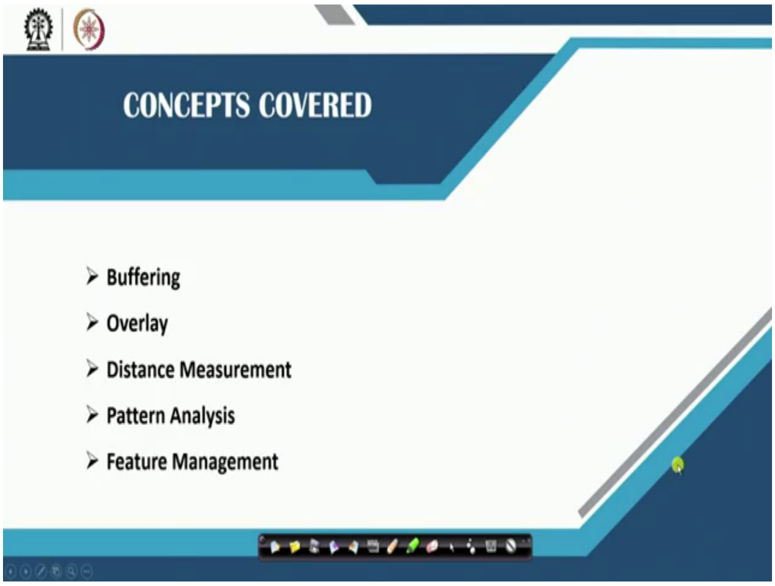
Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Module - 02
GIS functionality and spatial analysis in the urban planning
Lecture – 07
Vector Data Analysis

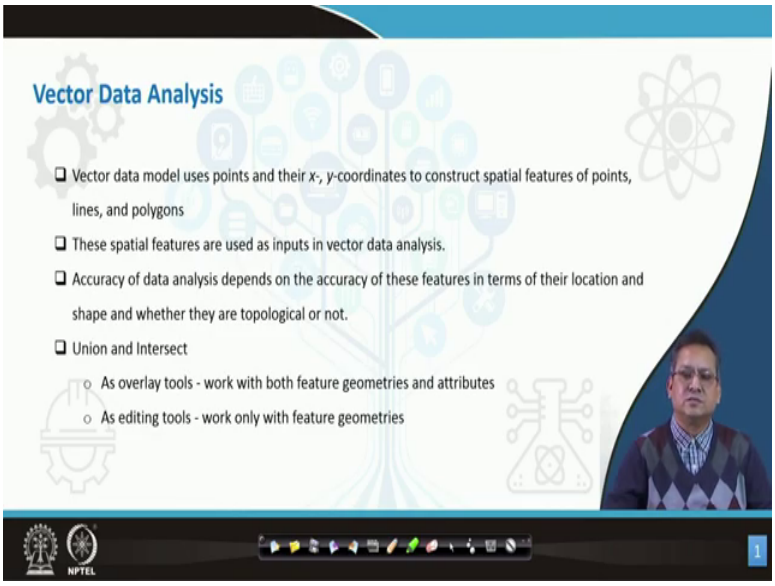
Welcome, back dear, students we are in the 7th lecture now. And, we are going to talk about vector data analysis. So, we have come to the module 2, but I still have to I mean tell you about the spatial statistics part so, that we will be covering in a future lecture. So, let us go through slides, the different lecture I mean the different aspects that we are going to cover today.

(Refer Slide Time: 00:37)



So, the first thing that we are going to cover is the buffering, the concept of buffering in GIS, we are going to look into overlay operations and we are going to look into the distance measurement part. Now, this particular presentation on your vector analysis has been broken up into two lectures, lecture 7 and lecture 8. So, in lecture 8 we are going to cover the point pattern analysis, we are going to cover the feature management part.

(Refer Slide Time: 01:10)

