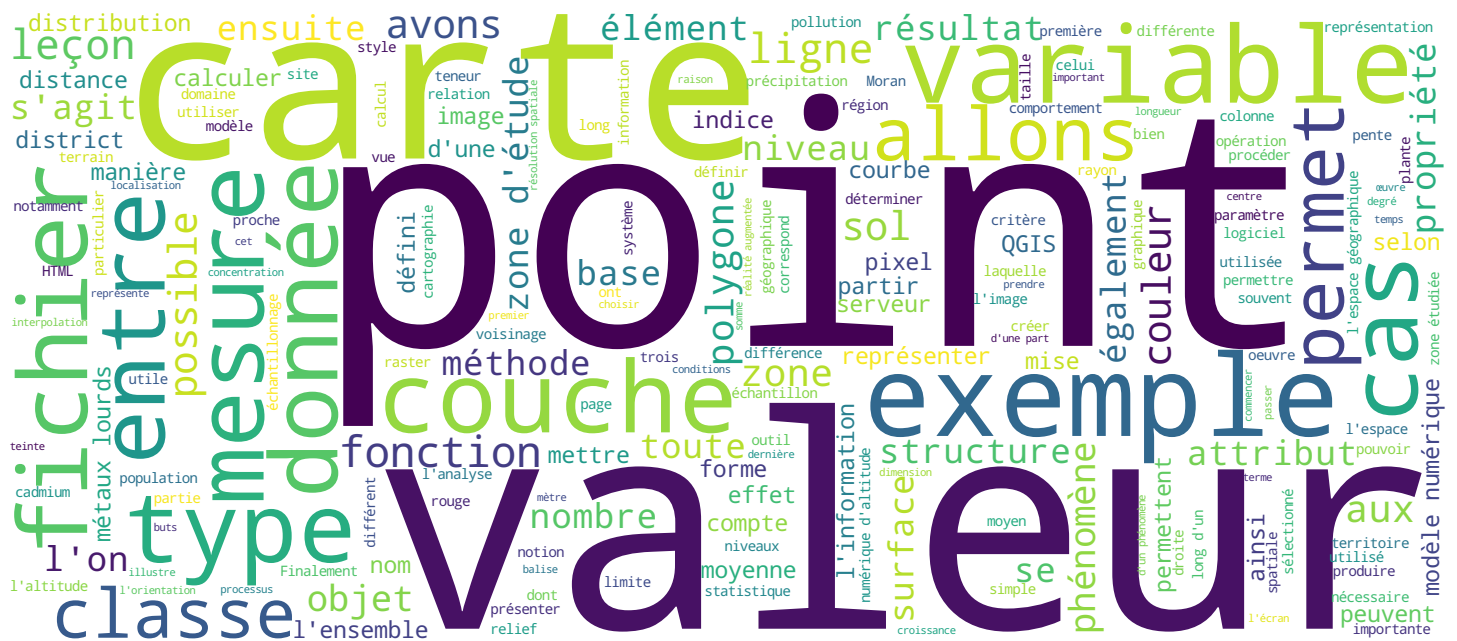


Phénomènes spatiaux continus - Echantillonnage

Introduction aux systèmes d'information géographique

Stéphane Joost, Marc Soutter, Fernand Kouamé, Amadou Sall



Search MOOC



Video



Phénomènes spatiaux continus - Echantillonnage



Buts de la leçon

- Présenter différentes procédures d'échantillonnage spatial
- Décrire les propriétés désirables d'un domaine à échantillonner

Après cette leçon vous serez capables

- D'appliquer une procédure d'échantillonnage spatial aléatoire ou systématique, selon un transect et selon une grille régulière

