. So, these could be related variables and when we are talking about these related variables, we have to see that there has to be some kind of a dependence between these variables. The other secondary variables should have an impact like we are talking about the elevation the hill and the elevation. So, we know the orography and the different I mean the elevation in the hill has an impact on the rainfall.

So, if we use it as a secondary data, then I mean it would be useful for predicting the rainfall parameters the rainfall data while we are trying to interpolate it. So, we should use such kind of data sets which has some kind of interdependence or some kind of association your ah. We use this co-kriging approach when our major data points, the primary data that we have is sparse. We do not have lot of data points, I mean the density of the data points is not very high, it is sparse. But our secondary data points are abundant say when we have a digital elevation model, we have data point at I mean very closely packed grid data points and we can have rainfall data which are sparse, which could be distributed in a given geographical space.

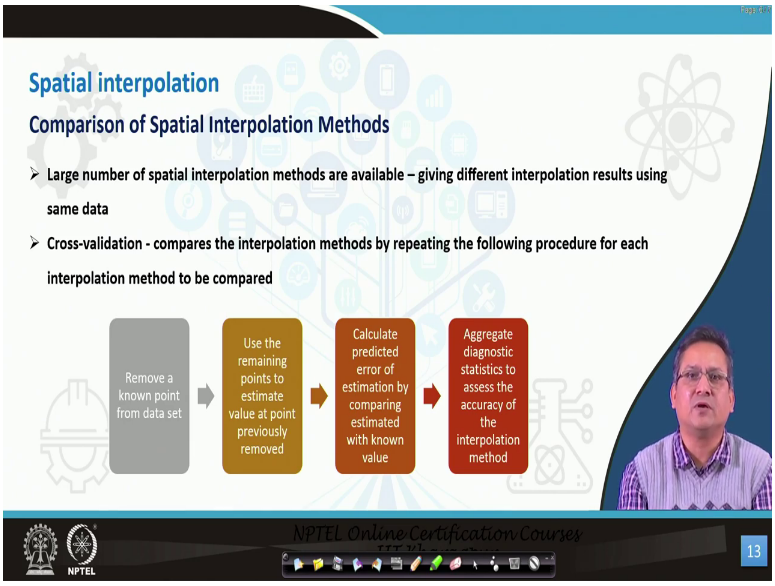
So, in such conditions, we use this cokriging approach ah. So, I mean if we have I mean I have already been talking about this that if your data variables, I mean the variables that you are having that is the primary variable or the secondary variable are correlated if there is a correlation between these variables input variables; then, the prediction values would be the it would be improved in such cases. So, I mean we were talking about the example of precipitation along with the topographical data.

(Refer Slide Time: 28:11)



There are two methods for this co-kriging. So, one is the uni ordinary cokriging and one is the universal kriging.

(Refer Slide Time: 28:21)



Now, when we are doing this interpolation methods, we are not really, I mean sure that which method should be used. So, we need to compare the interpolation methods. So, those of you who are into research or you are looking into projects and you would like to take up this spatial interpolation for interpolating some kind of spatial data values, so you can look into a body of research literature which is available.

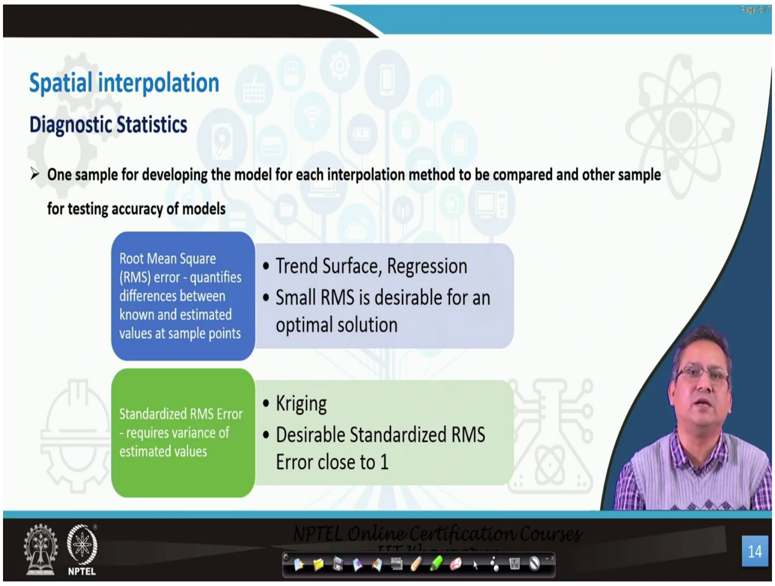
So, you can look into Google scholar and you can see that there are a lot of authors who have compared the different spatial interpolation methods. They have done a comparison of the spatial interpolation methods and they have tried to do an analysis of which method is better.