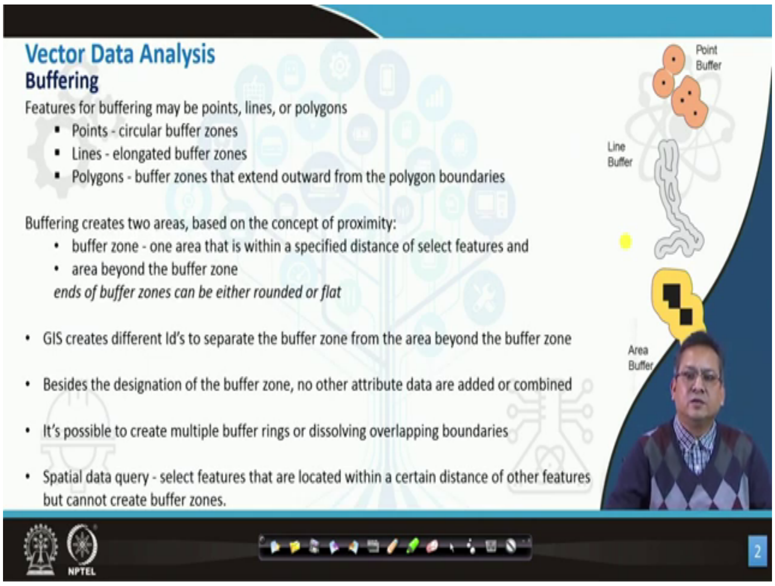


So, I mean talking about the vector data analysis; your we have seen that the vector data model uses points, the coordinates of the points, lines and the polygons. And, these are used as input when we are going to do any sort of modeling or we when we are going to any do any sort of analysis using these vector features. So, the accuracy of the data that we are digitalizing or that we have I mean aggregated from different sources has to be accurate, because otherwise these errors would carry on in the analysis stage.

So, the locational accuracy, accuracy in terms of the shape and the topological relationships needs to be taken care of these are very important things. Now, talking about the different type of data analysis, basic analysis would use union and intersect operation; it is similar to a union and intersect operations in your Boolean algebra that we do in maths. So, the overlay tool works with both the geometric as well as the attribute features; whenever we are doing this union and intersect operations through the overlay modules.

Then it would use both the features, I mean feature of the geometry and the attribute. The attribute tag that you have with the GIS data like point, polygon or your line data you would have attributes. So, both these are used whenever we are using the union and intersection operation as overlay tools. Then this union and intersection operation is also possible in I mean, you can incorporate it in the editing tools to the editing tools, but when you are operating the union or intercept algorithms through the editing tool, you have to bear in mind that it only works with the geometric features and it does not work with the attribute features.

(Refer Slide Time: 03:19)



So, let us see the idea of buffering. So, we can see that the buffering I mean there are three examples here; that we have the point buffer and we have circles around the buffer. In this case the buffers has been dissolved, the boundaries between this differential buffers they have been dissolved.

So, similarly, we can see the buffer from of a line your coverage and then we have a area buffer I mean, if you have a theme as area and we try to create a polygon buffer. So, this is how the operation the end operation of the buffering would be. Now, in this case we can see the end of the line buffer is a circular entity, but it can also be I mean, ended as a line as a flat buffer.

So, we will have a look at it while we are discussing about buffering. So, as we have seen that the points would have a circular buffer in the example, we have seen that points by default the loci of buffer would be a circle, lines would have elongated buffers, buffer zones and the polygon would the buffer zone would extend outward. Now, the buffering creates two areas it is based on the concept of proximity. So, I mean one area is within specified distance, which is the buffer zone and the area beyond the buffered zone is another entity.

So, we will have two polygons; one is the buffer zone entity like, if you have this polygon which is a buffer it will have a polygon entity and there will be a polygon entity created for your geographical extents and the polygon which is outside this buffer polygon.

So, you have two polygons you would have two polygons. One is within the specified distance of the selected area and the another is beyond the buffered zone, that we have just seen I mean in case of point buffer. Now, as I was talking earlier the end of the buffer zones could be either rounded or flat depending on your application. So, suppose you want to take a buffer zone of a highway, you want to find out some effects of pollution I mean, how far it is I mean impacting along the sides of the highway. So, you can create buffer zones which could be either rounded or flats, I mean flat buffers.

Now, the GIS would create different IDs to separate the buffer and the beyond the area beyond the buffer zone. We said it would the buffering would create two features two I mean areas. So, one is within the buffer area and another is beyond the buffer zone. So, this would have two different IDs, it would have two separate IDs from the for the area which is within the buffer zone and for the area, which is beyond the buffer zone. Now, your I mean we are designated the designating the buffer zone, but beyond this no other attribute will be added or combined, when we are creating the buffer zones.