Introduction aux systèmes d'information géographique

Bienvenue à cette leçon qui porte sur l'échantillonnage de variables continues. Ce type de variables est défini en tout point de l'espace et il est très important de faire un échantillonnage représentatif. Un phénomène est dit continu s'il est défini en tout point de l'espace géographique et que ses propriétés varient localement de manière graduelle et structurée. L'altitude, l'humidité du sol ou sa teneur en métaux lourds en sont des exemples. Et comme l'a laissé entendre Fernand Kouamé dans son introduction ce type de données doit être acquis par échantillonnage. Une fois les échantillons collectés, on va procéder par inférence spatiale en utilisant l'interpolation pour étendre les données ponctuelles à l'ensemble de la zone étudiée. Mais la mise en oeuvre des opérations d'échantillonnage et d'interpolation est régie par des hypothèses et par des lois spécifiques que nous allons passer en revue. Les buts de cette leçon sont de transmettre les notions de base qui vous permettront de mettre en oeuvre les procédures d'échantillonnage le plus courantes. Nous présenterons ces procédures ainsi que les propriétés que l'échantillonnage est censé satisfaire de manière à vous permettre de les assimiler et de les appliquer judicieusement à vos propres données.

Notes

Summary



0m 30s

Phénomènes continus

Phénomènes continus

- Définis en tout point de l'espace géographique
- Variant localement de manière graduelle et structurée



Comme mentionné en introduction, les phénomènes continus sont définis en tout point de l'espace.

Notes

Summary



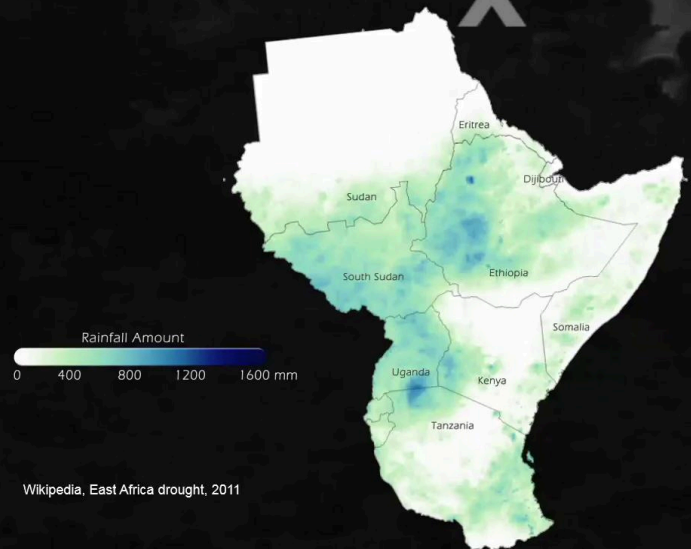
1m 51s

Phénomènes continus

Phénomènes continus

- Définis en tout point de l'espace géographique
- Variant localement de manière graduelle et structurée

Précipitations entre avril et juin 2011



Wikipedia, East Africa drought, 2011

Introduction aux systèmes d'information géographique

Ils peuvent varier localement, graduellement et de façon structurée, ce qui est le cas des précipitations qui sont illustrées sur cette carte de l'Afrique de l'est.

Notes

Summary



1m 59s

Phénomènes continus

Altitude



- 1858m
- 1547m
- 1155m
- 800m
- 372m

Wikipedia, Rochers de Naye, Frank H, 2009

Introduction aux systèmes d'information géographique

Ou de l'altitude, dont l'élévation graduelle est représentée schématiquement et de façon discrète sur ce panorama à l'est du lac Léman. Le vallon de Naye, ici sur la droite, va nous permettre d'illustrer un autre exemple de variable continue.

