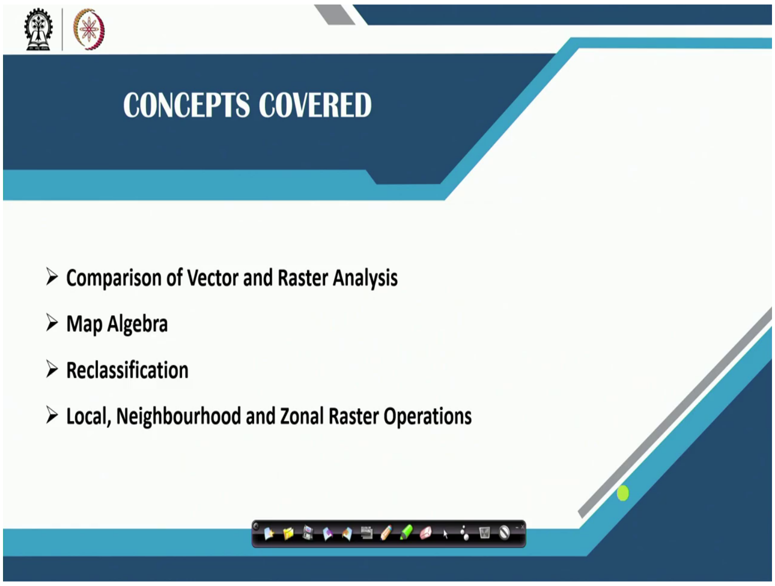
Geo Spatial Analysis in Urban Planning

Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture - 09
Raster Data Analysis

Welcome dear students. We are in the module 2 and in the lecture 9 where we are going to see the different types of Raster Data Analysis. So, this particular lecture will cover more than one lecture series one lecture. So, I mean we will have a follow up lecture also on the raster data modeling part.

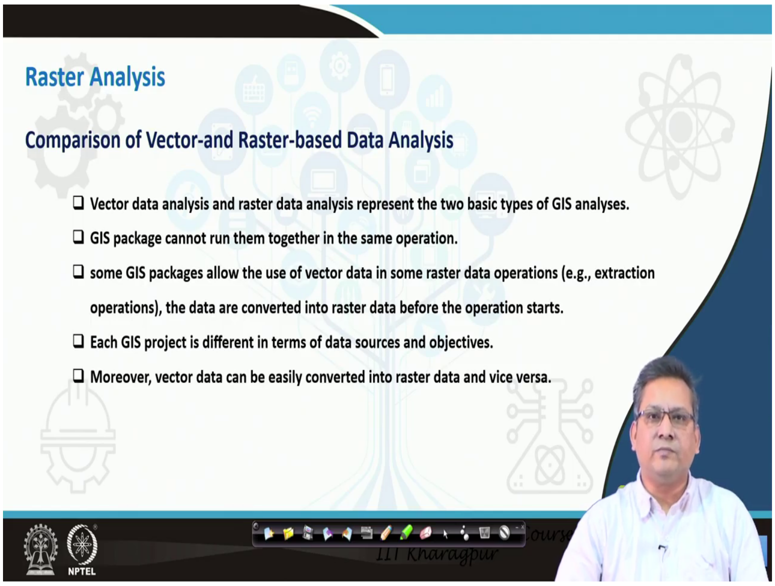
(Refer Slide Time: 00:41)



So, let us see what are the I mean topics which are going to be included or the concepts which are going to be included as a part of this lecture.

So, first we are going to do a comparison of the vector and the raster data analysis, then we are going to talk about map algebra. We are also going to talk about reclassification of GIS data specially raster data and we would look into the local neighborhood and the zonal raster operations.

(Refer Slide Time: 01:19)



So, let us see what is the, I mean how does the vector data analysis I mean come in comparison with the raster data analysis mode. So, in the vector data analysis, I mean we have seen that we can work out the data which is non discontinuous in nature. I mean it is not a discontinuous data set which can be represented as vector data analysis and a continuous data can be represented as raster data analysis.

So, these are the basic two types of data that we had talked about, so similarly we can do analysis on raster data as well as the vector data. But, the limitation of the GIS packages is that it is very difficult to do both the operations together in tandem. So, I mean we do different sets of operation on vector data and raster data and do it separately. And then collate the results or I mean there could be some inter intermediate steps in the modeling and which could be further used to model a given problem.

