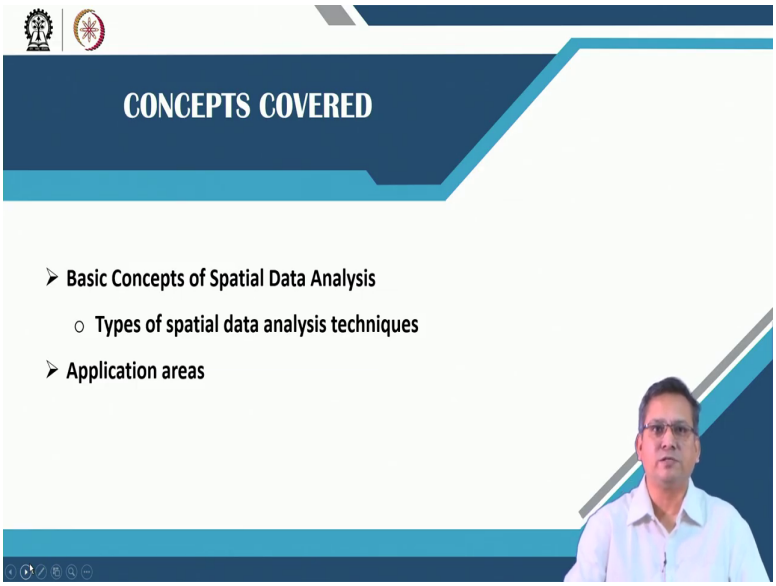
Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture - 06
Method of Spatial Analysis

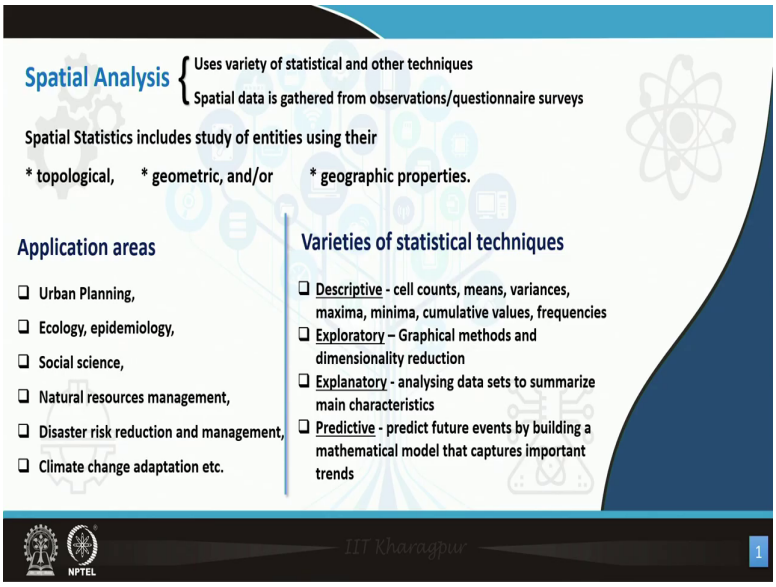
Dear student, welcome back to this course on Geo Spatial Analysis in Urban Planning we are starting the module 2 now, where we are going to introduce a spatial analysis in which the first lecture is on the Methods of Spatial Analysis.

(Refer Slide Time: 00:30)



So, the aspects that we are going to cover today is the basic concepts of spatial data analysis. So, in this we are going to cover the different types of data analysis techniques and we are going to look at a few areas where this can be applied.

(Refer Slide Time: 00:51)



So, going to the different methods of approach for spatial analysis it is mostly the statistical methods which are which were initially used for GIS analysis mostly for spatial purposes. So, for doing this analysis we need data so we get the data from different sources. So, it could be through some sensors, it could be through some manual observations, it could be through some questionnaire surveys and this data is collated in a digital format it is aggregated as a database and then we perform the GIS data analysis.

Now, the spatial statistics includes three types of I mean entities and their relationships amongst themselves. So, these relationships are of three natures mostly: topological, geometric and the geographic properties.

So, for the topological properties it is basically a set of rules and behaviors that model how points lines and polygons that is a GIS entities share the coincident geometry. Now, the point

said topology theory says that the topological relationships are invariant under homeomorphisms that is; their topology relationships are preserved even if we consider this objects to be rotated scaled or moved in a given GIS spatial domain.

Talking about the geometric properties these are basically unary features such as; length, area, perimeter example distance between two objects then we have the geographic properties which are basically location specific properties. Now there are different applications of the spatial analysis.