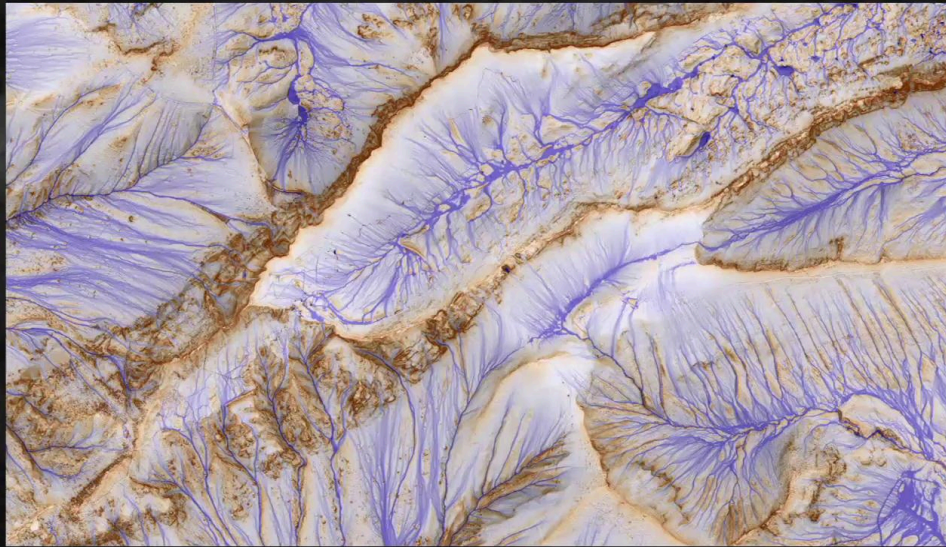
Notes

Summary



2m 05s

Phénomènes continus



Indice d'humidité du sol



Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S. (2015), Very high-resolution digital elevation models: are multi-scale derived variables ecologically relevant?. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.12427

Introduction aux systèmes d'information géographique

Il s'agit de l'humidité du sol, représenté par un indice morphométrique qui reproduit la concavité du terrain, calculé ici sur la base d'un modèle numérique d'altitude, d'une résolution spatiale de 1 m.

Notes

Summary



2m 20s

Echantillonnage

- Il n'est pas possible de mesurer une variable continue en tout point de l'espace (temps, coût)
- On mesure la variable d'intérêt pour quelques points jugés représentatifs de la zone d'étude : **échantillonnage**



géographique

Pour des raisons de temps et de coût, il n'est évidemment pas possible de mesurer une variable de ce type en tout point de l'espace. Mais à des fins analytiques, il est tout de même nécessaire de pouvoir disposer de ces valeurs. On va donc effectuer des mesures de la variable d'intérêt pour un nombre réduit de points qui sont jugés représentatifs. C'est ce qu'on appelle l'échantillonnage.

Notes

Summary

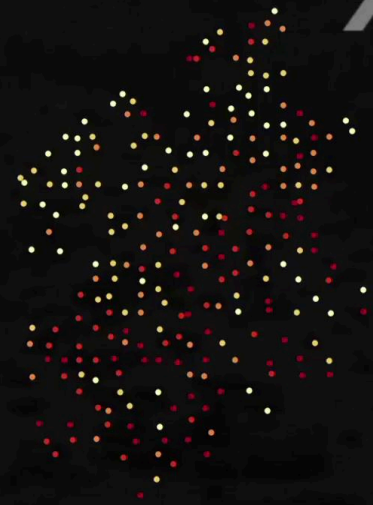


2m 40s

Echantillonnage

- Il n'est pas possible de mesurer une variable continue en tout point de l'espace (temps, coût)
- On mesure la variable d'intérêt pour quelques points jugés représentatifs de la zone d'étude : **échantillonnage**
- On déduit ensuite la valeur pour le reste de l'espace : **interpolation**

Concentration en Phosphore



Réseau d'observation des sols FRIBO | <http://www.fr.ch/sol/fr/pub/fribo.htm>

Introduction aux systèmes d'information géographique

Voici, par exemple, la concentration en phosphore mesurée dans le sol de 245 stations de mesure qui sont réparties sur le territoire du Canton de Fribourg en Suisse. Plus la teneur est élevée, plus foncé est le rouge du point qui représente la station de mesure réelle. Et à partir de cette distribution, on va déduire ou inférer la valeur pour le reste de l'espace géographique. C'est la démarche de régionalisation par interpolation.

Notes

Summary

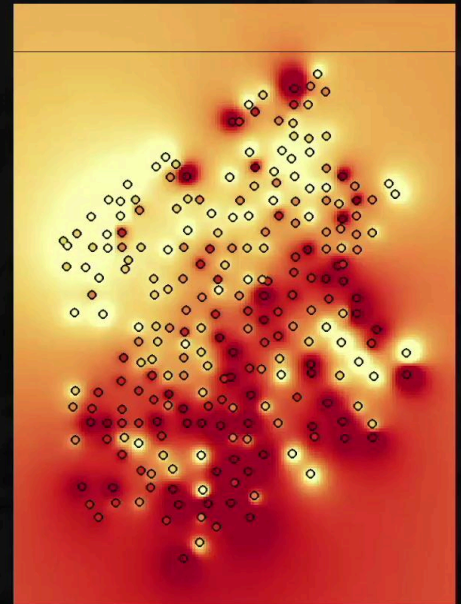
3m 01s



Echantillonnage

- Il n'est pas possible de mesurer une variable continue en tout point de l'espace (temps, coût)
- On mesure la variable d'intérêt pour quelques points jugés représentatifs de la zone d'étude : **échantillonnage**
- On déduit ensuite la valeur pour le reste de l'espace : **interpolation**

