



**BOOCs**  
**EPFL**

Systèmes d'Information  
Géographique – Partie 2  
Stéphane Joost, Marc Soutter,  
Fernand Kouamé et Amadou Sall

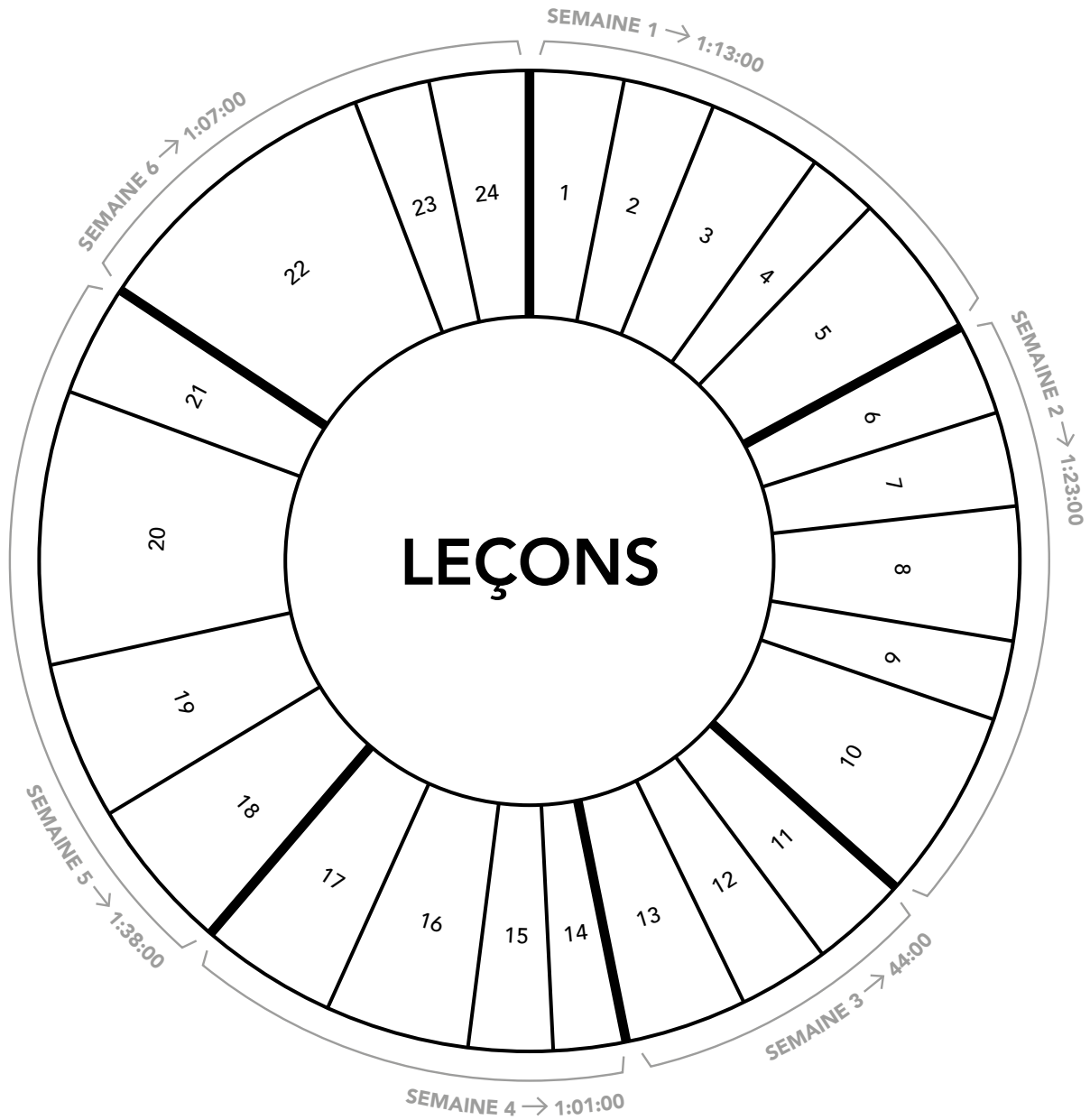


# SYSTÈMES D'INFORMATION GÉOGRAPHIQUE

## Partie 2



**Stéphane Joost, Marc Soutter,  
Fernand Kouamé et Amadou Sall**





## CONTENU

### SEMAINE 1

1. Le logiciel GeoDa	4
2. Variables discrètes – propriétés géométriques	8
3. Arrangement et voisinage spatiaux	12
4. Autocorrélation et dépendance spatiale	15
5. Autocorrélation spatiale globale	17

### SEMAINE 2

6. Échantillonnage	20
7. Interpolation I	22
8. Interpolation II – Krigeage	25
9. Isovaleurs	30
10. Étude de cas – Pollution de la lagune Ebrié à Abidjan	32

### SEMAINE 3

11. Modèles numériques d'altitude – Éléments de base	35
12. Modèles numériques d'altitude – Variables dérivées I	38
13. Modèles numériques d'altitude – Variables dérivées II	42

### SEMAINE 4

14. Interactions entre couches de données: Raster – Vecteur	44
15. Interactions entre couches de données: Raster – Raster	45
16. Interactions entre couches de données: Vecteur – Vecteur	47
17. Étude de cas – Le risque d'inondation à Niamey	50

### SEMAINE 5

18. Les composantes et fonctions de la carte	52
19. Les variables visuelles	55
20. Attributs thématiques et mise en classes	58
21. Habillage et production de cartes thématiques	62

### SEMAINE 6

22. Cartographie dynamique et interactive sur internet	71
23. Autres formes de représentation: la troisième dimension	71
24. Autres formes de représentation: SIG et réalité augmentée	73

# 1. LE LOGICIEL GEODA

GeoDa est un logiciel libre permettant la manipulation de données spatiales. L'objectif de cette leçon est de se familiariser avec ce logiciel. À la fin de cette leçon, vous serez capable de manipuler des données sur GeoDa, de fabriquer des cartes statistiques thématiques et de créer des graphiques liés aux cartes.

## INTRODUCTION À GEODA

Le logiciel GeoDa a été initialement développé par le Prof. Luc Anselin, il est maintenant maintenu et distribué par le Center for Spatial Data Science, the University of Chicago. Il a été conçu pour l'analyse exploratoire de donnée spatiale vectorielle en combinant les vues thématiques, graphiques et de statistiques spatiales.

La version 1.6.7, utilisée dans ce MOOC, est téléchargeable sur [http://geodacenter.github.io/download\\_windows.html](http://geodacenter.github.io/download_windows.html). Des références sont disponibles sur le site du GeoDa Center (<http://geodacenter.github.io/documentation.html>) ainsi que dans l'article du Prof. Anselin qui décrit le logiciel et ses fondamentaux. (L. Anselin, I. Syabri et Y. Kho (2005), GeoDa: An Introduction to Spatial Data Analysis, *Geographical Analysis* 38(1), pp. 5-22) La fenêtre visible sur la figure 1 présente l'interface utilisateur du logiciel GeoDa.

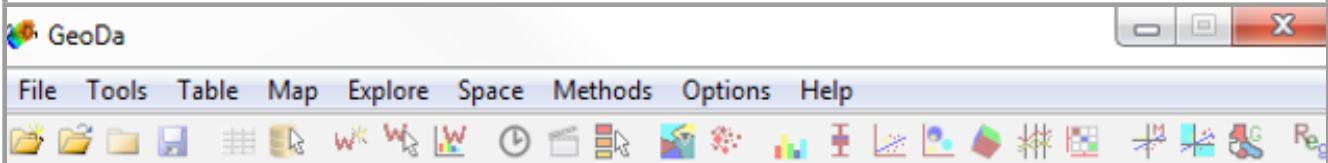


FIGURE 1

2:12

13:00

Interface utilisateur de GeoDa.

## GESTION DES DONNÉES

Pour importer des données, cliquer sur **File >> New Project**. Différentes sources de données peuvent être importées à partir de fichiers géoréférencés (ESRI Shapefile, GeoJSON, etc.), non géoréférencés (KML, DBF, CSV, Excel, etc.), de bases de données ou de serveurs Web.

Dans cet exemple nous importons un fichier Shapefile contenant des données sur les districts des Seychelles. Le projet peut être sauvegardé en tout temps en cliquant sur l'icône de la disquette dans la barre d'outils.

Des outils de zoom et de sélection sont disponibles dans le menu **Options**. Les quatre premières icônes de l'interface utilisateur correspondent à la création, l'ouverture et l'enregistrement de sets de données. La cinquième ouvre la table d'attribut du fichier vectoriel. En sélectionnant une unité spatiale dans la table d'attribut, celle-ci est simultanément sélectionnée sur la carte.

Il est possible d'opérer des sélections conditionnelles sur les attributs en cliquant sur **Table >> Selection Tool**. Il est également possible d'ajouter des colonnes ou faire des calculs sur les attributs en cliquant sur **Table >> Variable Calculation**. Les unités spatiales peuvent être ordonnées selon leurs attributs en double-cliquant sur l'en-tête de la colonne de l'attribut en question.



## CARTES THÉMATIQUES

Les cartes thématiques sont la fonction principale offerte par GeoDa. Elles permettent d'explorer et d'analyser la distribution spatiale d'un jeu de donnée. Différentes cartes peuvent être créées à partir du menu **Map**. Il est ainsi possible d'afficher la distribution d'un attribut (figures 2.2 et 2.3), leur écart à la moyenne (figures 2.5 et 2.6) ou encore de créer une carte bivariée en cliquant sur **Cartogram** dans le menu **Map**. La figure 2 en montre plusieurs exemples.

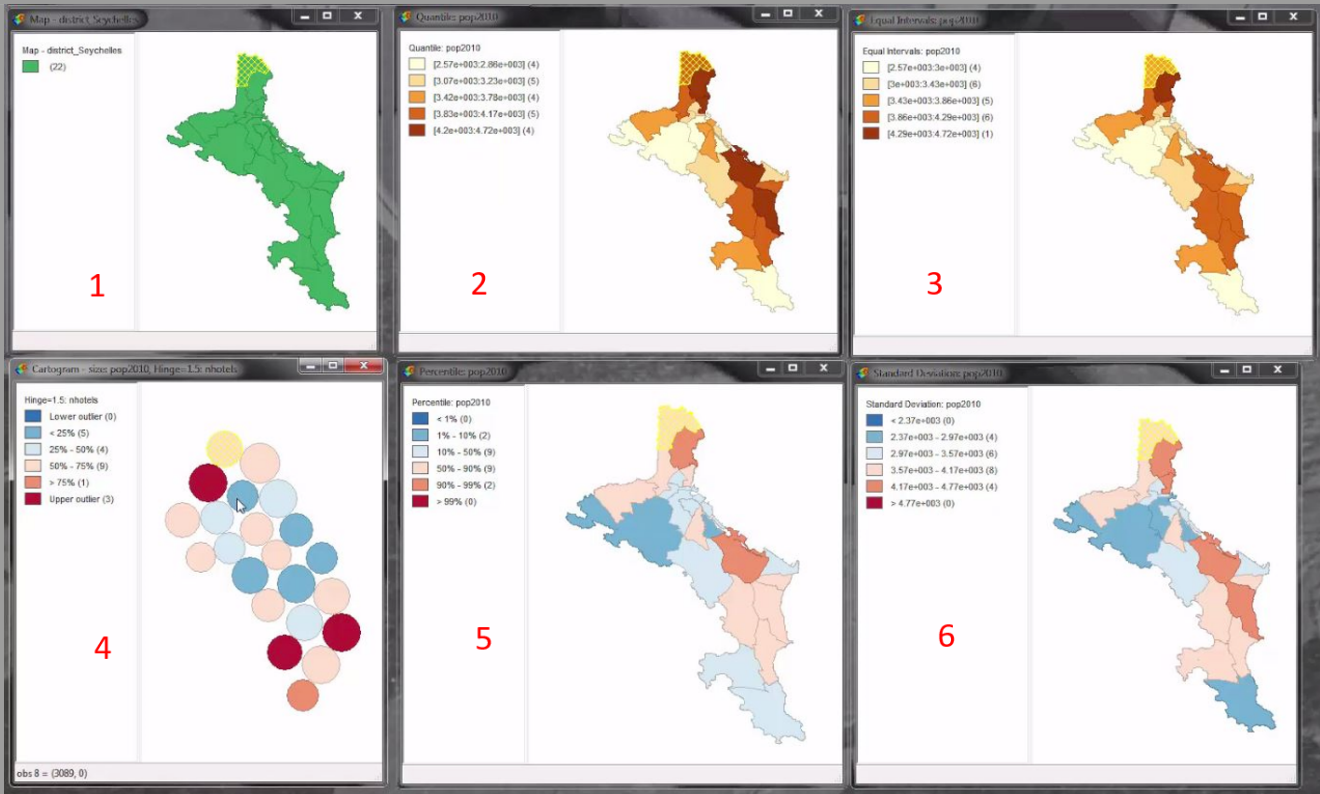


FIGURE 2

7:33

13:00

Exemples de cartes thématiques sur la population en 2010: 1) Carte initiale; 2) cartes en 5 quantiles; 3) cartes en 5 intervalles égaux; 4) carte bivariée avec cercles proportionnels (taille des cercles en fonction de la population et couleur des cercles en fonction du nombre d'hôtels); 5) carte des percentiles; 6) carte de l'écart-type.

### BOX PLOT, HISTOGRAMME, DIAGRAMME DE DISPERSION

Dans GeoDa, les attributs peuvent également être visualisés au moyen de graphiques et de diagrammes statistiques. Ces outils permettent de mettre en évidence les tendances intrinsèques aux attributs ou entre les attributs.

Dans le menu déroulant de l'onglet **Explore**, plusieurs graphiques et diagrammes statistiques sont présentés. Ils permettent d'explorer la distribution statistique d'un attribut à l'aide d'un histogramme ou d'un box plot et sa carte associée (box map) et d'en déterminer les comportements atypiques. La relation entre deux attributs peut être visualisée à l'aide d'un diagramme de dispersion (scatter plot). Les paramètres statistiques sont présents sous chaque graphique et diagramme.

L'interaction dynamique ou **brushing** de GeoDa permet de sélectionner des entités et de faire le lien entre leurs différentes représentations (diagrammes, cartes, etc.). Il permet également de calculer les paramètres statistiques dans un sous-ensemble d'unités spatiales. La figure 3 montre différents graphiques et diagrammes statistiques disponibles dans GeoDa.

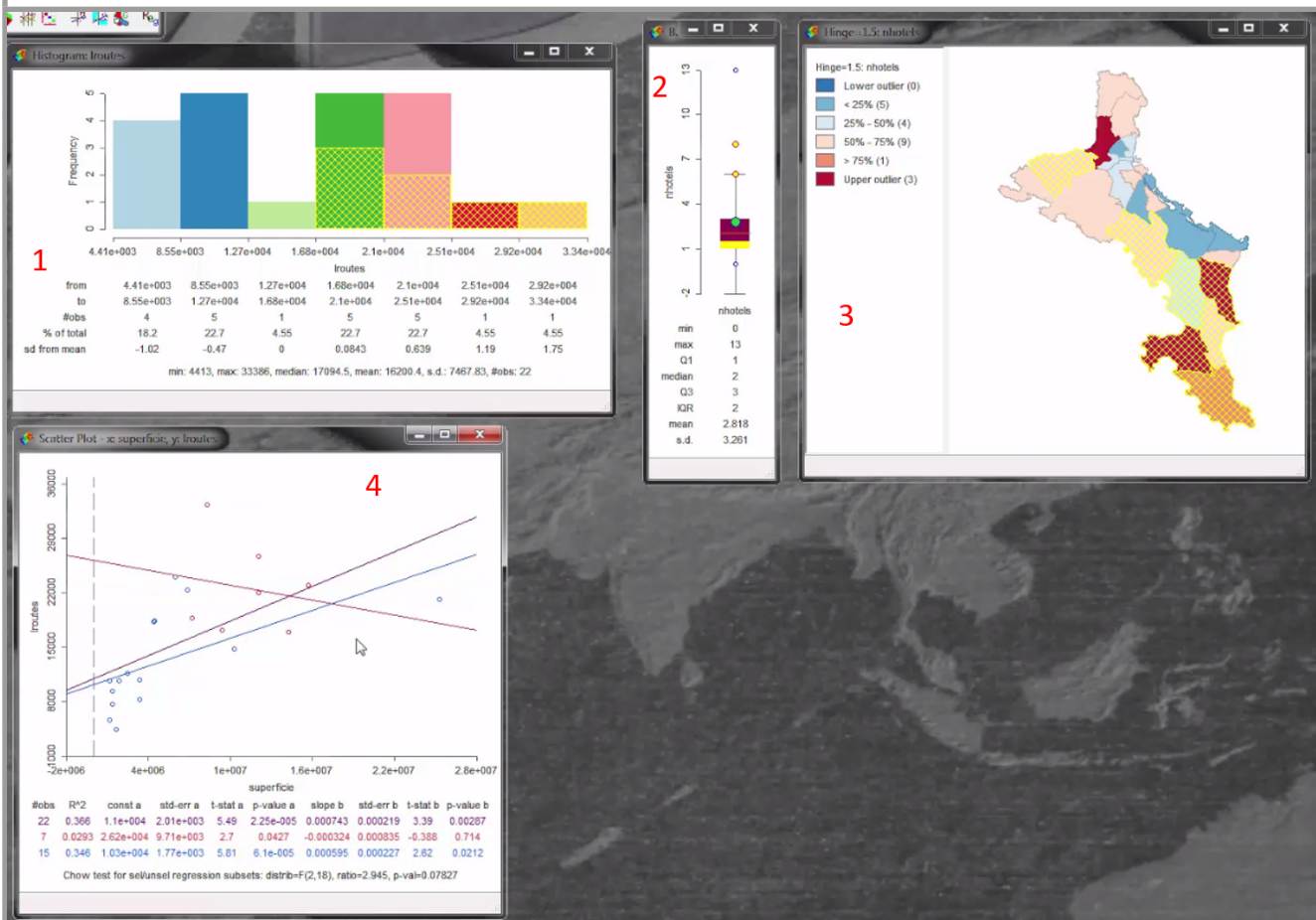


FIGURE 3

8:19

13:00

Exemple de graphiques et de diagrammes statistiques pour des données des districts aux Seychelle: 1) histogramme de la longueur des routes; 2) box plot du nombre d'hôtels par district; 3) box map associé au box plot; 4) diagramme de dispersion (Scatter Plot) entre la superficie des districts et la longueur totale de ses routes.

## STATISTIQUES SPATIALES

GeoDa offre la possibilité de calculer des statistiques spatiales comme l'autocorrélation spatiale globale et locale ainsi que des régressions linéaires spatiales. Pour le calcul de l'autocorrélation spatiale globale, il est nécessaire d'établir un fichier de pondération (pour le créer: [Tools >> Weights](#)). GeoDa propose plusieurs schémas de pondération, comme les  $K$  plus proches voisins ou une fonction de la distance. Les outils de calcul de l'autocorrélation globale, locale ou bivariée sont disponibles sous l'onglet [Space](#) de l'interface utilisateur.