So, I have seen lot of research papers for I mean interpolation of soil I mean data values for temperature precipitation and other parameters, wherein the researchers they have compared different spatial interpolation methods. The results I mean are different for these different spatial interpolation methods.

So, what we do is we do a cross validation by repeating or following the procedures for each interpolation method that needs to be compared. So, what we do is we remove in this process of cross validation, we remove a known point from the existing data set and whatever remaining points are there, we use those remaining points to estimate the value at that point, wherein we have removed the data.

And then, we try to calculate the error yeah error of estimation by comparing the estimation estimated value in the second step that we had said. In the second step, we calculate the estimated error value between the I mean estimated value with the known value and we can have aggregate diagnostic statistics of the different points to find out the accuracy of the of the interpolation method.

(Refer Slide Time: 30:45)

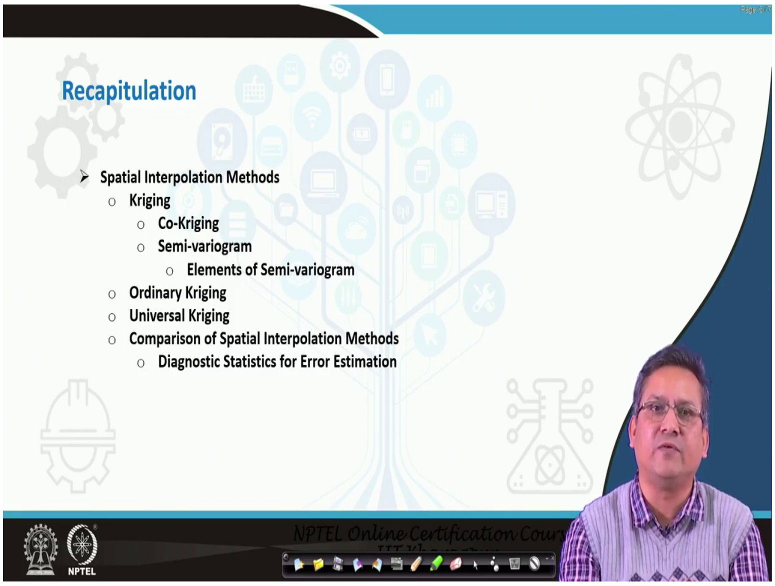


So, we can do this kind of an exercise for different types of interpolation methods. So, I mean, we can have one sample for one set of sample points for developing the interpolation method say suppose you have known data points with the known z values. So, you can say take 50 percent you can create an interpolated surface, using 50 percent of the known data point values; the other 50 percent of the data point values, you can use it for testing the accuracy of the different models. I mean when we are talking about the different spatial interpolation models. So, you can see which data interpolation model is more accurate.

So, we have two measures of accuracy. So, first one is the root mean square error which quantifies the difference between the known and the estimated values at these sample points. So, we can fit a trend surface or we can do a regression. We can find out the I mean RMS and try to meaning minimize it to see that the solution is optimal.

And the second method for diagnostic statistics is the standardized RMS Error ah, which I mean estimates the variance I mean which requires the variance of the estimated values. So, I mean this is applied on in kriging, I mean we can do the standardized RMS as an output of your kriging data, kriging interpolation and we can find out that standardized RMS Error and we should try to see that it is close to 1. That should be the desirable value for standardized RMS Error.

(Refer Slide Time: 32:37)



So, recapitulation of what we had covered today. We had talked about spatial interpolation method. First we had talked about kriging, then we had seen the concept of co-kriging and how we use secondary data points when your primary data points are sparse in nature. So, you can use secondary or multiple data points, I mean multiple layers which are correlated and which have a bearing on the primary data values which we are basically I mean interpolating. So, we can use the concept of co-kriging.

We have seen the concept of semi variogram and how we can calculate the semi variance; how distance and direction has an impact on spatial interpolation and we had seen the elements of semi variogram. Then, we had seen two different concepts of kriging, two different approaches to kriging; first is the ordinary kriging and second is the universal kriging. And then, we had compared, we had seen how we can compare the different spatial interpolation methods. Finally, this comparison could be done by running diagnostic statistics for calculating the error in the estimated values of your I mean between the estimated values and the actual values.

So, we said we can take up a sample of data for creating an estimated value and we can create the we can use the other data points which we have not used for estimation or interpolation for doing a cross check. So, we had looked into two measures of your errors to identify the error.

So, thanks and we shall be looking into different methods of Network Analysis in our subsequent lecture.

Thank you so much.