Concentration en Phosphore



Réseau d'observation des sols FRIBO | <http://www.fr.ch/sol/fr/pub/fribo.htm>

Introduction aux systèmes d'information géographique

On peut appliquer cette approche uniquement sous l'hypothèse qu'il est possible par inférence d'estimer avec une précision acceptable les valeurs prises par la variable en d'autres points de l'espace considéré.

Notes

Summary



3m 28s

Propriétés de l'échantillonnage

- Représentativité spatiale
- Théorie des sondages: relation entre les propriétés des échantillons et celles de l'ensemble de la zone étudiée
- Quand peut-on attribuer à toute la zone les propriétés de l'échantillon?
 - Limites de la zone d'étude circonscrites selon des critères précis
 - Répartition des échantillons homogène
 - Densité optimale



tion géographique

L'échantillonnage n'a de sens que s'il produit une connaissance valable pour l'ensemble de la zone étudiée. Sa qualité première est donc d'être représentatif. L'échantillonnage relève de la théorie des sondages, le problème est d'étudier dans toute situation les relations entre les propriétés de l'échantillon et celle de l'ensemble de la population qui correspond dans notre cas à toute la zone étudiée. Alors, dans quelle mesure peut-on attribuer à toute la zone les propriétés de l'échantillon ? Voici les conditions les plus importantes. Un échantillon est considéré optimal s'il satisfait les deux conditions suivantes : d'une part, les limites de la zone d'étude ont été circonscrites selon des critères très précis et d'autre part, la répartition des échantillons doit être homogène sur l'ensemble de la zone d'étude et de densité optimale, c'est-à-dire suffisante pour représenter la, ou les qualités recherchées.

Notes

Summary



3m 35s

Domaine

Avant de procéder à l'échantillonnage et à l'interpolation, il faut définir le **domaine**:

- Zone d'étude où le phénomène est censé exister et présenter un comportement homogène (**domaine délimité**)



ographie

Mais dans la pratique, ces conditions sont souvent remplies selon une procédure itérative. Les limites sont progressivement fixées selon les résultats des sondages ou des premières mesures. La distribution des échantillons peut être arbitraire au départ et non uniformément répartie et on pourra progressivement densifier l'échantillonnage en effectuant des prélèvements à des endroits qui sont jugés pertinents mais cela n'est pas possible pour tous les phénomènes analysés. Une notion importante qu'il convient d'utiliser est celle de domaine. Le domaine recouvre la zone d'étude dans laquelle le phénomène est censé exister et présenter un comportement similaire et homogène.

Notes

Summary

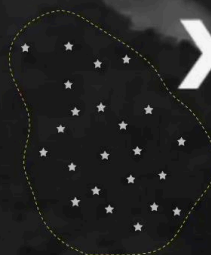


4m 30s

Domaine

Avant de procéder à l'échantillonnage et à l'interpolation, il faut définir le **domaine**:

- Zone d'étude où le phénomène est censé exister et présenter un comportement homogène (**domaine délimité**)
- Les limites du domaine ne sont pas toujours connues (**domaine ouvert**)
- Si la zone d'étude présente de l'hétérogénéité, il convient de la diviser en **sous-domaines** (**stratification de la région d'étude**)



Introduction aux systèmes d'information géographique

Parfois, les limites d'un domaine sont connues et l'on parlera alors de domaine délimité. Mais généralement, elles ne le sont pas ou alors seulement de manière approximative et l'on parlera de domaine ouvert. Cette distinction est importante dans la mesure où une loi d'interpolation va pouvoir être appliquée à l'ensemble d'un domaine que l'on suppose homogène d'un comportement d'un phénomène. Si l'on observe de l'hétérogénéité, il conviendra de diviser la zone initiale en plusieurs sous-domaines, dont chacun est censé présenter un comportement homogène de sorte que la même loi d'interpolation puisse être utilisée. On procède ainsi à une stratification de la région d'étude.