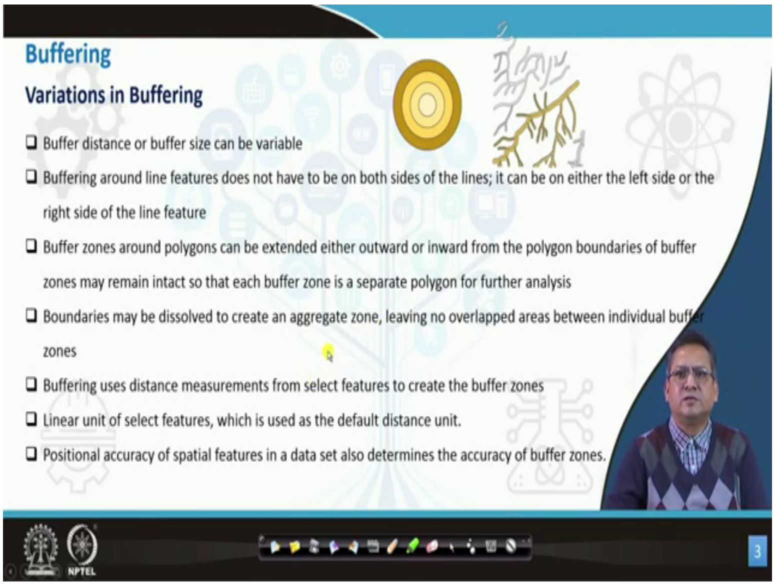
It is I mean here you can see that, we have only one possible ring around the point; wherein we have specified a singular buffer distance and this I mean polygon entity is created around the point input point feature. So, it is also possible for us to create multiple buffer. So, in a way if you want to see the prices along the highway, if you have a line I mean entity and you are creating a buffer you would like to see so, then number of houses within this particular distance.

And we would like to count the numbers so, you can create a buffer zone within say 500 meters, within say one kilometer, within two kilometers and we can find out how many houses or what is the population residing along this particular width of the highway.

Another example, could be your I mean the I mean the distance, the buffer distance from the coastal lines which will give you the coastal zone. So, you can create a distance buffer from the highest elevation line, along the coast and you can create the buffer zones; I mean wherein you can come off with different suggestions like conservation zones or active recreation zones. So, these are some of the examples of buffers that we use in the context of urban planning. Now, we can also do these queries I mean like; I was telling that how many houses are there, what is the population residing within a buffer distance.

So, we can do queries for a certain distance for facilities or features located within a certain distance, but cannot create the buffer zones hm.

(Refer Slide Time: 08:47)



So, there are variations in the buffer. So, your I mean you can create different distances for the buffer size and it can be variable like; when we are creating polycentric buffers I mean you have multiple buffer zones. So, you can have different distances while you are specifying the distance from the point or the I mean specific feature you can specify different buffer sizes for the different rings or the different zones. So, those sizes could be variable.

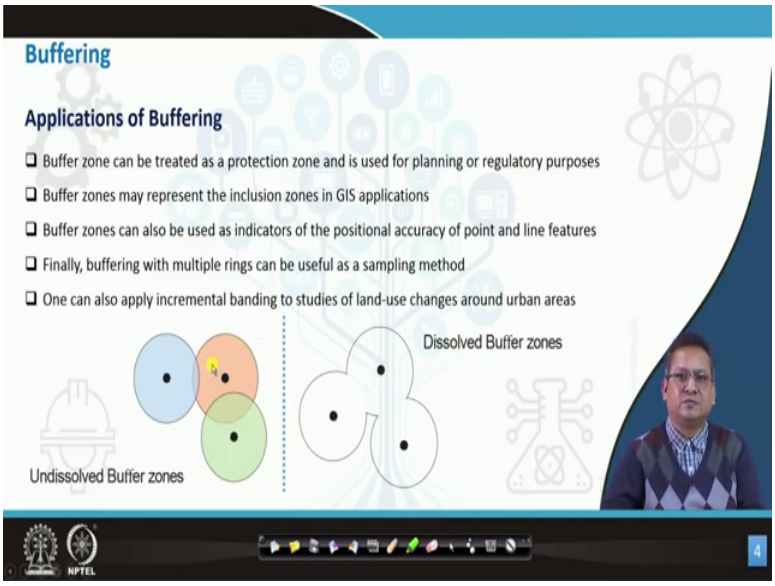
Second thing is whenever we are drawing a buffer or we are creating a buffer along a line feature, this buffer need not be balanced on both sides. It need not be equal on both sides it could be either on the left side or on the right side of the line feature. So, we can choose whether I mean we want to create the buffer by taking the road as a central I mean feature or we want to displace the buffer to the right or the left. Now, because you may have certain queries whenever you are doing some urban planning, you may have certain queries that you want to find out some features within a particular distance on the right or the left of the road segments or say canals or say railway lines. So, if it is a linear feature.