So, I mean we would be looking into the aspects of urban planning applications there are also applications in the field of ecology and epidemiology, then we have applications in the field of social science, we have applications in the field of natural resource management, we also can generate applications in the field of disaster risk reduction and management, we can look for different applications in the field of climate change adaptation etcetera; I mean these are related to ecological urban planning aspects so these are more pertinent for us.

So, there are varieties of your statistical techniques which were used for when we do the spatial statistics and do the analysis. So, first is the descriptive statistics in which we talk about cell counts, means variances, minima, maxima, cumulative values and frequencies. So, then what we have is the exploratory data analysis where in we have graphical methods and basically methods for dimensionality reduction.

So, in this graphical methods we have different approaches like the box plot which basically represents the minimum value the first quartile the median, the third quartile and the maximum value. Now, the next method the graphical method is the histogram. So, in this method we plot the frequency of score occurrences in a continuous data set that is divided into bins or classes. The next type of exploratory methods specially graphical method is multi vari chart where in it graphically displays the patterns of variation that is variation within a group, within a subgroup, between subgroups or overtime.

The next type of exploratory data analysis graphical method is run chart in which basically it is used to determine whether the central tendency of a process is changing by displaying

observed data in a time sequence that is we try to see the temporal changes in the central tendency of the data how the mean varies or the median or the mode varies so we try to see it in a temporal scenario.

The next method the graphical method is Pareto chart. So, in this method it helps in decision making used for the selection of a limited number of tasks that produce significant overall effects. The next method is scatter plot in which we plot the data points on a horizontal and vertical axis along the abscissa and the ordinate and to show that how these two ordinates and abscissa this variables are related to each other. So, the scatter of the data basically would give us an idea regarding how the data is related

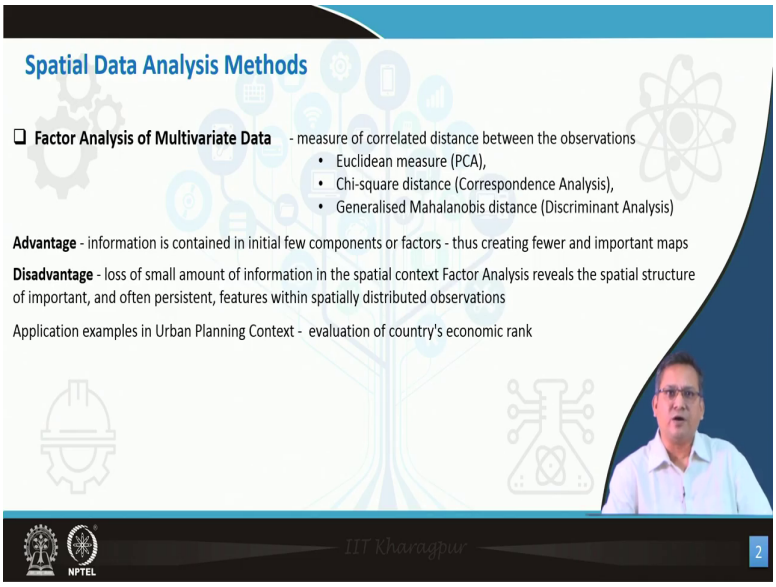
The next graphical approach is the stem and leaf plot which shows frequency with which certain classes of values occur and visualization of the shape of the distribution. So, I mean this gives us an understanding of the data distribution. The next graphical method is the odds ratio so which quantifies the strength of the association between two events.

So, talking about dimensionality reduction we have different approaches and one such approach is known as principal component analysis. Then we have the explanatory method of statistical approach which basically analyses data set to summarize the main characteristics of the data.

So, I mean we look at the different characteristics of the data using this explanatory approach and we identify variables that have a scientifically meaningful and statistically significant relationship with an outcome. Now, the test of the theoretical hypothesis I mean is done using this exploratory models to determine whether each relationship is statistically significant.

Now, the steps involved in this exploratory statistical method include fitting potentially theoretically important predictors next we what we do is check for the statistical significance in the next step we evaluate the effect size and finally, we run the diagnostics. So, now let us look at the predictive models. So, basically these models are used to predict future events, by building mathematical models that capture important trends.

