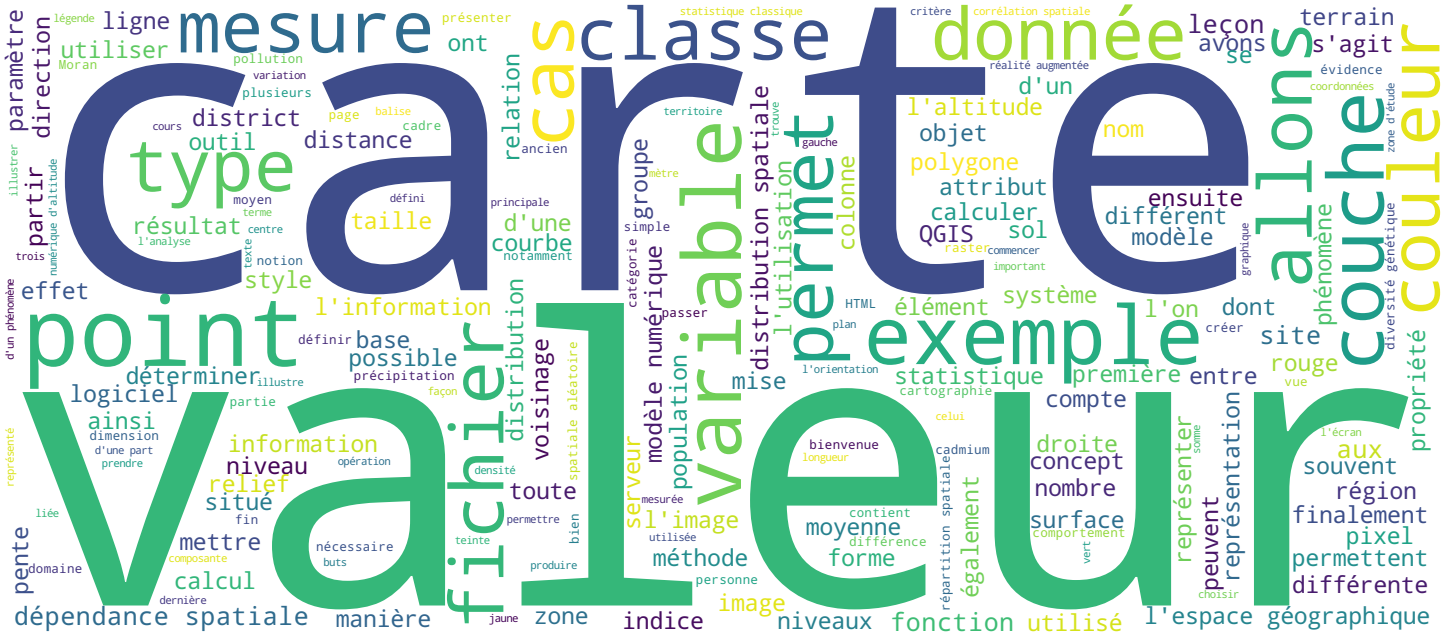
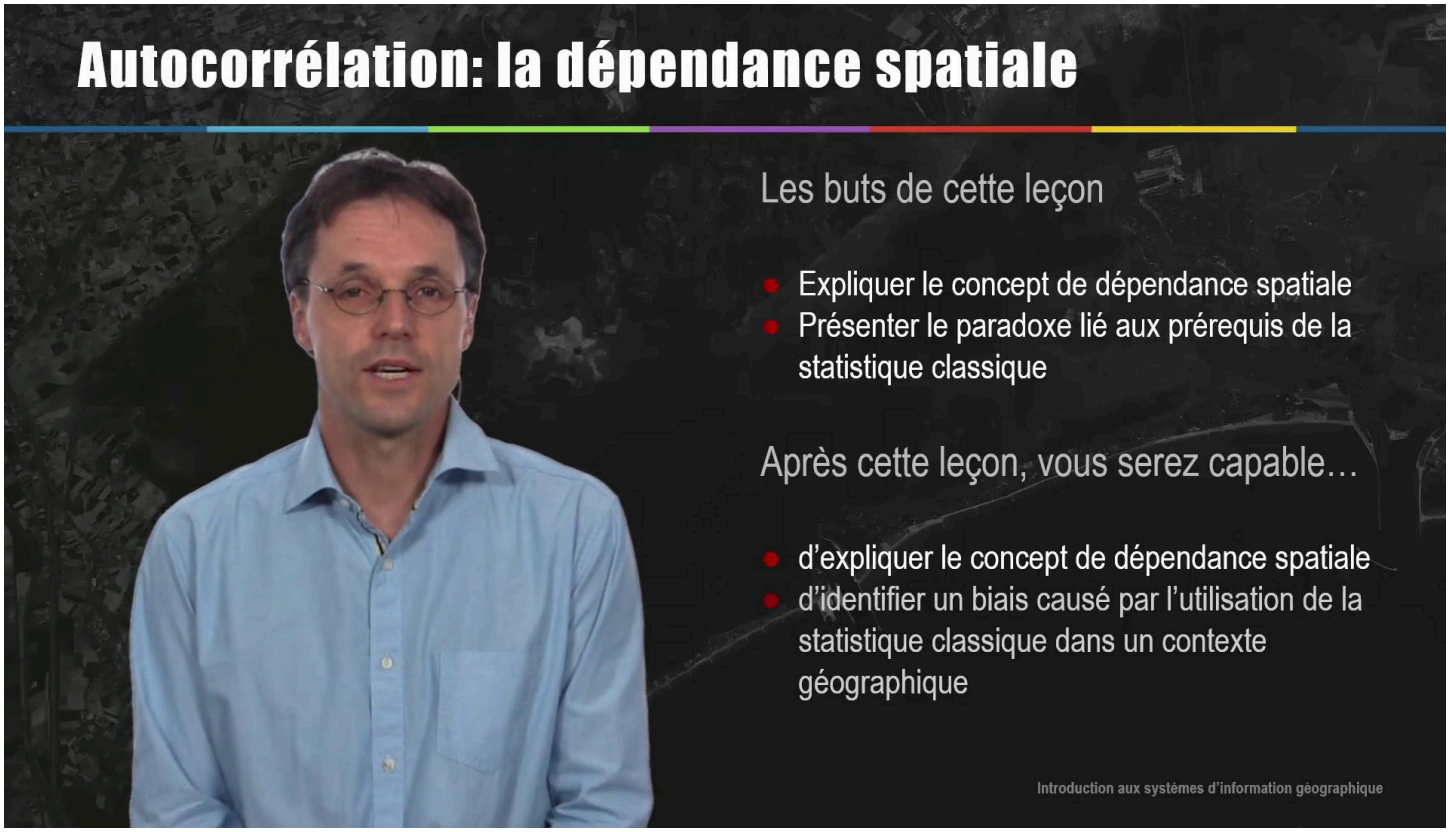