Pour calculer l'autocorrélation spatiale globale, cliquer sur [Space >> Univariate Moran's I](#) puis sélectionner l'attribut désiré. La significativité du  $I$  de Moran est obtenue en effectuant des permutations aléatoires. Pour cela, cliquer droit sur le graphique puis sélectionner [Randomisation >> 999 permutations](#). On obtient un histogramme des  $I$  de Moran de toutes les permutations. La barre jaune correspond au  $I$  de Moran observé et la barre grise à la moyenne obtenue à partir des permutations. Les barres bleues représentent les valeurs qui dépassent le  $I$  de Moran observé. Sur l'exemple de la [figure 4](#), le  $I$  de Moran n'est pas significatif car de nombreuses permutations donnent des valeurs de  $I$  de Moran supérieures au  $I$  de Moran observé.

Pour calculer les régressions linéaires, cliquer sur [Methods >> Regression](#). Cet outil permet d'exprimer la relation entre un attribut et un ou plusieurs autres attributs. Les résultats obtenus nous donnent les paramètres de la régression, les valeurs des tests statistiques et l'erreur du modèle due à la dépendance spatiale entre les échantillons.

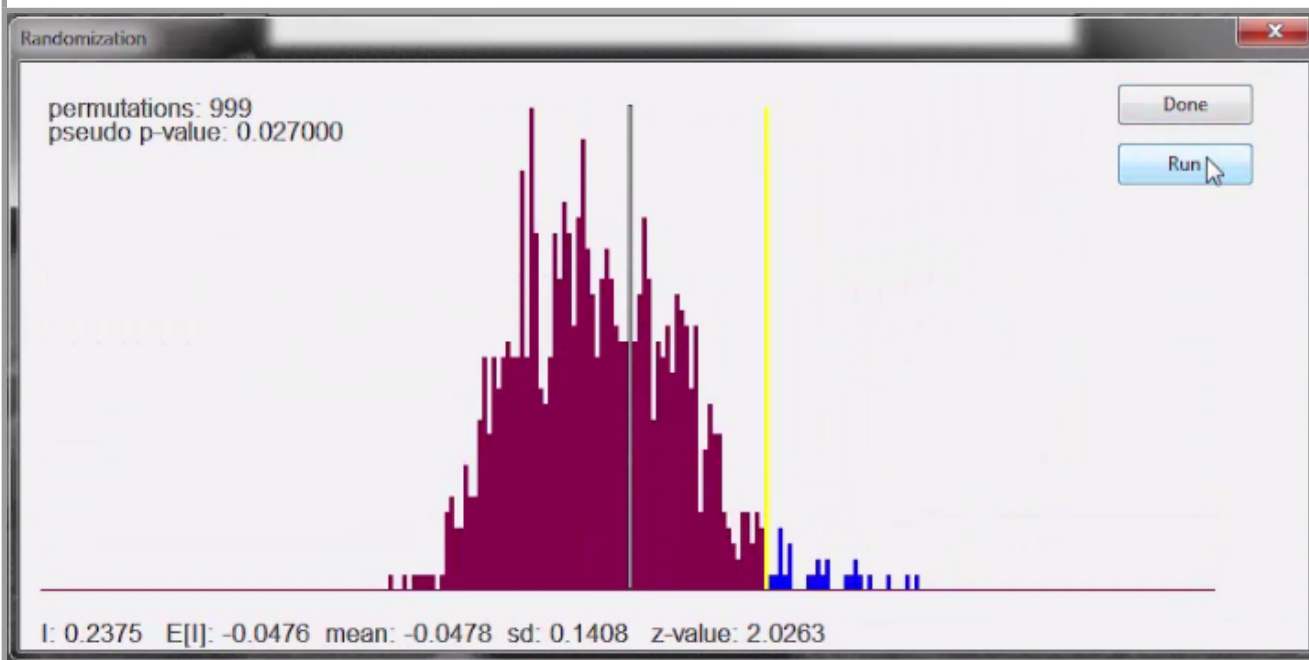


FIGURE 4

9:31

13:00

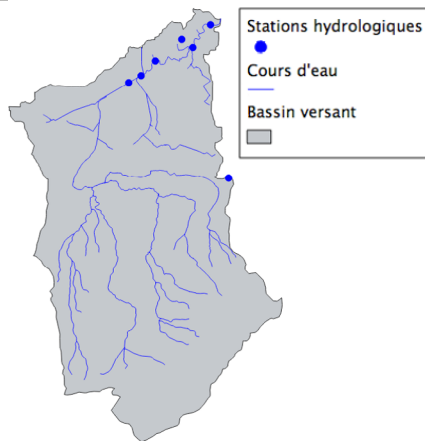
Exemple d'un histogramme de  $I$  de Moran pour 999 permutations.

## 2. VARIABLES DISCRÈTES – PROPRIÉTÉS GÉOMÉTRIQUES

La distribution spatiale des phénomènes se déroulant à la surface de la Terre peut être séparée en deux groupes.

- La **distribution spatiale continue** décrite par des *variables continues* concerne les phénomènes définis en tout point de l'espace (température, altitude, etc.) et modélisable par une surface continue.
- La **distribution spatiale discrète** décrite par des *variables discrètes* concerne les phénomènes définis par des unités spatiales ponctuelles, linéaires ou surfaciques c'est-à-dire les phénomènes spatiaux discontinus.

Cette leçon présente les différents types de variables géographiques discrètes ainsi que leurs propriétés. À la fin de cette leçon, vous saurez également comment utiliser les propriétés de ces variables discrètes pour calculer des indices de formes, de localisations et de tailles.



### VARIABLES DISCRÈTES

Les variables discrètes permettent de représenter les phénomènes spatiaux discontinus. Un phénomène est défini discontinu si la distribution de ses propriétés thématiques est spatialement discontinue. Il existe trois types d'entités spatiales capables de représenter ces phénomènes :

- les objets ponctuels (points, centres);
- les lignes (ou multilignes);
- les polygones.

FIGURE 1

2:42

13:00

Exemple d'entités spatiales. Les points représentent les stations hydrologiques, les cours d'eau sont représentés par des lignes et le bassin-versant par un polygone.

### OBJETS PONCTUELS (POINTS)

Lorsqu'il n'est pas possible de représenter le contour exact d'un objet à une échelle donnée ou que son contour n'est pas important, l'objet est représenté par un objet ponctuel en mode vecteur ou une région ponctuelle en mode raster. L'unique propriété géométrique du point est sa localisation. Les points sont donc de dimension 0 (0D). Leur localisation est décrite par un couple de coordonnées (X, Y) dans un système de référence donné.

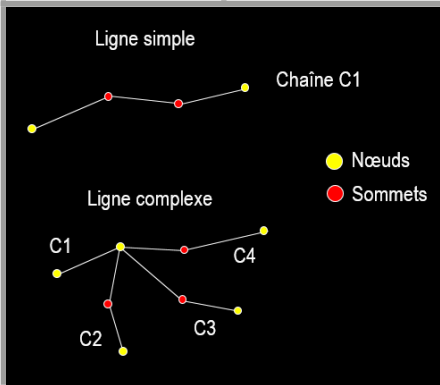


FIGURE 2

4:46

13:00

Ligne simple et multiligne.

### LIGNES

Les lignes sont utiles pour représenter des éléments linéaires dont la largeur exacte ne peut être représentée à une échelle donnée ou que sa largeur n'est pas utile à cette même échelle. Les réseaux routiers en sont un exemple bien connu. Les lignes sont appelées objets linéaires en mode vecteur et régions linéaires en mode raster.

Il existe deux types de lignes (figure 2). La *ligne simple* composée d'une seule chaîne, de deux extrémités et d'un ou plusieurs sommets. La *multiligne* (ou ligne complexe) composée de plusieurs chaînes, de plus de deux nœuds et d'un ou plusieurs sommets. La multiligne est considérée soit comme un ensemble de tronçons individuels connecté soit comme une entité unique appelée réseau.

De dimension 1D, les lignes possèdent trois propriétés géométriques.

- La *localisation* définie par le centre moyen (centre de gravité géométrique) et calculée à partir des sommets et des nœuds de la ligne brisée en mode vecteur et des coordonnées X, Y de toutes les cellules constituant la ligne en mode raster (figure 3.1).
- La *longueur* (taille) définie par la somme des longueurs des segments de la ligne (figure 3.2). En mode raster, les types de déplacements sont additionnés pour calculer la longueur. Ainsi un déplacement du centre d'un pixel vers le centre d'un autre pixel vaut racine de deux si le déplacement est diagonal et 1 si le déplacement est horizontal ou vertical.
- La *forme* (sinuosité) définie par le rapport entre la longueur totale et la distance entre les extrémités. En mode raster, la distance entre les extrémités est définie par le plus court chemin entre les deux pixels des extrémités. (figure 3.3)

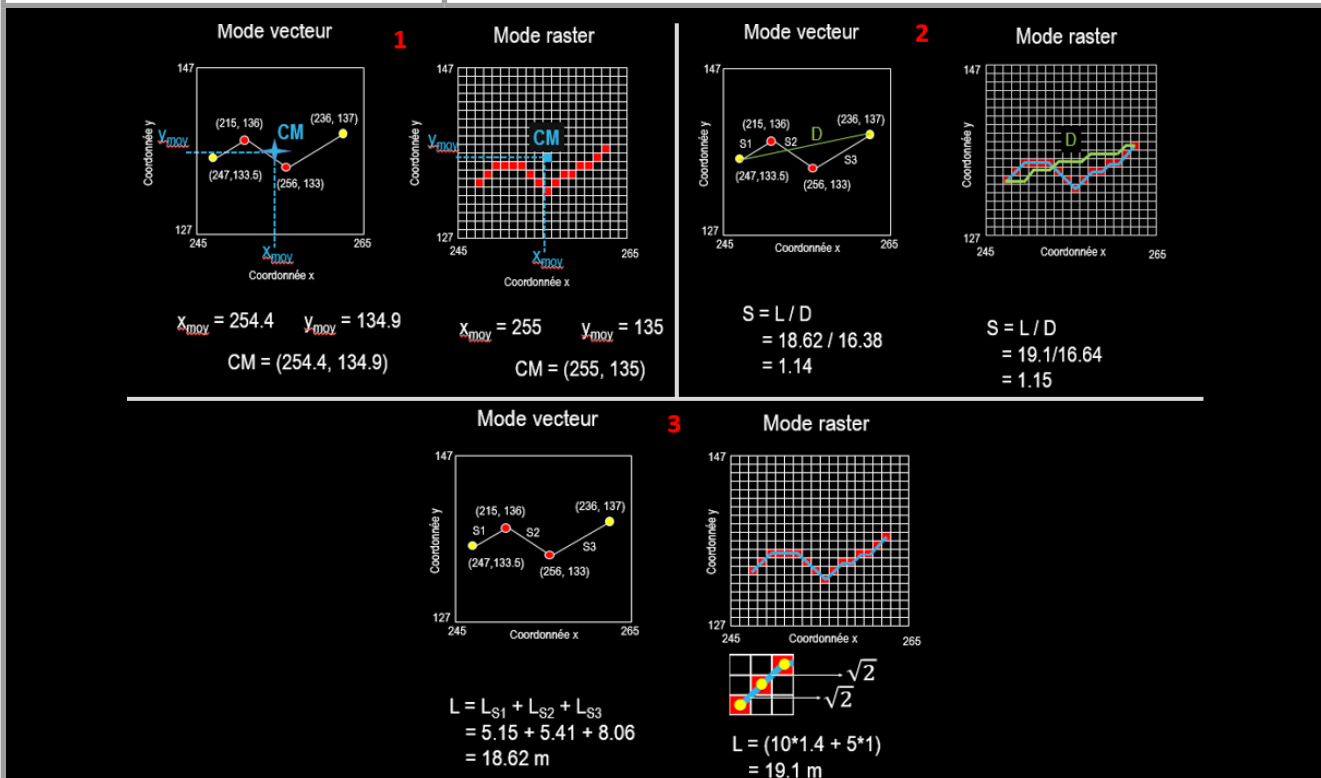


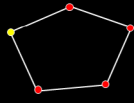
FIGURE 3

5:28

13:00



Polygone simple



Polygone complexe

● Nœuds  
● Sommets

FIGURE 4

7:50

13:00

Polygones simples et complexes.

## POLYGONES

Les polygones (ou régions zonales en mode raster) sont généralement utilisés pour représenter des surfaces (régions administratives, plans de quartier, etc.). On définit une entité spatiale linéaire comme étant un polygone lorsqu'il y a fermeture de la ligne brisée. C'est-à-dire lorsque la première paire de coordonnées est répétée. Il existe deux types de polygones. Le *polygone simple* composé d'une seule unité spatiale et le *polygone complexe* composé de plusieurs unités spatiales discontinues (figure 4).

De dimension 2D, les polygones possèdent trois propriétés géométriques.

- La *localisation* définie par le centre de masse (centre de gravité géométrique) et calculée à partir des sommets et des nœuds de la ligne brisée fermée en mode vecteur et des coordonnées X, Y de toutes les cellules constituant le polygone en mode raster (figure 5.1).
- La *taille* définie par le périmètre et l'aire. Le *périmètre* correspond à la somme des longueurs des segments de la ligne brisée fermée en mode vecteur tandis qu'en mode raster il est calculé à l'aide de la métrique de Manhattan (figure 5.2) (périmètre égal aux nombres de faces extérieures des pixels du polygone) ou comme pour les lignes, en calculant la distance entre les centres de chaque pixel extérieur (figure 5.3). L'*aire* est égale, en mode vecteur, à la somme des aires des triangles qui composent le polygone. En mode raster, l'aire est égale au nombre total de pixels de la région zonale (figure 5.4).
- La *forme* est définie par des indices de forme. Ces indices permettent de comparer les objets sans s'inquiéter de l'échelle ou de la taille. Cinq indices de compacité sont présentés dans la figure 6. Ces indices expriment la compacité relative d'un polygone par rapport à la forme compacte d'un cercle. La valeur maximale de compacité est égale à 1. Il existe une variété d'indices permettant de caractériser un polygone. L'indice de Gravelius (figure 6.6), par exemple, est souvent utilisé pour décrire la compacité des bassins-versants.

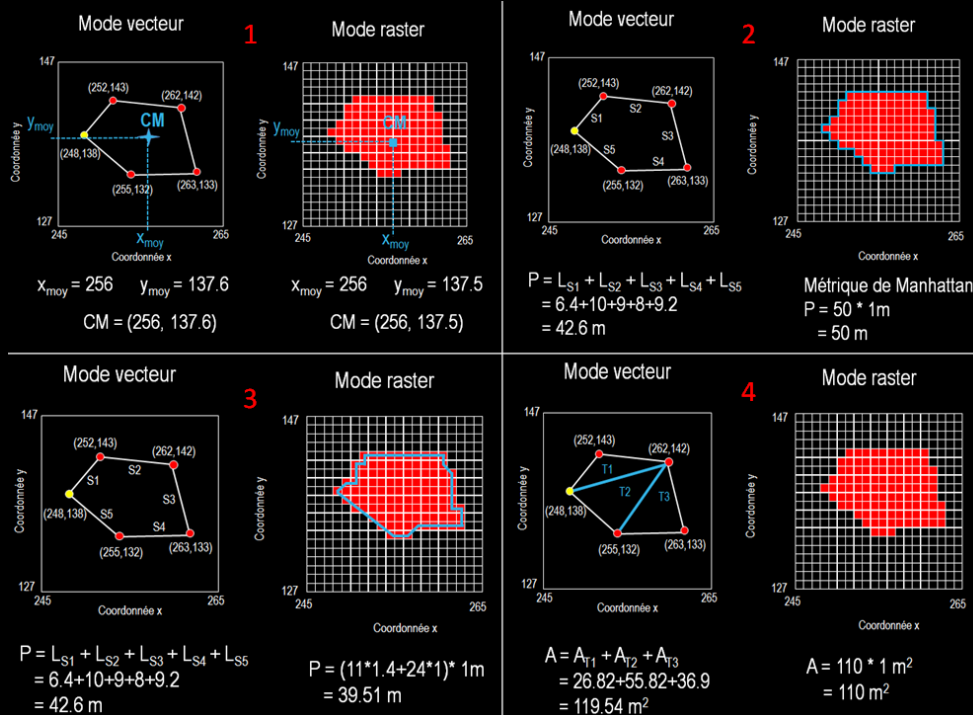
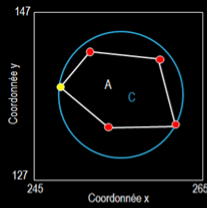


FIGURE 5

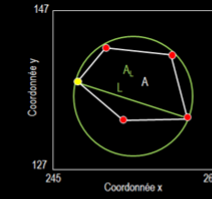
8:36

13:00



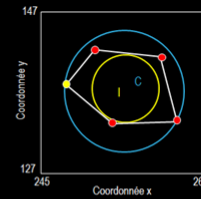
Indice 1 :  $S_{A,C} = A/C$

A = aire objet  
C = aire cercle circonscrit



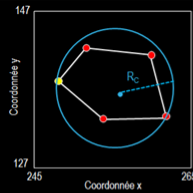
Indice 2 :  $S_{A,L} = A/A_L$

A = Aire de l'objet  
L = Axe majeur  
 $A_L = \text{Aire d'un cercle de diamètre } L = \pi (0.5 L)^2$



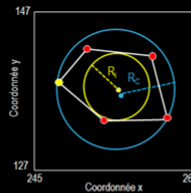
Indice 3 :  $S_{I,C} = I/C$

I = Aire du cercle inscrit  
C = Aire du cercle circonscrit



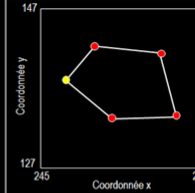
Indice 4 :  $SR_{A,C} = R_A / R_C$

$R_A$  = Rayon d'un cercle ayant la même aire que l'objet  
 $R_C$  = Rayon du cercle circonscrit



Indice 5 :  $SR_{I,C} = R_I / R_C$

$R_I$  = Rayon du cercle inscrit  
 $R_C$  = Rayon du cercle circonscrit



Indice de Gravelius K

$$K_G = P / 2\sqrt{\pi A}$$

P = Périmètre  
A = Aire

FIGURE 6



### 3. ARRANGEMENT ET VOISINAGE SPATIAUX

Il est possible de décrire les propriétés d'un ensemble d'objets spatiaux et la relation entre ces groupes d'objets en s'intéressant à leurs arrangements et leurs voisinages spatiaux.

Cette leçon présente les différentes notions et indicateurs d'arrangement et de voisinage spatiaux afin d'être capable de choisir l'indicateur le plus pertinent pour une situation donnée et d'en calculer sa valeur.

#### GÉNÉRALITÉS

Un **arrangement spatial** est l'organisation relative des objets spatiaux les uns par rapport aux autres. Le **voisinage spatial** est l'ensemble des relations spatiales permettant de situer les objets les uns par rapport aux autres. C'est une notion d'éloignement entre les objets spatiaux.

La nature de l'espace géographique pour un certain phénomène donné influence considérablement la notion d'arrangement et de voisinage. Ainsi une distance dépendra du moyen de déplacement utilisé et des caractéristiques du trajet si la pente et la qualité de la route sont prises en compte. Pour cette leçon, nous considérons l'espace géographique comme étant plan, homogène et isotrope. Nous ne considérons pas la dimension thématique. Les voisinages et arrangements des différents types d'entités spatiales sont donc décrits par trois critères: la *position*, la *dispersion* et la *proximité*.