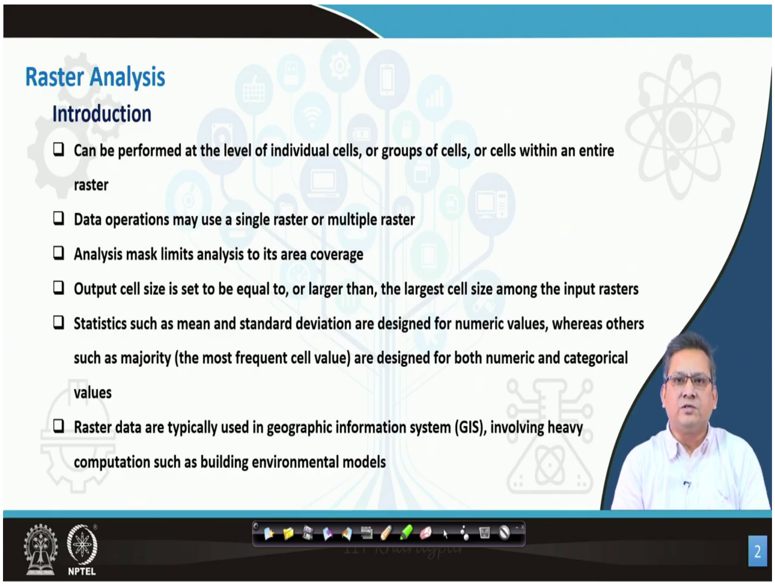
Now, your some of the; I mean vector data can also be used in raster data operations some of the packages would allow that say suppose we have a raster data. And we want to extract the data which is within the confines of an urban administrative area. So, we may have a vector data boundary of the administrative area and we would like to extract the raster data say, suppose you have a satellite data of Google image or say IRS resource at data.

So, you can subset that data based on the vector data or the profile of that particular area that you may have, so you can subset that. So I mean this data can be converted into the raster data. So, the output would be a raster data set, and this would be used further in the operation.

So, this is an example where in the vector data set as well as raster data set can be used only for extraction of the data. I mean when you have some kind of a area of operation or area of interest, you can extract that area. Now, I mean we have different types of requirements regarding your data modeling and we may have different objectives; we also have different sources of the data.

So, vector data have and raster data though they cannot be processed together, but it is possible for us to convert the vector data set into a raster data set and vice versa. I mean it is possible to interconvert the data in most of the software’s.

(Refer Slide Time: 04:14)



Now, talking about your raster data analysis, it is can be performed either at a individual level or it can also be performed on groups of cells or I mean the cells the total number of cells within a given raster image. I mean it could be at a very local level, it could be at a zonal level or never would level or it could be at a global level when we are talking about the entire raster data set.

Also I mean this analysis can be used on single layer, I mean we can run a raster analysis on a single layer or we whenever we are having multiple layers also we can run the analysis. It also I mean we can limit the analysis. We were talking about the I mean zone of interest or the area of interest say suppose we are talking about an urban area; and we may have data which extends beyond the urban area.

So, we can mask our analysis and limit this analysis only to the area in area of interest. So, I mean we can also look into the output cell size, the size that is the spatial dimensions of the cells or the pixels in the raster data. So, it could be either set equal to or larger than the input raster data.

Ok, I mean you can set the output cell size, you have the flexibility of choosing you can resample the output cell size. And, it could be either of the same size, it could be size which is smaller than the input cell size or cell size which is larger than the input cell size.

So, there could be several modeling requirements where in the input data resolution may be different for all the different layers. And, you may have to come down to a unit unique cell size for all the different layers. So, in those cases we go for these kind of operations where in we set the output cell size. We can also I mean workout the mean or the standard deviations statistical parameters.

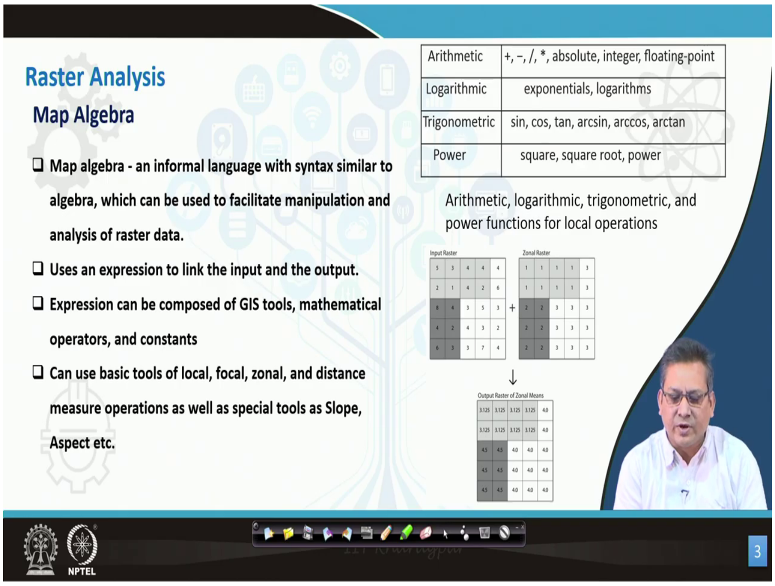
I mean we can also do majority filters or I mean we can work out the majority in a given neighborhood and we can work on both the numerical as well as categorical values when we are doing the raster analysis we can consider. I mean basically this raster data would be an array of numbers as we had talked about when we were talking about the raster data types.

So, this analysis could be taken up in a treating the numbers as numeral values or it could be also treated as categorical values. So, whenever we are dealing with say land use classification, the outcome would be cells of you similar values which would be treated as categorical data set. But, the input data values would be say satellite data in such cases you can have a high resolution satellite data and they would be examples of the numeric data types.