Notes

Summary



5m 09s

Types d'échantillonnage



Aléatoire



Régulier



Grille



Transects



Réseau ou lignes de structure

Systématique: un schéma de prélèvement est établi préalablement

Introduction aux systèmes d'information géographique

Il existe plusieurs façons de procéder à un échantillonnage spatial qui dépendent principalement de la connaissance préalable dont nous disposons sur le comportement du phénomène étudié ou inversement d'absence de l'information préalable. Dans ce dernier cas, on appliquera soit un échantillonnage de type aléatoire pour lequel aucun schéma spécifique n'est appliqué ou alors systématique, selon des variantes du semis de points réguliers ou de la grille régulière ou encore des transects ou des lignes de structure.

Notes

Summary

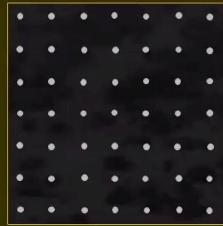


5m 57s

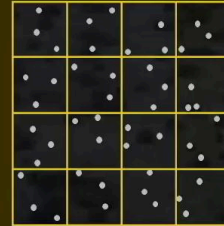
Types d'échantillonnage



Aléatoire



Régulier



Grille

Exploratoire



Transects



Réseau ou lignes de structure

Introduction aux systèmes d'information géographique

Pour les phénomènes invisibles, comme pour une contamination du sol par exemple on suit fréquemment une procédure exploratoire de manière à appréhender progressivement le comportement d'un phénomène avant de passer à une analyse complète. Par exemple, l'échantillonnage peut se faire selon un schéma régulier comme le transect ou alors de manière aléatoire mais avec une densification progressive des points de mesure.

Notes

Summary



6m 28s

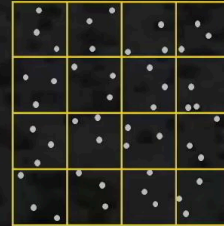
Types d'échantillonnage



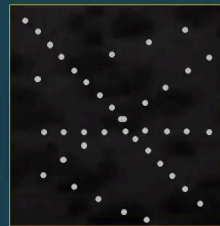
Aléatoire



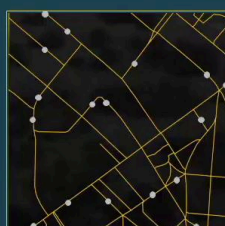
Régulier



Grille



Transects



Réseau ou lignes de structure

Connaissance *a priori* du phénomène

Introduction aux systèmes d'information géographique

Les procédures d'échantillonnage menées le long d'un ou de plusieurs transects ou le long de lignes de structure sont des procédures qui requièrent une connaissance préalable du phénomène. Par exemple, on étudiera certains processus biologiques liés à la croissance des plantes le long d'un transect altitudinal. En effet, on sait que l'altitude affecte la durée de la période de croissance.