Then I mean, you can have the buffer only on one side not on either sides. So, buffer zones along the around the polygon, it can extend outward or it can also extend inwards so, if you have a polygon you can create buffer zones both outside or inside as you desire. And, each buffer zone would be a separate polygon and can subset it or save it as a different feature and use it for further analysis. So, I mean we can also dissolve the boundaries as we had seen earlier in that circular profile buffer of the point coverage wherein the boundaries had been dissolved

In this particular case, we had seen that the boundaries has been dissolved between the buffer areas. I mean your buffering distance measurements from select features, I mean we can create the buffer zones, I mean use by using distance measures and this distance measures would have the units same as that of the unit of the feature; by input feature. If it is in meters then the buffer distance that we are specifying would by default be in terms of meters, if it is in kilometers or decimal degrees or feet and inches, I mean the buffer the output buffer would have similar units.

So, I mean you have to be careful while you are I mean creating the buffers and you have to look into the unit of the entities. Now, the positional accuracy of the spatial feature would determine, I mean as we had said earlier that it would the positional accuracy of the input features goes into identify the accuracy of the buffer elements as well.

(Refer Slide Time: 12:01)

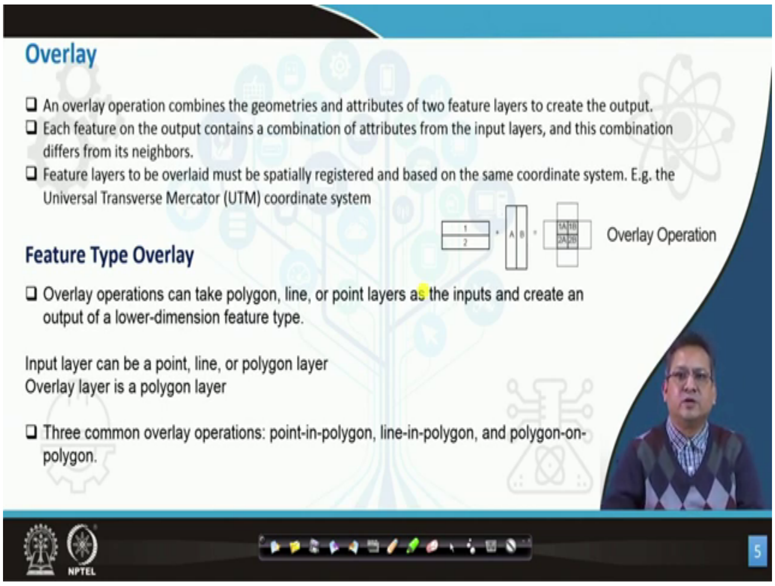


Now, the buffer zones I mean; how do we would apply it in terms of planning. So, we can use it as a protection zone or as for planning or regulatory purposes. So, in conservation whenever we have historical monuments, we generally have a buffer zone or a regulatory zone around the monument. So, we can create buffers and we can specify the urban use in that in the particular in that particular zone.

Now, buffer zones also may have inclusion zones, it may represent like we are talking about the exclusion zone zones, there also could be inclusion zones which could be represented by the buffer. Now, it could also be used as indicator of the positional accuracy of the point and the line feature and buffer multiple bufferings can be used as a sampling method. So, suppose you want to do a sampling; a random sampling or a stratified random sampling of some data points. So, you can use buffering to create partitions for different zones from a particular feature and do samples. I mean you can you can take sample points. So, I mean buffer can be used for that purpose.

So, I mean it can be applied to incremental banding to studies for land use changes, around the urban areas. So, whenever you have say peri urban growth or say urban sprawl, you can create bands or buffers to see that; what is the status of growth in the different bands? I mean what is the nature of growth and what is the status of growth in the different bands of buffers. So, we are talking about the undissolved buffers and this is how a dissolved buffer would look like.

(Refer Slide Time: 14:09)