#### OBJETS PONCTUELS

Les deux indices de position qui caractérisent l'arrangement et le voisinage spatial des objets ponctuels sont:

- Le *centre moyen* (figure 1.1), défini par les coordonnées moyennes en X et Y de tous les objets ponctuels considérés.
- Le *centre médian* (figure 1.1), défini par les coordonnées médianes en X et Y de tous les points. Le centre médian est moins influencé par les points de coordonnées extrêmes. Il est donc plus robuste que la moyenne.

La dispersion statistique des objets ponctuels est explorée en utilisant:

- L'*écart-type* des coordonnées X ( $S_x = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}}$ ) et Y ( $S_y = \sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}}$ ) de tous les points (figure 1.2).
- La *distance standard* ( $DS = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}$ ). La distance standard donne un indice unique pour les deux axes de coordonnées X et Y (figure 1.3).

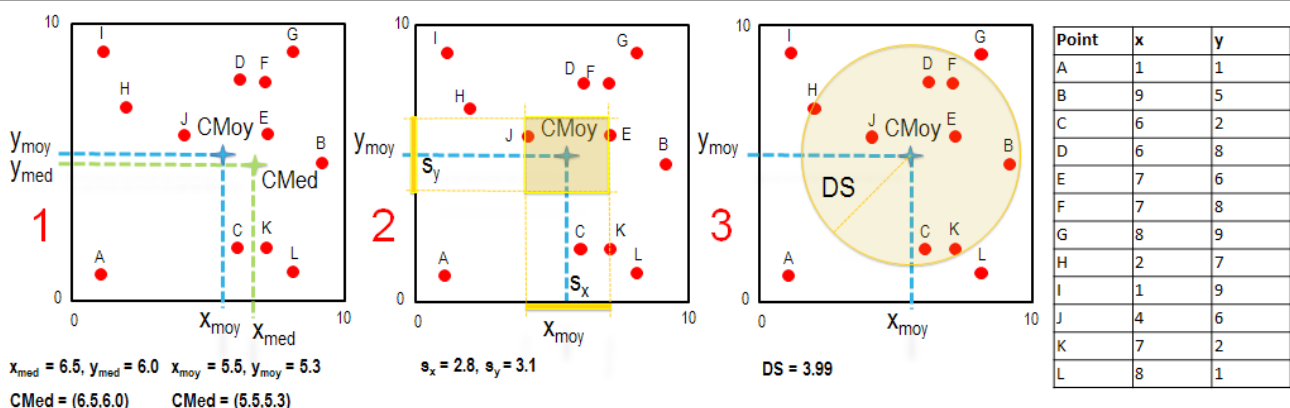


FIGURE 1

4:00

16:00

Il est souvent nécessaire d'effectuer un échantillonnage de points représentatifs dans l'espace plutôt que de mesurer un phénomène continu (ou discontinu) en tout point de l'espace. La distribution de cet échantillon de points est, ensuite, caractérisée en utilisant des indices d'arrangement spatial. L'indice  $R$  (ou indice du plus proche voisin), par exemple, compare une distribution observée de points à une distribution aléatoire. L'indice  $R$  est le rapport entre la distance moyenne au plus proche voisin et la distance moyenne pour une distribution aléatoire.

$$R = \frac{\bar{d}}{d_{al}} \text{ où } d = \sum_i d_i \text{ et } d_{al} = \frac{1}{2\sqrt{\frac{n}{S}}}$$

$S$  représente l'aire de la zone d'étude et  $n$  est le nombre de points. La distribution est dite groupée lorsque l'indice  $R$  vaut 0, elle est aléatoire pour  $R$  égal à 1 (figure 2).

La relation de voisinage des points permet également de caractériser spatialement la répartition d'un nuage de points. Cette relation peut être étudiée soit en attribuant à chaque pixel une couleur exprimant la distance au point le plus proche (en mode image). Soit par le découpage de l'espace en Polygones de Thiessen (en mode objet). La construction de ces polygones nécessite d'établir un réseau triangulé (TIN) en reliant les points entre eux pour former des mailles triangulaires. Le TIN vérifie le critère de Delaunay lorsque le cercle circonscrit de chaque triangle ne contient aucun autre point.

Les intersections des médiatrices de chaque côté des triangles forment les sommets des polygones de Thiessen. Ces polygones permettent de définir les zones dans lesquelles les points de l'espace sont plus proches d'un objet considéré que de tous les autres objets.

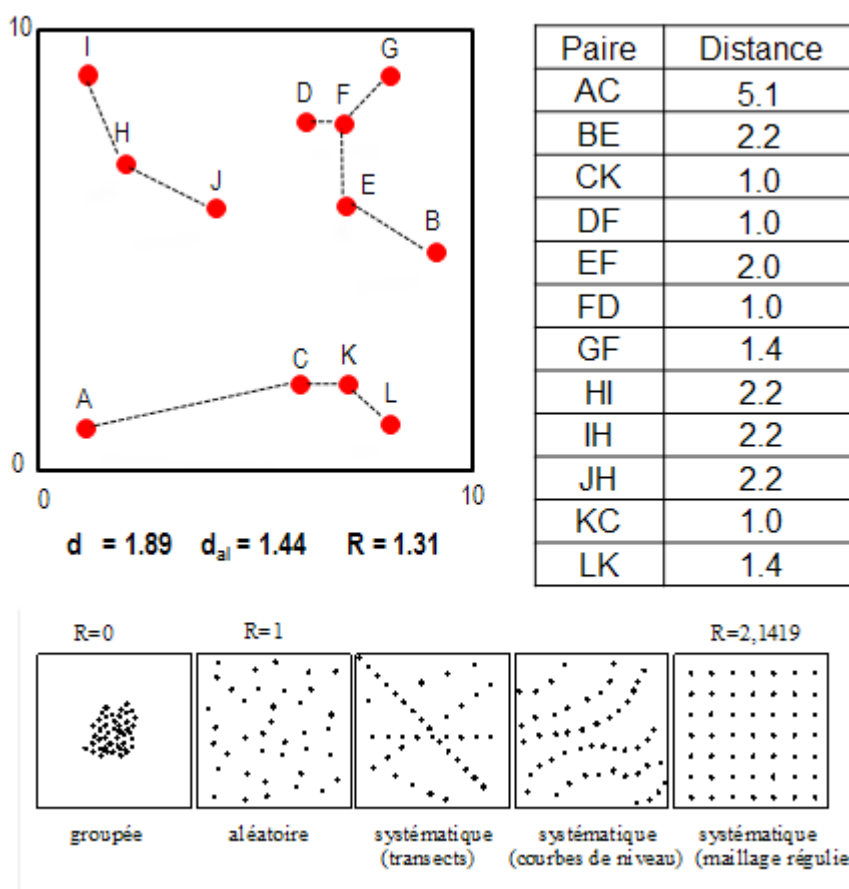


FIGURE 2

**OBJETS LINÉAIRES**

Les objets linéaires sont généralement représentatifs d'un réseau (hydrographique, routier, etc.). Il existe de nombreux indices d'arrangement et de voisinage spatial. Voici quelques indices et descripteurs.

- La taille moyenne ( $T_m = \frac{L}{C}$ ), définie par le rapport entre la longueur totale des chaînes ( $L$ ) et le nombre de chaînes ( $C$ ).
- La densité spatiale ( $D = \frac{L}{A}$ ), définie par le rapport entre la longueur totale des chaînes ( $L$ ) et l'aire de la zone étudiée ( $A$ ).
- L'indice de connexité ( $C = \frac{c}{0.5n*(n-1)}$ ), représente la complexité d'un réseau où  $c$  est égal aux nombres de chaînes et  $n$  représente le nombre de nœuds.
- L'indice de distance minimale est semblable à celui défini pour les objets ponctuels. En mode image, chaque pixel se voit attribuer une couleur qui exprime la distance aux pixels du segment le plus proche. En mode objet, les zones de plus grande proximité sont construites à partir des polygones de Thiessen. Comme pour les objets ponctuels, il est nécessaire de construire un réseau triangulaire irrégulier (TIN) à partir des nœuds pour ensuite déterminer le sommet des polygones de Thiessen.

**OBJETS ZONAUX**

Très utilisés en analyse du paysage, les indices d'arrangement (ou de structure) des objets surfaciques ou zonaux peuvent être divisés en trois catégories. Les indices de type *Patch* qui s'intéressent à chacun des objets surfaciques de la zone étudiée, les indices de type *Class* qui considèrent l'ensemble des objets zonaux d'une même catégorie et les indices de type *Landscape* qui se concentrent sur l'ensemble de la région étudiée. Les indices présentent peu de différence entre le mode raster et le mode objet. L'objet zonal devient une région zonale.

L'indice de distance minimale de type patch est défini de la même manière que pour les objets linéaires ou ponctuels. Il existe de nombreux indices de type *Class* qui se base sur l'aire, le périmètre des objets et le nombre d'objets. Voici quelques indices.

- L'indice de taille moyenne des surfaces d'une même classe (*Mean Patch Size*, MPS) est défini par l'aire de la classe ( $a_c$ ) divisée par le nombre d'objets de la même classe (NP).  $MPS = \frac{a_c}{NP}$ .
- L'indice de densité des contours (*Edge density*, ED) définit par le rapport entre la somme des périmètres des objets appartenant à une classe  $i$  et la somme des aires des objets appartenant à la même classe.

$$ED = \frac{\sum_j P_{ij}}{\sum_j a_{ij}}, \text{ où } P_{ij} \text{ est le périmètre de l'objet } j \text{ appartenant à la classe } i \text{ et } a_{ij} \text{ est l'aire de l'objet } j \text{ appartenant à la classe } i.$$

- L'indice de fractal moyen (*Mean Patch Fractal Dimension*, MPFD), plus complexe, est construit sur la base de dimension fractale ( $D$ ) de chaque patch. La dimension fractale est la façon qu'a une surface irrégulière de remplir l'espace.

$$MPFD = \frac{1}{n_i} \sum_j \frac{2 \ln(P_{ij})}{\ln(a_{ij})}, \text{ où } n_i \text{ est le nombre d'objets de la classe } i.$$

Les indices de structure des objets surfaciques de type *Landscape* décrivent la variabilité des propriétés thématiques dans l'ensemble de la zone d'étude. Voici quelques exemples couramment utilisés.

- L'indice de majorité ( $MOD = v_{f_{\max}}$ ), défini par la valeur pour laquelle l'aire est majoritaire dans toute la région ( $v_{f_{\max}}$ ).
- L'indice de diversité de Shannon ( $SHDI = -\sum_i (p_i * \ln(p_i))$ ), aussi appelé entropie, se base sur la proportion de la surface de la région d'étude occupée par la classe  $i$  ( $p_i$ ). La valeur de l'indice augmente ou diminue en fonction du nombre de catégories dans la zone étudiée ainsi que de la tendance à l'uniformité des aires de chaque catégorie.
- L'indice de régularité de Shannon ( $SHEI = \frac{\sum_i (p_i * \ln(p_i))}{\ln(c)}$ ), se base sur le rapport de la proportion de la surface de la région d'étude occupée par la classe  $i$  ( $p_i$ ) et le nombre de classes ( $c$ ). Les valeurs de l'indice varient entre 0 et 1. Un indice proche de zéro indique que les aires de chacune des catégories sont différentes tandis que lorsqu'il vaut 1 les aires sont égales.

## 4. AUTOCORRÉLATION ET DÉPENDANCE SPATIALE

Le concept de dépendance spatiale permet de déterminer dans quelle mesure la valeur de l'attribut d'un objet spatial dépend de sa position dans l'espace géographique. Dans cette leçon, vous apprendrez ce qu'est le concept de dépendance spatiale. Vous serez en mesure d'expliquer ce concept ainsi que d'identifier le biais causé par l'utilisation de la statistique classique dans un contexte spatial.

### LA PREMIÈRE LOI DE LA GÉOGRAPHIE

En 1970, le géographe quantitativiste Waldo Tobler décrit le concept de dépendance spatiale en exprimant ce qu'on appelle aujourd'hui la première loi de la géographie. La **première loi de la géographie** affirme que *tout interagit avec tout dans l'espace géographique, mais deux objets proches ont plus de chances de le faire que deux objets éloignés.*

Dans l'exemple de la figure 1, il est possible de faire la distinction entre les supporters de l'équipe du Brésil de ceux de la Croatie. Une personne vêtue de jaune ou de rouge et blanc a plus de chance d'interagir avec une personne habillée de la même couleur qu'elle. Ainsi, l'appartenance à un même groupe détermine la distribution spatiale de ces personnes. Et la dépendance spatiale est visible spatialement grâce à la couleur des t-shirts.

Les outils permettant de mesurer la dépendance spatiale sont basés sur la comparaison entre une répartition spatiale observée et une répartition spatiale aléatoire. Ils permettent de quantifier la régularité spatiale d'un phénomène, de déterminer le rayon d'action de la dépendance spatiale et de différencier une distribution spatiale observée d'une distribution aléatoire.



FIGURE 1

2:31

10:00



### NEUTRALITÉ DE L'ESPACE GÉOGRAPHIQUE

La quantification de la dépendance spatiale fait face à un paradoxe. En effet, selon la première loi de la géographie, des phénomènes tels que la température de l'air ou la densité de population ne sont pas distribués au hasard dans l'espace géographique. Or la mesure de la structure spatiale de tels phénomènes nécessite l'utilisation d'outils de statistique classique qui requièrent que les échantillons soient indépendants et aléatoirement distribués.

Les outils de la statistique classique ne conviennent donc pas dans le contexte géospatial. Leur utilisation est basée sur l'hypothèse d'un espace géographique neutre. Dans cet espace, la localisation d'observations ne doit, théoriquement, pas influencer leurs attributs. Néanmoins, comme il n'existe pas d'alternative à l'utilisation des outils de la statistique classique, il est nécessaire d'admettre des biais induits par leur utilisation avec des données géographiques et de choisir des jeux de données qui respectent les prérequis de la statistique.

La régression linéaire présente ce type de biais. Les observations nécessaires au calcul de la régression doivent être sélectionnées selon une procédure aléatoire. En effet, si les observations sont spatialement dépendantes, les valeurs estimées seront biaisées pour toute la zone d'étude. Des valeurs extrêmes situées dans des sous-régions influenceront les valeurs prédites sur toute la région étudiée. De même, une forte corrélation entre les attributs d'échantillons situés dans une petite sous-région influencera également toute la zone d'étude. Il est donc nécessaire que les calculs statistiques soient faits sur une sélection aléatoire des objets constituant les jeux de données géoréférencés.

## 5. AUTOCORRÉLATION SPATIALE GLOBALE

La régularité et la structure d'un phénomène spatial compte tenu du voisinage peuvent être quantifiées en mesurant l'autocorrélation spatiale. Dans cette leçon, l'indice d'autocorrélation spatiale – le I de Moran – sera présenté et vous apprendrez les notions nécessaires à son calcul et à son interprétation en tant que coefficient de régression. Vous serez aussi capable de choisir un mode de pondération spatiale avec pertinence. Un *indicateur global d'autocorrélation* décrit l'arrangement spatial d'entités géographiques sur l'intégralité d'une zone d'étude selon les valeurs d'un attribut donné. Il existe plusieurs indicateurs (C de Geary, K de Ripley, etc.), mais le I de Moran est l'indice le plus utilisé.

### RELATIONS DE VOISINAGE

Pour calculer l'autocorrélation spatiale globale, il est impératif de prendre en compte le voisinage des objets spatiaux considérés. L'**autocorrélation spatiale globale** compare le comportement d'un objet spatial avec le comportement de tous ses voisins sur l'intégralité de la zone étudiée. Elle est quantifiée par le calcul d'une corrélation entre les mesures géographiquement voisines d'un phénomène mesuré. La définition du voisinage d'un objet spatial dépend de l'attribut considéré et du type d'objet considéré. Voici certains critères de définition du voisinage présent dans le logiciel GeoDa.

- Objets ponctuels :
  - critères de distance avec noyau fixe (figure 1);
  - critères de proximité avec noyau variable ( $k$  plus proches voisins) (figure 1).
- Objets surfaciques (polygones) :
  - Relation de contiguïté de type Queen. Le voisinage correspond aux polygones qui touchent (au minimum une paire de coordonnées en commun) le polygone considéré (figure 2.1).
  - Relation de contiguïté de type Rook. Le voisinage inclut tous les polygones dont au moins un côté est commun au polygone d'intérêt (figure 2.2).

Les relations de contiguïté peuvent être d'ordre supérieur à 1. Dans ce cas-là, l'ordre définit jusqu'à quelle relation la contiguïté s'exerce. Par exemple, une relation de contiguïté d'ordre 2 indique que l'on considère les voisins (selon Queen ou Rook) des polygones d'ordre 1. Il est ensuite possible de considérer ou non les polygones d'ordre inférieur suivant le besoin.

Pour les objets linéaires et surfaciques, il est également possible d'utiliser des critères de distance ou de proximité en utilisant les coordonnées des centres de gravité géométriques des objets.

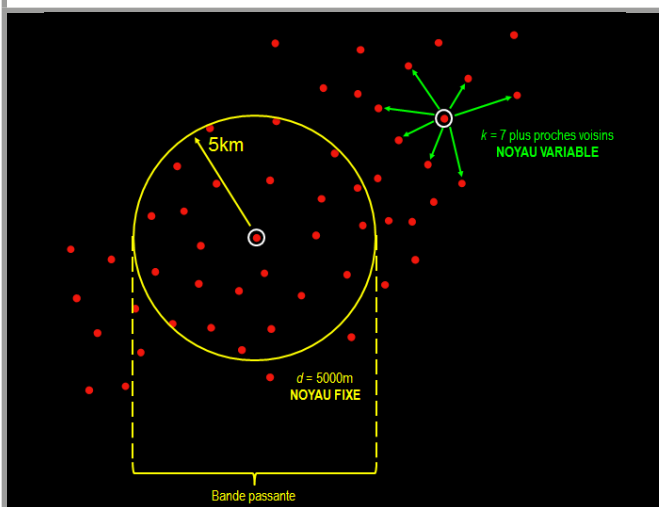


FIGURE 1

4:32

21:00

Exemple de critères de définition du voisinage dans le cas d'objets ponctuels.

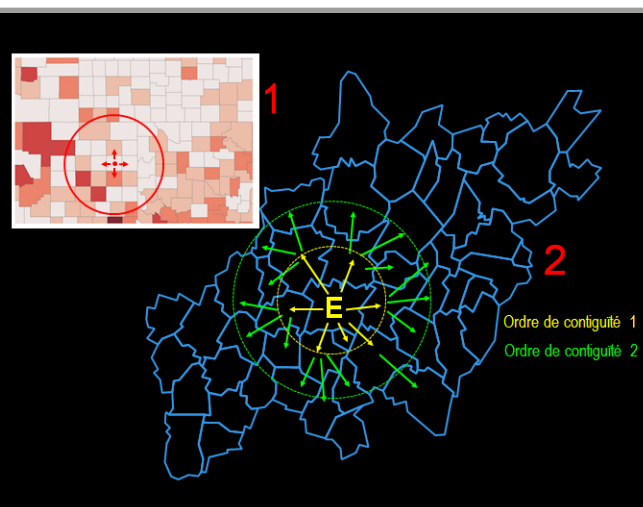


FIGURE 2

5:56

21:00

Exemple de relation de contiguïté: 1) relation de type Rook; 2) relation de type Queen avec deux ordres de contiguïté.

