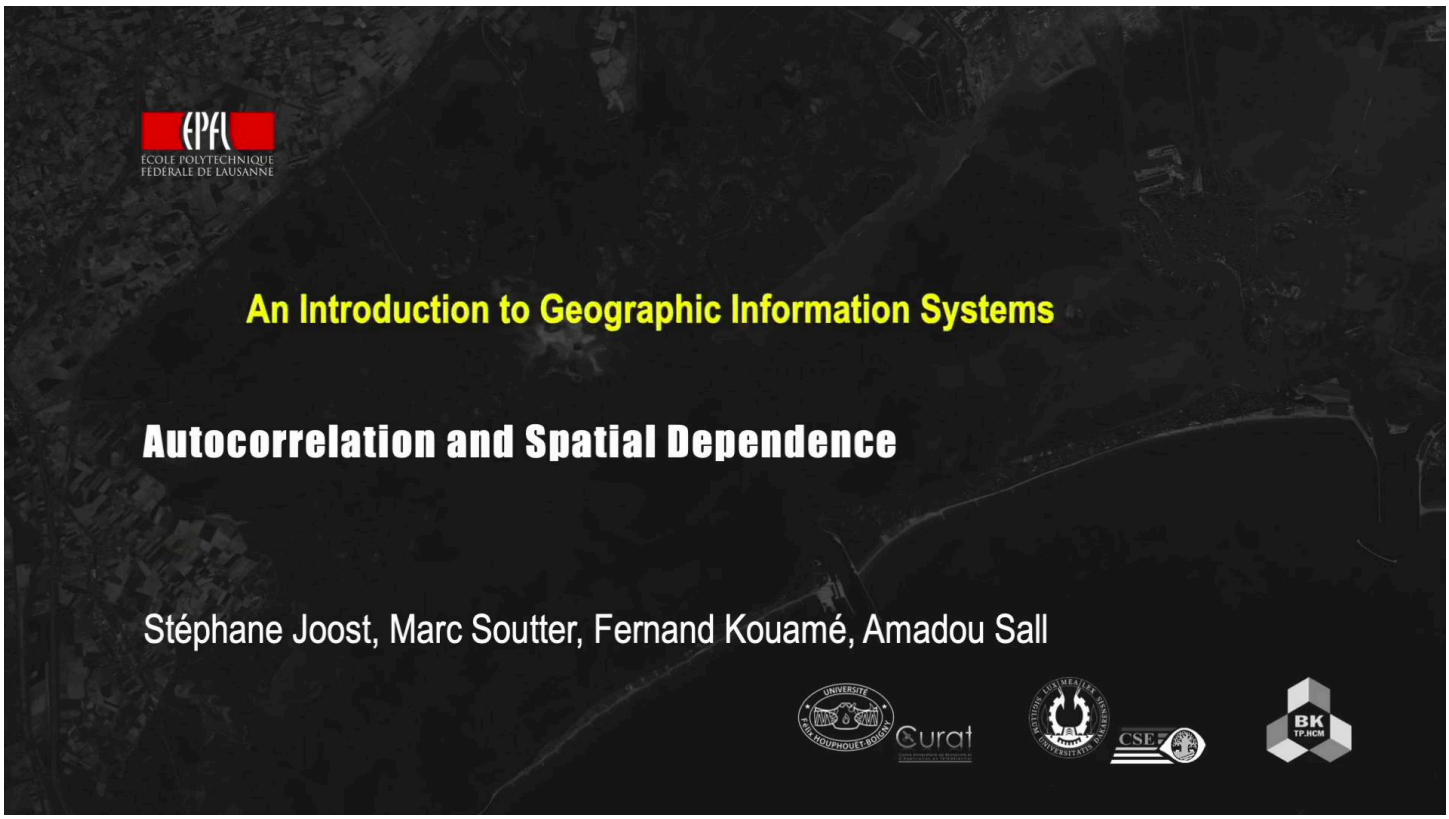
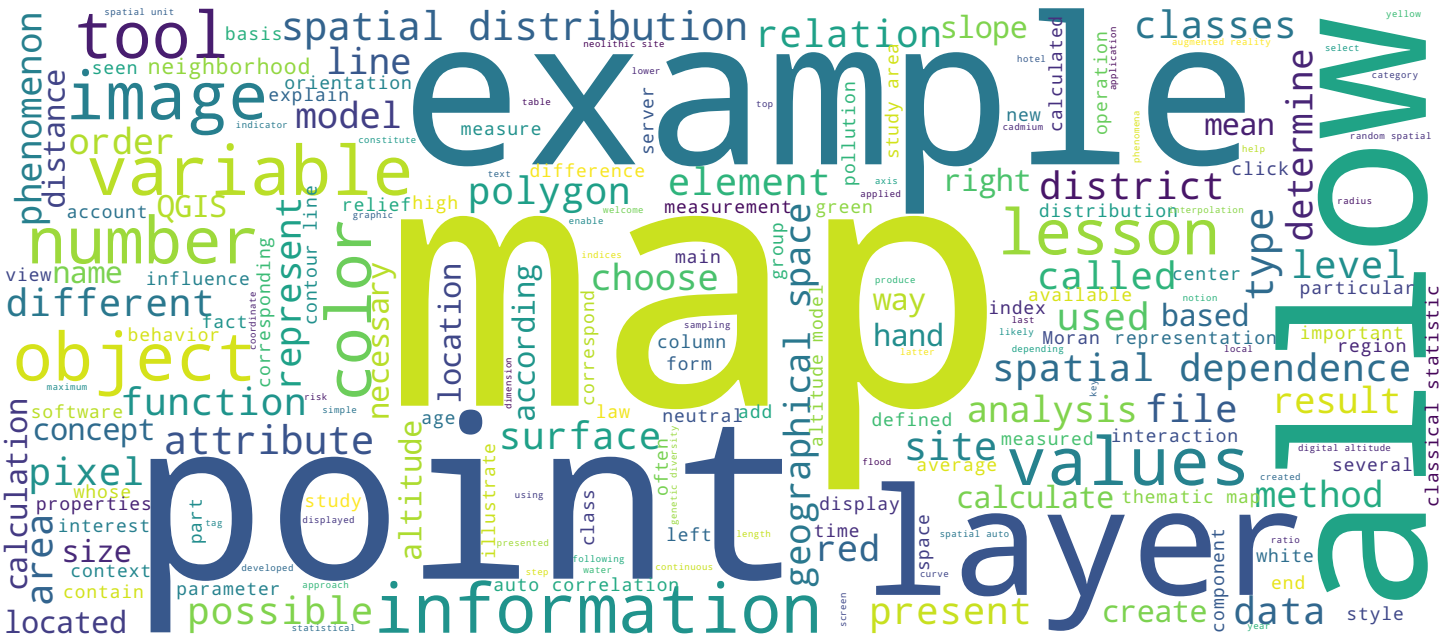
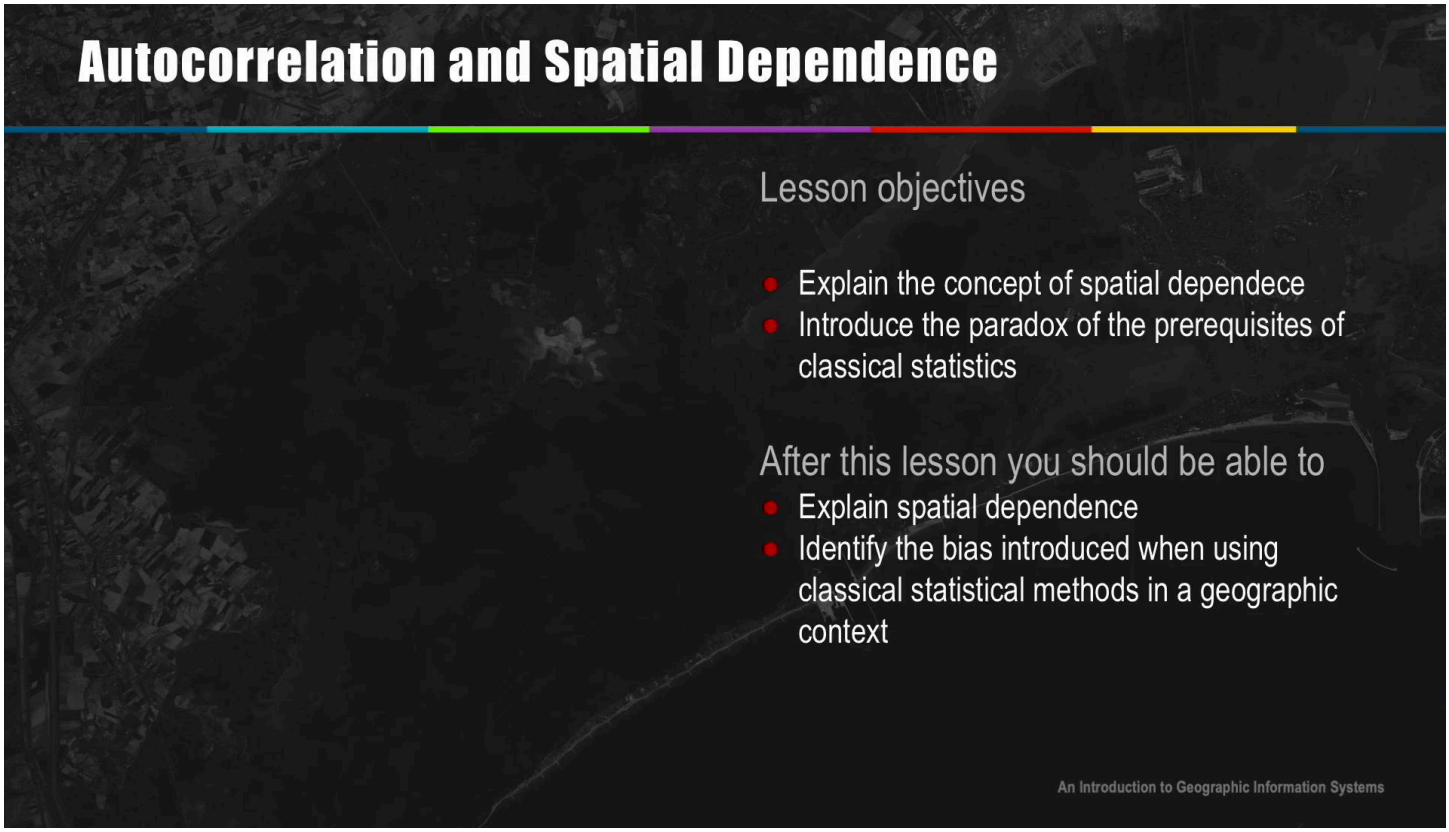