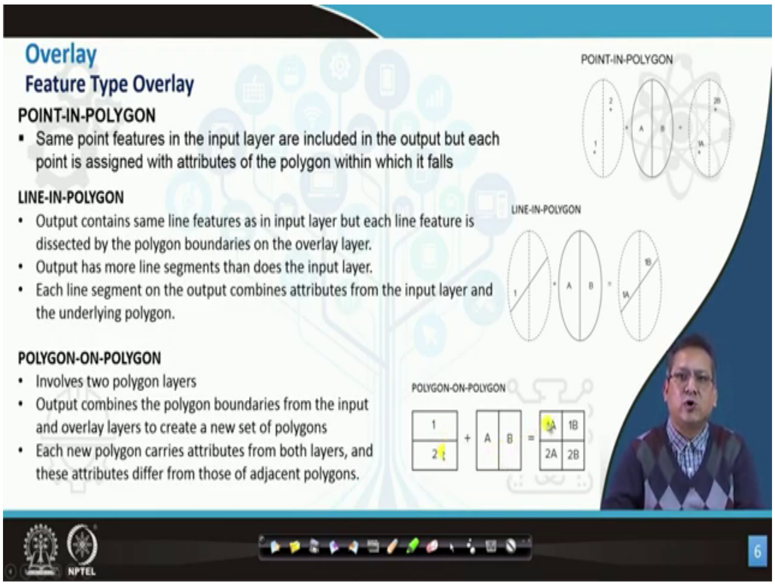
Now, talking about the overlay operation; overlay operation it would combine the geometry and the attributes and it creates the output. Now, whenever you have this combination of attribute from the input layer and this would differ each feature, I mean would have its combination of attributes, it would have some list of attributes multiple attributes and it would differ from the neighboring attributes.

So, I mean it has to be the layers that you are overlaying on top of each other, whenever you have multiple layers that you are overlaying; they have to be spatially co registered. I mean there should be a line on top of each other. So, we had earlier talked about different types of coordinates systems and the projection systems.

So, we have to fall back on geo referencing these spatial themes and see that they are overlying on top of each other, for this operation of overlay to happen. Now, you can see this operation overlay, wherein there you have two input inputs. There are two polygon coverages wherein you have this polygon, which is horizontal in nature having ids one and two and this polygon has two entities A and B. And, when you do an overlay operation you can see that the geometry and the attribute both get edited. So, you have 1A, 1B, 2A and 2B in the output and you have 4 boxes in the intersect region overlay region.

This overlay operations can take either polygons, lines or point layer as input and the output would be I mean lower dimension feature type. So, I mean if you are I mean taking point, line or polygon layer as an input, the overlay is a polygon layer. So, your output would be a polygon dissected by a line or entity. So, it would have resulting to polygons. Now, there are three common overlay operations as we had said that this I mean the input layer could be point line or polygon. So, there could be three types of operations which can be performed. First is the point in polygon method, second is the line in polygon method and the third is the polygon on polygon method.

(Refer Slide Time: 16:45)



So, talking about the first one, the point in polygon method so, we see that there are two points out here point one and point two. And, there are two polygons; the polygon has entity A and entity B. So, this polygon when you are overlaying it is shown as dotted. So, actually in this layer you have only point 1 and point 2. Since these are geometrically co referenced the overlay the outlay of the I mean the second theme, that is the overlaying polygon is shown as dotted in this particular diagram.

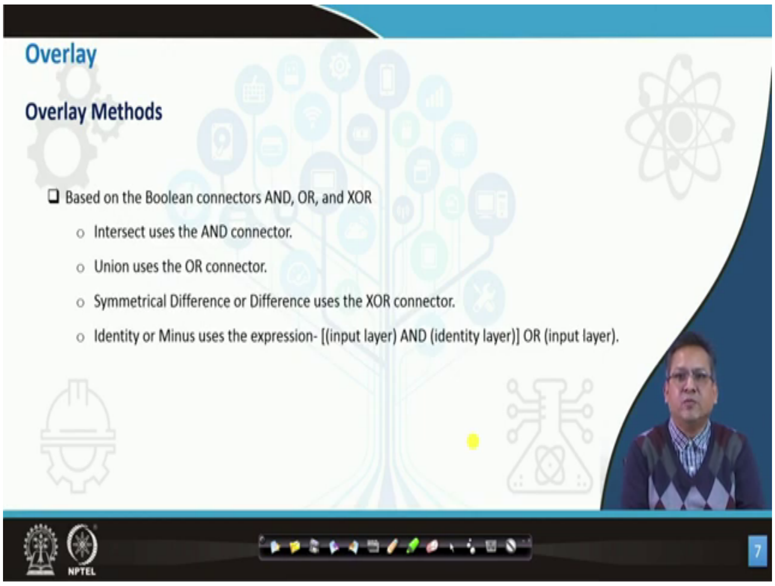
So, in the output you can see that this point entity is very higher the values of the point entity as well as the polygon. And, you would not have any polygon entity in the outcome. So, you will have two point entities; having attributes of both the point as well as the polygon. So, here we see that the same point features as the input layer are included in the output, but each point is assigned at the attribute of the polygon. Now, the second operation which is the line in polygon method so, we can see that the output contains the same line features as the input, but each feature is dissecting the polygon boundary on the overlay layer.