(Refer Slide Time: 08:43)



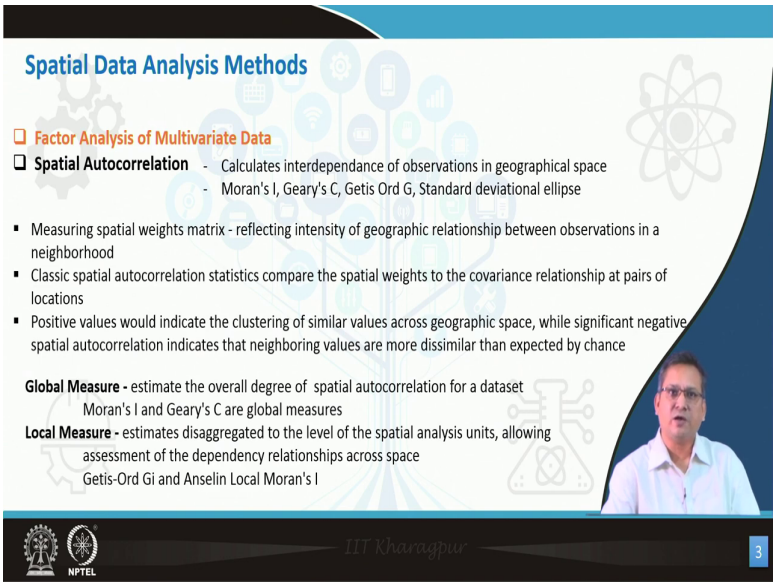
So, let us look at the different methods of spatial analysis. So, first method is the factor analysis of multivariate data which basically measures the correlated distance between the sets of observations that we may have. So, in this particular method we have few techniques the first is the Euclidean measure which is the principal component analysis measure and which measures the distance I am talking about distance we can also measure the chi square distance using correspondence analysis we can also have Mahalanobis distance which is done in the discriminant analysis.

So, the advantage of the factor analysis method is that I mean; when we have a multitude of data sets I mean we have I mean lots of attributes corrected with a I mean given in a given model what we can do is we can basically sew out the important components which basically

contribute and we can sew out the noise and we can remove the redundant components. So, we can create fewer components and we can create important maps.

Now, the most I mean the disadvantages of this method is; loss of small amount of information in terms of spatial content. So, we have seen that factor analysis reveals the spatial structure of important and other persistent features within spatially distributed observations. So, I mean there are lot of important advantages of doing this factor analysis of the multivariate data basically we can reduce the components and I mean we can apply this type of models say for example, evaluation of countries economic rank based on lot of different types of parameters.

(Refer Slide Time: 10:57)



So, the next approach is the spatial autocorrelation approach which basically I mean calculates the interdependence of the observation sets in a given geo spatial I mean or a geographical

space. So, in this we have different types of measures such as; Moran's I, Geary's C, Getis Ord Gi, standard deviation ellipse.

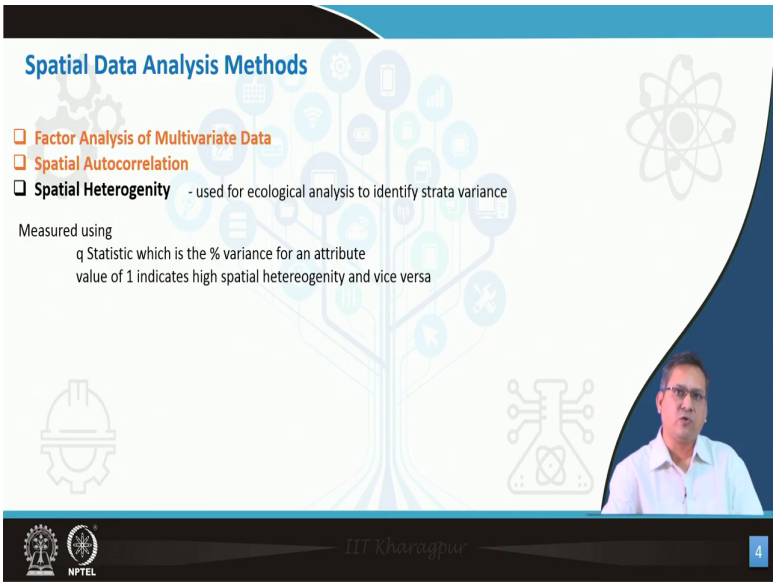
So, I mean what we do is we measure the spatial weights matrix wherein it reflects, the intensity of geographic relationships between observations in a given neighborhood. So, we have classical spatial autocorrelation statistics and we compare the spatial weights by creating the covariance matrix, covariance relationships at the pair of the locations so it is done using covariance matrix.