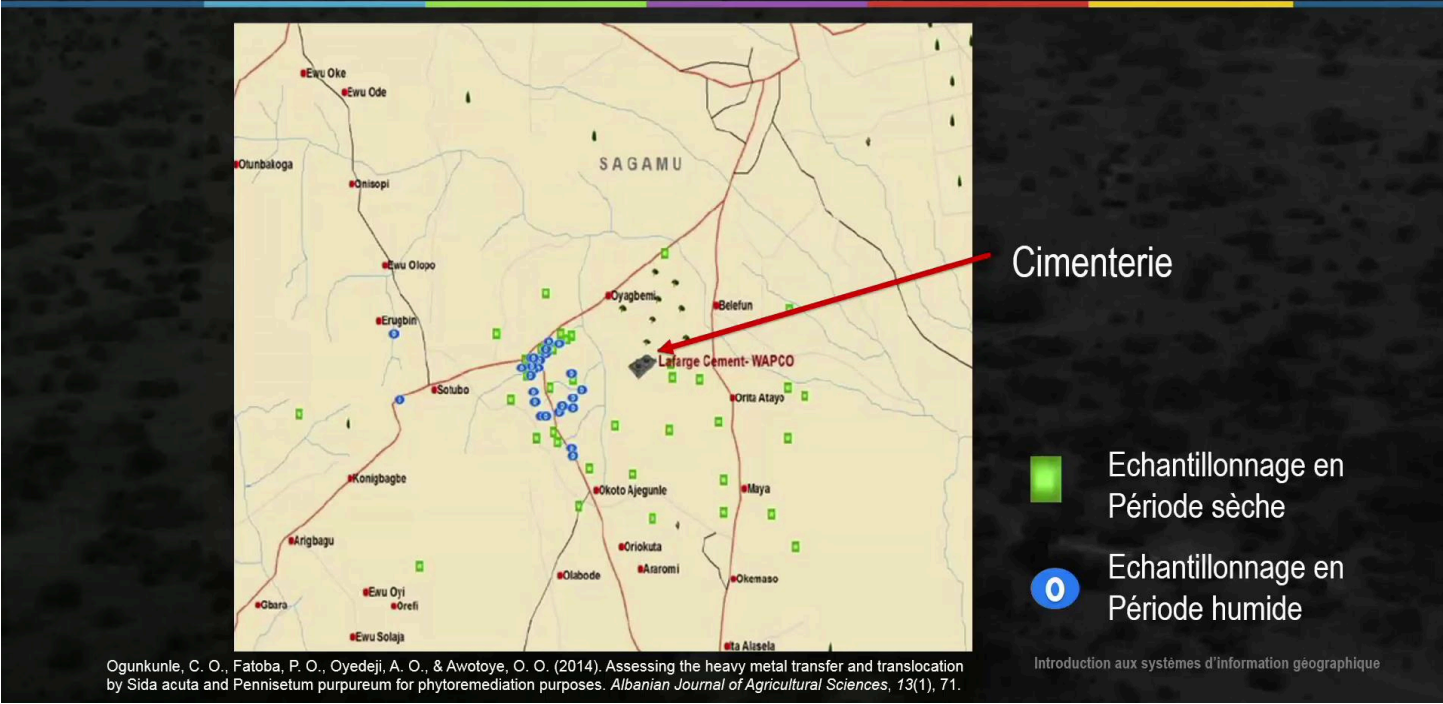
Notes

Summary



6m 54s

Aléatoire



Nous présentons ici un exemple qui illustre la procédure d'échantillonnage mise en œuvre dans le cas de l'étude de la concentration en métaux lourds, comme le plomb, le cuivre ou le zinc, dans les sols situés aux environs d'une usine de production de ciment au sud-ouest du Nigéria. En effet, en fonction du degré de pureté des ingrédients des métaux lourds entrent dans la composition de certains ciments. Pour évaluer l'importance de cette pollution les auteurs ont prélevé deux plantes qui ont la capacité d'accumuler les métaux lourds avec 15 cm de sol superficiel en 64 points et ceci dans un rayon de 3 km autour de l'usine. La principale contrainte était temporelle puisqu'il était nécessaire d'obtenir des données représentatives pour la saison sèche et pour la saison des pluies. Par contre, la localisation des points d'échantillonnage a été déterminée de façon totalement aléatoire.

16 of 22

Aléatoire



J.M.Garg - Own work, GFDL,
<https://commons.wikimedia.org/w/index.php?curid=5332796>



Forest & Kim Starr, CC BY 3.0,
<https://commons.wikimedia.org/w/index.php?curid=6178217>

Introduction aux systèmes d'information géographique

Les résultats de l'étude montrent que les deux plantes en question *Sida acuta*, qui est considérée comme une mauvaise herbe et *Pennisetum purpureum* aussi appelée herbe à éléphant, constituent de bons accumulateurs en particulier de chrome et de cadmium et qu'elles pourraient être utilisées dans des opérations de remédiation des sols pollués.

Notes

Summary



8m 06s

Réseau et transects



Morris Pügner –
Own work, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=332686>

Introduction aux systèmes d'information géographique

Voici maintenant un exemple qui illustre une procédure d'échantillonnage réalisée le long d'une ligne de structure soit ici d'une rivière. L'expansion des activités agricoles le long de la rivière Mara, entre le Kenya et la Tanzanie, a entraîné de nombreux changements de couverture du sol. Les forêts et prairies de savane ont été converties en terres agricoles. Plusieurs études ont montré que ces changements pouvaient induire une plus grande érosion du sol, une réduction en nutriments et une modification de la structure du sol.