Now, raster data also can be georeferenced. I mean because when we are doing this analysis we would need to have all the data in a unique frame work specially reference framework and we had talked about projection systems, we had talked about geodesy when we were talking about introduction to geodesy.

So, we can basically georeference this raster data sets before we do such an operation. And whenever we are trying to build some models, we need to see that the rasters are basically collocated. I mean they are coregistered with the other data sets.

(Refer Slide Time: 08:25)



Now, talking about the map algebra we can work on the operations I mean algebraic operations on raster data sets. It could be data which come as layers which come as I mean different input layers.

So, this map algebra, it would facilitate to analyze the raster data set the input data set, and it could be used as operators or expressions to link both the input and the output model. Now, your these operations whenever we are having this operations and we have expressions for this operations so, the expressions could be a sequence of expressions the you could have more than one expression.

So, you can have sequence of expression and these are composed of mathematical operators. These expressions are generally composed of mathematical operators that you would find in a calculator I mean most of your basic calculators and then we may also include some constant values.

Now, we can also use the tools basic tools such as the local, the focal the zonal or the distance measure. I mean how far we are going to do this analysis say, suppose we are going to apply a given model. So, from on a given point so, how far we are going to use that model or it could be awaited model that is if we take a distance function as we move away from that point. The impact or the weight of that particular algorithm, it reduces overtime.

So, those kind of operations are used in your interpolation algorithms when we talked about it we will we shall see it. So, we can look into the operations. We can do the basic operations in GIS in a local, in a focal mode, or in a zonal mode.

So, in the local mode we are just talking about each and every individual pixel in the focal mode, we may have a set of 3 by 3 windows. I mean a matrix having 3 elements 3 cross 3 elements that is 3 rows and 3 columns, and the focal element you are the zonal function could be a bigger area.

So, we shall talk about it, so these I mean operations are used; I mean these measures are used with these tools are used when we are calculating say metrics such as slope or aspect. So, here is an example of the different types of operators arithmetic operators, or say mathematical operators that we have access to in most of the GIS database.

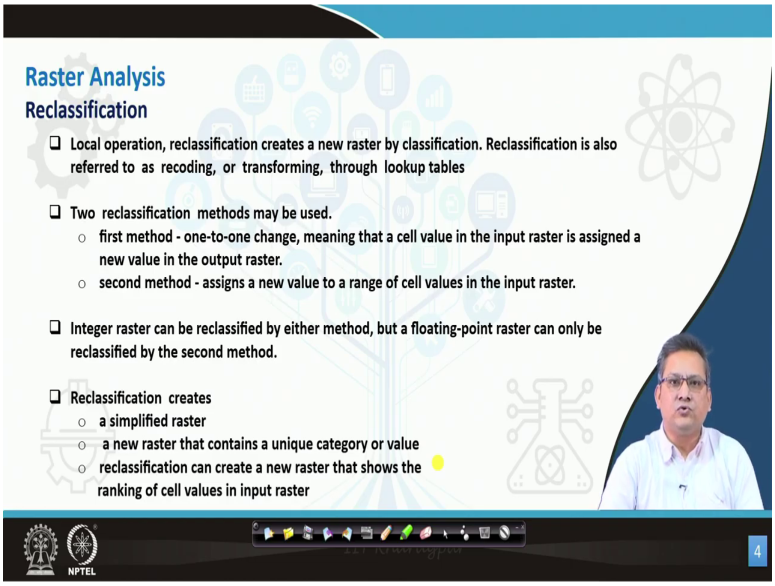
Now, data software's; now some of the software's also have python scripting integration where in you can run your own codes you can run your python script and run it on the softwares'. So, I mean you can do it for RGIS or in open source tools you can run a python script you can code your own python script and run it on QGIS platform. So, we see examples of your arithmetic operators, basic arithmetic operators, logarithmic operators, trigonometric operators, and power operators. So, these apart from this we may also have logical operators, Boolean operators etcetera. So, here you see you have an example where in you have a input raster layer, and then you have a zonal raster layer.

So, in this you see you are trying to work out the mean values of the pixels. I mean wherever you have this zone 1 in the zonal raster so, in that top area you can see there are 6 values; there are 6 values which have your value 1 which is highlighted in gray. So, what we do is in the input raster in that particular zone we do an average take an average of the pixel values and I mean write it in this particular zone.

So, you can find out that you get the average values of this input layers and you are basically dividing your entire region into different zones are running this particular analysis. Similarly the second zone that we see is shown in deep gray color, so again you have 6 pixels and corresponding to it in the input you have pixel values such as 8, 4, 4, 2 and 6 and 3.

So, in the output you can calculate the mean for this particular zone tw2 and it is shown in the output raster of the zonal means. So, this is how basically we can run data analysis using a zonal operator. So, the third one is I mean layer is shown white in color, and then you have the corresponding values which are averaged. So, you get the mean in all these pixels which are shown in the output image.