And when we have positive values of this autocorrelation values it indicates the clustering of similar values in a geographical space and when we have negative autocorrelation spatial autocorrelation values it indicates that I mean more of dissimilar values in a given region than I mean similar values than the expected by chance basically.

So, we have two measures we can do global assessment of spatial autocorrelation. So, where in we do an overall assessment of the spatial autocorrelation for a data set in this we have two I mean measures which are primarily being used; one is the Moran's I and the next one is Geary's C then we have local measures which are estimates of disaggregated data and I mean this analysis is done at a disaggregated level of spatial units and it allows for assessment of dependency relationships across space.

So, some of the measures which use which are used for local measure assessment for spatial autocorrelation are Getis-Ord Gi and Anselin Local Moran's I.

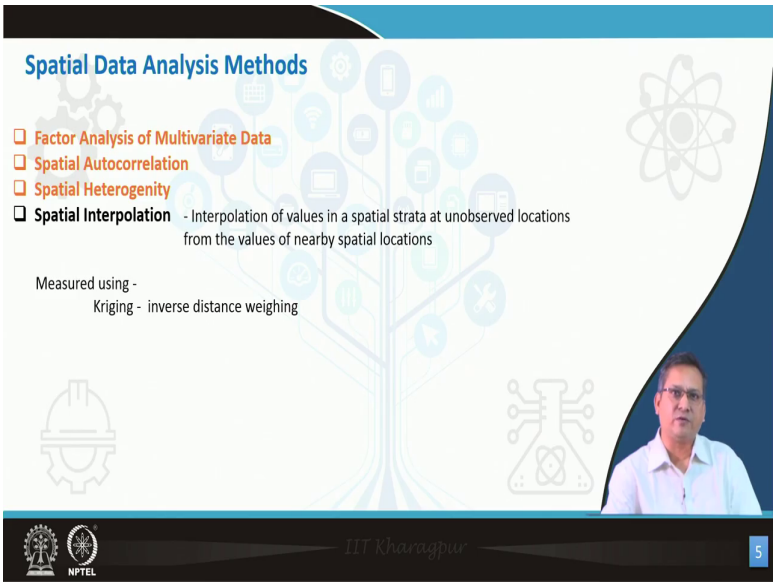
(Refer Slide Time: 13:26)



So, we will be discussing in detail about these topics and probably seeing one or two applications in GIS. So, I mean we are just going through cursorily through the different methods or approaches for spatial data analysis. So, the next one that we discuss is the spatial heterogeneity.

So, this is used for ecological analysis to basically identify the variance in the given spatial strata. So, there are measures like q statistics which measures the percentage variance for an attribute and in case we have value of 1 it generally would indicate very high level of spatial heterogeneity and if we have 0 value I mean it would indicate otherwise that is low level of spatial heterogeneity.

(Refer Slide Time: 14:18)



So, the next method that we would be discussing with the is the spatial interpolation method. So, in this value in this method what we do is we have discrete set of observations you have observations dispersed in space as point data values as point locations and we try to interpolate the intermediate values or the values of the nearby spatial locations.

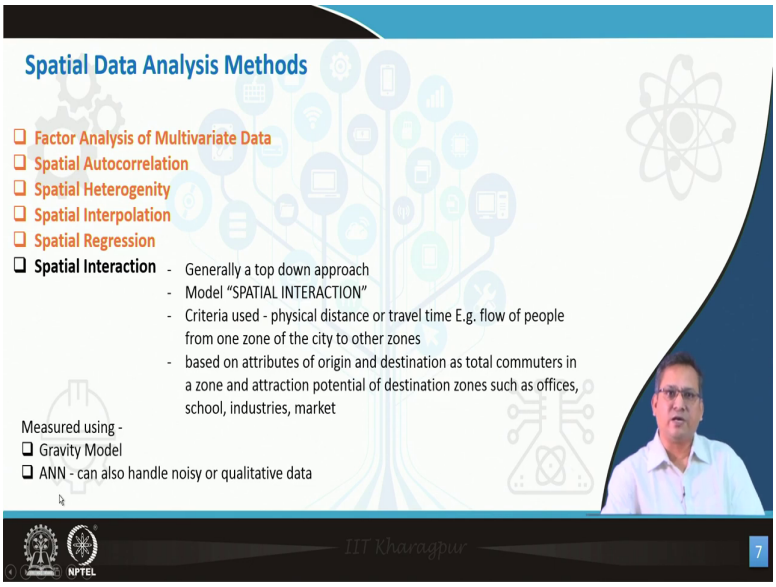
So, I mean; the most prominently and commonly used method is known as kriging and it is a very useful tool and the method that is used for kriging is the inverse distance weight method weighing method which is used in kriging. So, basically it uses a spatial lag relationship using symmetric and random components.