Search MOOC

Video

EPFL



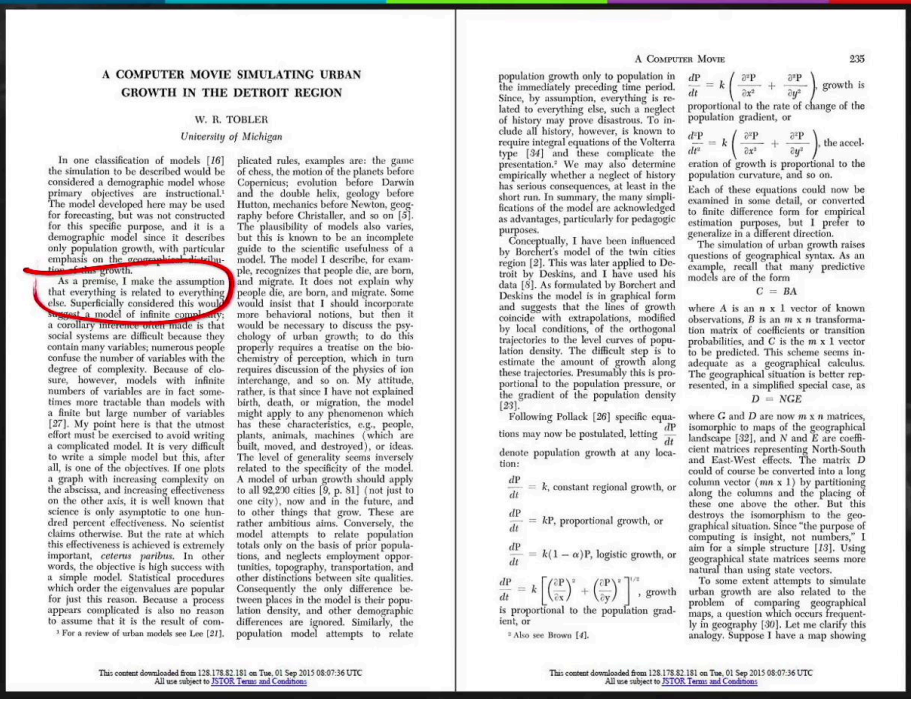
Hello and welcome to this lesson which introduces the concept of spatial auto-correlation. We will mainly talk about spatial dependance, so determine to what extent the value taken by the attribute of an object depends on its geographical position or not This can be the case of the measured temperature on the surface of the leaves of a plant for example. The objectives of this lesson are to explain the concept of spatial dependence and to present the paradox linked to the use of the tools of the classical statistic in a geographical context. The information presented here will allow you to assimilate the concept of spatial dependance, which is a fundamental concept for the measurement of spatial auto-correlation and to recognize the bias induced by the use of classical statistics in this context.