(Refer Slide Time: 13:57)



Now, let us see how the reclassification can be done. Now reclassification is a process where in a new raster layer would be created, based on some rules. So, you can have say I mean input pixels ranging from different having different ranges. So, within a particular given range you can reclassify this data as one specific data. So, I mean what we can do is we can create lookup tables in some of the software's, it is abbreviated as LUT files. Earlier it gives to be done in your most of the GIS tools and software's.

So, we can create a lookup tables, so that we would be able to reclassify this data whatever data is available as continues data sets. It can be reclassified, and grouped into different categories.

Now there are two different reclassification methods which can be used one is one to one change whenever there is a cell value in the input raster it can be assigned to a new value.

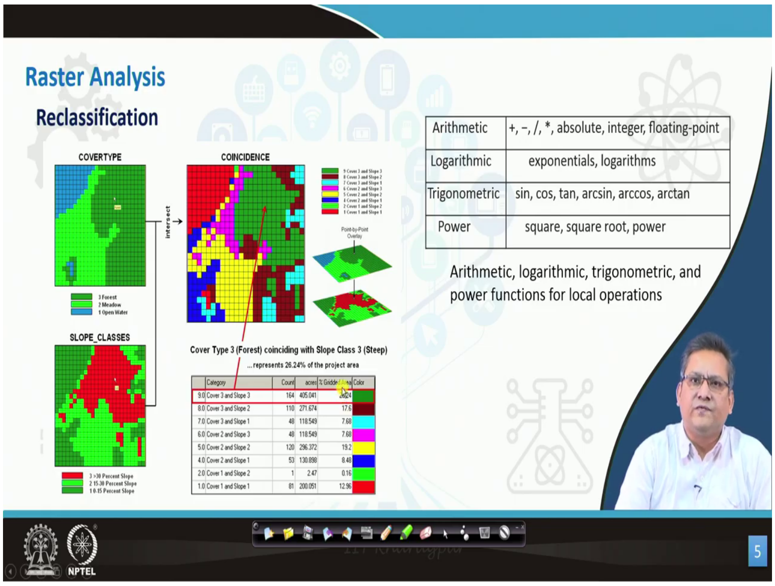
Whenever you want to have a output raster, you can take of one pixel at a time and assign output pixel value to that I mean raster entire raster image.So, that is the first method in which we do a one to one change.

Now the second method it assigns a new value to a range of the cell values of the input raster. So, I mean we would take a range of cell values, there could be like we have seen in the earlier example we had seen a zone. So, in a given zone, we can we can take values which have a specific range of cell values in the input raster. And, we can assign a new value for that particular range.

Now, your I mean the integer raster which is a input can be we can reclassified by either of the methods. But if you have a floating point raster I mean we are talking about an integer raster and floating point raster in this case. So, I mean if you have a floating-point raster, it can only be classified by the second method the first method in that case is difficult to use. So, we can create raster's which are integer which may have integer values or we also can create raster's which may have float values.

So, we shall look into details the classification methods in the follow up lecture to this particular lecture. So, the reclassification would create a simplified raster, it would generalize the data values, and create a simple raster and it would also contain unique category or values. So, you may have some few categories of I mean of or unique values in the output raster. Now, it would show you can thus create the ranking in the cell values.

(Refer Slide Time: 17:17)



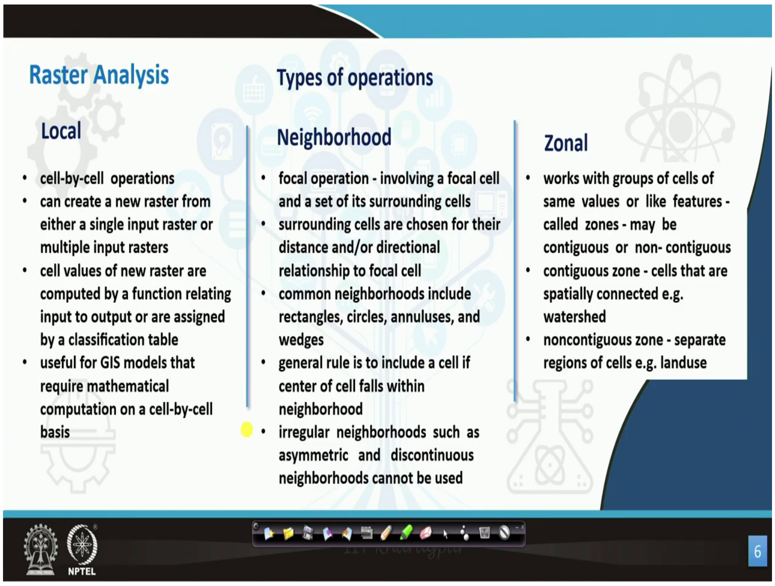
So, whenever we are having this kind of input image, so you can see you have the cover type and the slope in these two maps. So, in this case you have the different land use categories of forest which is categorized as deep green. You have a meadow which is categorized as light green and one is the open water body.