(Refer Slide Time: 15:16)

The next method that we are going to discuss is the spatial regression. Now you are already familiar about regression I mean we do regression in statistics. So, we can also similarly do spatial I mean regression amongst dependent and independent variables which are collocated in a geographical space.

So, we can do a geographically weighted regression that is parameters which are close to a given location would have a higher impact on a independent variable; on the dependent variable. So, we can do a geographically weighted regression which would basically generate parameters which are disaggregated by the spatial units of analysis.

(Refer Slide Time: 16:05)

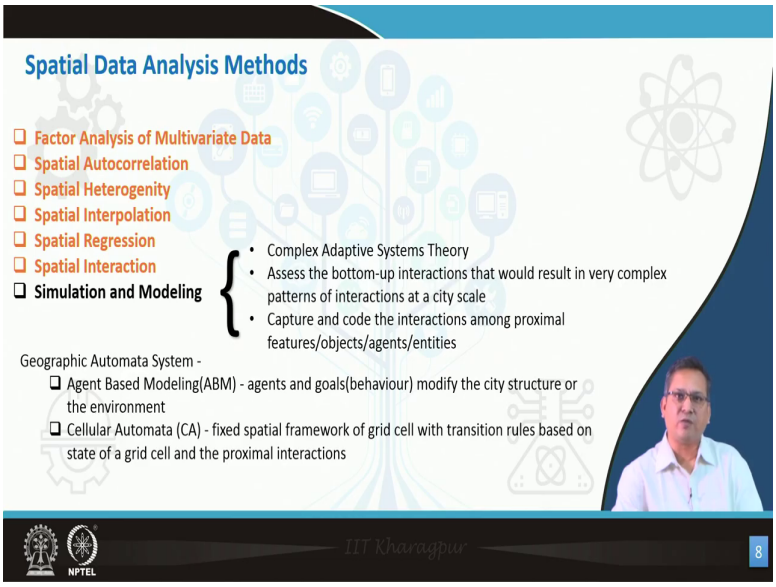


The next one is the spatial interaction models which are which is basically a top down approach. So, in this we model spatial interaction among the different entities or variables and we can calculate it calculate the different types of criteria one such criteria could be the physical distance it could be a cost distance or travel time so I mean we generally use it for the flow of people characterize the flow of people from one zone of the city to other zones. Now this are based on attributes of the origin as well as the destination.

So, as in I mean if you are talking about the I mean attributes of the origin we would have to estimate the total number of commuters if we are talking about the transportation problem total commuters in a given zone that is how I mean the zone where the I mean which would generate the traffic and then we would have the attraction potential of the destination zone such as offices, schools, industries or markets.

Now, there are different ways of doing this spatial interaction models modeling one such method is gravity model which is extensively used in planning; urban planning and we also have soft computing methods such as artificial neural networks which can also handle noisy data or even qualitative data wherein, for gravity model we generally work with quantitative data sets.

(Refer Slide Time: 17:45)



The next approach that we are going to talk about when we are talking about spatial data analysis methods is simulation and modeling approach. So, in this approach I mean we have a theory which is known as complex adaptive systems theory which talks about modeling agents and basically talks about adaptive systems which keeps on changing because of this, behaviors.

So, I mean we had seen the spatial interaction models which are basically top down model. So, these type of models the simulation and modeling approaches using complex adaptive system