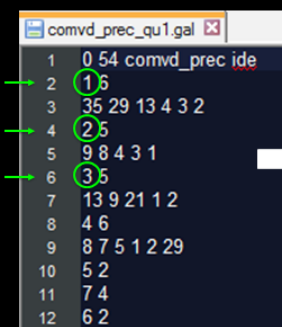
## PONDÉRATION SPATIALE

Le schéma de pondération spatiale est défini à partir des relations de voisinage. Dans le logiciel GeoDa, ce schéma de pondération est exprimé par la création d'un fichier de pondération spatiale dans lequel le nombre total et le numéro des voisins de chaque polygone sont inscrits.

## I DE MORAN

Le *I* de Moran ( $I_m$ ) exprime la corrélation entre l'attribut d'un objet géographique ( $Z$ ) et la moyenne de cet attribut au sein d'un voisinage défini ( $\bar{Z}$ ). Il est défini par le rapport entre la covariance d'une variable et sa variance sur toute la zone d'étude. Le fichier de pondération est utilisé pour calculer la moyenne de l'attribut dans le voisinage défini. Dans l'exemple de la figure 3, la première colonne (*ide*) présente le numéro d'identification du polynôme étudié. La deuxième colonne contient la pondération, ici les numéros d'identification des voisins du polynôme étudié. La colonne *C* exprime le nombre de voisins.  $Z$  est la valeur de l'attribut étudié et  $\bar{Z}$  est la valeur moyenne des attributs des voisins.

Luc Anselin, propose l'interprétation du *I* de Moran en tant que coefficient de régression. En appliquant une régression de la variable indépendante (variable étudiée,  $Z$ ) sur la variable dépendante (moyenne de la variable étudiée dans son voisinage,  $\bar{Z}$ ) il est possible de visualiser la régression dans un scatter plot bivarié et d'en déduire la pente de la droite de régression, qui est égale au *I* de Moran. Un *I* de Moran élevé pour une variable étudiée et un voisinage défini indique qu'il y a autocorrélation spatiale, c'est-à-dire que la variable étudiée est spatialement dépendante pour un voisinage considéré.



```

comvnd_prec_qu1.gal
1 0 54 comvnd_prec ide
2 1 6
3 35 29 13 4 3 2
4 2 5
5 9 8 4 3 1
6 3 5
7 13 9 21 1 2
8 4 6
9 8 7 5 1 2 29
10 5 2
11 7 4
12 6 2
  
```

ide	voisins ( $\rightarrow \omega$ )	C	Z	$\bar{Z}$
1	35, 29, 13, 4, 3, 2	6	10291.14	10244.17
2	9, 8, 4, 3, 1	5	10166.64	10085.33
3	13, 9, 21, 1, 2	5	10494.71	10334.34
...	...	...	...	...

$$I_m = \frac{\text{Covariance}}{\text{Variance}} = \frac{1/C \sum \omega_{i,j} (z_i - \bar{z})(z_j - \bar{z})}{1/n \sum (z_i - \bar{z})^2} = \frac{n \sum (z_i - \bar{z})(z_j - \bar{z})}{C \sum (z_i - \bar{z})^2}$$

Où  $n$ : nombre d'unités spatiales;  $C$ : nombre de voisins ou de connexions;  $z_i$ : valeur de la variable pour l'unité  $i$ ;  $z_j$ : valeur de la variable pour l'unité  $j$ ;  $\omega_{i,j}$ : poids de la connexion, 1 si adjacent, 0 autrement

La valeur de  $I$  varie entre +1 (corrélation positive totale) et -1 (corrélation négative totale); 0 signifie absence d'autocorrélation, pas de dépendance spatiale, ou encore espace géographique neutre

FIGURE 3

10:15

21:00



## ÉVALUER LA SIGNIFICATIVITÉ

Une valeur élevée de  $I$  de Moran n'est pas nécessairement significative. En effet, il se pourrait aussi que l'agencement spatial d'une variable soit dû au hasard. Dans ce cas, la valeur élevée du  $I$  de Moran ne permettrait pas de dire que la variable étudiée est dépendante spatialement. Il est donc important d'évaluer la significativité du  $I$  de Moran. Pour évaluer sa significativité, il suffit de comparer le  $I$  de Moran observé avec des milliers de  $I$  de Moran calculés après chaque permutation aléatoire des valeurs entre toutes les autres localisations possibles dans la zone étudiée.

Dans l'exemple du logiciel GeoDa de la figure 4, le  $I$  de Moran de la moyenne mensuelle de la somme des précipitations de 54 communes a été calculé et 999 permutations de ces valeurs de précipitation ont permis de calculer 999  $I$  de Moran. Les valeurs obtenues sont présentées sous forme d'histogramme. L'espace neutre est représenté par la distribution aléatoire autour de la moyenne tandis que le  $I$  de Moran observé est indiqué en jaune. Dans cette figure, le  $I$  de Moran observé se situe en dehors de la distribution aléatoire, il est donc significatif.

La significativité du  $I$  de Moran peut aussi être exprimée numériquement par la probabilité de rejeter l'hypothèse nulle. L'hypothèse nulle affirme que la valeur observée est aléatoire. La  $p$ -valeur est la valeur limite de rejet et dépend du nombre de permutations effectué. Une  $p$ -valeur faible indique que le risque de rejeter l'hypothèse nulle alors qu'elle est vraie est faible et donc que la valeur observée est significativement différente d'une distribution aléatoire.

$$p\text{-valeur} = \frac{\text{Nb } I_{al} \geq I_{obs} + 1}{\text{Nb permutations} + 1}$$

$$\text{ou } \frac{\text{Nb } I_{al} \leq I_{obs} + 1}{\text{Nb permutations} + 1}$$

$$\text{Ici, la } p\text{-valeur} = \frac{0+1}{999+1} = 0.001$$

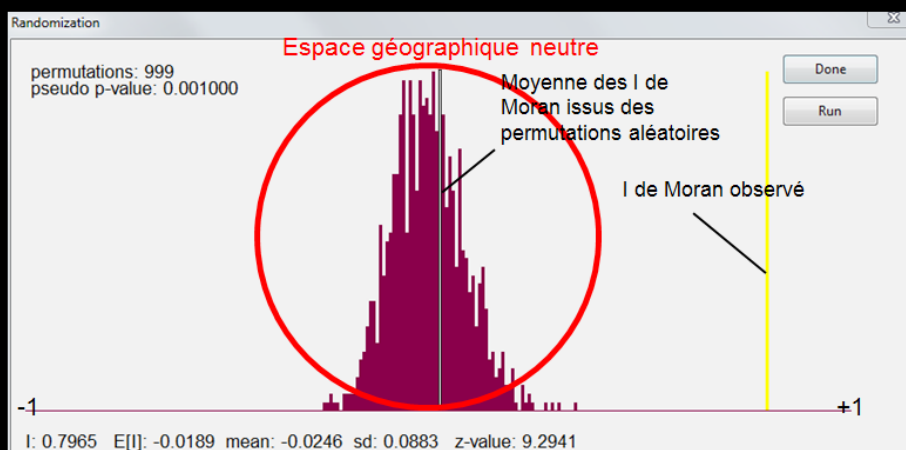


FIGURE 3

15:03

21:00

Exemple d'évaluation de la significativité du  $I$  de Moran par permutations aléatoires.

## 6. ÉCHANTILLONNAGE

L'altitude, l'humidité du sol ou la température à la surface du sol sont tous des phénomènes continus. Ce type de données doit être acquis par échantillonnage. Dans cette leçon, les différentes procédures d'échantillonnage spatial ainsi que ses propriétés désirables seront présentées. Ces informations permettront d'appliquer une procédure d'échantillonnage de manière pertinente.

### PHÉNOMÈNES CONTINUS

Les phénomènes continus (précipitation, altitude, humidité du sol, etc.) sont définis en tout point de l'espace et varient localement de manière graduelle et structurée.

### PROPRIÉTÉS DE L'ÉCHANTILLONNAGE

L'*échantillonnage* est la démarche qui consiste à mesurer une variable d'intérêt en quelques points représentatifs d'une zone d'étude. L'*interpolation* est la méthode permettant de déduire la valeur en tout point de l'espace à partir des points d'échantillonnage. L'interpolation est applicable si l'estimation des valeurs prise par une variable en d'autres points de l'espace considéré possède une précision acceptable.

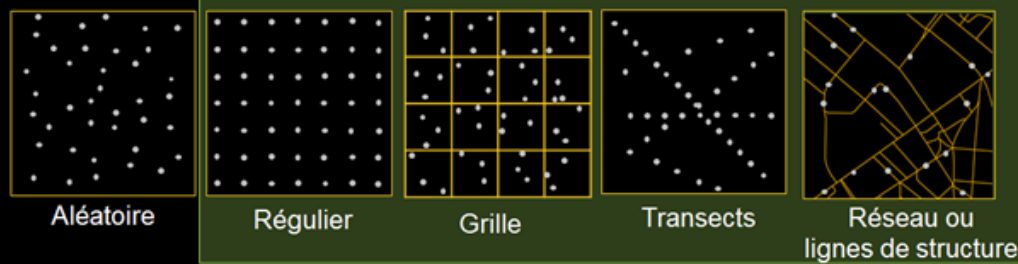
La première qualité d'un échantillonnage est celle d'être représentatif de tout l'espace considéré. Un échantillon est optimal s'il satisfait les conditions suivantes :

- Les limites de la zone d'étude sont circonscrites selon des critères précis.
- La répartition des échantillons est homogène et de densité optimale sur l'ensemble du territoire étudié.

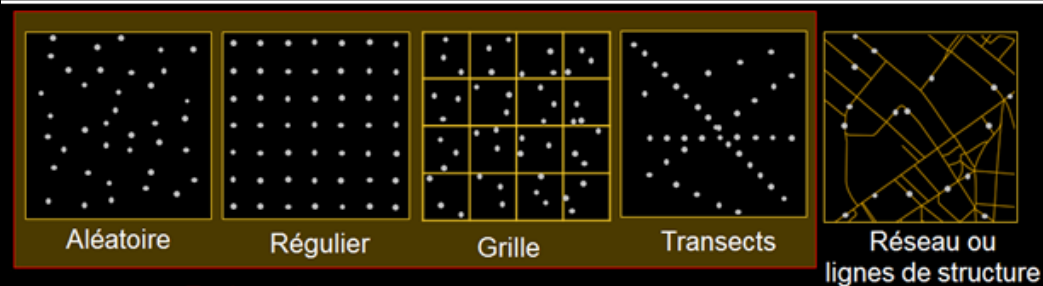
Dans la pratique, les limites de la zone d'étude sont souvent fixées de manière progressive en s'appuyant sur les résultats précédents et la distribution des échantillons de départ est souvent arbitrairement choisie pour ensuite être complétée par de futurs échantillonnages dans des endroits jugés judicieux. Dans cette optique, l'utilisation de la notion de domaines est pertinente. Un *domaine* est une partie de la zone d'étude dans laquelle un phénomène est censé exister, présenter un comportement similaire et homogène. Le domaine est *délimité* lorsque ses limites sont connues et *ouvert* lorsque ses limites sont inconnues ou approximatives. Une même loi d'interpolation pourra donc être appliquée à l'ensemble d'un domaine ou à un sous-domaine lorsqu'une zone d'étude présente de l'hétérogénéité.

### TYPES D'ÉCHANTILLONNAGE

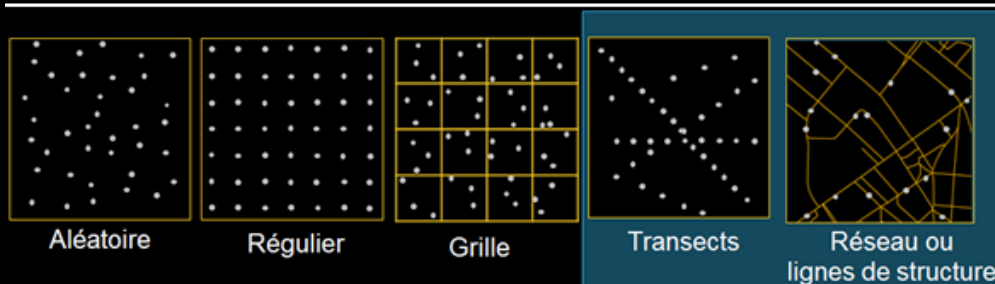
La quantité d'information préalable disponible sur le comportement d'un phénomène étudié influence fortement le choix du type d'échantillonnage. En absence d'information préalable, deux types d'échantillonnages peuvent être appliqués. L'échantillonnage aléatoire ou l'échantillonnage systématique (régulier, grille, transects ou lignes de structures). En présence d'information préalable, il est possible de mettre en place un échantillonnage le long de transects ou de lignes de structures (le long d'une rivière par exemple). Un échantillonnage exploratoire est mis en place afin de connaître progressivement le comportement d'un phénomène, lorsque celui-ci est invisible.



**Systématique:** un schéma de prélèvement est établi préalablement



**Exploratoire**



**Connaissance a priori du phénomène**

FIGURE 1

6:25

13:00

Types d'échantillonnage.

### QUALITÉ DE L'ÉCHANTILLON

Un échantillon spatial est de qualité acceptable s'il représente géométriquement et thématiquement le comportement du phénomène étudié. L'obtention d'une densité optimale d'échantillonnage (densité de points) permet de garantir une représentativité acceptable de la distribution. L'échantillonnage d'une zone d'étude doit respecter certaines contraintes en vue de garantir une représentativité acceptable de la distribution et de garantir un processus de régionalisation par interpolation uniquement. Ces contraintes sont d'une part l'obtention d'une densité optimale de points d'échantillonnage et la constitution d'un ensemble de points d'échantillons extérieurs à la zone étudiée. Des échantillons extérieurs à la zone d'étude, pris dans une couronne suffisamment large, permettent d'éviter des effets de bords lors de l'interpolation.

Thématiquement, les échantillons doivent représenter de la meilleure manière possible la dynamique des valeurs présente dans la région d'étude. On considère donc également la dimension temporelle afin que les valeurs thématiques soient correctement représentatives du comportement d'un phénomène pour une période de temps définie. Dans les cas où le phénomène présente une forte variabilité (température de l'air par exemple), il est possible de calculer un indice de tendance centrale pour exprimer le comportement général de la variable.

## 7. INTERPOLATION I

À partir de l'échantillonnage, il est possible, par interpolation, d'estimer la valeur d'une variable en tout point de la zone étudiée. La méthode d'interpolation permet donc de généraliser l'information discrète acquise sur le territoire. L'interpolation est une prédiction justifiée d'une valeur de la variable en un point non mesuré (figure 1). Dans cette leçon, le fonctionnement et la mise en pratique des principales méthodes d'interpolation déterministes seront expliquées.

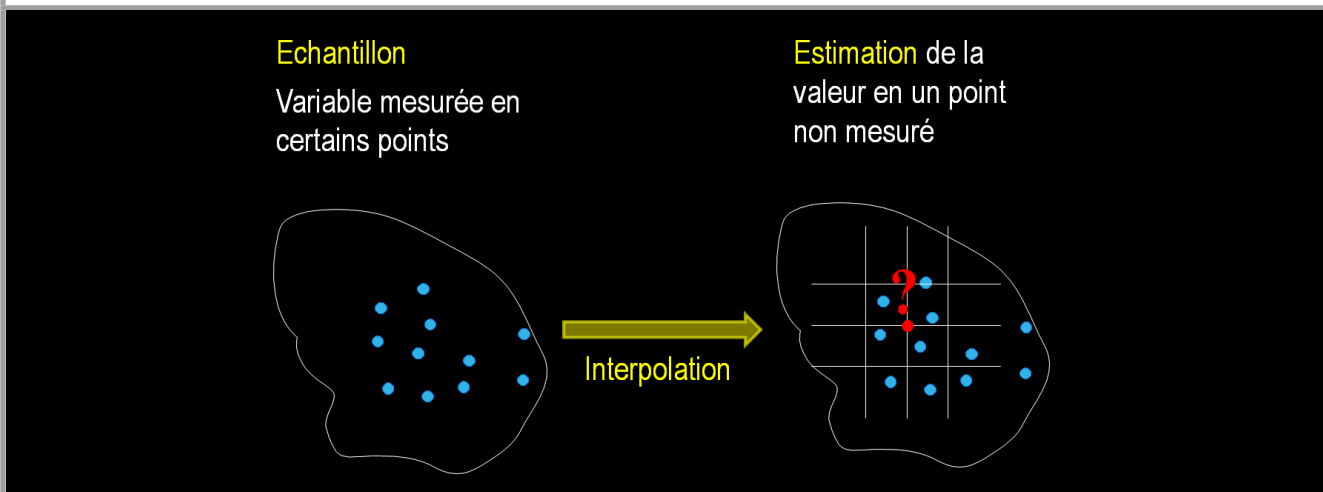


FIGURE 1

1:12

13:00

L'interpolation : de l'échantillon à l'estimation.

### MÉTHODES DÉTERMINISTES

Les **méthodes d'interpolation déterministes** sont des méthodes d'interpolation dont aucune étude statistique n'est effectuée lors de leur application. Parmi ces méthodes déterministes, il existe deux groupes distincts :

- les méthodes globales ; pour lesquelles l'estimation du phénomène est construite en utilisant tous les points du jeu de données ;
- les méthodes locales ; pour lesquelles l'estimation du phénomène se base sur un nombre limité de points d'appui au voisinage du point à estimer et pour lesquelles il est possible d'attribuer des poids spécifiques à chaque point de mesures.

### MÉTHODES GLOBALES

Une méthode globale d'interpolation analyse les tendances de distribution spatiale d'un phénomène dans un domaine donné. Un polynôme de degrés 1 ou supérieur à 1 peut, par exemple, être utilisé pour représenter la distribution spatiale du phénomène sur l'ensemble du domaine. Pour cela, il est nécessaire de trouver le polynôme qui minimise les écarts aux points de mesures en faisant intervenir la méthode des moindres carrés. Il en résulte une surface de tendance représentative du comportement moyen du phénomène dans le plan X, Y.

Les méthodes locales sont souvent préférées aux méthodes globales. En effet, les modèles polynomiaux de degrés inférieurs à 4 sont généralement trop simples pour fournir une approximation réaliste et ceux de degrés supérieurs sont trop compliqués à calculer. De plus, il est admis, par hypothèse, que la dépendance spatiale est locale et non globale.

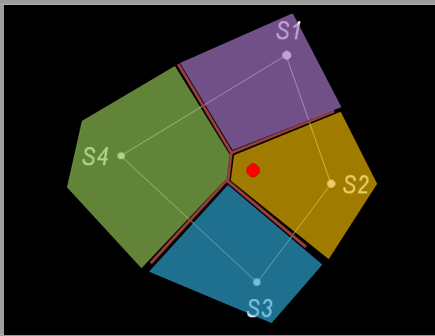


FIGURE 2

5:07

13:00

Interpolation par la méthode du plus proche voisin. Le point rouge prend la valeur du point S2.