So, when we are intersecting these two classes of the slope class and the land use cover or the land cover class; so, in the slope class we correspondingly see that they are the class 3 pertains to a class where in the slope is greater than 30 percent. And the class 2 that is light green in color has slope ranging from say 15 to 30 percent. And, the last one that is your deep green has slope class for lands which are having slope of 0 to 15 percent.

Now, when we intersect this two layers, we would have a output categories of layers and in this case you can see the I mean output cover has got land use cover type 3 that is forest. And, slope cover type 3 that is have a slope having greater than 30 percent.

So, you can see that we had started with three categories in the input cover and again three categories in the slope classes. So, in all I mean when they intersect it creates 9 classes as the outputs. Now, we can in this case again we can use different types of operators as we had talked about in the earlier slide, we can use different types of arithmetic operator.

(Refer Slide Time: 19:10)



Now, let us see the types of operations we were talking about the local function the neighborhood operators or the focal operators and the zonal operators. So, in the local operator it is a cell by cell operation I mean in our last example we had seen that we can use your float data sets, and we can only use cell by cell operation in that case. So, I mean they can we use generally local operators when we are having such kind of data sets.

So, it would create new raster I mean data from either multiple input rasters or even a single raster whenever we are running a local operator. Now, this can be computed using different functions or algorithms or equations relating the output and the to the input, and can be assigned by a classification table or a lookup table. We were talking about lookup table earlier. So, I mean we can assign the output values either by using a function or algorithm or an equation or using a classification tables that is we may have certain range of data values and, it can be grouped or classified to a particular a value of the output pixel.

So, this is how basically your local operators work, and your local operators would be useful for models specially when we you have mathematical computation on a cell by cell basis. I mean you are not running of application or an algorithm for a range of cells, but on each and individual cells. So, in those cases, the local operators would be very useful and for most of the simulations we use cell by cell operations.

Now, there are also cases where we use the focal functions or the focal operator. So, it would involve a focal cell and a set of surrounding cell, so we can denote the surrounding cell as neighborhoods. So, the neighborhood can be the matrix of pixel surrounding a particular cell. So, in case we have a cell, we can have the three rows and three columns, so we would have eight surrounding pixels. So, when we are talking about cellular automata model for simulation of land use change or future land use simulation, we talk about different types of neighborhood functions.

So, one such neighborhood function that we talk about is a Moore neighborhood. So, there are different types of neighborhood which we can use and that size of the neighborhood elements can vary depending on your analysis. So, the surrounding cells I mean to the focal cell would be chosen based either on a distance based criteria or it could be a directional relationship to the focal cell to the central cell. Now, we can have these neighborhoods as rectangles, we can have these neighborhoods as circles, we can create neighborhoods having shapes of an annulus or as wedges.

Now, the general rule when we are trying to include a cell, I mean when we are doing this kind of neighborhood based operators. So, if a cell falls within the neighborhood it needs to be included in the cell. So, some amount of resampling may be required I mean we have talked about your by cubic interpolation different types of interpolation. So, some amount of interpolation could be required or we can also use the nearest neighborhood interpolation. So, whenever pixel falls within a neighborhood, so such type of operators would be used. Now, irregular or asymmetric neighborhoods are discontinuous neighborhoods cannot be used.

I mean whenever we are talking about neighborhood we are talking about a continuous set of pixel which are I mean in the immediate neighborhood of a focal cell. So, if we are if we want to include cells which are out layers which are asymmetric in nature or discontinuous neighborhoods at a distance away from the neighborhood, then, it cannot be used for such type of neighborhood based operators. Talking about the zonal operation it works with groups of cells having similar values or having similar set of features and we can call them as zones. So, these zones could be contiguous or noncontiguous, I mean they could be asymmetric or discontinuous neighbors neighborhoods or it could be continuous types.

So, in the zonal operator we can use both the types of data whenever we have say irregular neighborhoods or regular neighborhoods; both the types of neighborhoods can be used. We are talking about contiguous zones where in the cells would be connected. Say suppose, we are talking about a watershed when we are doing a hydrological analysis, we want to do a runoff analysis. So, in that case the cells would be connected to each other in a watershed. So, that is an example of a contiguous zone and there we can run a zonal analysis.

Now, in the noncontiguous zone we can have separate regions of cells like we can talk about a land use data where in I mean just in the last slide we had seen an example of land use. So, you can have multiple parts of land having similar values in a non contiguous way. So, that is an example of a noncontiguous area, so these kind of data sets can be used in the zonal operators. So, a recap of what we had talked about say for the local operators, we do we are running a cell by cell operation.