Search MOOC

Video

EPFL



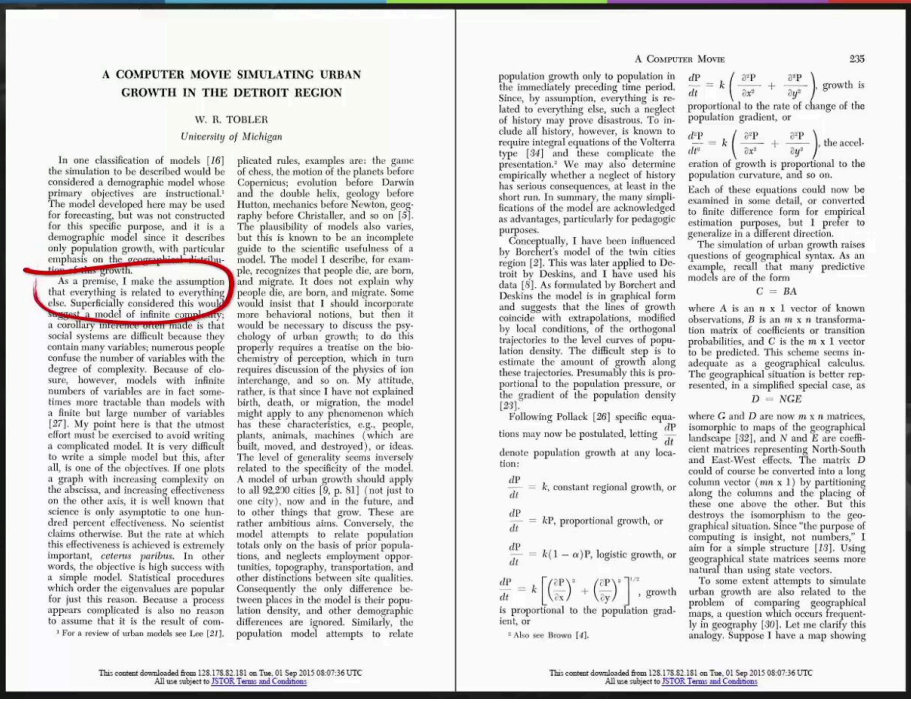
Bonjour et bienvenue dans cette leçon qui introduit le concept d'auto-corrélation spatiale. Nous allons principalement parler de dépendance spatiale, soit de déterminer dans quelle mesure la valeur prise par l'attribut d'un objet dépend de sa position géographique ou pas. Cela peut être le cas de la température mesurée à la surface des feuilles d'une plante par exemple. Les buts de cette leçon sont d'expliquer le concept de dépendance spatiale et de présenter le paradoxe lié à l'utilisation des outils de la statistique classique dans un contexte géographique. Les informations présentées ici vous permettront d'assimiler le concept de dépendance spatiale, qui est un concept fondamental pour la mesure de l'auto-corrélation spatiale, et de reconnaître le biais induit par l'utilisation des statistiques classiques dans ce contexte.

Notes

Summary

0m 31s

Dépendance spatiale



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 geographical distribution of that 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 92,200 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»

Introduction aux systèmes d'information géographique

C'est un géographe quantitativiste américano-suisse, appelé Waldo Tobler, qui a le mieux décrit ce concept en énonçant ce qu'on appelle la première loi de la géographie dans un article écrit en 1970. Selon cette loi, tout interagit avec tout dans l'espace géographique, mais deux objets proches ont plus de chance de le faire que deux objets éloignés.

Notes

Summary

1m 33s

Dépendance spatiale

Introduction aux systèmes d'information géographique

Pour illustrer cette notion de dépendance spatiale, voici un premier exemple lié au football.