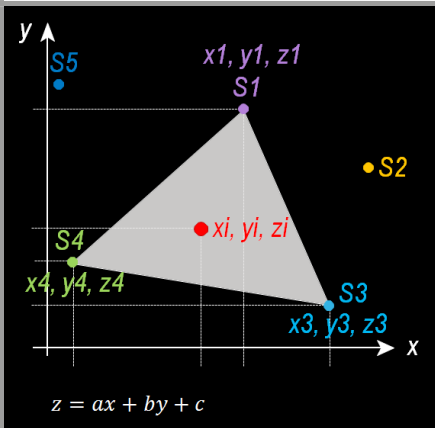


FIGURE 3

5:55

13:00

Interpolation par la méthode TIN. La valeur du point rouge est estimée en résolvant l'équation à trois inconnues.

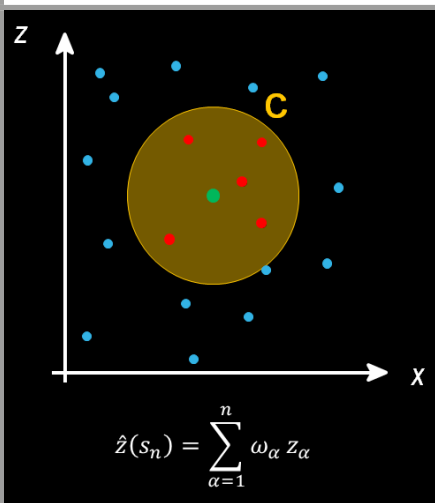


FIGURE 4

9:28

13:00

Interpolation par moyenne mobile pondérée. Le point vert est estimé par une moyenne pondérée des points de son voisinage.

## MÉTHODES LOCALES

Les méthodes locales ont la particularité de permettre la pondération des points de mesures. Cette pondération peut, par exemple, être attribuée en fonction de la distance entre les points mesurés et le point à estimer ou en fonction de la qualité de mesure des points. Choisir une méthode locale d'interpolation dépend principalement de la densité de l'information à disposition, de la précision souhaitée, de l'objectif principal de l'interpolation et de la connaissance préalable du phénomène.

Voici quelques méthodes locales d'interpolation :

- La méthode du plus proche voisin. Dans cette méthode, la valeur estimée en un point quelconque est égale à la valeur du point mesuré le plus proche. Chaque point de mesure possède une zone d'influence spécifique dans laquelle tous les points sont plus proches de lui que de tous les autres. Ces zones sont déterminées par les polygones de Thiessen créés à partir des médiatrices des segments. Tout point à estimer à l'intérieur d'un polygone prend donc la valeur du point correspondant à la zone d'influence. Cette méthode discrétise le phénomène normalement spatialement continu (figure 2).
- La méthode basée sur un réseau de triangles (TIN = *Triangulated Irregular Network*). Le TIN est créé en reliant les points d'appui entre eux et en respectant le critère de Delaunay. La variation du paramètre à l'intérieur des triangles est supposée linéaire et l'interpolation est réalisée en établissant l'équation à trois inconnues du plan formé par chacune des faces du triangle. Si  $z$  est la valeur associée au phénomène et  $a, b$  et  $c$  les inconnues. L'équation  $z = ax + by + c$  est résolue en utilisant les trois sommets du triangle. À partir de là, tout point contenu dans le plan formé par les faces du triangle peut être estimé (figure 3).
- La méthode de la fonction B-spline calcule un polynôme du troisième degré pour quatre points successifs contenus dans une fenêtre mobile. Dans le cas à une dimension, pour chaque intervalle, l'évolution de la variable à prédire ( $z(x)$ ) est simulée au moyen d'un polynôme de degrés 3 ( $p(x)$ ). La fenêtre mobile glisse vers le prochain point et un nouveau polynôme est estimé. Le polynôme  $p(x)$  est calculé en respectant des contraintes à la jonction entre deux polynômes. Une contrainte liée à l'intervalle dans laquelle  $x$  est compris et une contrainte de continuité entre les segments. Lorsqu'un point d'inflexion se trouve à la jonction entre deux polynômes, la continuité est validée par égalité approximative des dérivées secondes.
- La moyenne mobile pondérée (*Inverse Distance Weighting, IDW*) s'appuie sur l'intuition que l'influence d'un point éloigné est moins importante qu'un point rapproché. Elle permet donc d'estimer une variable en s'appuyant sur quelques valeurs de point mesurées dans un voisinage en leur attribuant un poids. Pour mettre en place cette méthode, il est nécessaire de définir certains paramètres tels que la taille du voisinage, la pondération de chaque point et l'orientation (anisotropie). Dans le cadre déterministe, les valeurs des paramètres sont définies par l'expérience de l'analyste pour un phénomène donné. Dans l'exemple de la figure 4, le voisinage (C) est constitué de 5 points et la valeur prédite ( $\hat{z}(S_n) = \sum_{\alpha=1}^n \omega_{\alpha} z_{\alpha}$ ) est égale à la moyenne pondérée des 5 points d'appui considérés. Le poids ( $\omega_{\alpha}$ ) est souvent exprimé en fonction de l'inverse de la distance élevée à une certaine puissance. Plus la puissance est grande plus l'effet local est accentué.



Bien qu'il ne soit pas possible d'estimer directement l'incertitude des valeurs estimées par ces méthodes, il est possible de procéder par *validation croisée*. En partageant aléatoirement l'échantillon en deux groupes, l'un destiné à l'interpolation et l'autre comme groupe test, une comparaison peut être établie entre la valeur interpolée et la valeur mesurée en chaque point d'échantillon du groupe test. L'incertitude est ensuite déduite en calculant la différence des moyennes quadratiques des distributions. Pour augmenter le nombre de tests et améliorer la qualité de l'incertitude, l'opération (partage aléatoire de l'échantillon, comparaison et estimation de l'incertitude) est répétée plusieurs fois. Cette méthode de validation croisée permet d'évaluer l'incertitude globale et locale des valeurs interpolées.

Les méthodes d'interpolation citées plus haut sont toutes empiriques et leurs résultats sont influencés par la subjectivité et l'expérience de l'analyste. Plus l'information manque sur le phénomène étudié plus le choix de la fonction d'interpolation est arbitraire. Ces incertitudes engendrent souvent des coûts supplémentaires. Cette conséquence fâcheuse a motivé de nombreux chercheurs à trouver d'autres méthodes d'interpolation qui ont abouti au concept de variables régionalisées.

## 8. INTERPOLATION II – KRIGEAGE

L'incertitude des résultats obtenus par les méthodes d'interpolation déterministes locales a poussé les spécialistes et chercheurs à approfondir les connaissances sur la distribution spatiale. Dès les années 1940, Daniel Krige et Georges Matheron ont développé le concept de variable régionalisée. Dans cette leçon, la notion de variable régionalisée en géostatistique, l'analyse variographique et le fonctionnement du krigeage seront présentés et expliqués. Il sera ensuite possible d'appliquer une procédure d'analyse variographique et d'interpoler un jeu de mesures par krigeage ordinaire.

### INTRODUCTION À LA GÉOSTATISTIQUE

La géostatistique s'est développée avec les travaux de Kolmogorov et Wiener au cours des années 1930 et 1940 sur l'étude des fonctions aléatoires. Mais les premiers travaux ont été attribués à Danie Krige, un ingénieur des mines sud-africain. Ses travaux portaient sur la corrélation entre deux points échantillonnés en fonction de la distance qui les sépare. Il est le premier à proposer le *variogramme* comme représentation de cette corrélation. Georges Matheron, un mathématicien français, reprend les travaux de Krige et la notion de fonction aléatoire de Kolmogorov et Wiener pour établir les fondements théoriques de la géostatistique, dans les années 1960. Il introduit le concept de *variable régionalisée* pour décrire la nature particulière d'une variable caractérisant un phénomène spatial continu.

Le *variogramme* et la notion de *variable régionalisée* sont les outils de base des techniques d'interpolation par krigeage. Le **variogramme** est un outil de géostatistique qui met en évidence des valeurs selon la distance qui les sépare. D'après Matheron, « **une variable régionalisée** est une fonction de l'espace  $Z(x, y) = Z(s)$  dont la valeur varie avec une certaine apparence de continuité, sans qu'il soit, en général, possible d'en représenter la variation par une loi mathématique extrapolable » (Matheron, 1965, 1970).

Pour bien comprendre ces deux outils, il est possible d'appliquer une procédure progressive d'analyse variographique à un jeu de données empiriques. La première étape consiste à cartographier le domaine échantillonné et à calculer des indicateurs statistiques descriptifs (moyenne, max, etc.) pour permettre de visualiser les propriétés des valeurs mesurées. Cette étape permet de déceler les incohérences ou les valeurs aberrantes pouvant provenir d'erreurs de mesure.

### NUÉE VARIOGRAPHIQUE

Les statistiques descriptives vues précédemment ne permettent pas de caractériser le comportement local d'une variable d'intérêt. Pour combler cette lacune et mettre en évidence le comportement local et spatial de la variable, il est nécessaire de faire intervenir le principe de nuée variographique. La **nuée variographique** est un outil capable d'analyser la différence de valeur entre les paires de points du domaine (figure 1). Pour utiliser cet outil, il est nécessaire de créer un indicateur  $\gamma^*(h)$  où  $h$  est la distance entre deux points de mesure et  $\gamma^*(h)$  est définie comme des demi-écarts quadratiques. C'est-à-dire le carré de la différence entre deux points divisé par deux.

$$\gamma^*(h_{\alpha,\beta}) = \frac{(z(S_\alpha) - z(S_\beta))^2}{2}$$

avec

$$h_{\alpha,\beta} = |S_\alpha - S_\beta| = h_{\beta,\alpha}$$

où  $\alpha$  et  $\beta$  sont deux points de mesure.  $S_\alpha$  et  $S_\beta$  sont les vecteurs position des points échantillonnés  $\alpha$  et  $\beta$ .  $z(S_\alpha)$  est la valeur de la variable du point  $\alpha$ .

Les valeurs de  $\gamma^*(h)$  peuvent être visualisées au travers d'un diagramme de dispersion de  $\gamma^*(h)$  par rapport à  $h$ . Ce diagramme est omnidirectionnel et porte le nom de *nuée variographique*.



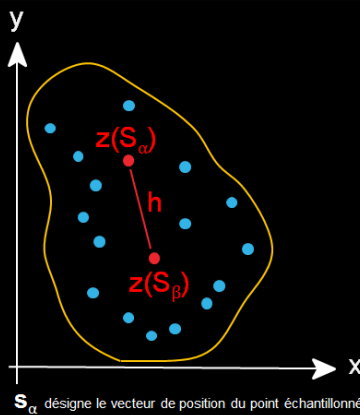


FIGURE 1

6:02

19:00

Différences de valeurs entre paires de points.

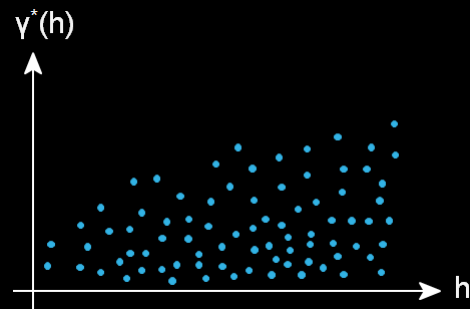


FIGURE 2

6:19

19:00

Exemple de nuée variographique.

### VARIOGRAMME EXPÉRIMENTAL

Le grand nombre de paires formées par la nuée variographique ne permet pas de décrire aisément la dispersion des demi-écartés quadratiques. Néanmoins, la segmentation du domaine en classe de distance, à la manière d'un histogramme, permet de simplifier la visualisation. Pour chaque pas (lag)  $\Delta h$  choisi, et pour chaque intervalle  $h(0)$  et  $h(n)$  correspondante, la moyenne des valeurs de  $\gamma^*(h)$  est calculée. À chaque classe de distance correspond une certaine moyenne et cette dispersion quadratique ( $\gamma(h)$ ) est appelée *variogramme expérimental* (figure 3).

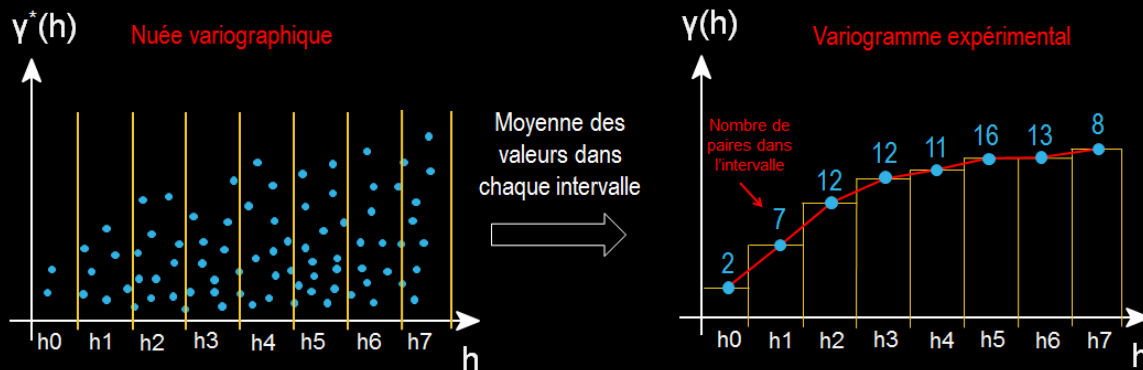


FIGURE 3

7:41

19:00

De la nuée variographique au variogramme expérimental.

Pour que la moyenne à l'intérieur des classes soit statistiquement représentative, il est important que le nombre de paires de chaque classe soit adéquat. Plus le nombre de paires est élevé, plus la moyenne est représentative. Il est possible d'ajuster le nombre de paires dans les classes en ajustant le pas.

L'interprétation du variogramme expérimental est facilitée en reliant les points des moyennes par une ligne. Trois paramètres importants permettent l'interprétation du variogramme (figure 4).

- La *portée*: distance à partir de laquelle il n'y a plus d'autocorrélation.
- Le *palier*: partie plane du variogramme, censée apparaître au-delà de la portée.
- La *pépite*: valeur non nulle à  $\gamma(0)$ . Proportionnelle à l'incertitude de la valeur mesurée.

En théorie, la hauteur du palier correspond à la variance de la variable si la moyenne de la variable est constante. Si cette hypothèse est respectée alors une valeur  $\gamma(h(x))$  correspond à l'incertitude d'une valeur estimée ( $z(s)$ ) et éloignée d'une distance  $h(x)$  du point d'appui le plus proche.

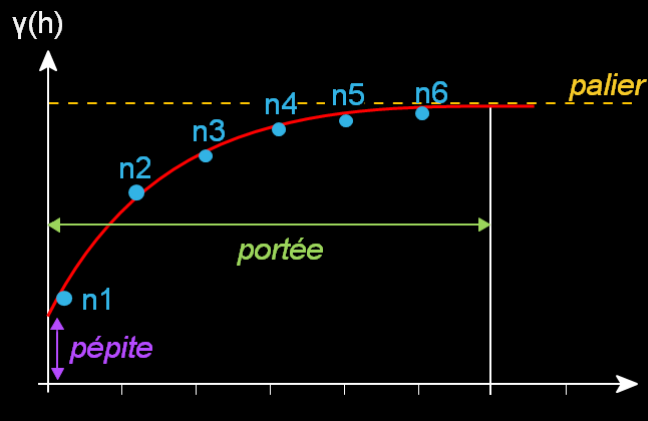


FIGURE 4

8:28

19:00

Paramètres du variogramme expérimental.

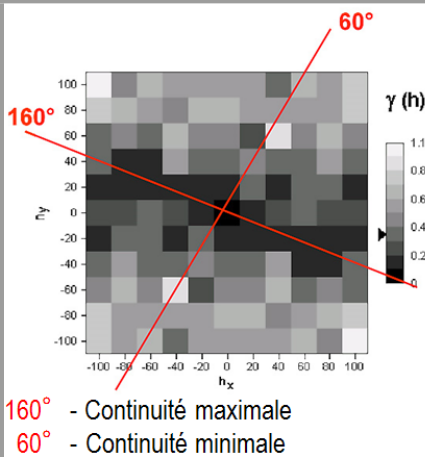


FIGURE 5

10:40

19:00

Variogramme surfacique. Les principales directions d'anisotropie sont mises en évidence.

### VARIOGRAMME SURFACIQUE ET DIRECTIONNEL

Jusqu'à présent, la nuée variographique et le variogramme expérimental ne prenaient pas en compte l'orientation des paires de points. Un phénomène est dit *anisotrope* s'il présente un comportement différent selon l'orientation. Le relief d'une vallée est anisotrope. En effet, si l'on considère l'altitude dans la direction du cours d'eau, ou perpendiculairement à celui-ci, son comportement est différent.

Pour prendre en compte l'anisotropie, il faut tenir compte de l'orientation des paires de points en représentant la nuée variographique ou le variogramme selon les composantes  $h_x$  et  $h_y$  de la distance  $h$  et en projetant les valeurs de  $\gamma(h)$  perpendiculairement au plan  $h_x, h_y$ . Il en résulte un variogramme bivarié appelé variogramme surfacique (figure 5). Un variogramme surfacique dont les valeurs sont uniformément réparties autour du centre indique que la variable étudiée est isotrope et les isovaleurs forment des cercles concentriques. Lorsque les isovaleurs forment des ellipses concentriques, il y a anisotropie géométrique et le palier est le même dans toutes les directions. Lorsque le palier diffère selon les directions alors il y a une anisotropie zonale.

Lorsqu'une anisotropie est détectée, les variogrammes dans la direction observée sont calculés en définissant une tolérance angulaire afin d'éviter un nombre trop faible de paires.

### VARIOGRAMME THÉORIQUE

Le variogramme expérimental permet d'exprimer la dépendance entre deux valeurs mesurées en fonction de la distance qui les sépare. Cette dépendance peut être utilisée pour pondérer les mesures dans une procédure d'interpolation. Pour mettre en œuvre une interpolation par krigeage, il est nécessaire d'interpréter les variables spatiales échantillonnées en tant que variable aléatoire. Néanmoins, l'information fournie par un variogramme expérimental ne permet pas de résoudre les contraintes fixées aux variables aléatoires pour la résolution des équations de krigeage. Il est donc primordial d'utiliser un variogramme théorique.

Le variogramme théorique peut être obtenu en ajustant un nombre restreint de fonctions (fonctions autorisées (figure 6)) ou une combinaison de celles-ci à un variogramme expérimental. Cette opération est généralement proposée par les logiciels de variographie. Néanmoins, il est important de préciser que même en cas d'anisotropie, un seul variogramme est défini pour l'ensemble du domaine et pour toutes les directions.

Les fonctions autorisées sont :

- le modèle de la pépité pure ( $\gamma(h) = \text{constant}$ );
- le modèle linéaire;
- le modèle exponentiel;
- le modèle gaussien;
- le modèle sphérique.