Notes

Summary

0m 31s

Spatial Dependence



A COMPUTER MOVIE SIMULATING URBAN GROWTH IN THE DETROIT REGION

W. R. TOBLER

University of Michigan

In one classification of models [16] the simulation to be described would be considered a demographic model whose primary objectives are instructional.[1] The model developed here may be used for forecasting, but was not constructed for this specific purpose, and it is a demographic model since it describes only population growth, with particular emphasis on the geographic distribution of the growth.

As a premise, I make the assumption that everything is related to everything else. Superficially considered this would suggest a model of infinite complexity, a corollary inference often made is that social systems are difficult because they contain many variables; numerous people confuse the number of variables with the degree of complexity. Because of closure, however, models with infinite numbers of variables are in fact sometimes more tractable than models with a finite but large number of variables [27]. My point here is that the utmost effort must be exercised to avoid writing a complicated model. It is very difficult to write a simple model but this, after all, is one of the objectives. If one plots a graph with increasing complexity on the abscissa, and increasing effectiveness on the other axis, it is well known that science is only asymptotic to one hundred percent effectiveness. No scientist claims otherwise. But the rate at which this effectiveness is achieved is extremely important, *ceteris paribus*. In other words, the objective is high success with a simple model. Statistical procedures which order the eigenvalues are popular for just this reason. Because a process appears complicated is also no reason to assume that it is the result of complicated rules, examples are: the game of chess, the motion of the planets before Copernicus; evolution before Darwin and the double helix, geology before Hutton, mechanics before Newton, geography before Christaller, and so on [5]. The plausibility of models also varies, but this is known to be an incomplete guide to the scientific usefulness of a model. The model I describe, for example, recognizes that people die, are born, and migrate. It does not explain why people die, are born, and migrate. Some would insist that I should incorporate more behavioral notions, but then it would be necessary to discuss the psychology of urban growth; to do this properly requires a treatise on the biochemistry of perception, which in turn requires discussion of the physics of ion interchange, and so on. My attitude, rather, is that since I have not explained birth, death, or migration, the model might apply to any phenomenon which has these characteristics, e.g., people, plants, animals, machines (which are built, moved, and destroyed), or ideas. The level of generality seems inversely related to the specificity of the model. A model of urban growth should apply to all 92,290 cities [9, p. 81] (not just to one city), now and in the future, and to other things that grow. These are rather ambitious aims. Conversely, the model attempts to relate population totals only on the basis of prior populations, and neglects employment opportunities, topography, transportation, and other distinctions between site qualities. Consequently the only difference between places in the model is their population density, and other demographic differences are ignored. Similarly, the population model attempts to relate

[1] For a review of urban models see Lee [21].

population growth only to population in the immediately preceding time period. Since, by assumption, everything is related to everything else, such a neglect of history may prove disastrous. To include all history, however, is known to require integral equations of the Volterra type [34] and these complicate the presentation.[2] We may also determine empirically whether a neglect of history has serious consequences, at least in the short run. In summary, the many simplifications of the model are acknowledged as advantages, particularly for pedagogic purposes.

Conceptually, I have been influenced by Borchert's model of the twin cities region [2]. This was later applied to Detroit by Deskins, and I have used his data [5]. As formulated by Borchert and Deskins the model is in graphical form and suggests that the lines of growth coincide with extrapolations, modified by local conditions, of the orthogonal trajectories to the level curves of population density. The difficult step is to estimate the amount of growth along these trajectories. Presumably this is proportional to the population pressure, or the gradient of the population density [23].

Following Pollack [26] specific equations may now be postulated, letting $\frac{dP}{dt}$ denote population growth at any location:

$\frac{dP}{dt} = k$, constant regional growth, or

$\frac{dP}{dt} = kP$, proportional growth, or

$\frac{dP}{dt} = k(1-\alpha)P$, logistic growth, or

$\frac{dP}{dt} = k\left[\left(\frac{\partial P}{\partial x}\right)^2 + \left(\frac{\partial P}{\partial y}\right)^2\right]^{1/2}$, growth is proportional to the population gradient, or

[2] Also see Brown [4].

$\frac{dP}{dt} = k\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2}\right)$, growth is proportional to the rate of change of the population gradient, or

$\frac{d^2P}{dt^2} = k\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2}\right)$, the acceleration of growth is proportional to the population curvature, and so on.

Each of these equations could now be examined in some detail, or converted to finite difference form for empirical estimation purposes, but I prefer to generalize in a different direction.

The simulation of urban growth raises questions of geographical syntax. As an example, recall that many predictive models are of the form

$$C = BA$$

where A is an n x 1 vector of known observations, B is an m x n transformation matrix of coefficients or transition probabilities, and C is the m x 1 vector to be predicted. This scheme seems inadequate as a geographical calculus. The geographical situation is better represented, in a simplified special case, as

$$D = NGE$$

where G and D are now m x n matrices, isomorphic to maps of the geographical landscape [32], and N and E are coefficient matrices representing North-South and East-West effects. The matrix D could of course be converted into a long column vector (mn x 1) by partitioning along the columns and the placing of these one above the other. But this destroys the isomorphism to the geographical situation. Since "the purpose of computing is insight, not numbers," I aim for a simple structure [13]. Using geographical state matrices seems more natural than using state vectors.

To some extent attempts to simulate urban growth are also related to the problem of comparing geographical maps, a question which occurs frequently in geography [30]. Let me clarify this analogy. Suppose I have a map showing

«As a premise, I make the assumption that everything is related to everything else»

An Introduction to Geographic Information Systems

It is a quantitative geographer, a Swiss-American, named Waldo Tobler who best described this concept by enunciating what is called the first law of geography in an article written in 1970. According to this law, everything interacts with everything in the geographical space, but two close objects are more likely to do so than two distant objects.