So, you can see here that there is a line feature which is coded as one and then you have a polygon feature, which has two polygon entities A and B and when we are overlaying this two polygons what happens is you would get two polygons in the output and you would have, you would have line entity and this line entity would be divided into two segments having the attributes of both the input entity and the input polygon. So, you have that the line segment divided into two parts 1A and 1B. This is the idea of the line which is one and this and this part of the polygon was A so, when you are intersecting it this line segment up to this point becomes 1A and the rest of the line segment becomes 1B.

So, we see that in the output it combines the attributes from the both the input layers and the underlying polygon. Now, the third one which is the polygon on polygon overlay the output when we have two input polygon layers, one input layer and one overlay layer the output would combine the polygon boundary from the input and the overlays to create new set of polygons. So, unlike your point operation or line operation wherein the outcome is only points or lines, when you are intersecting polygons we see that outcome would become polygons and it would take attributes from both the polygons, I mean the input polygon as well as the overlay polygon.

So, if we consider this as the overlay polygon having two polygons entities A and B, and the input polygon has two polygon entities one and two, the output of 4 polygon entities 1A, 1B, 2A and 2B when we are doing the overlay operation.

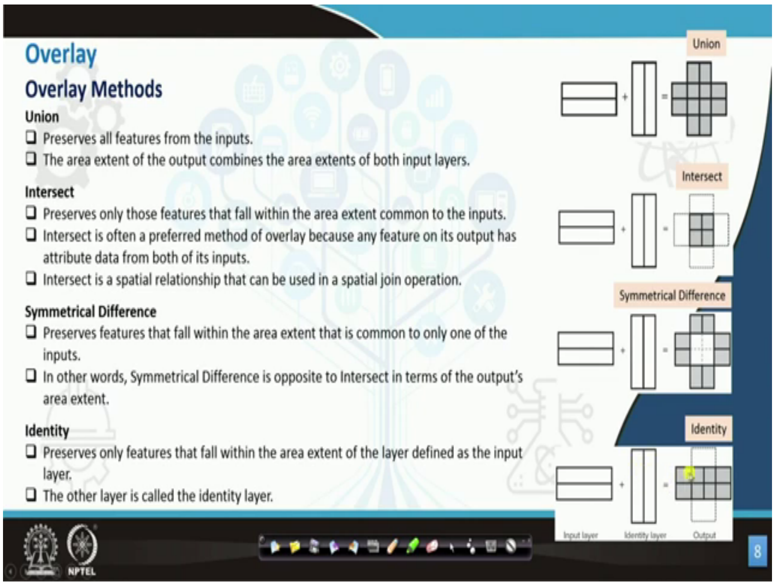
(Refer Slide Time: 20:05)



Now, these overlay operations are based on the Boolean connectors; like AND, OR and XOR. So, the intersect uses the AND connector. Union would use the OR connector and symmetric difference or difference uses the XOR connector. Now, identity or minus uses the expression input layer and identity layer or input layer.

So, this is how we would have the software functioning at the back end whenever you want to create a identity layer or you want to create a symmetric difference, you want to create a union or an intersection operation. So basically, it is based on the overlay functions or the Boolean operators.

(Refer Slide Time: 20:49)



Now, the when we are using the union feature; it would preserve all features from the input. And, the area extent of the output it would combine the area extent of the input features. So, suppose if we have two polygon entities whenever we are having the overlay, you can see the boundary extents would be these two particular these 4 corners for this particular rectangle, for this polygon particular polygon entity. And, for this you it would be a vertical extent, but in the output you see this extent shifts over here, this point this on the right, this on the bottom right and this on the bottom left.