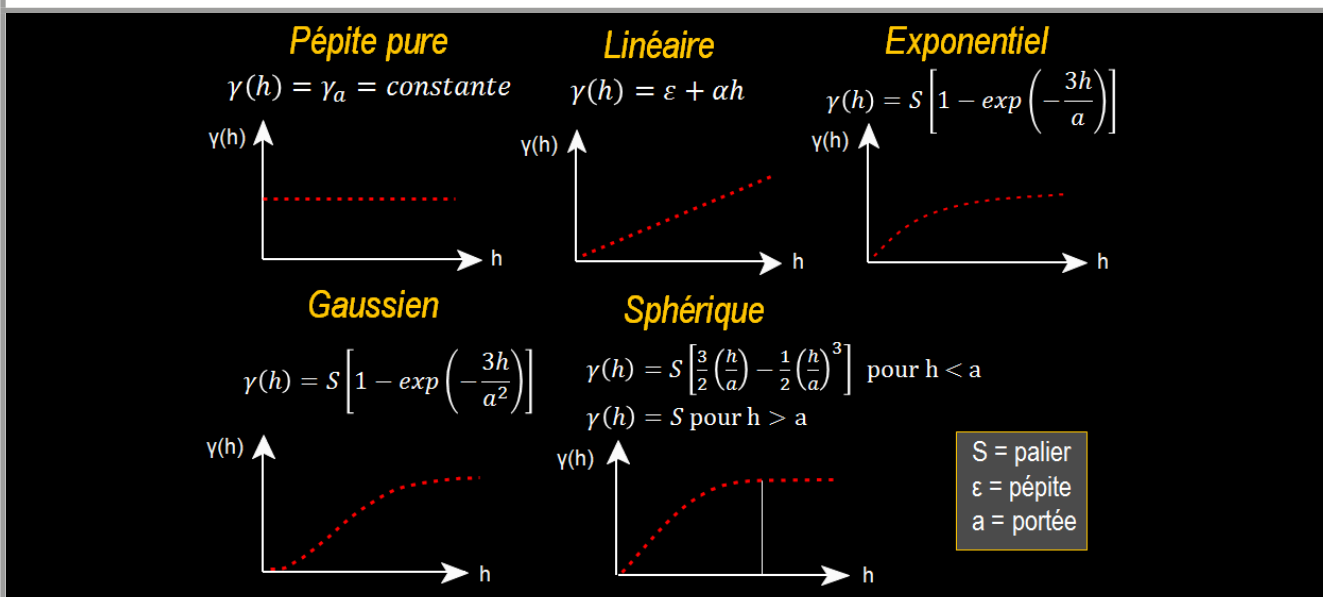


FIGURE 6

13:30

19:00

**KRIGEAGE**

L'objectif du krigeage est de trouver la meilleure estimation possible d'une variable régionalisée  $Z(s)$  à partir des valeurs des échantillons disponibles. En théorie, un estimateur est bon si la moyenne des écarts entre les valeurs estimées et mesurées est nulle (sans biais,  $E[\hat{Z}(S_0) - Z(S_0)] = 0$ ) et que la variance des écarts entre les valeurs estimées et mesurées est minimale (optimal,  $E[(\hat{Z}(S_0) - Z(S_0))^2] = \min$ ).

Le krigeage ordinaire est utilisé dans le cas d'un domaine échantillonné dont la moyenne est constante et l'analyse du variogramme montre que les hypothèses de stationnarités sont respectées (l'espérance mathématique des résidus est nulle et la variance ne dépend que de  $h$ ). Dans ce cas, l'estimation d'une variable par krigeage ordinaire se base sur une pondération des valeurs mesurées. Il est un cas particulier de la moyenne mobile pondérée. Une valeur estimée est obtenue en résolvant l'équation suivante :

$$\hat{Z}(S_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(S_{\alpha})$$

où  $S_0$  est la position où a lieu l'interpolation,  $S_{\alpha}$  est la position d'une donnée mesurée et  $\lambda_{\alpha}$  est le poids calculé à partir du variogramme.

Le calcul des poids ( $\lambda_{\alpha}$ ) se base sur les propriétés du krigeage. La somme des poids doit être égale à 1. En effet, si :

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1$$

Alors,

$$E \left[ \sum_{\alpha=1}^n \lambda_{\alpha} Z(S_{\alpha}) - Z(S_0) \sum_{\alpha=1}^n \lambda_{\alpha} \right] = 0$$

$$\sum_{\alpha=1}^n \lambda_{\alpha} [Z(S_{\alpha}) - Z(S_0)] = 0$$

La condition d'un estimateur sans biais est donc respectée. La condition d'optimalité est aussi respectée en minimisant la variance estimée.

$$-\gamma(h) - \sum \sum \lambda_{\alpha} \lambda_{\beta} \gamma(h_{\alpha, \beta}) + 2 \sum_{i=1}^n \lambda_{\alpha} \gamma(h_{\alpha, \beta}) = \min$$

Le minimum de variance estimée est atteint lorsque la dérivée par rapport au poids est nulle. Pour résoudre cette équation, il est donc nécessaire de calculer sa dérivée. La solution finale est obtenue par une méthode se basant sur les paramètres de Lagrange, mais son développement n'est pas présenté dans le cadre de ce cours.

Le krigeage est une méthode robuste car ses paramètres sont produits par le variogramme et les conditions de biais nul et d'optimalité permettent de définir une pondération liée à la distance et donnent la possibilité de connaître l'incertitude liée aux points interpolés.

## 9. ISOVALEURS

Une *isoligne* ou *isoplèthe* est une ligne qui relie des points d'égale valeur d'un phénomène continu. Elle est le plus souvent stockée en mode image et permet de mettre en évidence sur une carte les zones d'égale valeur. Dans cette leçon, le concept d'isolignes et les principes de calcul en mode image seront présentés. Ces principes permettront de produire des lignes d'isovaleurs avec QGIS.

### EXEMPLE D'ISOLIGNES

Il existe de nombreux exemples d'isolignes en voici quelques-uns.

- Les **courbes de niveau** constituées des points du relief situé à la même altitude et intersectant le plan horizontal. Sur une carte, des courbes de niveau rapprochées indiquent une pente raide tandis qu'une pente douce présente des courbes plus espacées entre-elles. Il existe deux types de courbes de niveau sur une carte, la *courbe directrice* ou courbe maîtresse (figure 1.1) représentée par un trait épais et continu avec indication de l'altitude et la *courbe standard* (figure 1.2). L'**équidistance** est la distance verticale qui sépare deux courbes de niveau. Les courbes de niveau permettent aussi de visualiser la topographie à partir d'un modèle à deux dimensions.
- Les courbes d'**isobares** reliant les points de même pression atmosphérique. Ces courbes permettent de mettre en évidence les zones de dépression et d'anticyclone et leurs déplacements dans le temps sur les cartes météorologiques.
- Les courbes d'**isothermes** reliant les points de même température. La région arctique est définie par le secteur où la température moyenne pendant le mois le plus chaud est en dessous de 10 °C (figure 2).
- Les courbes d'**isohyètes** reliant les points de même précipitation utilisées sur les cartes hydrologiques et météorologiques.

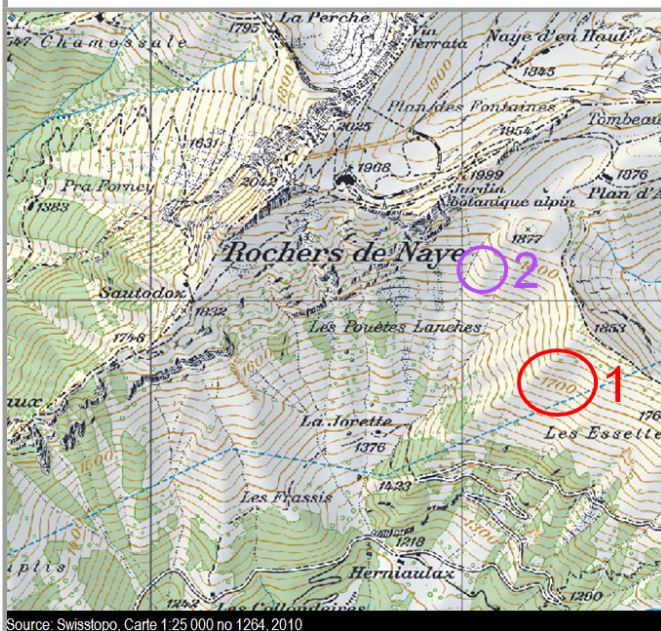


FIGURE 1

2:03

11:00

Exemple de courbes de niveau sur une carte au 1:25 000.  
1) Courbe directrice; 2) Courbes standards.



FIGURE 2

3:52

11:00

Région arctique définie par l'isotherme 10°C.

Un exemple d'isolignes obtenues par interpolation peut être vu dans la figure 3. La résolution spatiale des pixels est de 1100 m, l'interpolation par krigeage a utilisé un variogramme théorique gaussien et les dix plus proches voisins ont été pris en compte.

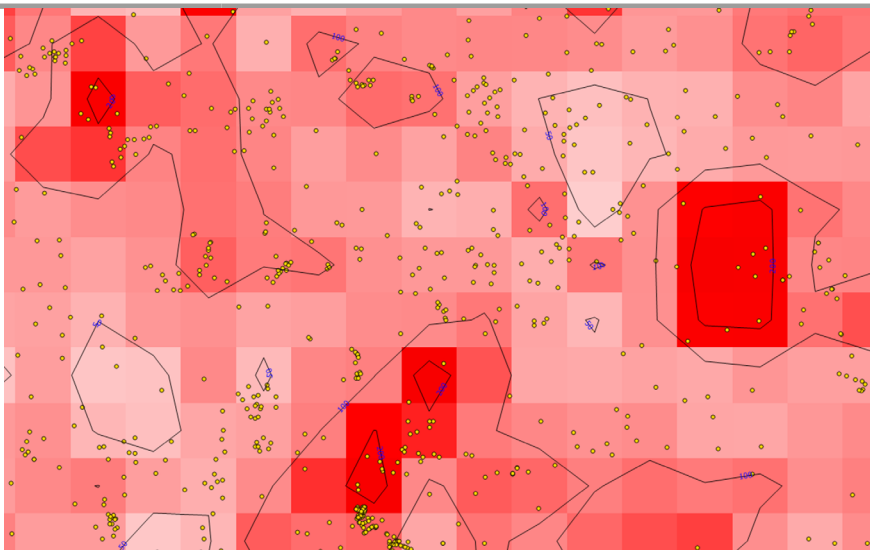


FIGURE 3

4:15

11:00

Géostatistique sur la concentration de Radon en sous-sol d'habitation.

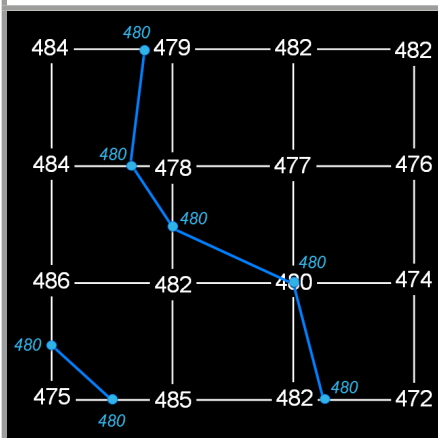


FIGURE 4

6:20

11:00

Création de courbes d'isovaleurs par interpolation linéaire à partir d'une grille de valeurs.

### ISOLIGNES PAR INTERPOLATION LINÉAIRE

Les courbes d'isovaleurs sont générées par interpolation linéaire à partir d'un modèle de grilles dont les valeurs ont été mesurées ou interpolées. Il est également possible d'utiliser un modèle triangulé.

L'interpolation linéaire est une méthode simple pour estimer une valeur prise par une fonction continue entre deux points connus. En utilisant la fonction affine  $f(x) = ax + b$  passant par deux points connus, il est possible de tracer des lignes d'isovaleurs. Dans l'exemple de la figure 4, une courbe d'isovaleur pour l'altitude 480 est créée en interpolant linéairement le long des segments de la grille. L'emplacement de la valeur 480 est défini en résolvant l'équation affine passant par les nœuds. Ces emplacements sont ensuite reliés entre eux pour former la courbe d'isovaleurs. Une résolution supérieure de la grille améliore la précision de la courbe.

### EXEMPLE SUR QGIS

À partir du modèle numérique de terrain au format raster (résolution spatiale de 50cm) de la région des Rochers de Naye en Suisse, il est possible d'extraire des courbes de niveau.

Pour extraire des courbes de niveau espacées de 100m ou 20m, cliquer sur le menu **Raster >> Extraction >> Création de contours**. Dans la fenêtre qui s'ouvre, choisir le fichier source correspondant au modèle numérique de terrain. Sous **fichier de sortie**, spécifier le nom et l'emplacement d'enregistrement du fichier de résultats. Sous **intervalle entre les lignes de contours**, spécifier l'espacement (ici 100m ou 20m). Stocker les isovaleurs en donnant un nom d'attribut. Pour finir, cliquer sur OK.

À la fin de cette démarche, les courbes d'isovaleurs sont visibles sur l'écran. Les propriétés de la couche d'isovaleurs permettent de définir une couleur aux courbes et de les étiqueter avec les valeurs d'altitudes.



## 10. ÉTUDE DE CAS – POLLUTION DE LA LAGUNE EBRIÉ À ABIDJAN

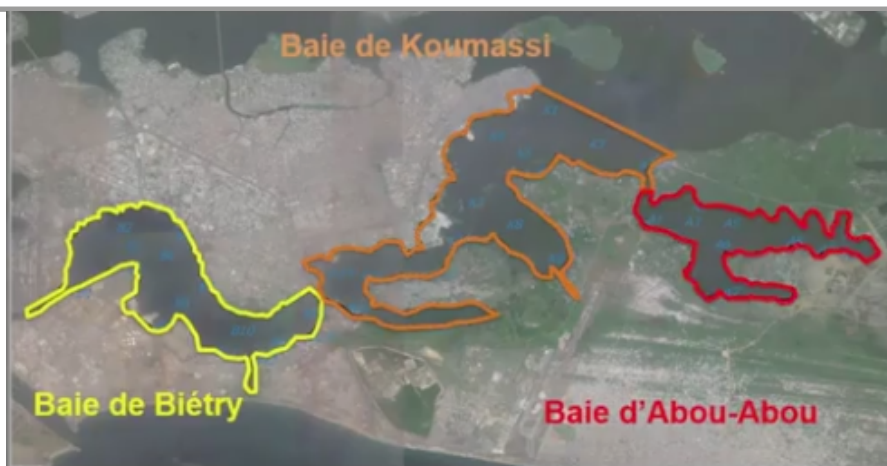
Cette étude de cas permet d'illustrer à la fois l'utilisation des SIG dans l'élaboration de cartes de pollutions des eaux par des métaux lourds et des hydrocarbures, mais aussi l'utilisation de ces cartes comme outil d'aide à la décision.

### PROBLÉMATIQUE

La lagune Ebrié subit depuis plusieurs années une dégradation accrue de son milieu aquatique. Surnommée autrefois la perle des lagunes, elle est aujourd'hui soumise à une forte pollution physique, chimique, organique et microbiologique. La charge polluante majeure est celle des eaux résiduelles des déchets domestiques. Bien que des mesures aient déjà été prises, la situation exige toutefois une intervention concrète afin d'inverser la tendance et restaurer ce plan d'eau.

La pollution chimique qui est au cœur de cette étude est principalement due aux rejets de polluants industriels riches en métaux lourds (cadmiums, plombs, mercures, manganèse et cuivre) et en hydrocarbures. Les déversements d'eaux usées non traitées, les pesticides et les polluants agricoles aggravent aussi la situation. Cette pollution affecte et menace les domaines liés à l'eau (biodiversité, santé, chaîne alimentaire, tourisme, pêche, etc.).

Les SIG sont de précieux outils capables de combiner une base de données d'analyse chimique avec des données cartographiques pour déterminer plus précisément les sites vulnérables et sensibiliser les décideurs.



### LA LAGUNE EBRIÉ

Le site d'étude de la lagune Ebrié traverse Abidjan d'Ouest en Est. La lagune Ebrié fait face à une forte urbanisation depuis les années 1960 et toutes ses rives font l'objet d'un usage particulier tant agricole qu'urbain ou industriel. L'étude se concentre sur trois baies spécifiques qui font toutes face à une forte pollution (figure 1). La baie de Biétry entourée d'industries lourdes (raffineries, savonneries, brasseries, etc.). La baie de Koumassi entourée de petites industries telles que des entreprises de fabrication de PVC ou de transformation du bois. La baie d'Abou Abou, en région rurale, entourée de mangrove et de plantations de cocotiers.

FIGURE 1

6:12

27:00



## MÉTHODOLOGIE

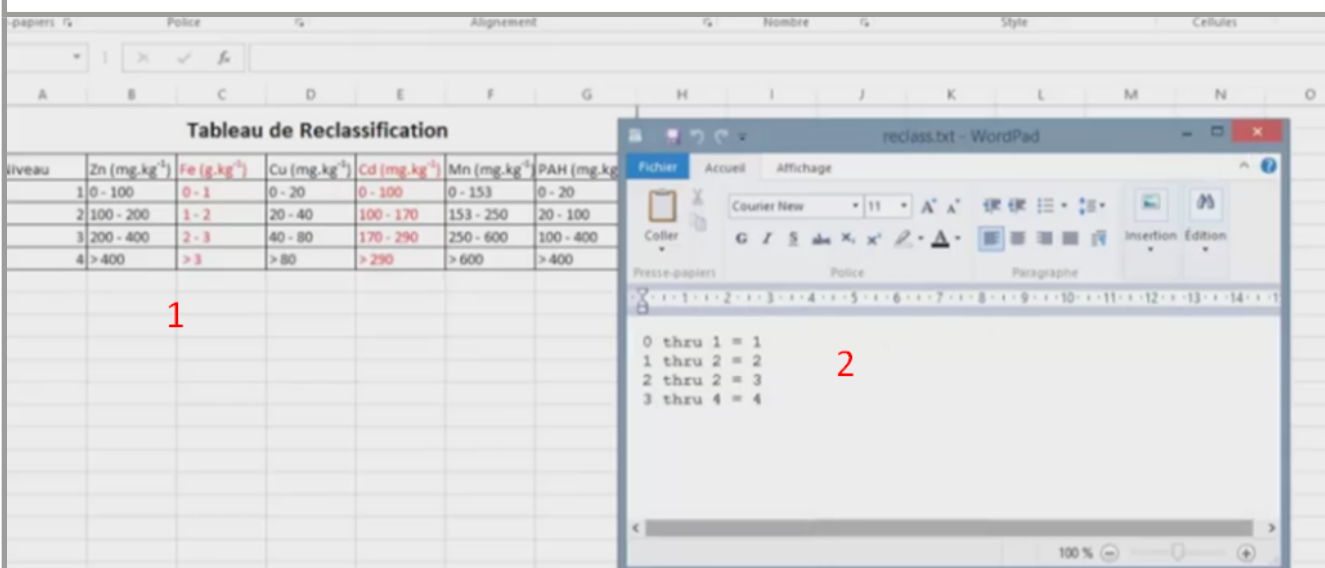
Les trois étapes importantes pour l'étude et la création de cartes appropriées sont l'acquisition des données, l'élaboration des cartes de pollution et l'aide à la décision.