Notes

Summary

1m 55s

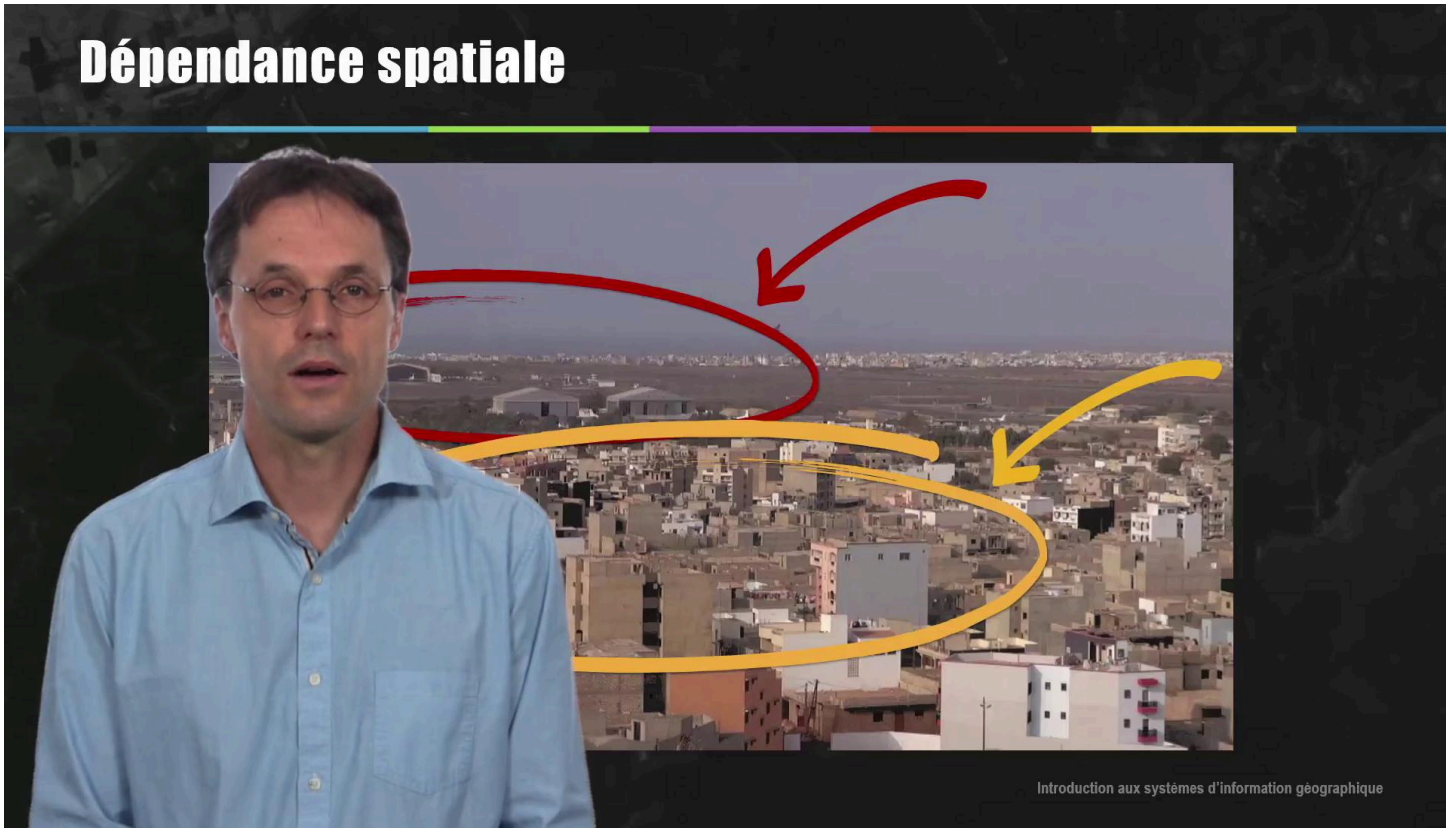
Dépendance spatiale

Nous allons porter notre attention sur les supporters brésiliens et croates lors du match d'ouverture de la coupe du monde 2014 entre le Brésil et la Croatie. Sur cette image, un individu vêtu de jaune a plus de chance d'interagir avec une autre personne portant la même couleur. Et de la même façon, une personne en rouge et blanc a plus de chance d'interagir avec une autre personne en rouge et blanc. L'appartenance à un même groupe a déterminé la distribution spatiale de ces personnes et la dépendance spatiale induite par cette appartenance est perceptible dans l'espace géographique grâce à la couleur des t-shirts. Un autre exemple, ici à Dakar, autour de l'aéroport international, que l'on voit à l'image. Les activités commerciales et logistiques qui sont liées à l'activité de l'aéroport sont regroupées dans ses alentours, alors que d'autres activités, comme le résidentiel par exemple que l'on voit au premier plan, sont regroupées ailleurs sur le territoire. Autour de l'aéroport, les bâtiments se ressemblent, on trouve des entrepôts ou des hangars de grande surface. Et c'est aussi le cas dans la zone résidentielle. Les maisons ont une apparence et une taille comparables.