Notes

Summary

1m 33s

Spatial Dependence

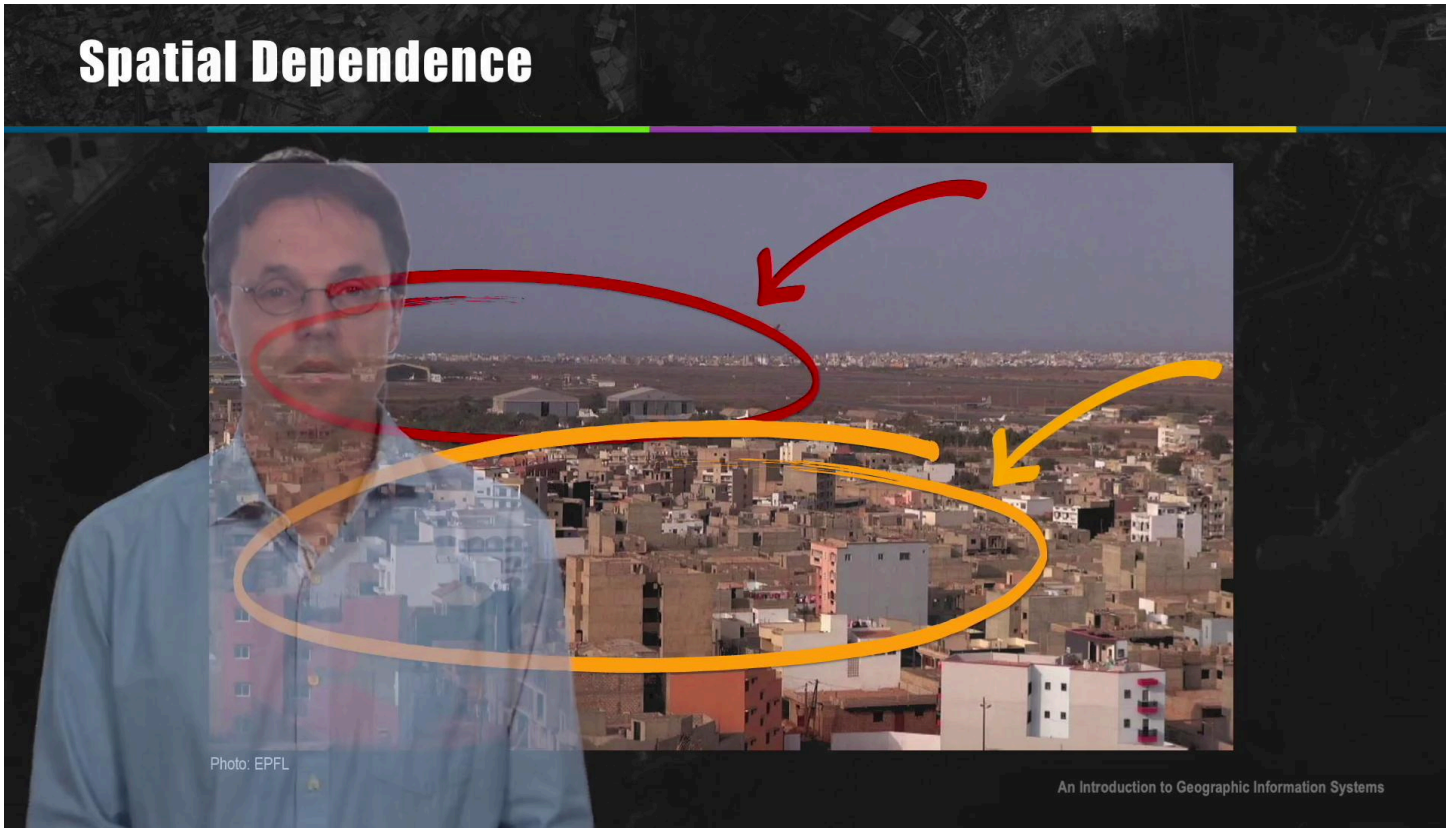
Photo: copa2014.gov.br / CC BY 3.0

To illustrate this notion of spatial dependence, here is a first example related to football. we will focus our attention on brazilian and croatian supporters during the opening match of the 2014 world cup between Brazil and Croatia. In this image, an individual dressed in yellow has more chance to interact with another person wearing the same color. And in the same way a person in red and white has a better chance of interacting with another person in red and white. The membership of a same group determined the spatial distribution of these people and the spatial dependence induced by this membership is perceptible in the geographical space thanks to the color of the t-shirts.

Notes

Summary

1m 55s



Another example, here in Dakar, around the international airport, which we see in the image. The commercial and logistical activities linked to the activity of the airport are grouped around it, whereas other activities, such as the residential area, for example, which we see in the foreground, are gathered elsewhere on the territory. Around the airport, the buildings are similar, there are large warehouses or hangars. And this is also the case in the residential area. The houses have a similar appearance and size. The spatial proximity favours the interaction between objects of the same category and the nature of the activity is betrayed by a similar appearance. The spatial dependence highlighted by the two examples which we have just seen can be measured with simple tools that you will learn to manipulate. The functioning of these tools is based on the comparison between an observed spatial distribution and a random spatial distribution.