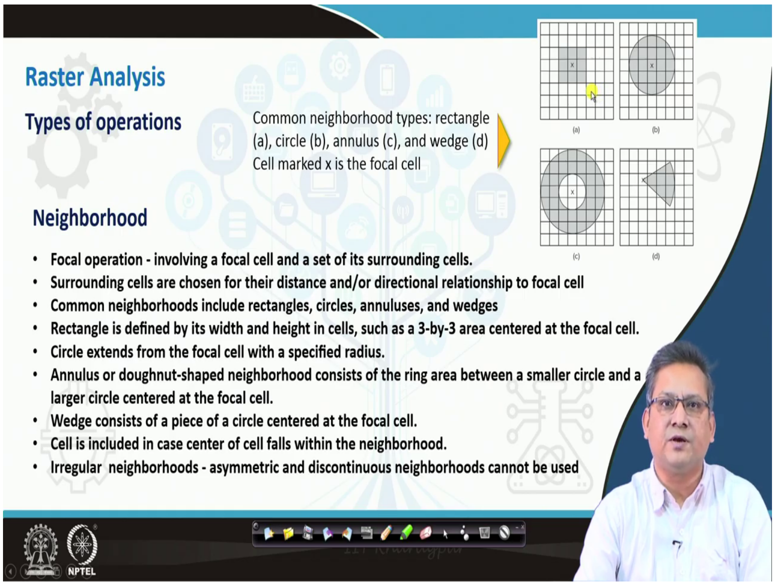
In the neighborhood, we are talking about contiguous zones working out contiguous zones and mostly neighborhoods that could be in the shape of rectangles, circles, annulus or a wedge. And, we said that we cannot used the irregular network neighborhoods or asymmetric or discontinuous neighborhood.

But, in the zonal operators when we are running with the zonal functions, we can use noncontiguous data sets as well as we can use the contiguous zones as well cells which are contiguous in nature or noncontiguous in nature can be run into the zonal analysis. Now, most of the zonal analysis is done on ecological issues. We have we can calculate the ecological metrics like patchiners, porosity, intense version just a position on a data set and it is run using the zonal operators.

(Refer Slide Time: 26:57)

Now, talking about the types of operator in the local operator we had said we are working on the cell by cell I mean operations. We can create a single input layer from multiple input rasters. And, I mean the cell values would be computed either by a function or we can use the classification table.

(Refer Slide Time: 27:20)



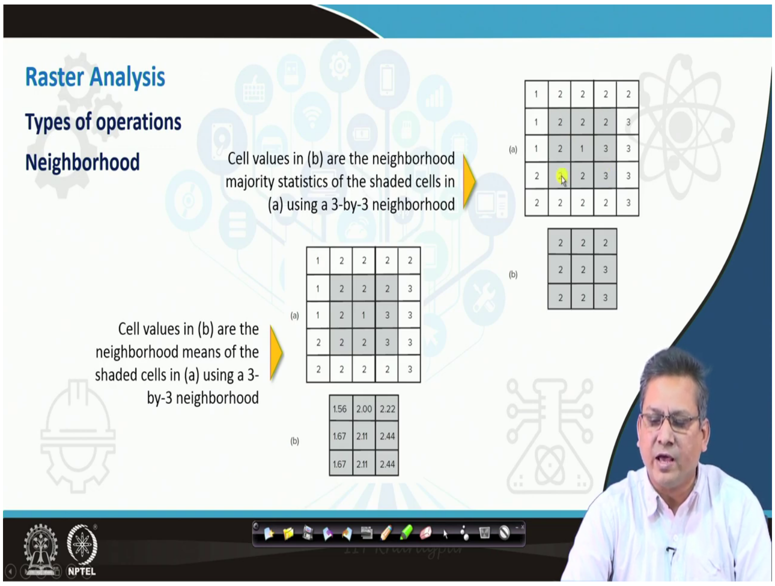
Now, for the neighborhood operation, we said we uses focal cell and set it and a set of the surrounding cells, the cells the matrix surrounding this focal cells.

So, this said the cell the surrounding cells can be chosen as a function of the distance and or the directional relationship to the focal cell. Now, we have also talked about the different types of neighborhoods. I mean we can a circle; for an example would have a specified radius and we can have a rectangle where in we can specify the width and the height of the cells.

So, say suppose we can have a 3 by 3 I mean number of cells and it would have a center focal cell. Now, we can also have your annulus shaped or doughnut shaped I mean neighborhood and I mean which might consist of a ring area between a smallest circle and a largest circle centered at the focal cell.

Now, the wedge would consist of a piece of a circle centered at the focal cell and the cell is included in case the center of the cell falls within the neighborhood that we had already talked about. We had talked about that irregular or asymmetric neighborhoods or discontinues neighborhoods cannot be used in such type of analysis. So, we can see the four different types of neighborhoods that we can work in this data type. This is the first is the circular one, then we also have the rectangular function and then we have this circular one, but with the I mean focal cell at the core. And, we also have a wedge shaped I mean neighborhood.

(Refer Slide Time: 29:28)



So, in this example, we can see for the neighborhood operation we have two examples where in the first one does majority statistics on the shaded cell. So, you can see when we have this particular cell, first in the first iteration when we are running a code having say iteration of different values of i and j in the program.