Notes

Summary

2m 00s

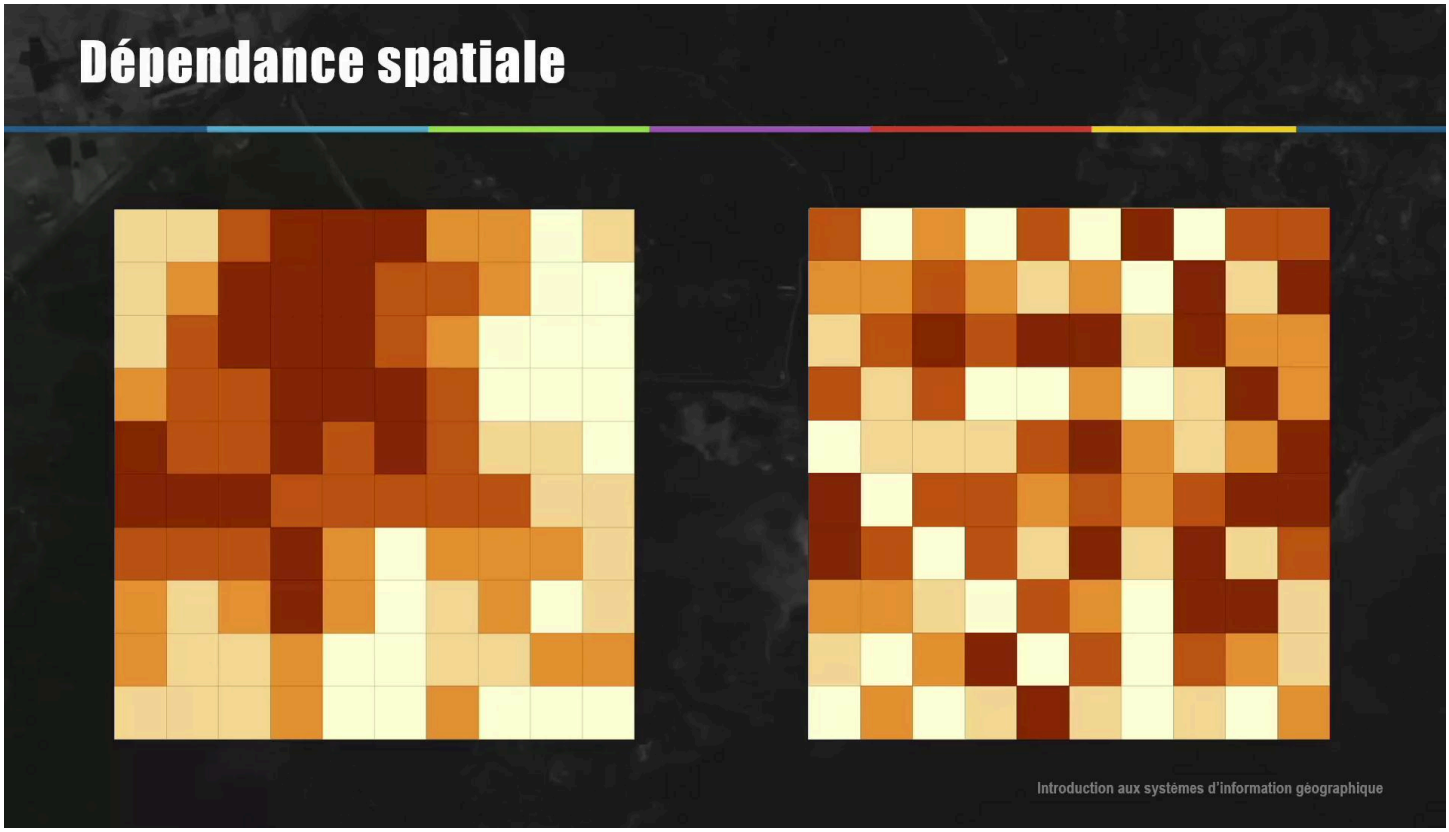


La proximité spatiale favorise l'interaction entre objets de la même catégorie et la nature de l'activité est trahie par une apparence similaire. La dépendance spatiale mise en évidence par les deux exemples que nous venons de voir peut être mesurée avec des outils simples que vous allez apprendre à manipuler. Le fonctionnement de ces outils est basé sur la comparaison entre une répartition spatiale observée et une répartition spatiale aléatoire.

Notes

Summary

3m 16s

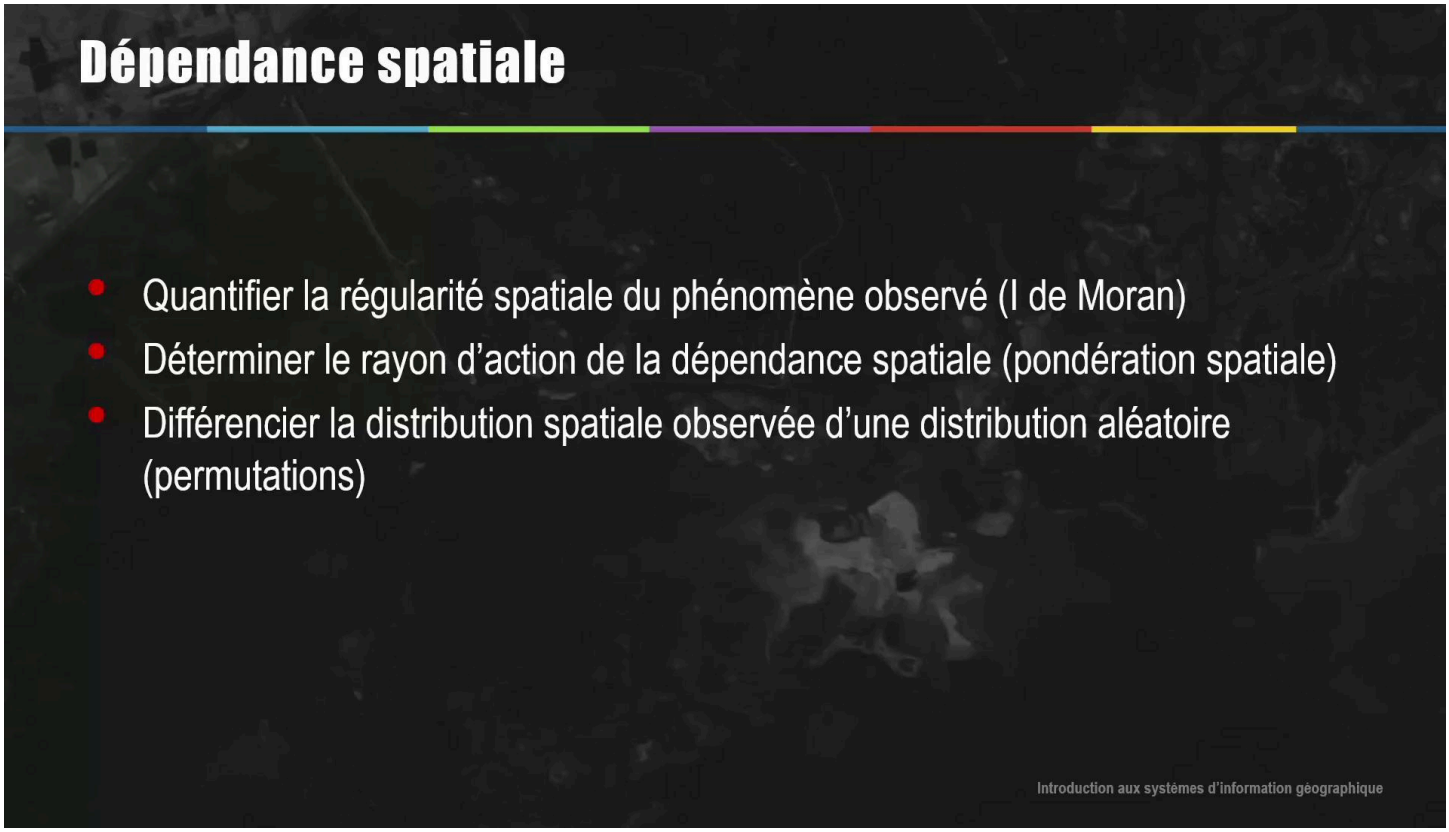


Sur cette grille régulière de 10 x 10 cellules, on a représenté la distribution spatiale d'un phénomène. On remarque un agencement particulier des valeurs affichées, ce qui dénote une certaine dépendance spatiale. A droite, nous avons illustré la répartition spatiale aléatoire des mêmes valeurs mais dans plusieurs configurations. À gauche, l'espace géographique n'est pas neutre. Celui-ci fixe, d'une manière ou d'une autre, certaines valeurs à des endroits précis. À droite, il est neutre. Toute localisation dans l'espace peut prendre toutes les valeurs possibles.