Notes

Summary

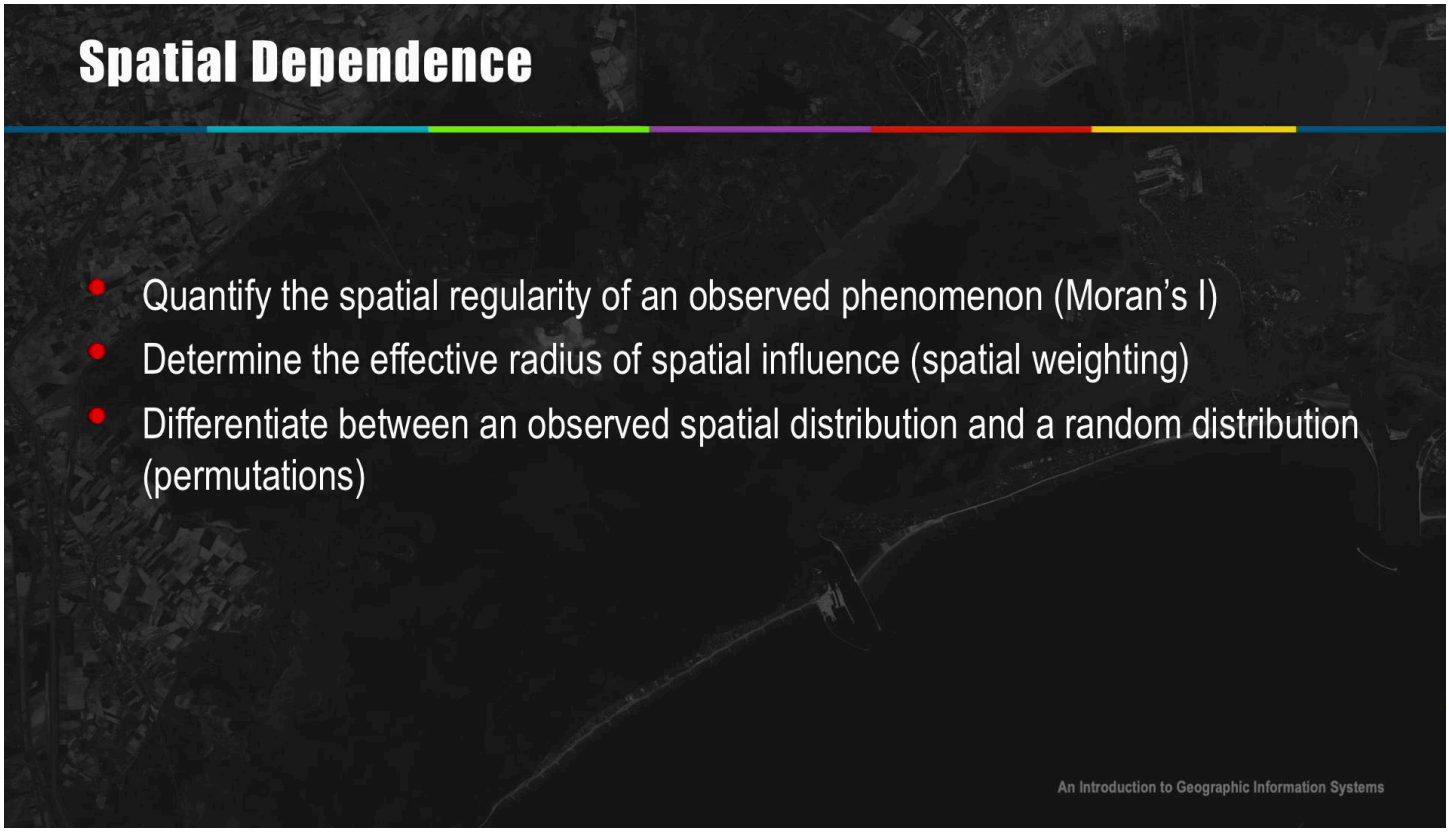
2m 40s

On this regular grid of 10 × 10 cells, we have represented the spatial distribution of a phenomenon. There is a particular arrangement of the displayed values, which indicates a certain spatial dependence. On the right, we have illustrated the random spatial distribution of the same values but in several configurations. On the left, the geographical space is not neutral. In some ways, it fixes certain values in specific places. On the right, it is neutral. Any location in space can take all possible values.

Notes

Summary

3m 43s

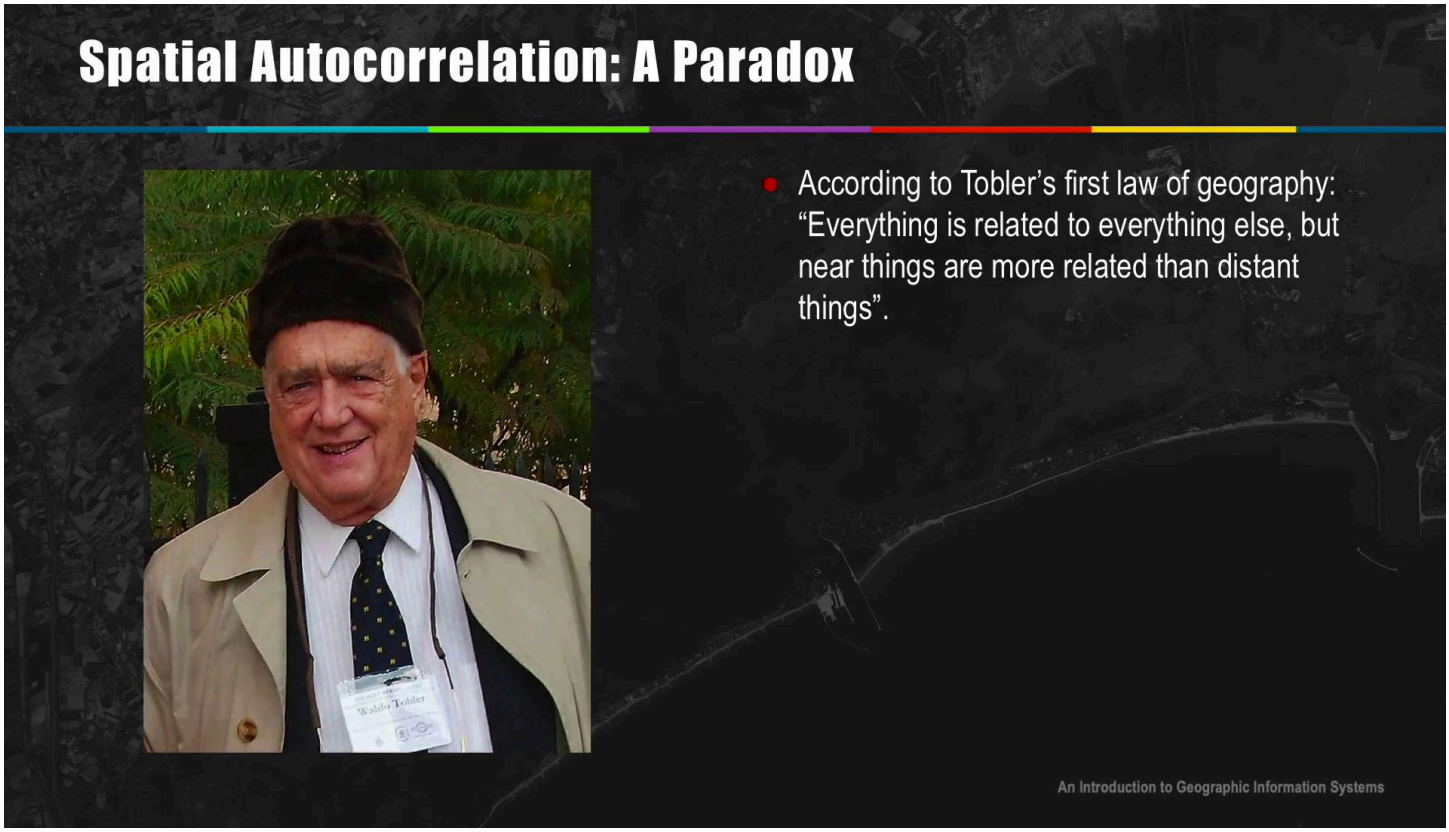


On this base, the measurement tools that we will use allow firstly to quantify the spatial regularity of a phenomenon then to determine the radius of action of the spatial dependence and finally to differentiate an observed spatial distribution from a random spatial distribution.

Notes

Summary

4m 23s



Let's consider real data now. This map shows 765 neolithic sites, they have been dated and their age is between 6,000 and 8,000 years before the present. The darker the green of the point, the older the neolithic site. We notice that the oldest sites are concentrated in the fertile crescent region and moving gradually towards the North-West, we find sites that are younger. This is an example of spatial dependence that illustrates human migration towards the North after the end of the last major ice age. We will reuse this example a little later. From the moment we want to quantify this spatial dependence, we are confronted with a paradox. Indeed, according to Tobler and the first law of geography, everything interacts with everything but close objects are more likely to do so than distant objects.