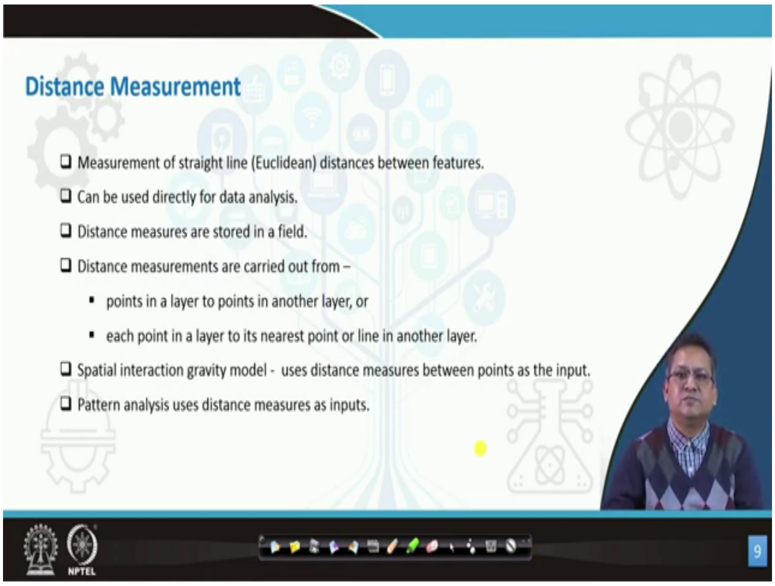
So, it combines the area extents of both the input and the output layer. Now, for a intersect operator it preserves those features, which fall within the area extent and which is common to both the inputs. So, we will see that as an image and what happens really when we are using the intersect operator. So, it is a generally a preferred method when we are doing an overlay, because if any feature is there on its output; its attribute data it would get from both the inputs. So, in case whenever you are doing an intersection, we would use the overlay method as it would get your attribute feature of both the input layers.

The input this intersect is basically a spatial relationship and it could be used in case whenever you want to do a joining operation for spatial join. So, you can use the intersect command. So, you can see that whenever you are intersecting, you are just returning the common boundaries that are there in the input your both the inputs input layers. Now, talking about the symmetric distance say difference, it would preserve the features which fall within the area extent that is common to only one of the inputs.

The symmetric difference in a way is opposite to the intersection, I mean intersection we have seen that you only return the area, which are common to both the areas. So, the symmetric difference is just the opposite of the intersect mode. So, and in terms of the output a area extent, I mean in the intersect operation you can see that the output area extent is only limited to the intersect area, but in their symmetric difference this output extents; I mean it incorporates the output extents of input extents of both the input layers basically.

So, talking about the identity it would preserve features that fall within the area extent of the layer defined as the input layer. And, the other layer is called as the identity layer. So, within that layer the input layer that entity is preserved, that zone is preserved and the I mean you will get the intersect only within that particular zone of the input wherein the other part is discarded, you can see the other part is discarded.

(Refer Slide Time: 24:23)

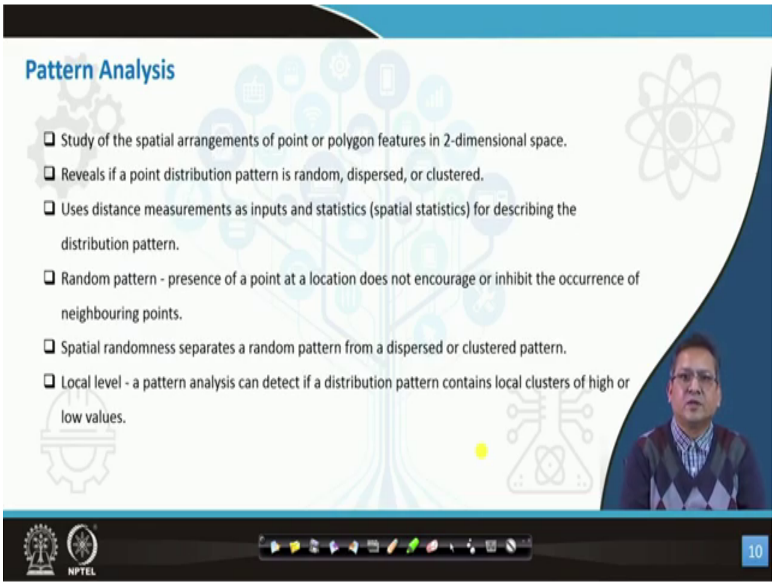


So, talking about the distance measurement part so, measurement I mean we generally use the Euclidean distance and this distance is used extensively in modeling approaches. So, suppose you are you want to do a pollution modeling in an urban area so, you have a chimney stack and you want to see, what is the I mean effect of pollution along a particular distance. So, you would have to do the distance measure along the straight line and it can be used directly for data analysis, this distance measurement whatever values you get from the GIS is directly used in the data analysis and it is stored in a attribute field.

So, you can carry out the distance measurement from points in a layer to points in another layer or it can also be done for each point in a layer to the nearest point or line in another layer. Now, we can use this distance measure when we are doing some spatial interaction model; like a gravity model when where we try two model the interaction between different zones in urban area. So, suppose there are few people who are who have started from the home and going for work.