Notes

Summary

3m 43s

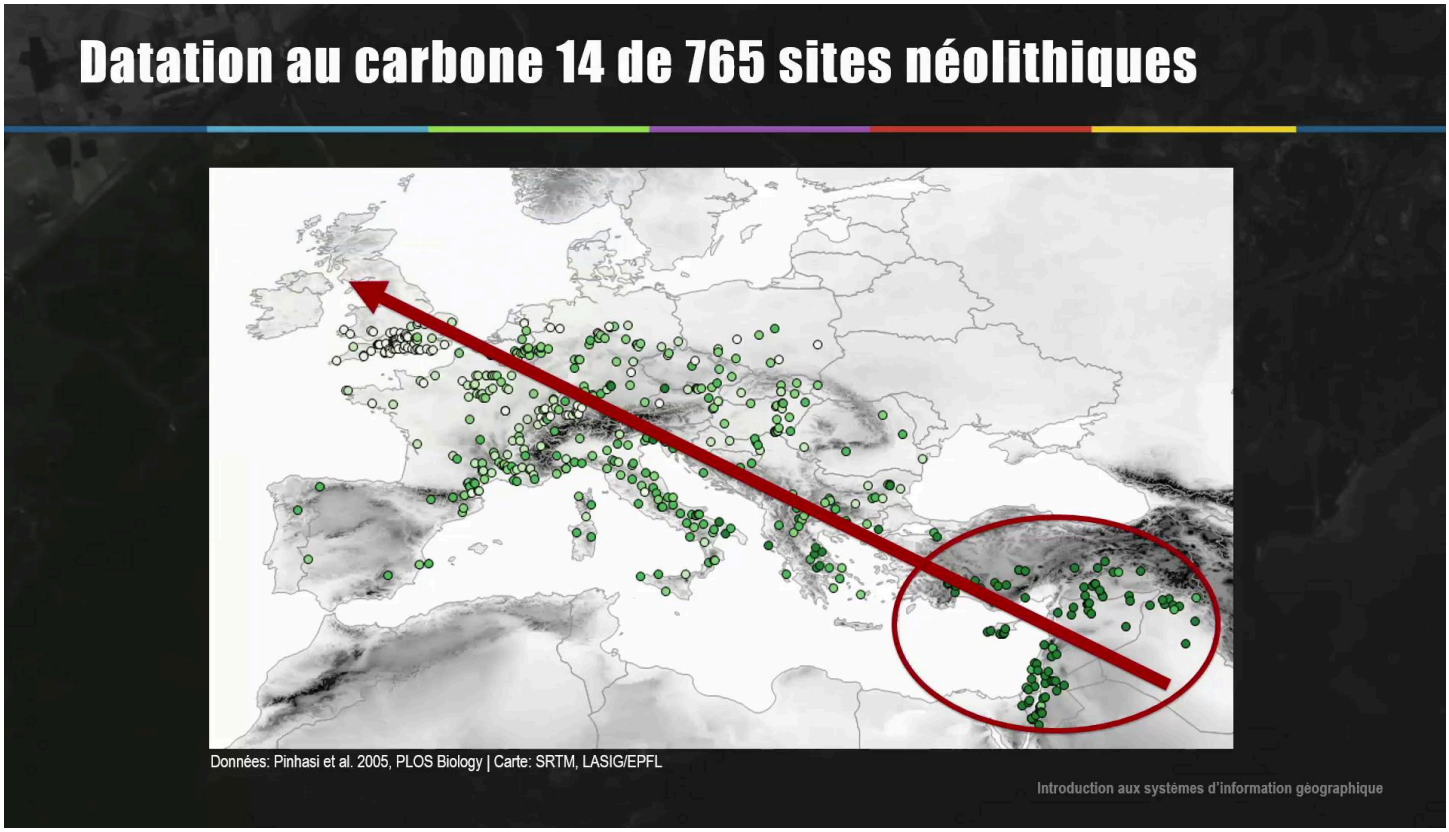


Sur cette base, les outils de mesure que nous allons utiliser permettent premièrement, de quantifier la régularité spatiale d'un phénomène, ensuite, de déterminer le rayon d'action de la dépendance spatiale et finalement, de différencier une distribution spatiale observée d'une distribution spatiale aléatoire.

Notes

Summary

4m 23s



Considérons des données réelles maintenant. Cette carte montre 765 sites néolithiques, ils ont été datés et leur âge est compris entre 6'000 et 8'000 ans avant le présent. Plus le vert du point est foncé, plus le site néolithique est ancien. On constate que les sites les plus anciens sont concentrés dans la région du croissant fertile et en se dirigeant progressivement vers le nord-ouest, on trouve des sites qui sont plus jeunes. C'est un exemple de dépendance spatiale qui illustre ici les migrations humaines en direction du nord après la fin de la dernière importante période glaciaire. Nous réutiliserons cet exemple un peu plus tard.

Notes

Summary

4m 43s

Mesure de l'autocorrélation spatiale: un paradoxe

- Selon Tobler et la première loi de la géographie «Tout interagit avec tout, mais les objets proches ont plus de chances de le faire que des objets éloignés»

Introduction aux systèmes d'information géographique

Dès le moment où l'on désire quantifier cette dépendance spatiale, on est confronté à un paradoxe. En effet, selon Tobler et la première loi de la géographie, tout interagit avec tout mais les objets proches ont plus de chance de le faire que des objets éloignés.