Notes

Summary

4m 43s

Spatial Autocorrelation: A Paradox

- Natural phenomena (air temperature), or socio-economic factors (population population) are not randomly distributed in geographic space

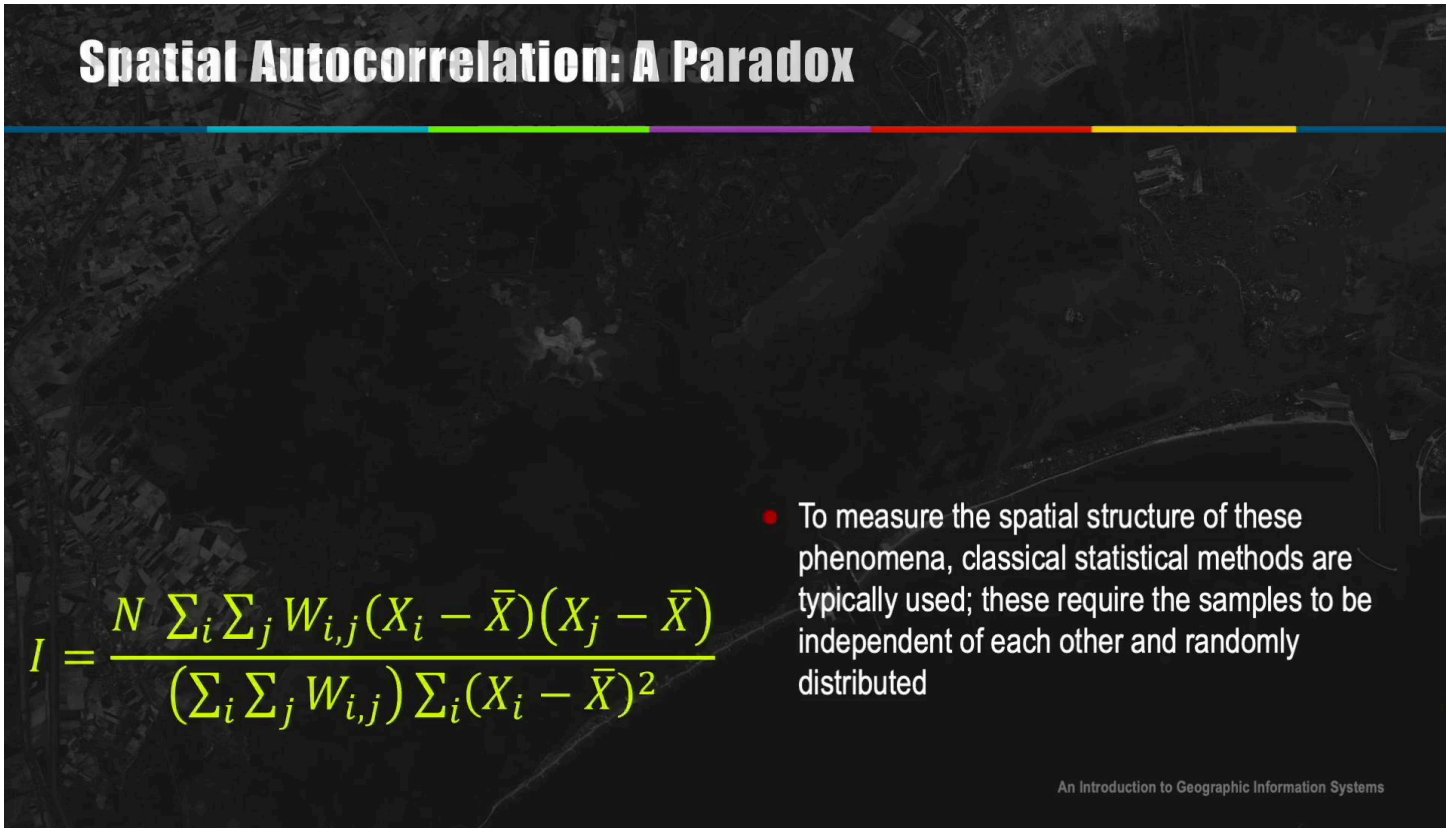
An Introduction to Geographic Information Systems

So natural phenomena, such as air temperature, or socio-demographic phenomena, such as the population density, are not randomly distributed in the geographical space.