So, we can model these kind of interactions and this is generally used in transportation modeling and (Refer Time: 25:55) transportation modeling. So, this kind of interactions can be modeled in GIS by using distance measure between the points or the centroids as the input features. Now, the pattern analysis also uses the distance measure as inputs.

(Refer Slide Time: 26:14)

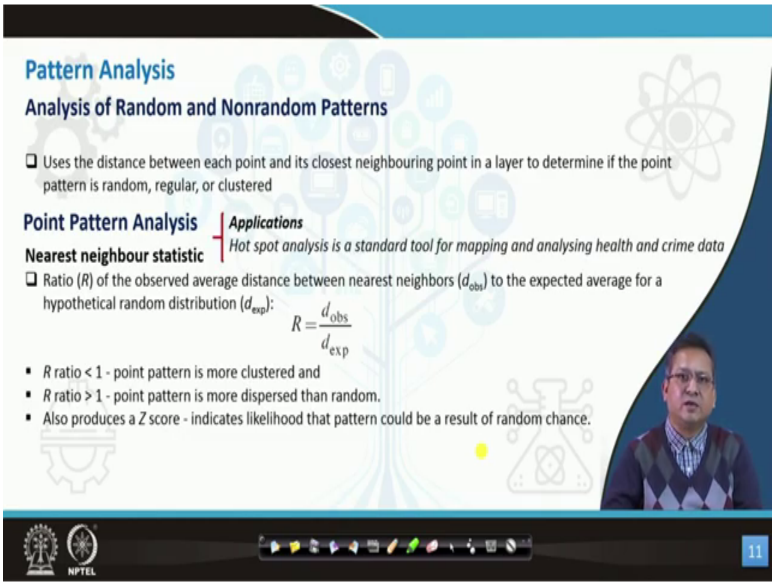


Now, talking about pattern analysis; we study in pattern analysis spatial arrangement of point features or either line features in a two dimensional space. So, this is extensively used in analysis where we want to find out; what is the nature of dispersion? So, whether it is random, whether it is dispersed or whether it is clustered. So, I mean we can identify it from the pattern analysis, the point pattern or polygon pattern analysis. So, it uses the distance of as inputs and the statistics that is a spatial statistics for describing the point pattern or the distribution of the point pattern.

So, I mean when we have a random pattern it shows that the presence of a point or a location it inhibits the occurrence of your neighboring points. So, basically there is no trend in the point, how they are located. So, they may not be a clustering. So, it would result in a random point that can be explained through the spatial statistics in the data. Now, spatial randomness it separates the random pattern from dispersed or clustered pattern. So, a randomness would separate a random dispersed pattern from a random clustered patterns.

So, I mean we were talking about the spatial statistics by which we can differentiate these types of patterns. In the local level also the pattern analysis can be done, I mean when we are talking about the spatial randomness, we are talking about the entire landscape of operation, but at a local level also you can do a spatial analysis pattern analysis. And, distributions would contain local clusters, it could be of either high value clusters or low value clusters.

(Refer Slide Time: 28:18)



You may have point points having different types of representating different types of entities and these could have different values, I mean high values or low values. So, that would give you the I mean randomness or non-randomness in terms of the clustering patterns. Now, for analyzing the random and nonrandom patterns, we use the distance between each point and the closest neighboring point in a given layer.

So, I mean to identify whether the pattern is random whether it is regular or whether it is clustered. Now, talking about point pattern analysis; we can do a neighbor nearest neighbor statistic. So, this statistic is represented as ratio I mean aggregated as R capital R, it is the observed average distance between the nearest neighbors to the expected average of the hypothetical random distribution. So, the I mean which is represented as d expected.

So, we calculate the this value R which the ratio of observed distances to the expected distance; random of the random distribution and if this pattern is found to be at this ratio R is found to be less than 1. Then we say that the point pattern has is more clustered and in comparison to if the ratio is greater than 1, which would indicate that the point pattern is more or less dispersed in nature than and then random. So, it would also produce a Z score, we are talking about spatial statistics.

So, in this analysis it produces a Z score as well which gives the likelihood that pattern would be the result I mean whether it is the result of random chance or it is I mean event, which has been done systematically. So, this point pattern analysis is applied for hot spot analysis which is a standard tool for mapping and analyzing say health data or crime data in a urban context. So, there we can create this kind of hot spot analysis based on the distribution of the pattern. So, we can use the nearest neighbor statistic and come up and find out what is the R value to find out whether the point is clustered or it is disposed, if you are using the point pattern analysis.

So, thank you and we look forward to the next lecture, which is a continuation of this particular lecture.

Thanks once again.