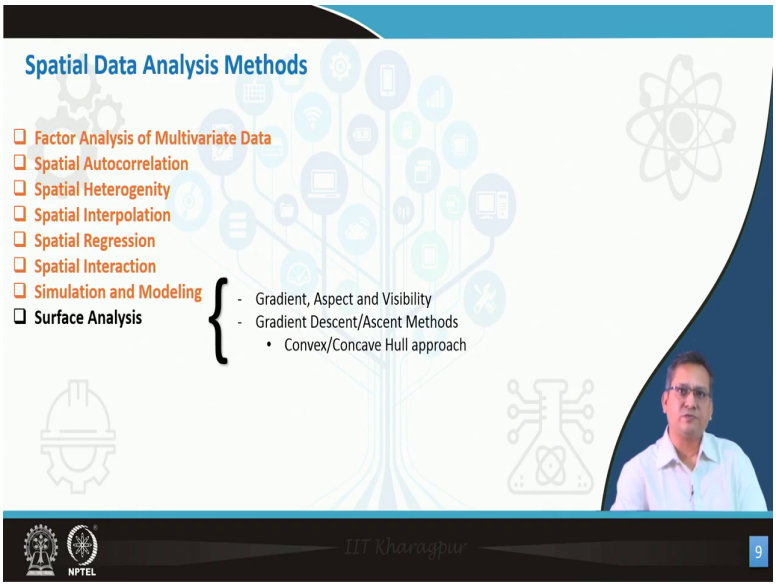
theory these models are basically bottom-up approach wherein, the interaction between the agents or the I mean different elements it would result in complex patterns of the in the system because of the interactions in a local scale or a city scale.

So, basically in this approach what we need to do is, we have to capture and code the interactions among proximal features that is neighboring features, objects, agents or entities we may call whatever we may like and different literature suggest different nomenclature. So, these are some of the terms which are very commonly used. So, the kind of framework that is used for simulation and modeling specially in a complex adaptive systems are agent based models I mean they are abbreviated as ABM models. So, basically we have agents and goals which are basically behavior and which modify the city structure or the agent or the environment that we were discussing about.

Another approach is known as cellular automata; I mean this is a very I mean good candidate for in the raster data analysis domain so this can be cellular automata modeling can be done in a fixed spatial reference framework of grid cells and we can I mean code the transition rules based on the state of the grid cell and the proximal interaction.

So, what we can do is we can do a historical assessment of the transition and we can frame rules and these rules can be applied in local context so as we have talked about that these are bottoms-up approach. So, we can see that basically it is the interactions of the agent which would have a bearing on the structure of the city or the environment of the city.

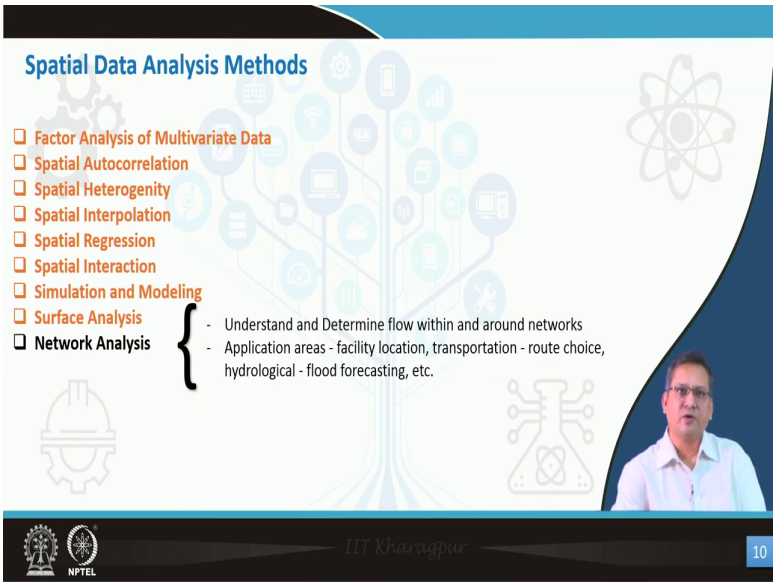
(Refer Slide Time: 20:36)



So, the next model that we are going to talk about is surface analysis. Now, surface analysis is a very interesting tool and it is very commonly used for gradient studies gradient based studies of how the slope changes aspect changes or I mean; what is the visibility then I mean we can have ascent or descent based methods I mean gradient descent of gradient ascent based methods to do this kind of analysis.