Notes

Summary

5m 32s

Mesure de l'autocorrélation spatiale: un paradoxe

- Les phénomènes naturels (température de l'air), ou socio-démographiques (densité de population) ne sont pas distribués au hasard dans l'espace géographique

Introduction aux systèmes d'information géographique

Donc les phénomènes naturels, comme la température de l'air, ou des phénomènes socio-démographiques, comme la densité de population, ne sont pas distribués au hasard dans l'espace géographique.

Notes

Summary

5m 49s

Mesure de l'autocorrélation spatiale: un paradoxe

$$I = \frac{N \sum_i \sum_j W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{\left(\sum_i \sum_j W_{i,j}\right) \sum_i (X_i - \bar{X})^2}$$

- Pour mesurer la structure spatiale de ces phénomènes, on doit utiliser des outils de la statistique classique qui requièrent l'indépendance entre les échantillons et une distribution aléatoire de ces échantillons

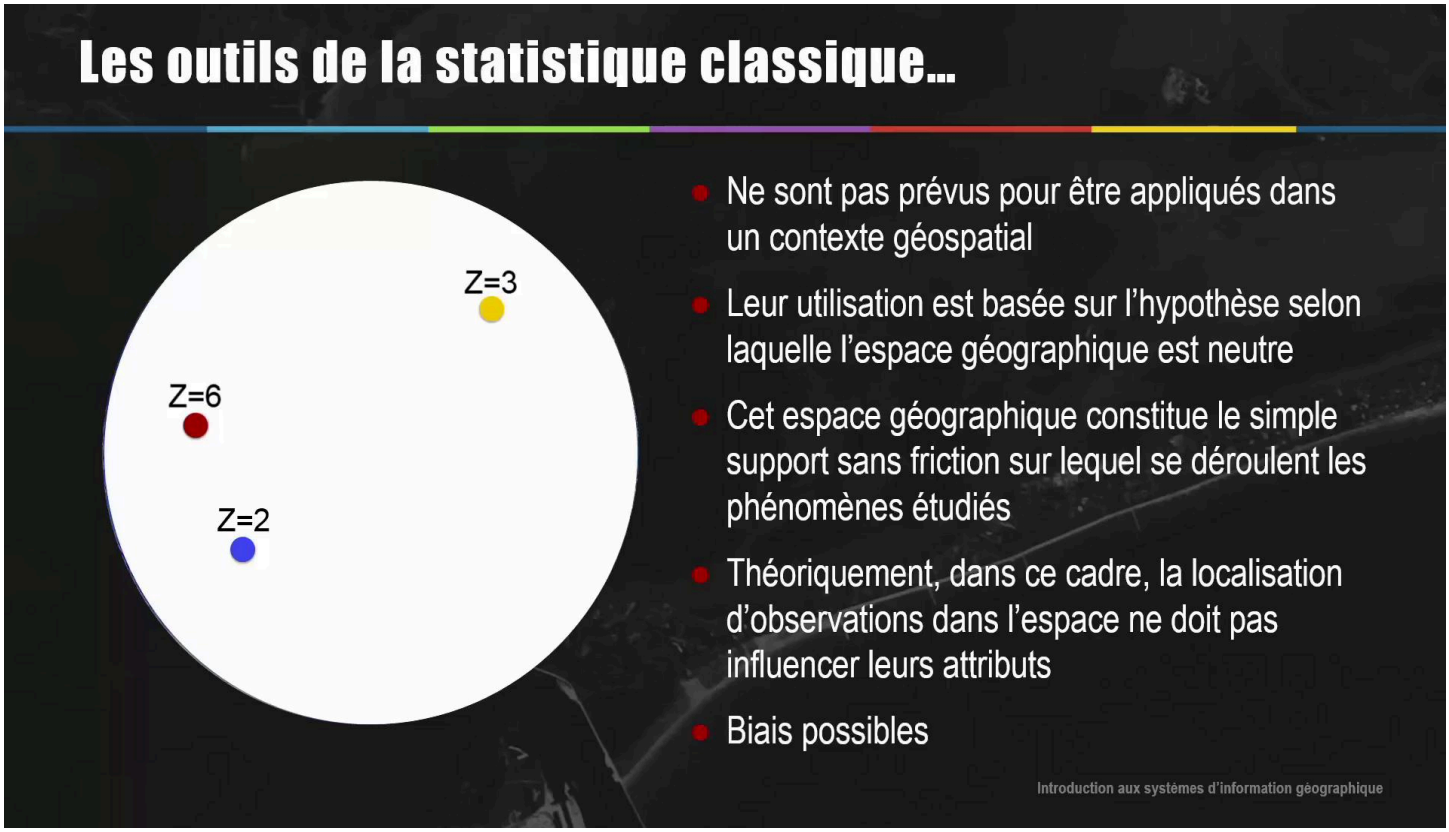
Introduction aux systèmes d'information géographique

Mais pour mesurer la structure spatiale de ces phénomènes, on doit utiliser des outils de la statistique classique. Et ces outils requièrent d'une part l'indépendance entre les échantillons et d'autre part une distribution aléatoire de ces derniers, il y a donc une contradiction.