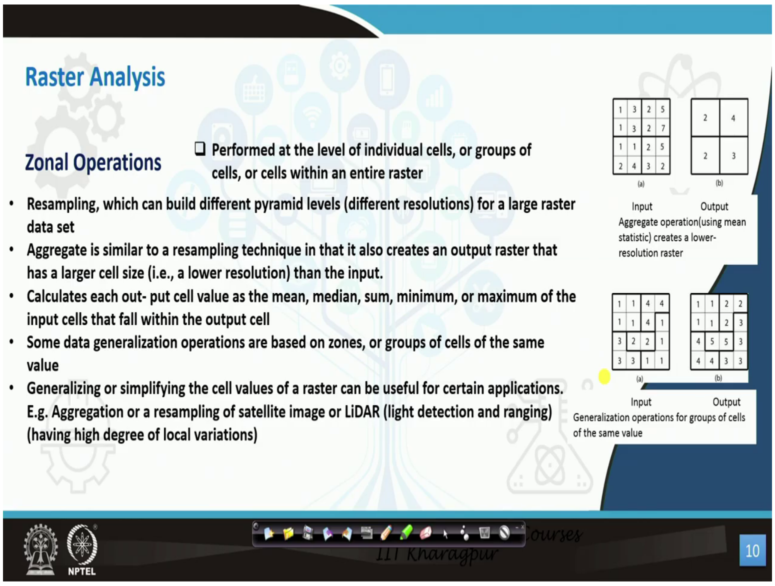
So, let us take this first I mean neighborhood which has these values of 1 2 and 2 and then we have this 3 by 3 matrix. So, in this 3 by 3 matrix what we do is we find out the majority of the numbers. So, in this first element you can see there are five values of two numbers including the focal area and four values of two numbers. So, the central value is replaced by a value of 2. Now, when we shift this value of i by one place to your right and we take the next three elements in the matrix where in we have the focal cell outlier. And, then we have this neighborhood of 3 by 3 elements. So, in this case again you can see number 2 dominates the number of alliteration of 2 is the highest. So, it are the central pixel value is again replace by 2. Now, let us take an example where we have this input pixel value is 1.

So, in this case what happens is you take this particular neighborhood and again you see that there are six values which are of value 2. So, we are doing a majority statistics in this neighborhood operation we are doing a majority statistic. So, this value of one in the output image is replaced by a value of 2.

So, similarly you can find out the majority values using a majority statistic you can from this input matrix, you can calculate the output matrix. Now one thing we should take a note is that that the output matrix will have a i minus two elements and j minus two elements. So, in the input matrix you can see there are 5 cross 5 elements. But in the output matrix, you have i minus 2 that is 5 minus 2 that is 3 cross 3 elements i minus 2 and j minus 2 elements.

So, similarly when we are talking about the neighborhood means so again the first matrix that we take is this particular matrix the first element. So, in this case we take all these values and the mean comes out to 1.56. So, this number 2 in the output cell is replaced by this value 1.56. So, similarly we keep on shifting by one cell to your right till we encounter the last matrix that is this matrix. And then we shift the next in the next iteration we shift it by one row to the I mean bottom. So, you can see that the mean values of the cells are encoded in the output.

(Refer Slide Time: 32:45)



So, this is one such operation neighborhood operation which can be worked out. Now, we have zonal functions also, so we can it can be perform at the level of individual cells or we can also have group of cells which could be contiguous or non contiguous. So, these kind of operations can be used when we are talking about I mean resampling.

So, we can for a raster big raster layer. We can resample this image into a different pyramid level having different resolutions. And, in those cases we use the zonal operations, now aggregation or aggregate function is also similar to the resampling technique. It creates an output layer which has larger cell than the input.

Now, it the this calculates each output cell values as it could be mean, or median, or sum, or minimum, or maximum value of the input cells that fall within the output cell. Now, some data would be generalized and these are based on the zones or groups of cell of the same value.

Now, if we see this example I mean we see an example of aggregation operation where we used this aggregation operation for your resampling the data. When we want to create a resampled image, we generally aggregate the data values.

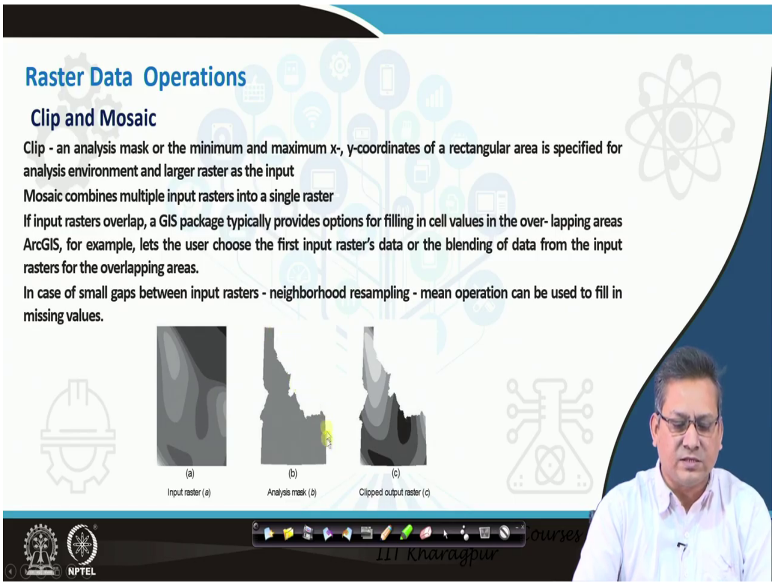
So, you can see we are aggregating this first pixel four pixels into one pixel and we are aggregating it using the mean statistic operator. So, the mean of 1 plus 3 plus 1 plus 3 comes out to two for this particular pixel. And similarly, for the other elements you have the mean values. So, these are examples of some of the aggregation operators using the mean values. Now, the second one that we had talked about is the generalize it generalization operator. So, in this operator we can see that there are groups of values, and in the output cells it would have similar values.