Notes

Summary

5m 49s

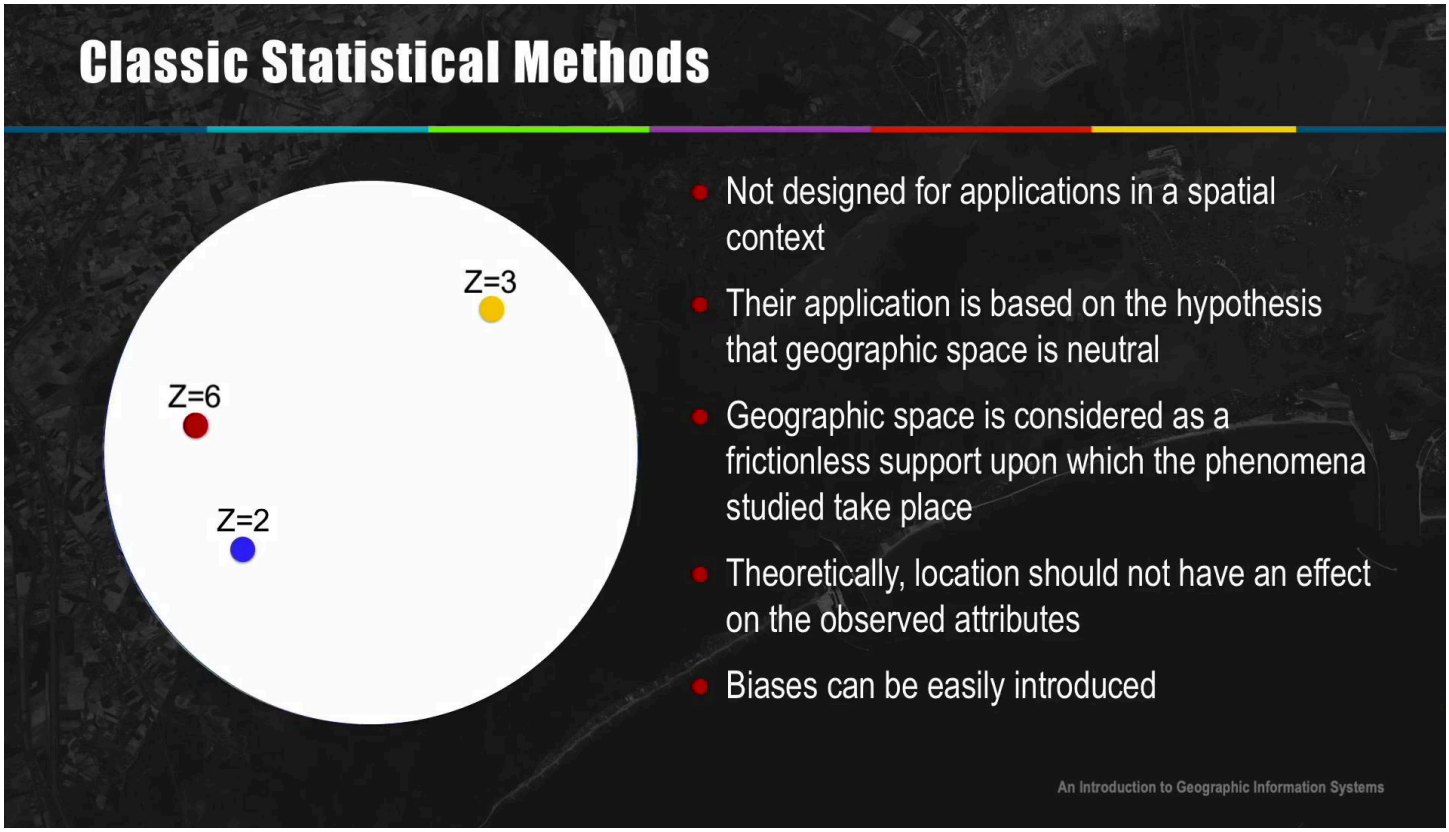


But to measure the spatial structure of these phenomena, we have to use tools of classical statistics. And these tools require on one hand the independence between the samples and on the other hand a random distribution of the latter, there is therefore a contradiction.

Notes

Summary

6m 02s



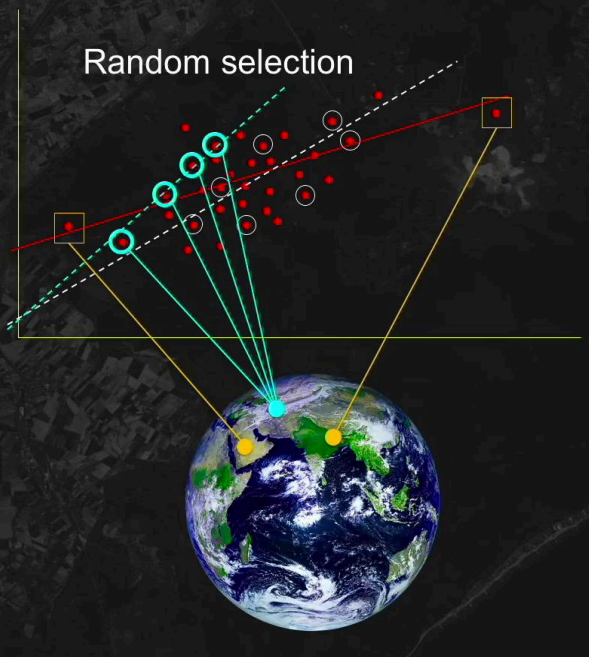
This contradiction is due to the fact that the tools of the classical statistics are not intended to be applied in a geospatial context. Their use is based on the assumption that geographical space is neutral This geographical area constitutes the simple support, without friction, on which the studied phenomena take place. Theoretically, in this context, the location of observations in space must not influence their attributes. However, since there are often no alternative tools, we should use them with the awareness of the biases induced by their use with geographic data and adapt the datasets to meet the statistical prerequisites.

Notes

Summary

6m 20s

An Example of Bias: Linear Regression



- Linear regression: calculated on randomly selected observations
- If the observations are spatially dependent, their predicted values will be biased for the entire study area
- Extreme values that are located in particular sub-regions can influence the predicted values for the entire territory
- A strong correlation between two sample attributes that are located in a small sub-region will have a significant effect on the entire study region

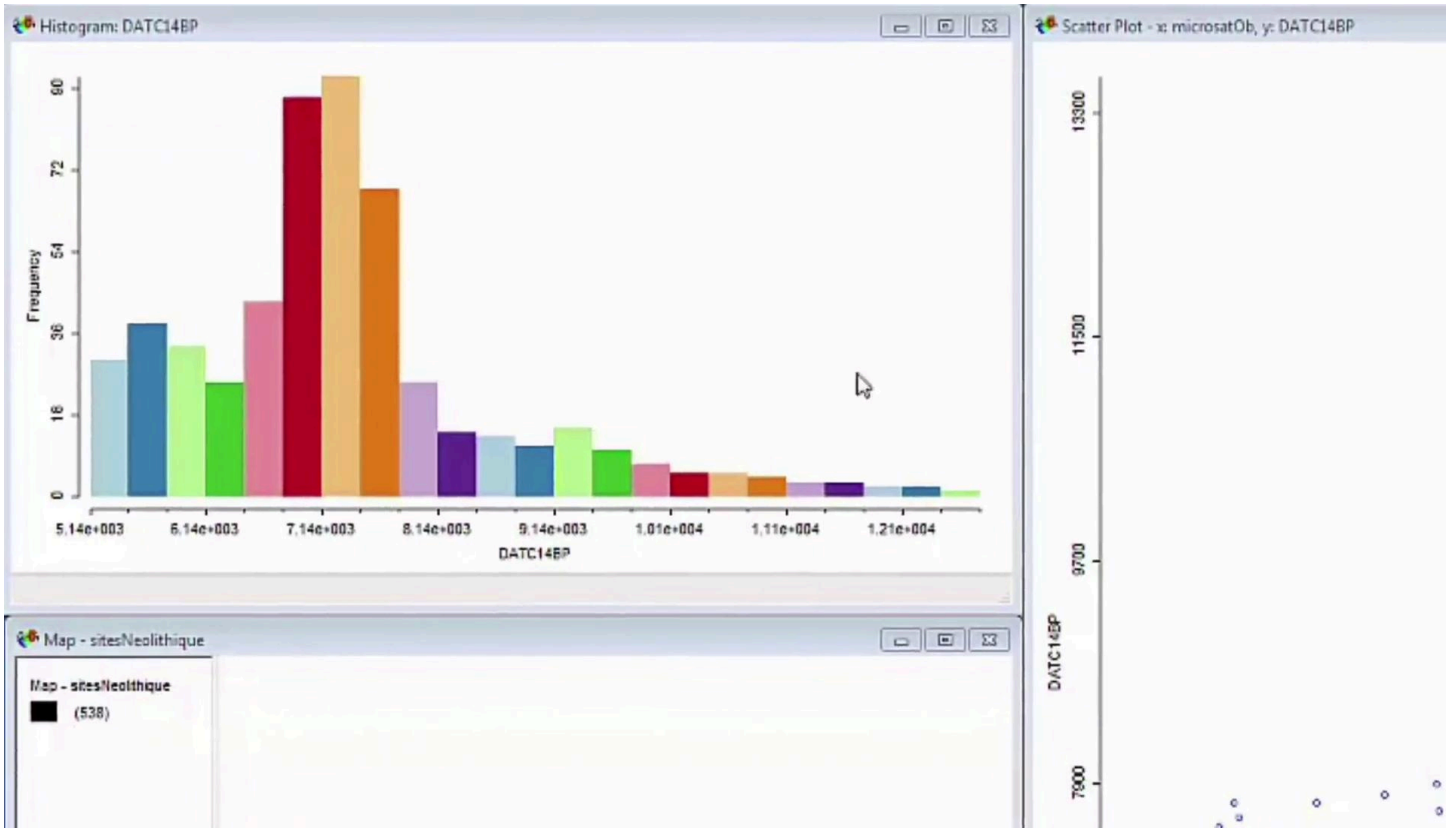
An Introduction to Geographic Information Systems