Notes

Summary

6m 02s



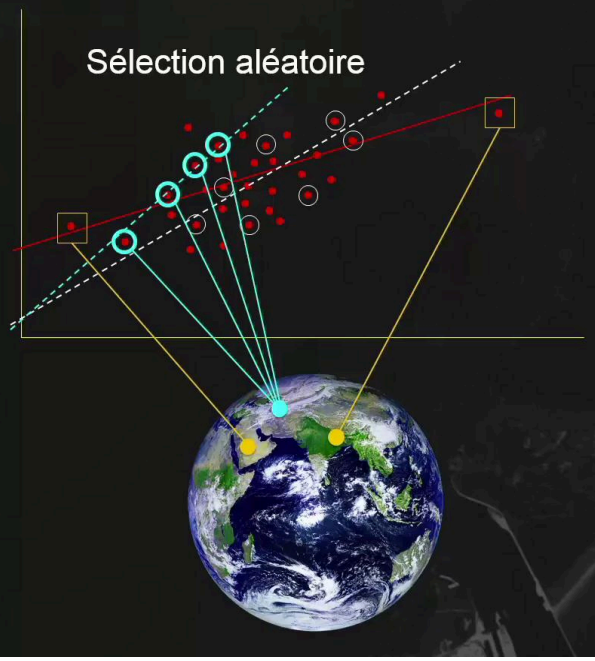
Cette contradiction est due au fait que les outils de la statistique classique ne sont pas prévus pour être appliqués dans un contexte géospatial. Leur utilisation est basée sur l'hypothèse selon laquelle l'espace géographique est neutre. Cet espace géographique constitue le simple support, sans friction, sur lequel se déroulent les phénomènes étudiés. Théoriquement, dans ce cadre, la localisation d'observations dans l'espace ne doit pas influencer leurs attributs. Mais comme il n'existe souvent pas d'outils alternatifs, on doit les utiliser en étant conscient des biais induits par leur utilisation avec des données géographiques et adapter les jeux de données pour respecter les prérequis de la statistique.

Notes

Summary

6m 20s

Un exemple de biais: la régression linéaire



- Régression linéaire : calculée avec des observations sélectionnées selon une procédure aléatoire
- Si les observations sont spatialement dépendantes, les valeurs estimées sont biaisées pour toute la zone d'étude
- Des valeurs exceptionnelles localisées dans des sous-regions particulières influencent les valeurs prédites sur tout le territoire analysé
- Une forte corrélation entre deux attributs d'échantillons situés dans une petite sous-région aura un effet sur toute la zone étudiée

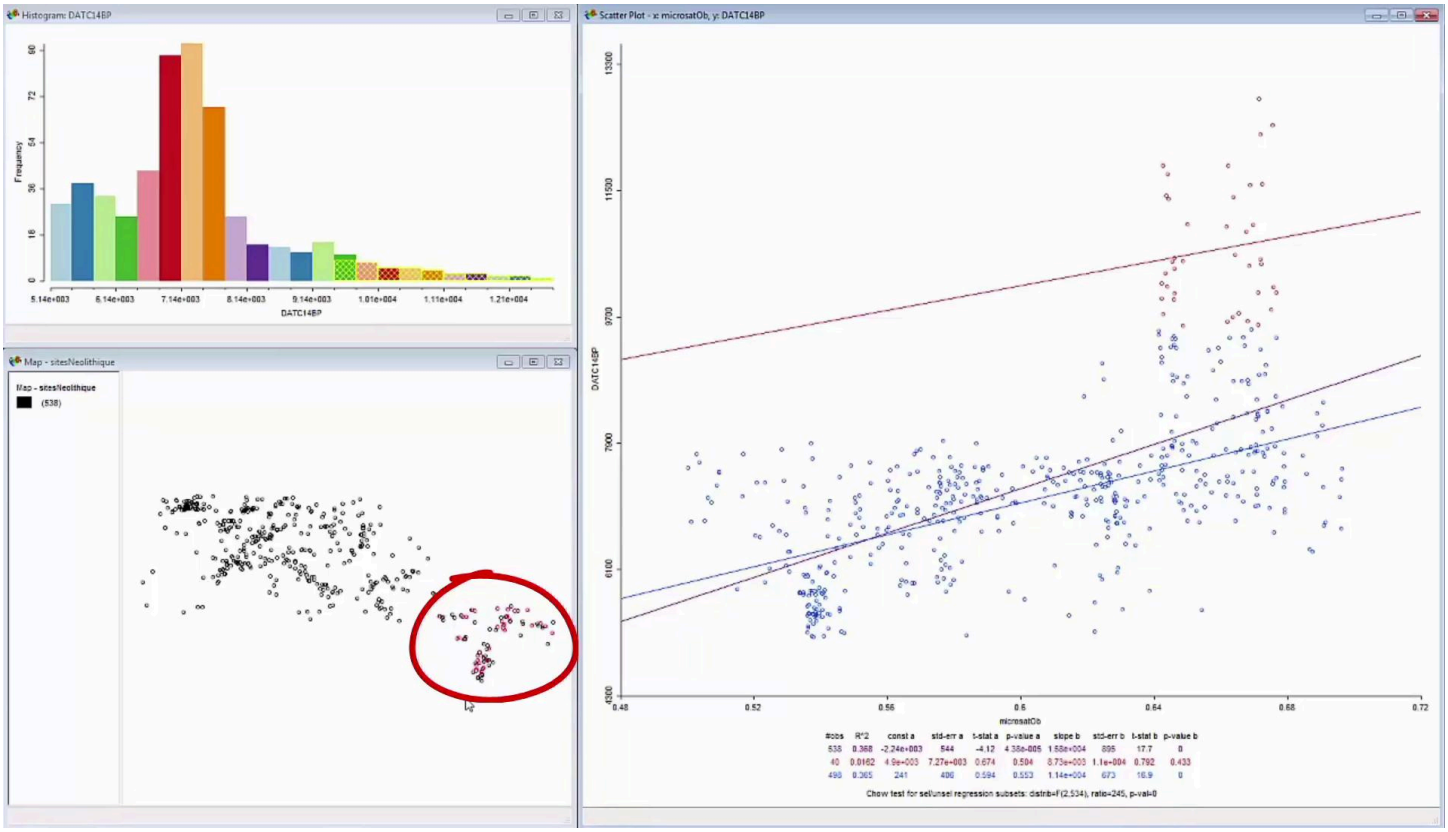
Introduction aux systèmes d'information géographique

Un bon exemple est la régression linéaire. Théoriquement, celle-ci doit être calculée avec des observations sélectionnées selon une procédure aléatoire. En effet, si les observations sont spatialement dépendantes, les valeurs estimées seront biaisées pour toute la zone d'étude, ceci parce que des valeurs exceptionnelles, localisées dans des sous-régions géographiques, vont influencer les valeurs prédites sur tout le territoire analysé. Ou encore, une forte corrélation entre deux attributs d'échantillons situés dans une même petite sous-région va influencer la mesure sur toute la zone étudiée. Nous allons maintenant passer à un exemple pratique pour illustrer ce type de biais.