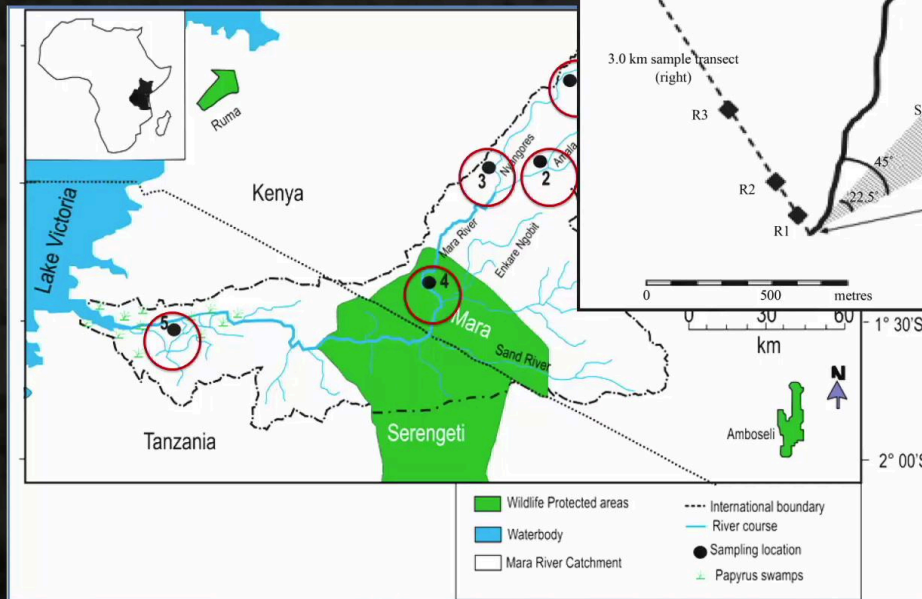
Notes

Summary



8m 26s

Réseau et transects



Matano, A. S., Kanangire, C. K., Anyona, D. N., Abuom, P. O., Gelder, F. B., Dida, G. O., ... & Ofulla, A. V. (2015). Effects of Land Use Change on Land Degradation Reflected by Soil Properties along Mara River, Kenya and Tanzania. *Open Journal of Soil Science*, 5(01), 20.

Introduction aux systèmes d'information géographique

L'étude présentée ici a donc été menée dans le but d'étudier l'impact des changements d'occupation du sol autour de la rivière Mara sur les qualités physico-chimiques des sols, comme la teneur en carbone, en azote, en phosphore, ou l'acidité du sol. Des échantillons de sol ont été collectés dans 5 sites différents choisis le long de la rivière, dans des zones présentant des types d'occupation du sol différents. Dans chaque site, 8 points d'échantillonnage ont ensuite été définis le long d'un transect, 4 de chaque côté de la rivière.

Notes

Summary

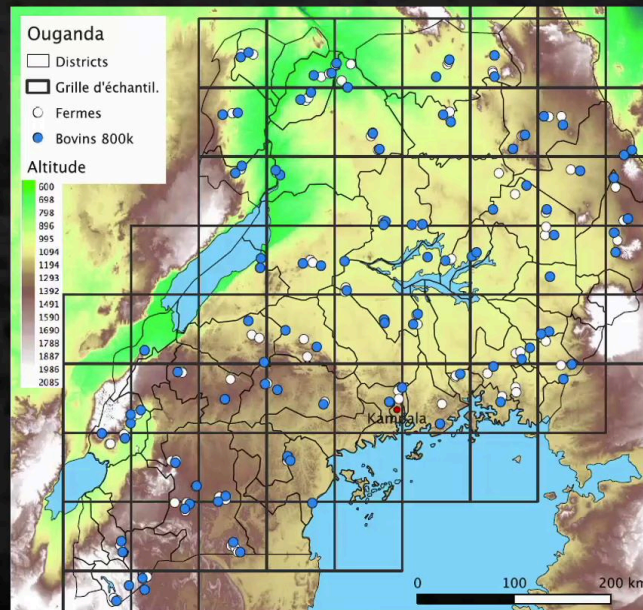


8m 55s

Grille



Ankole
Photo <http://nextgen.epfl.ch>



Zebu
Photo <http://nextgen.epfl.ch>



Stucki, S., Orozco-terWengel, P., Bruford, M. W., Colli, L., Masembe, C., Negrini, R., ... & Joost, S. (2014). High performance computation of landscape genomic models integrating local indices of spatial association. *arXiv preprint arXiv:1405.7658*.