L'échantillonnage des sédiments subsurfaciques (0-2 cm) a été entrepris à proximité des sites de rejets domestiques et d'activités industrielles dans les trois baies. Les échantillons ont ensuite été analysés pour mesurer leurs concentrations en zinc, fer, cuivre, cadmium et manganèse. La teneur en hydrocarbure a aussi été analysée. Chaque point d'échantillons possède une paire de coordonnées X, Y connue.

À partir de ces informations, il est possible d'établir des cartes de concentrations des différents polluants. La première étape consiste à géoréférencer la carte de la lagune Ebrié à l'aide de l'outil de géoréférencement de QGIS (système de projection WGS 84 zone 30 N) puis à digitaliser la zone d'étude. Désormais la zone digitalisée et les points d'échantillonnages et leurs concentrations en polluants sont disponibles et se superposent correctement. Il est maintenant possible de procéder à une interpolation entre les points de mesures. Cette étape est nécessaire pour permettre la visualisation de la tendance spatiale géochimique des polluants sur la zone d'étude. Une méthode de pondération par l'inverse de la distance (IDW) basée sur les trois plus proches voisins a été utilisée pour cette étude. L'outil d'interpolation est disponible dans QGIS sous l'onglet **Raster >> Interpolation**. Il suffit ensuite de sélectionner l'attribut à interpoler provenant de la couche des points d'échantillonnages, de préciser l'emprise sur laquelle aura lieu l'interpolation, le nom et l'emplacement d'enregistrement du fichier de sortie ainsi que la taille de l'échantillonnage (50 m ici)

Après l'interpolation des concentrations de polluants, il est nécessaire de procéder à une reclassification des valeurs extrapolées. Dans cette étude, 4 niveaux (classes) ont été définis. Les niveaux 1, 2, 3 et 4 expriment respectivement des teneurs très faibles, faibles, fortes et très fortes en métaux lourds. L'étape de la reclassification permet de créer des cartes plus facilement comparables entre elles. Pour procéder à cette étape, il est nécessaire d'exporter le raster produit par l'interpolation en format GRID à l'aide de l'extension GRASS dans QGIS. Ces fichiers GRID peuvent ensuite être combinés avec un fichier de règles de décisions pour réaliser une reclassification. Le fichier de règles de décisions est un fichier texte dans lequel apparaissent, sous forme codifiée, les règles de décisions (voir figure 2).

En appliquant un style à quatre intervalles égaux pour les nouvelles cartes après reclassification, les quatre classes sont visualisables et il est ainsi possible de comparer les teneurs en polluants à la surface des trois baies. Pour ne garder que la reclassification à l'intérieure de la lagune Ebrié et non sur le continent, il est indispensable de premièrement exporter les cartes provenant de la reclassification en mode image pour ensuite les découper selon la couche digitalisée de la zone d'étude (**Raster >> Extraction >> Découper**).



The image shows a screenshot of a spreadsheet titled "Tableau de Reclassification" and a text editor window titled "reclass.txt - WordPad".

**Tableau de Reclassification**

Niveau	Zn (mg.kg <sup>-1</sup> )	Fe (g.kg <sup>-1</sup> )	Cu (mg.kg <sup>-1</sup> )	Cd (mg.kg <sup>-1</sup> )	Mn (mg.kg <sup>-1</sup> )	PAH (mg.kg <sup>-1</sup> )
1	0 - 100	0 - 1	0 - 20	0 - 100	0 - 153	0 - 20
2	100 - 200	1 - 2	20 - 40	100 - 170	153 - 250	20 - 100
3	200 - 400	2 - 3	40 - 80	170 - 290	250 - 600	100 - 400
4	> 400	> 3	> 80	> 290	> 600	> 400

The text editor shows the following code transcription:

```

0 thru 1 = 1
1 thru 2 = 2
2 thru 2 = 3
3 thru 4 = 4

```

FIGURE 2

### RÉSULTATS ET AIDE À LA DÉCISION

L'objectif de l'élaboration de ces cartes est de permettre une meilleure prise de décision. Dans le cadre de cette étude, ces cartes permettent de recenser les zones qualifiées de très polluées, moyennement polluées ou peu polluées. Ces zones peuvent soit être observées séparément pour chacun des polluants ou alors au travers d'une combinaison des cartes Raster par addition des indices de pollution. Séparément, les cartes montrent que la Baie de Biétry présente une très forte teneur en fer et cuivre, une forte teneur en manganèse et une teneur moyenne en cadmium. La Baie de Koumassi présente une très forte teneur en fer, une forte teneur en cadmium, une teneur moyenne en cuivre et une faible teneur en manganèse. Pour finir, la Baie d'Abou Abou présente une très faible teneur en fer, cuivre et cadmium, mais une forte teneur en manganèse.

Les hydrocarbures sont présents en forte concentration dans la Baie de Biétry tandis que dans les baies de Koumassi et d'Abou Abou les teneurs sont faibles. La carte obtenue par addition des autres cartes individuelles montre que la Baie de Biétry et la Baie de Koumassi font face à une pollution élevée et la Baie d'Abou Abou présente une pollution faible (figure 3).

En conclusion, les outils SIG ont permis de mettre en évidence les zones de la lagune Ebréi fortement polluées en métaux lourds et hydrocarbures. Les cartes obtenues sont de précieux documents capables d'aider les autorités ivoiriennes dans leurs prises de décisions en rapport avec la réduction de la pollution, la restauration des écosystèmes et la sensibilisation de la population.

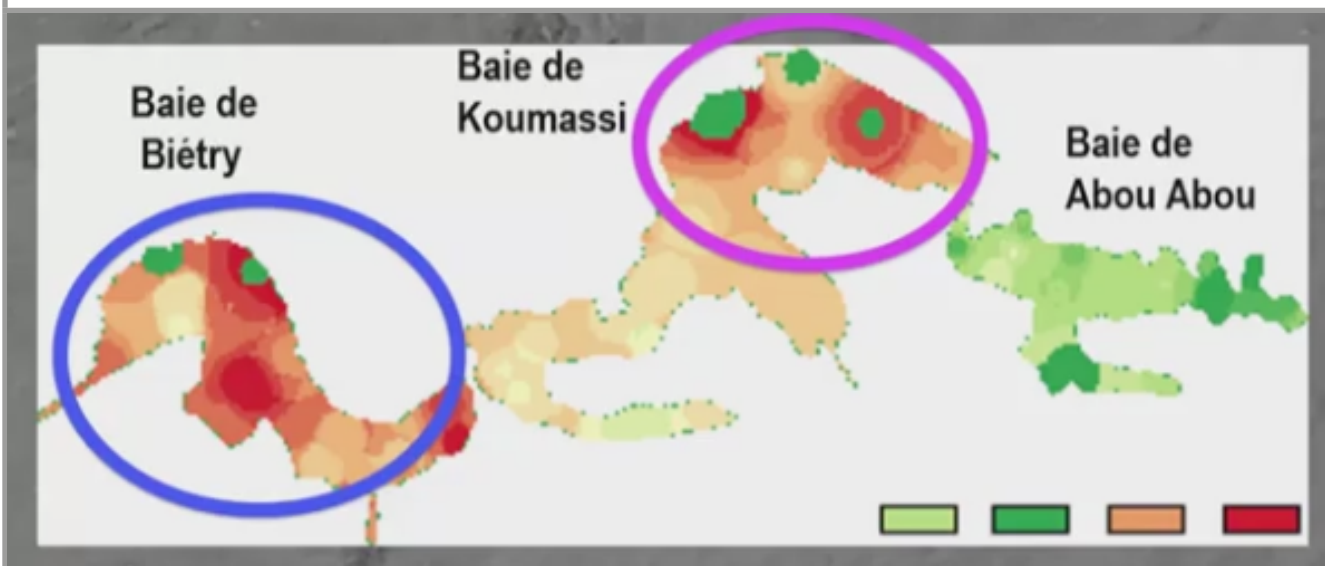


FIGURE 3

24:46

27:00

Carte des niveaux de pollution de la lagune Ebréi. Les zones rouges indiquent une forte pollution. Les zones vertes, une plus faible pollution.

## 11. MODÈLES NUMÉRIQUES D'ALTITUDE – ÉLÉMENTS DE BASE

Par définition, le **relief** représente la forme et les inégalités de la surface terrestre issues des mouvements géologiques et climatologiques survenus au cours de l'histoire terrestre. La mesure primaire utilisée pour caractériser le relief est l'altitude sur mer et un **modèle numérique d'altitude (MNA)** est un fichier qui contient des mesures d'altitude aux nœuds d'une grille régulière. Dans cette leçon, les principaux modes d'acquisition et de représentation de l'altitude seront présentés. Il sera ainsi possible d'expliquer le fonctionnement de modes d'acquisition de l'altitude ainsi que de maîtriser les méthodes de représentation du relief.

### LE RELIEF

Le relief est un paramètre important dans l'analyse du paysage ou pour l'exploitation des ressources liées au sol. Il constitue donc un élément important en analyse spatiale.

L'**altitude** est la mesure primaire du relief. Elle est définie par la distance verticale qui sépare un point du relief de la surface du géoïde terrestre. Elle est donc une variable continue. Le **géoïde** est la surface en gravité équipotentielle ajustée sur un niveau moyen de la surface de la mer. Bien que l'altitude soit une variable continue, il n'est pas réalisable de mesurer de manière exhaustive l'altitude sur toute la surface de la Terre. Il est donc nécessaire de créer un modèle dans lequel les altitudes sont mesurées de manière discontinue selon un mode d'échantillonnage régulier ou non. Ce modèle est appelé *modèle numérique d'altitude (MNA)* ou *modèle numérique de terrain (MNT)*.

### MODES D'ACQUISITION DE L'ALTITUDE

Les propriétés d'un MNA sont déterminées par son mode d'acquisition et par sa précision. Il existe trois principaux modes d'acquisition :

- le nivellement ;
- la photogrammétrie ;
- l'altimétrie par laser à balayage.

Le **nivellement** géométrique (nivellement par visée horizontale ou nivellement direct) est une technique qui permet de mesurer une différence d'altitude entre un *repère de nivellement* (point d'altitude connu) et un point d'altitude inconnu. Pour faire du nivellement, un appareil permettant de matérialiser une ligne de visée horizontale est nécessaire. Cet appareil, appelé niveau, est placé sur un trépied à mi-distance de deux mires tenues verticalement sur les deux points. La lecture successive, dans la lunette du niveau, de la hauteur verticale de chacun des points permet de calculer la dénivelée (différence entre les deux valeurs lues). L'ajout de la dénivelée à l'altitude du point connu fournit l'altitude du point inconnu. Pour minimiser les erreurs, la longueur d'une nivelée ne dépasse pas 80 mètres.

La **photogrammétrie** est née avec les débuts de l'imagerie aérienne vers la fin du XIX<sup>e</sup> siècle. Le principe de la photogrammétrie exploite le déplacement apparent d'un objet vu, sur des images, sous deux angles différents. Ce principe est géométriquement semblable au système visuel humain. Dans l'image formée par le cerveau, les différences de relations entre les objets se traduisent en une perception de profondeur. En photogrammétrie, deux images d'un même lieu prises avec deux angles de vue différents ont le même rôle que les deux rétines. Ces deux images forment un couple stéréoscopique par lequel il est possible de calculer la position en X, Y et Z des objets au sol. Dans l'exemple de la *figure 1*, deux objets ( $O_1$  et  $O_2$ ) dans l'espace réel sont perçus sur la même ligne de visée. Toutefois, la position relative des images (1 et 2) est connue et permet de déduire des relations angulaires. En connaissant les relations angulaires des triangles de sommet  $f_2$  et la distance  $d'$  sur l'image 2, il est possible de déduire la distance  $d$ . Des points de repère au sol facilement identifiables sur les images sont indispensables pour assurer la restitution photogramétrique. Cette mesure implique une connaissance parfaite de la position de prise de vue des images.

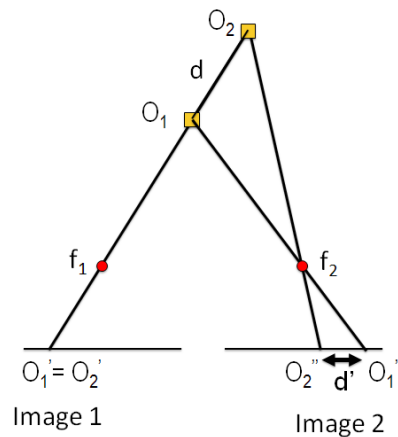


FIGURE 1

6:06

14:00

Relation angulaire du principe de photogrammétrie.  $f_1$  et  $f_2$  représentent les foyers photographiques des images 1 et 2. Les points  $O_1'$ ,  $O_2'$  et  $O_1''$ ,  $O_2''$  dans l'espace image sont les points homologues des objets  $O_1$  et  $O_2$ .

L'**altimétrie par laser à balayage** est un système électronique nommé **LiDAR** (*Light Detection And Ranging*) qui mesure le temps aller-retour de l'intensité d'un faisceau laser émis depuis une plateforme aéroportée. Pour calculer la position des éléments de la surface, il est nécessaire de connaître précisément l'altitude et la position de la plateforme aéroportée et le temps d'aller-retour des impulsions laser. La précision est obtenue par un système de GPS différentiel. Suivant la nature de la couverture du sol, plusieurs échos du signal émis peuvent être mesurés. Un système LiDAR est capable d'émettre plusieurs centaines de milliers d'impulsions sur lesquelles il est possible de mesurer plusieurs échos. Les points mesurés forment donc un nuage de points de haute résolution. Ces différents échos et la nature de la longueur d'onde du faisceau (proche infrarouge) permettent de distinguer le sol du couvert végétal. La partie supérieure des réflexions est nommée **Modèle Numérique de Surface (MNS)** et la partie inférieure conserve le nom de *modèle numérique de terrain* (figure 2).

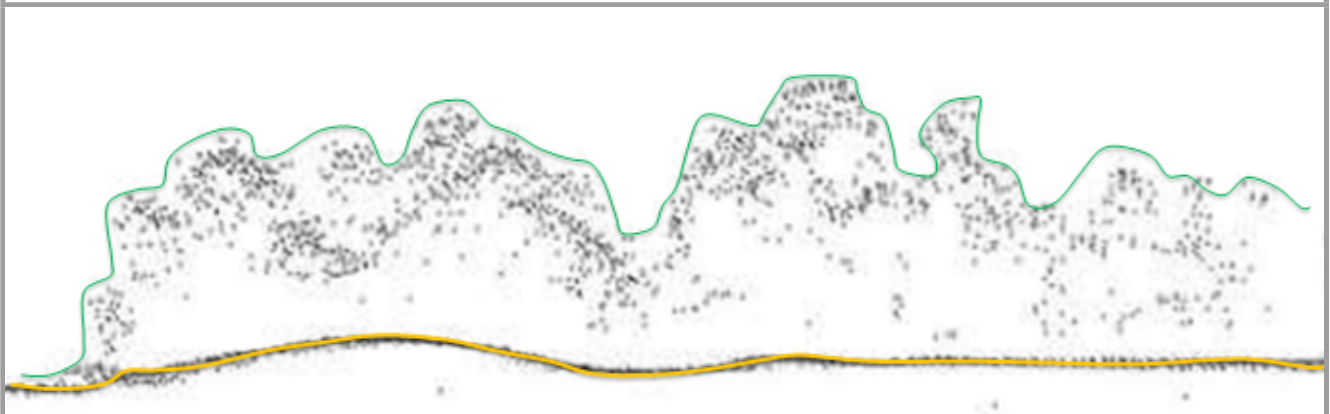


FIGURE 2

8:20

14:00

Profil vertical d'échos LiDAR. La partie supérieure (vert) représente le MNS et la partie inférieure (Orange) représente le MNT.



## REPRÉSENTATION DE L'ALTITUDE

Les techniques d'acquisition de l'altitude présentées ci-dessus permettent de créer un modèle numérique de terrain. Les modèles ont beaucoup évolué au cours de l'histoire. Le relief a d'abord été représenté par des hachures sur les anciennes cartes puis au travers d'ombrages et de courbes de niveaux pour les cartes actuelles. L'ombrage représente qualitativement l'effet d'un éclairage fictif dans la direction du sud-est du relief et l'ombre projetée associée. Bien que la représentation des ombrages et des courbes de niveau sur les cartes soient appréciée, le modèle associé n'est pas satisfaisant car le calcul de variables dérivées (pente, orientation, etc.) n'est pas évident.

L'altitude peut aussi être modélisée sous forme de semis de points irréguliers en mesurant ponctuellement et aléatoirement (ou le long des lignes de structures du relief) les valeurs d'altitude. Ce modèle peut être représenté sous forme de TIN (*triangulated irregular network*) lui donnant l'avantage de représenter des points spécifiques du relief. La précision de ce modèle dépend des appareils utilisés pour l'acquisition des points de mesures et de la densité de points mesurés.

L'altitude peut également être représentée par une grille régulière de points dont la résolution du modèle est définie par le côté des mailles. Les coordonnées de chaque point sont soit placées sur les nœuds de la grille ou au centre de la maille. Dans la plupart des cas, les points sont le produit d'une interpolation d'un semis de points irréguliers. La précision de ce type de modèle dépend donc d'une part de la procédure d'acquisition et de la procédure d'interpolation, mais aussi de la densité de points mesurés par cellule dans le cas des techniques LiDAR.

La technologie LiDAR permet de produire un modèle numérique d'altitude composé d'un modèle numérique de terrain et d'un modèle numérique de surface. La soustraction de ces deux modèles (MNS-MNT) permet l'obtention d'un *Modèle Numérique de Hauteur (MNH)*.

## 12. MODÈLES NUMÉRIQUES D'ALTITUDE – VARIABLES DÉRIVÉES I

Les modèles d'altitude permettent de produire des variables globales et locales de géomorphométrie caractérisant le relief. Dans cette leçon, les principaux indicateurs globaux et locaux du relief seront présentés et vous serez en mesure de les calculer.

### GÉOMORPHOMÉTRIE

La **géomorphométrie** est une discipline qui a pour objectif la détermination de variables capables de caractériser les formes du relief. Elle est d'approche globale lorsque les indicateurs statistiques caractérisant le relief sont calculés sur la zone d'étude dans son ensemble. Elle est d'approche locale lorsque ces indicateurs sont calculés en différents points du territoire. L'approche globale permet rapidement la comparaison entre les reliefs de différentes régions.

### INDICATEURS GLOBAUX

Il existe plusieurs indicateurs géomorphométriques globaux. Une partie de ceux-ci sont présentés ci-dessous.

- L'altitude maximale et minimale permet de résumer les propriétés d'un modèle numérique de terrain.
- La courbe hypsométrique exprime la répartition de la surface d'un bassin-versant en fonction de l'altitude entre les valeurs minimales et maximales. Elle permet d'estimer le comportement hydrologique et hydraulique d'un bassin-versant et de son système de drainage.
- Les classes moyennes d'altitude, de pente et les distances entre les sommets de même classe permettent de caractériser globalement la rugosité du relief.