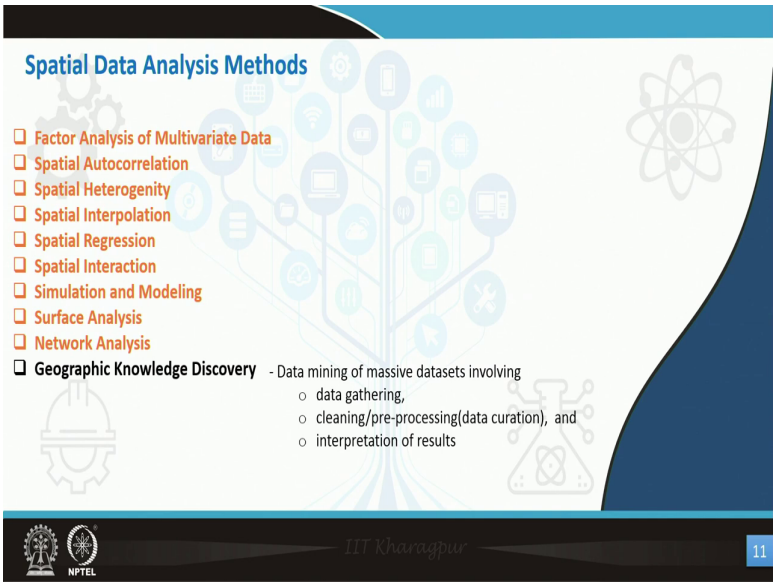
So, I mean there are convex or concave hull approaches which can be used for surface analysis, these are also used for your hydrological modeling so a part of hydrological modeling. So, some of the input comes from the surface analysis.

(Refer Slide Time: 21:14)



The next method that we are going to talk about is the network analysis. Wherein, I mean we understand and determine the flow in a network in a network or around the network. So, we can have different types of application of this kind of network analysis like we can talk about locating facilities in a given region, we can talk about the route choice or we can I mean do a flood forecasting using hydrological models. So, these are some of the applications of the network analytical tools in a spatial data analysis method.

(Refer Slide Time: 21:59)



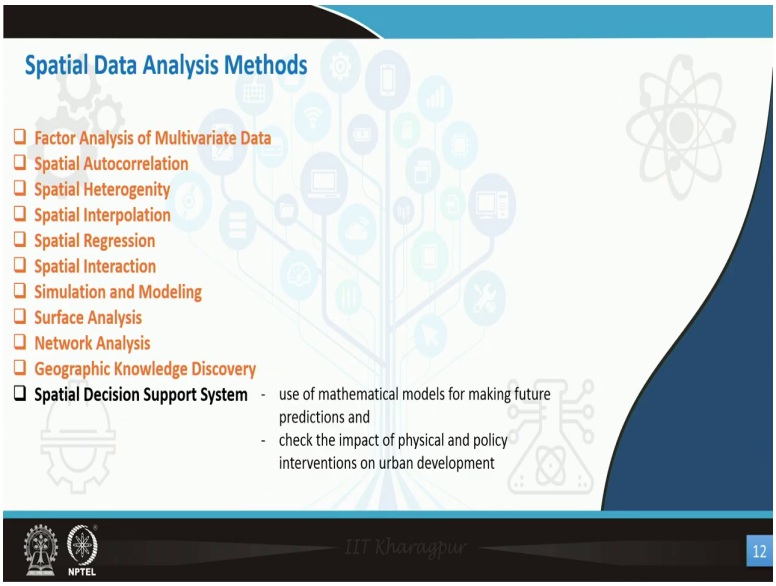
Then we have the interesting one which is known as the geographic knowledge discovery system which is the data mining of massive data sets involving data gathering, clearing pre processing data I mean preprocessing of the data which is also known as data curation and interpretation of the results.

Now, in this geographic knowledge discovery method is basically a human centered process of using computational methods and visualization to explore massive geo referenced digital data bases for novel and useful geographic knowledge. This basically generates non trivial solutions we have data mining tools like geo minor, spin and engines which are I mean in which have been synthesized when we are talking about this geographical geographic knowledge discovery systems.

This geo minor it was the first approach to mine geographic data from large spatial database and it focuses on the discovery of spatial association rules. The second one that we had talked about spin; it integrates several algorithms for spatial data mining which includes multi relational sub group discovery a rule induction, spatial cluster analysis and looks into scalability of algorithms and allows for a tight coupling with the database.

The next one is the engines which is inductive geographic information system and it is a prototypical GIS that processes an inductive learning capability. It can generate first order logic descriptions or geographic objects and it includes a training facility that allows the interactive selection of examples and counter examples of geographic concepts. So, these are some of the methods that we have gone through.