Introduction aux systèmes d'information géographique

Finalement, nous illustrons une campagne d'échantillonnage effectuée selon une grille régulière. Il s'agit de l'analyse de la diversité génétique et de la structure de populations de races de vache Ankolé et Zébu en Ouganda. Cette étude a été menée dans le cadre d'un projet européen appelé "NextGen" et dont l'un des buts était d'identifier les gènes impliqués dans les processus de résistance à certaines maladies tropicales comme la trypanosomiase. Pour assurer la plus grande représentativité spatiale possible à travers le pays 917 individus, soit environ 4 par localisation, ont été sélectionnés dans 229 fermes réparties sur le territoire de l'Ouganda. Chaque cellule contient 4 fermes, et le pas de la grille est d'environ 70 kilomètres. Les points bleus représentent des fermes où un animal a été sélectionné pour un génotypage spécifique à haute densité.

Notes

Summary

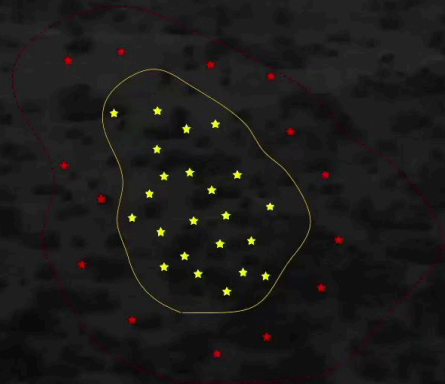


9m 27s

Qualité de l'échantillon



- Représentativité géométrique et thématique
- Densité de points et degré d'incertitude
- Anticiper la procédure de régionalisation à venir (faciliter l'interpolation)



Introduction aux systèmes d'information géographique

Nous avons déjà mentionné qu'un échantillon spatial doit impérativement représenter le comportement du phénomène étudié à la fois dans sa dimension géométrique et thématique. En ce qui concerne la géométrie, il est utile de souligner que la densité du semis de points affecte le degré d'incertitude lié à cette distribution jusqu'à l'obtention d'une taille optimale. Un autre aspect impliquant la géométrie concerne les contraintes liées à la procédure d'interpolation souvent appliquées aux mesures obtenues par échantillonnage et que nous aborderons dans les leçons suivantes. Il faut en effet garantir que le processus de régionalisation à l'intérieur de la zone d'étude, procède bien par interpolation et non pas par extrapolation. Pour s'en assurer, il est nécessaire de constituer un échantillon comportant des points situés à l'extérieur de la zone d'étude dans une couronne suffisamment large pour éviter les effets de bord.

Notes

Summary



10m 25s

Qualité de l'échantillon



- Représentativité géométrique et thématique
- Densité de points et degré d'incertitude
- Anticiper la procédure de régionalisation à venir (faciliter l'interpolation)
- Dynamique des valeurs mesurées dans le domaine d'étude
- Dimension temporelle:
 - Variabilité et indices de tendance centrale
 - Surcoût



D'un point de vue thématique maintenant les valeurs mesurées doivent dans la mesure du possible rendre compte de la dynamique des valeurs présentes dans la région d'étude, ce qui n'est pas aisé pour des variables invisibles. Finalement, on doit aussi considérer la représentativité de l'échantillonnage dans sa dimension temporelle de façon à s'assurer que les valeurs thématiques expriment bien le comportement du phénomène pour une période de temps définie. Dans certains cas présentant une variabilité importante on devra alors produire des indices de tendance centrale rendant compte du comportement général de la variable mesurée comme des moyennes mensuelles de température de l'air par exemple. Suivant le phénomène analysé, la prise en compte de la dimension temporelle dans l'échantillonnage peut représenter un surcoût important.

Notes

Summary



11m 20s