So, in the first one it is clubbed as 1 in the second group which is 4 if the output is given as 2. Now, the third group which is the group 1 if the output is given as 3 the fourth group which is the input group as 3 is given as 4 and the group which is 2 is given a output value of 5.

So, these are some of the generalization operators this is how the generalization happens. And these are mostly used when we are trying to do a classification of the cell if we are running a classification algorithm in that case we run the zonal operators. So, it can be used for certain applications for example, if we are doing an aggregation or resampling of satellite data, or say LiDAR data where in you may have different reflectance values it could be continuous data values very similar.

But, closed data values not exactly same data values. So, in those cases we would be using the generalization operators we could run the generalization operators. And these are basically examples of data sets where in there is a high degree of local variation.

(Refer Slide Time: 36:27)



Now, we have the different types of operators, so like in vector data we had done a clip and join operator.

So, here also in raster we can run a clip operator where in we may have an analysis mask, and we can or the minimum or the maximum coordinates. So, it could be either a rectangular area where in you specify. The area of interest giving the I mean the domain of that particular x and the y-coordinates, and we can also run using this kind of data analysis such as clip operation using a mask. So, this mask we said could be a raster layer or it could be a vector layer as well.

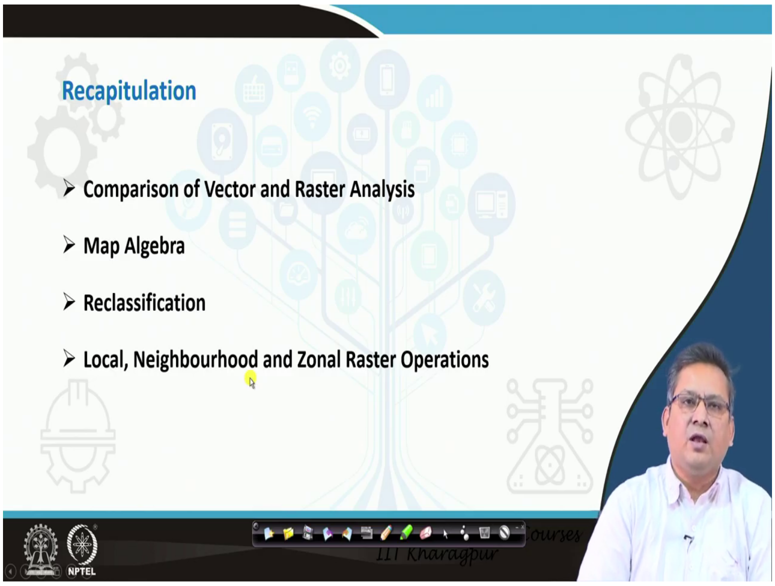
Now, there is another operation which is Mosaic which we generally use when we say we may have different say topographic maps. And we would like to combine it together join the latitude longitude coordinates join it together and Mosaic or make a bigger map. So, similarly it could be done on a data like your satellite images where in your area of interest or you are the city where you are analyzing the data may not fall on one data set.

But, it may lie within the immediate I mean images. So, what we can do is we can Mosaic we can join all these images together into a single input raster. And we can clip this data and take out our area of interest using the clip operator. Now, these operators are known as known by different names in different packages

So, we can in some cases we can also see that there could be some small gaps between the input raster's. I mean they may not be continuous data sets, so in those cases what we do is we do a mean operation a neighborhood resampling to fill in the missing data values. So, this is an example of clip operation I mean you can see the input data in which it is a continuous I mean it is a group data.

And the second data that that is the analysis mask. So, when you this is your area of interest you want to only retain data specific to your area of the interest you run a clip operation and you will get an output image having the clipped area of your input data.

(Refer Slide Time: 39:03)



Now, recapitulation of what we have done in this I mean learnt in this particular lecture is we have first compared the vector and the raster data analysis techniques. Second is we had talked about the map algebra and the different kind of operators that can be used then we had talked about reclassification methods and finally, we had talked about different types of operators at local, neighborhood and the zonal level.

So, later on when we develop this operation we will be when we develop applications, we will be using this concepts of your vector or say raster analysis when we are applying it on some grading some kind of analysis urban analysis. So, we shall look for some of those examples in our later I mean lectures. So, thanks and I mean we shall again continue with this raster data analysis in the next lecture.

Thank you.