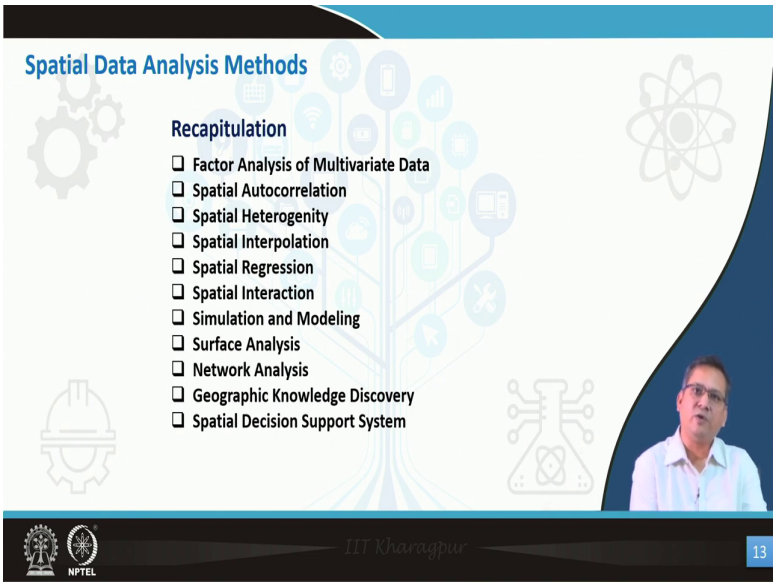
(Refer Slide Time: 24:24)



The next one is the decision support system which might have ingredients of some of the methods that we have already discussed I mean which are listed above which is shown here in orange color.

So, I mean we can integrate these systems into a decision support system and we can have mathematical models for making future predictions and check the impact of physical policy interventions in a urban context so I mean; how these interventions would impact the urban development specifically, the physical and the policy interventions can be studied using this kind of decision support systems.

(Refer Slide Time: 25:15)



So, a recapitulation of the methods that we had discussed today first we had talked about the factor analysis of multivariate data where in we have different types of data and we talk about dimensionality reduction. So, we can have major components and we can only make

significant important maps reduce the number of I mean we can compress the amount of information that we have into significant components.

Now, second we have talked about spatial autocorrelation methods wherein we have different attributes or I mean we may have themes I mean it could be geo spatial themes. So, we can find out what is the autocorrelation between the different themes and is there if there is a spatial connotation to it I mean if there is a trend then we have the spatial heterogeneity where in it is extensively used in ecological models. So, we talk about the dispersion of specific attributes in a geographical landscape.

We have talked about spatial interpolation method which we had say kriging is a I mean method where in we said it we can use geographically rated kriging to interpolate data values when we have say I mean we do not have continuous data sets when we have discrete data set or data observations in a geographical space we can use such kind of methods.

We had talked about spatial regression in which we try to find out the dependence of I mean independent variable on the dependent variables and we can derive a equation to model a given phenomena.

So, most of the I mean applications are mostly in the social domain of spatial regression methods. So, there are different methods of spatial regression as well so we will discuss about that later on. Then we had talked about spatial interaction methods and models we had talked about simulation and modeling approach.

So, I mean we had talked about different types of data models where in we had talked about top down approach wherein we are talking about spatial interaction and we had talked about bottoms-up approach where we were talking about simulation and modeling.

So, we had talked about agent based models wherein agents would be interacting among themselves in a geographical space and it would have a impact on the structure of the given city or it could be on the environment or something else probably. So, then we had talked about surface analysis where in we see the nature of the surface which is geo coded as grid

cells as continuous data values this could be data observations from remotely sensed sensors and this data could be available as raster data.

So, we already know that raster data are I mean it is lattice or grid of cells. So, each cell would have some data value and we can analyze the slope the aspect and I mean the visibility studies using the surface analysis methods then we had talked about network analysis. So, we can model the flow within a given network we can locate facilities in a given network I mean we can find out the optimum location we can do a routing of the I mean of goods in a given network. So, there it could be a different types of network oriented analysis in a geographical space.

Then we have knowledge discovery geographic knowledge discovery methods; in which we had discussed extensively about how we can create I mean information I mean; tangible information from huge data sets among us big data sets I mean so this comes in the domain of big data analysis and we said that the solution is always non trivial it may not be unique so I mean we may work with pertinent data sets and we can I mean we can see which are I mean important for the data analysis.

Then we have the spatial decision support systems; in which we said we can aggregate this different methods and create a decision support system and we can basically I mean have scenarios or assessment of policy approaches or physical planning measures.

And what impact it I may have on the resulting urban scale. So, these are some of the spatial data analysis methods which are very important in the context of urban planning. So, we would be looking into the details of some of these models and methods through examples in the subsequent lectures.

Thank you for being with me.