A good example is the linear regression. Theoretically, it must be calculated with observations selected according to a random procedure. Indeed, if the observations are spatially dependent, the estimated values will be biased for the entire study area, because exceptional values, located in geographical subregions, will influence the predicted values on all the analyzed territory. Or furthermore, a strong correlation between two sample attributes located in a same small subregion will influence the measurement on all the studied area. We will now see a practical example to illustrate this type of bias.

Notes

Summary

7m 06s

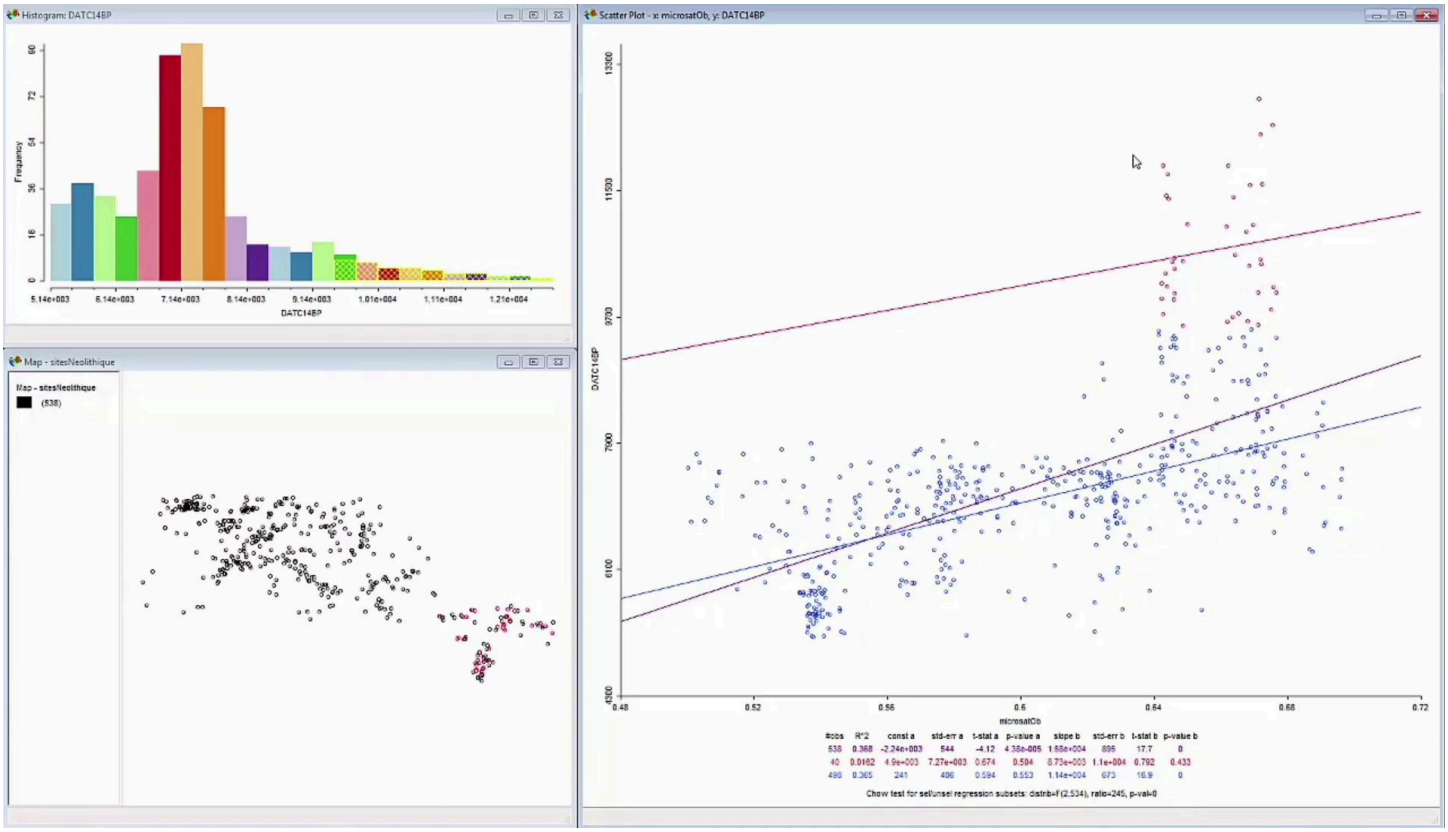


Let's take again the example of the 765 neolithic sites. In the GeoDa software, we created three views.

Notes

Summary

7m 52s



At the top, on the left, we have the histogram of the age class distribution of the sites. At the bottom, on the left, the spatial distribution of the 765 sites. And on the right, a graph that illustrates the relation between the age of the sites on the Y axis and a genetic diversity variable on the X axis that characterizes goat populations sampled around the sites at the beginning of the 21st century. This relationship is of interest because it confirms the hypothesis that the neolithic populations of the fertile crescent began to migrate towards the North-West from the end of the last great ice age. The lower the age of the sites, the lower the genetic diversity, since the animal populations have progressively fragmented and the reproduction has subsequently taken place between related individuals. The regression line shows that the older the age of the sites, the higher the genetic diversity, but this relation is strongly influenced by a group of sites concentrated in the Middle East. Indeed, if this group of sites is removed from the calculation, the slope of the regression decreases noticeably. The values predicted by the model across Europe are largely influenced by about forty points, all located in the same subregion.

Notes

Summary

8m 01s