Notes

Summary

7m 06s

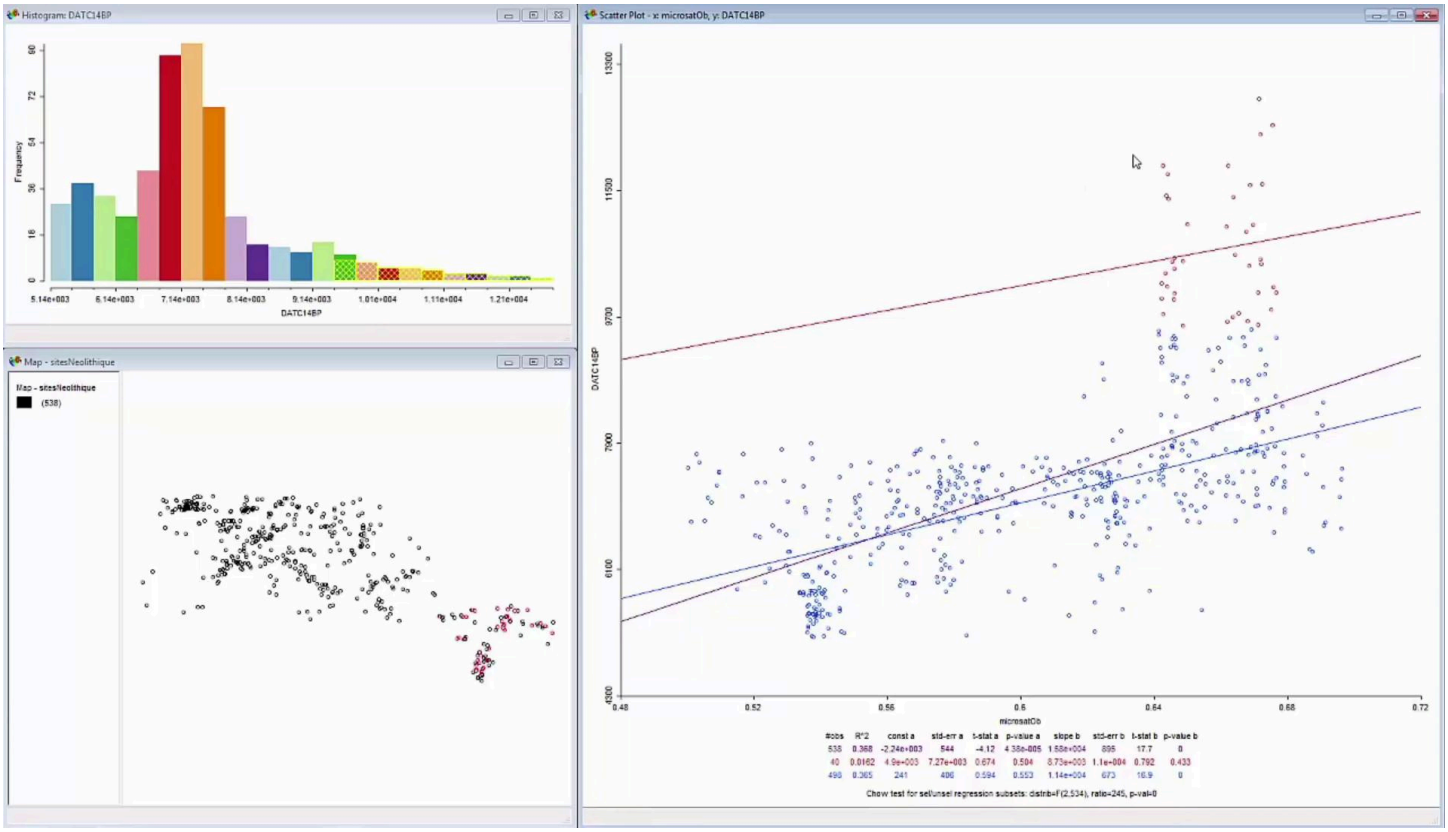


Reprenons l'exemple des 765 sites néolithiques. Dans le logiciel GeoDa, nous avons créé trois vues. En haut, à gauche, nous avons l'histogramme de la distribution des classes d'âge des sites. En bas, à gauche, la distribution spatiale des 765 sites. Et sur la droite, un graphe qui illustre la relation entre l'âge des sites en ordonnée et une variable de diversité génétique en abscisse qui caractérise des populations de chèvres échantillonnées aux alentours des sites au début du XXIe siècle. Cette relation présente un intérêt puisqu'elle permet de confirmer l'hypothèse selon laquelle les populations néolithiques du croissant fertile ont commencé à migrer en direction du nord-ouest à partir de la fin de la dernière grande période glaciaire. Moins l'âge des sites est avancé, moins la diversité génétique est élevée puisque les populations animales se sont progressivement fragmentées et que la reproduction a ensuite eu lieu entre individus apparentés. La droite de régression montre bien que plus l'âge des sites est ancien, plus la diversité génétique est élevée, mais cette relation est fortement influencée par un groupe de sites concentrés au Moyen-Orient.

Notes

Summary

7m 52s



En effet, si on retire du calcul ce groupe de sites, la pente de la régression diminue sensiblement. Les valeurs prédites par le modèle à travers toute l'Europe sont largement influencées par une quarantaine de points, tous situés dans la même sous-région.

Notes

Summary

9m 11s