### INDICATEURS LOCAUX

L'approche locale de la géomorphométrie est semblable à une procédure de filtrage du modèle numérique de terrain par une fenêtre mobile dont le résultat est une variable capable de décrire une propriété locale. La fenêtre mobile est centrée sur le pixel d'intérêt dont la nouvelle valeur est calculée à partir de son voisinage. L'approche locale dépend donc de la taille et de la forme de la fenêtre mobile. L'*effet de bord* décrit le problème rencontré par la fenêtre mobile dans les bords de la zone d'étude (figure 1). En effet, si l'information en dehors de la zone d'étude est manquante, il n'est pas possible de calculer de manière correcte la valeur des pixels en bordure interne de la zone étudiée. Ce problème est facilement résolu en prévoyant un modèle numérique d'altitude plus grand que la zone d'étude. La taille de la fenêtre mobile introduit également un effet contextuel.

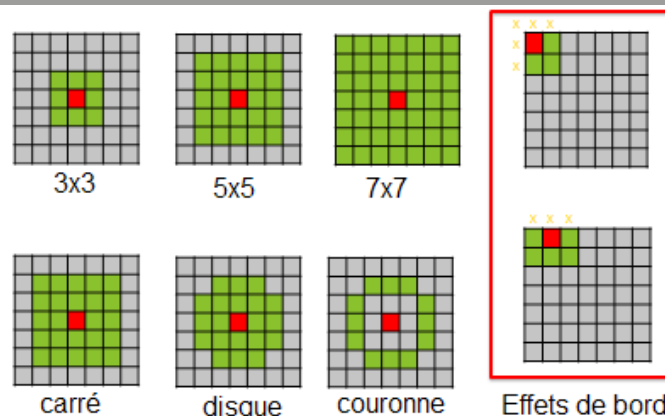


FIGURE 1



### PENTE, ORIENTATION ET COURBURE

La *pente* est, par définition, l'inclinaison d'une surface par rapport au plan horizontal. L'*orientation* est une direction déterminée par rapport aux points cardinaux. Mathématiquement, la pente et l'orientation sont parfaitement définies en un point lorsque le gradient de la surface peut être calculé. Pour une surface continue non discrète, la pente est définie par le gradient.

$$Pente = grad(Z) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}, \text{ où } Z = f(x, y)$$

En mode discret (mode image), la pente et l'orientation sont définies pour chaque pixel et son calcul se base sur le rapport entre l'altitude et la distance horizontale selon le voisinage pris en compte. Voici plusieurs algorithmes possibles pour le calcul de la pente et de l'orientation.

L'algorithme de la *différence maximale aux voisins* permet de définir la pente par  $\arctan\left(\max\left(\frac{\Delta z}{d}\right)\right)$ . Un exemple d'application pour une fenêtre 3 x 3 est présenté dans la figure 2.

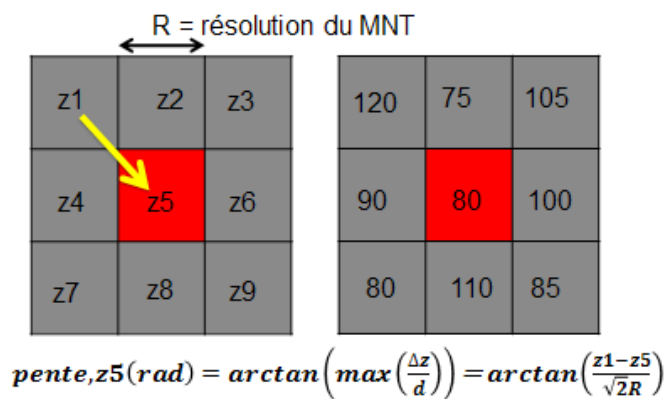


FIGURE 2

5:50

13:00

Exemple d'application de la différence maximale aux voisins (fenêtre 3 x 3).

L'algorithme du *Gradient ligne-colonne* consiste à identifier la différence maximale entre le pixel d'intérêt et un autre pixel situé sur la même ligne et la même colonne. Ces différences permettent de calculer les valeurs  $H$  et  $G$ . La pente est ainsi définie par  $\arctan(\sqrt{G^2 + H^2})$  et l'orientation par  $\pi - \arctan\left(\frac{H}{G}\right) + \frac{\pi}{2} \left(\frac{G}{|G|}\right)$ . Un exemple d'application pour une fenêtre 3 x 3 est présenté dans la figure 3.

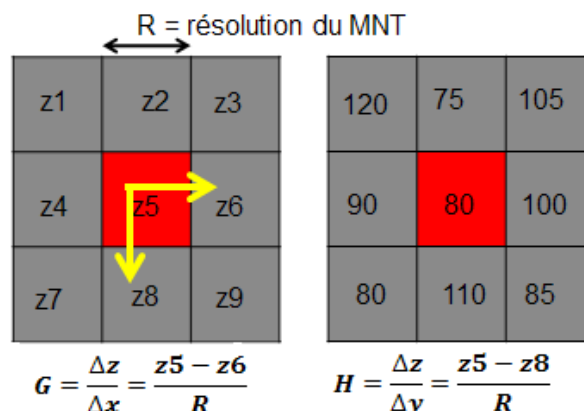


FIGURE 3

6:38

13:00

Exemple d'application du Gradient ligne-colonne (fenêtre 3 x 3).

L'algorithme le plus utilisé est celui de Zevenbergen et Thorne (1987). Le gradient de ligne ( $G$ ) et le gradient de colonne ( $H$ ) pour lesquels la distance horizontale est égale à deux fois la résolution spatiale (voir figure 4). La pente est ensuite définie par  $\arctan(\sqrt{G^2 + H^2})$  et l'orientation par  $\pi - \arctan\left(\frac{H}{G}\right) + \frac{\pi}{2} \left(\frac{G}{|G|}\right)$ . Selon Zevenbergen et Thorne, les paramètres  $H$  et  $G$  font partie d'une équation à 9 paramètres ( $A-I$ ) décrivant la surface qui passe par les altitudes de la fenêtre mobile.

$$Z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Fxy + Gx + Hy + I$$

Il est possible de résoudre cette équation en passant par des polynômes de Lagrange. Les solutions des 9 paramètres sont décrites ici.

$$A = \frac{\left[ \frac{Z_1 + Z_3 + Z_7 + Z_9}{4} - \frac{Z_2 + Z_4 + Z_6 + Z_8}{2} + Z_5 \right]}{R^4}$$

$$B = \frac{\left[ \frac{Z_1 + Z_3 - Z_7 - Z_9}{4} - \frac{Z_2 - Z_8}{2} \right]}{R^3}$$

$$C = \frac{\left[ \frac{-Z_1 + Z_3 - Z_7 + Z_9}{4} + \frac{Z_4 - Z_6}{2} \right]}{R^3}$$

$$D = \frac{\left[ \frac{Z_4 + Z_6}{2} - Z_5 \right]}{R^2}$$

$$E = \frac{\left[ \frac{Z_2 + Z_8}{2} - Z_5 \right]}{R^2}$$

$$F = \frac{-Z_1 + Z_3 + Z_7 - Z_9}{4R^2}$$

$$G = \frac{-z_4 + z_6}{2R}$$

$$H = \frac{z_2 - z_8}{2R}$$

$$I = z_5$$

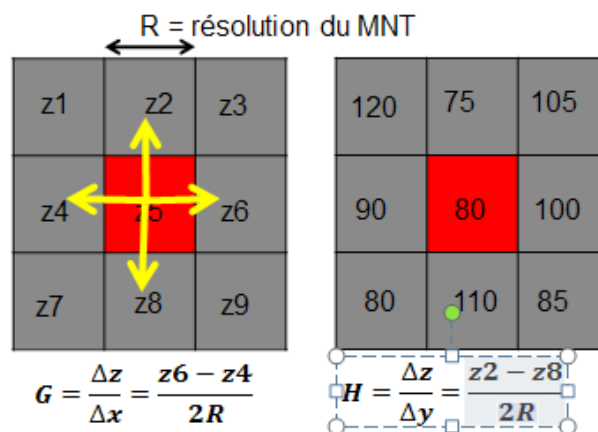


FIGURE 4

L'algorithme de Horn se base sur les 8 plus proches voisins. Les voisins les plus proches parmi les 8 ont un poids de deux (voir figure 5) dans le calcul de  $H$  et  $G$ . La pente est ensuite définie par  $\arctan(\sqrt{G^2 + H^2})$  et l'orientation par  $\pi - \arctan\left(\frac{H}{G}\right) + \frac{\pi}{2} \left(\frac{G}{|G|}\right)$ .

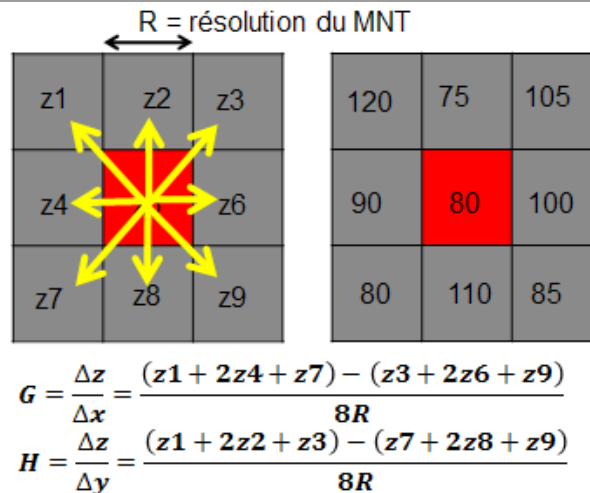


FIGURE 5

9:10

13:00

Exemple d'application de l'algorithme de Horn (fenêtre  $3 \times 3$ ).

Une surface est dite convexe ou concave lorsqu'elle est entièrement située du même côté d'un plan tangent. La courbure est positive lorsque la surface est convexe et négative lorsque celle-ci est concave.

La courbure permet de caractériser les phénomènes d'érosion. Une pente convexe subit une érosion plus forte qu'une pente concave. Lors de glissement de terrain, la convexité et la concavité permettent de décrire le relief du glissement dans sa configuration horizontale et verticale à l'aide d'indicateurs de courbure verticale et horizontale.

Le modèle de Zevenbergen et Thorne est également utilisé pour mesurer la courbure. Les paramètres  $G$ ,  $H$ ,  $D$ ,  $E$  et  $F$  décrits plus haut sont utilisés pour calculer la courbure verticale (en direction de la pente) et horizontale (direction transversale à la pente) (voir figure 6).

La taille de la fenêtre mobile influence la cartographie de la pente, de l'orientation et de la courbure. Une fenêtre plus grande produit un lissage de la pente et de l'orientation et a un effet structurant sur la concavité et la convexité.

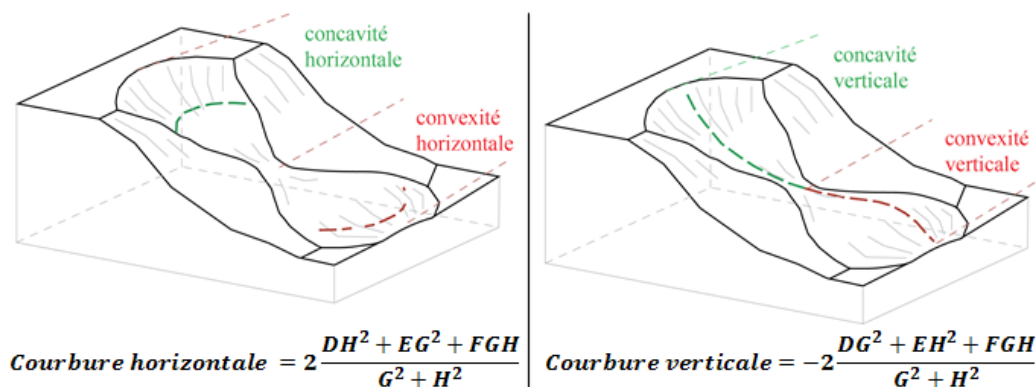


FIGURE 6

11:03

13:00

Courbures horizontales et verticales.

## 13. MODÈLES NUMÉRIQUES D'ALTITUDE – VARIABLES DÉRIVÉES II

En plus des variables de base présentées à la leçon précédente, les MNA peuvent produire d'autres variables très utilisées par les spécialistes. Dans cette leçon, ces nouvelles variables thématiques, leurs applications-types et leur mise en œuvre sont présentées.

### LIGNES DE VUE

Une ligne de vue est une ligne droite reliant deux points et constituée de tous les pixels situés sur son tracé. Le principe de ligne de vue est utilisé dans le cadre de nombreuses techniques.

L'**analyse de visibilité** est une technique qui permet de déterminer les zones visibles depuis un point de vue donné. Cette technique se base sur l'analyse de toutes les lignes de vue par rapport au relief. L'analyse de visibilité est souvent utilisée dans l'analyse du paysage et permet de définir, par exemple, les zones dans lesquelles une nouvelle construction (bâtiments, éoliennes, etc.) sera visible.

L'**ombre portée** est une technique qui permet de griser les zones d'orientation opposées à celle de la source lumineuse et présentant une inclinaison supérieure à celle du Soleil selon l'intensité de la pente. L'**ombre projetée** est l'ombre réelle produite par le relief faisant écran aux rayons du soleil. Ces deux techniques se basent sur l'analyse des lignes de vue provenant de la position de la source lumineuse.

Le taux d'ensoleillement est calculé en utilisant la technique de l'ombre projetée. L'ensoleillement reçu par chaque maille du MNA dépend de la hauteur du Soleil, de la transparence de l'atmosphère, de la pente, de l'orientation et de la ligne d'horizon. En se basant sur la définition d'une grille d'horizon de 360 degrés pour chaque pixel, il est possible de calculer l'énergie potentielle reçue du soleil pour une période de temps donnée. Cette technique permet, par exemple, de créer des cartes de l'énergie solaire potentielle cumulée pendant une journée.

### HYDROLOGIE

Une **ligne de drainage** est le chemin théorique qu'une goutte d'eau tombant sur le relief est censée parcourir jusqu'à l'exutoire du bassin-versant. En supposant une surface lisse et imperméable, les lignes de drainage sont obtenues en calculant la pente en chaque point puis à partir de chaque point haut (ligne de crête) les lignes d'écoulements sont définies de proche en proche en suivant la ligne de plus forte pente. La **densité de drainage** est déduite de l'ensemble des lignes de drainage. Elle est la longueur totale des cours d'eau d'un bassin-versant.

Il existe plusieurs méthodes connues pour la détection automatique des bassins-versants à l'aide d'un MNA. Le bassin-versant peut être modélisé en déterminant les lignes de crêtes à partir d'un exutoire en suivant de proche en proche les points d'altitude les plus élevés. Il est également possible d'agrèger les points voisins les plus élevés à partir de l'exutoire. Bien que ces méthodes soient simples, elles ne sont pas idéales et présentent des erreurs importantes dans le cas de minima locaux ou de cols en forme de selle de cheval. Vincent et Soille, en 1991, ont proposé une méthode permettant de s'affranchir des problèmes expliqués plus haut. La **figure 1** permet d'expliquer cette méthode pour un MNA à une dimension et contenant 10 points d'altitude. Les points (pixels) sont numérotés de 1 à N selon leur ordre initial depuis un bord et un diagramme altitude-fréquence est établi en notant pour chaque point son numéro d'identification. L'algorithme de Vincent et Soille identifie le ou les pixels d'une même altitude en commençant par la plus basse. Si ces pixels ne sont pas adjacents, ils forment le premier point d'un bassin-versant. S'ils le sont alors ils appartiennent à un bassin-versant déjà existant. Lorsque deux bassins se rejoignent, ils forment une ligne de crête.

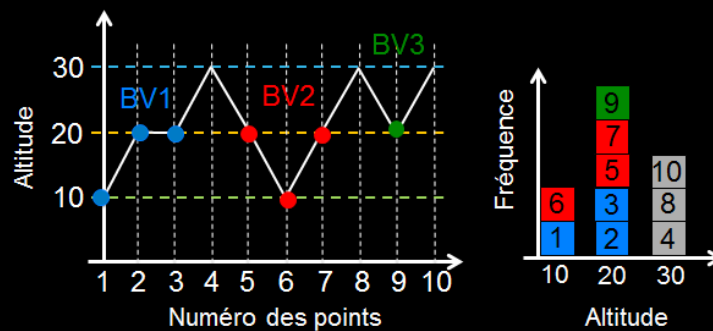


FIGURE 1

9:00

17:00

Exemple de la méthode de Vincent et Soille pour un MNA à une dimension et contenant 10 points d'altitude.

## DANGERS NATURELS

Les deux exemples qui suivent ont fait l'objet d'une publication scientifique.

Le premier exemple concerne l'identification de zones inondables à Akure au Nigeria. Cette ville compte environ 495 000 habitants et connaît une saison des pluies de 7 mois entre avril et octobre avec de fortes précipitations. La ville a été contrainte, en raison d'une forte demande en terrains à bâtir, de définir les zones inondables. Dans un premier temps, l'étude a utilisé des images satellites Landsat pour classifier et mettre en évidence les changements de l'utilisation des sols en zones urbaines entre 2002 et 2011 qui ont engendré une réduction du potentiel d'absorption des sols. Les auteurs ont également utilisé un modèle numérique de terrain SRTM (*Shuttle Radar Topography Mission*) à 90 mètres de résolution pour calculer la pente et identifier les lignes de drainage et ils ont intégré les données de précipitation moyenne annuelle de la zone d'étude. Toutes ces informations ont ensuite été pondérées en fonction de leur probabilité de contribuer à une inondation et combinées au moyen d'une approche multicritère pour créer une carte de risque.

Le deuxième exemple concerne l'élaboration d'une carte de risque d'éboulements rocheux dans le bassin-versant de Dades au Maroc. Différents critères ont été pris en compte pour l'élaboration de cette carte. Il s'agit de la pente, calculée à partir d'un modèle d'altitude SRTM à 90 mètres de résolution, des cartes géologiques, digitalisées afin de mettre en évidence les zones de couches géologiques sensibles aux mouvements et de l'information des fractures tectoniques, obtenue par des levés de terrain et au travers d'images satellites. Les caractéristiques de pentes, de sensibilité aux mouvements et de fracturation ont été combinées afin de produire des cartes de risques. La carte finale montre cinq classes d'aléas allant de très faible à très élevé. 25% de la surface du bassin-versant de Dades est soumise un risque élevé.

## SANTÉ PUBLIQUE

L'étendue de l'utilisation des modèles numériques d'altitudes est vaste. Voici un exemple de son utilisation dans le domaine de la santé publique. Cette étude sur la schistosomiase, une maladie liée à l'eau très répandue en Afrique subsaharienne, a été publiée dans *PLOS Neglected Tropical Diseases* en 2015. Cette maladie est véhiculée par des escargots d'eau douce, qui agissent comme hôtes intermédiaires. Le but de l'étude est de définir l'habitat potentiel du vecteur et d'identifier les zones à risques de transmission de la maladie. La persistance de l'eau stagnante à la surface de la Terre a été calculée à partir d'image satellite RapidEye à 6.5 m de résolution. La température de l'eau influençant la mortalité de l'escargot a été dérivée de la bande thermique des images satellites. La couverture végétale a été dérivée des images satellites à l'aide d'un indice de végétation. Pour finir, la profondeur de l'eau et la vitesse du courant ont été calculées à partir d'un modèle numérique d'altitude ASTER à 30 mètres de résolution. Un indice d'habitat potentiel a été établi en utilisant une fonction multicritère additive et cet indice a permis d'estimer le risque sur la zone d'étude. La carte finale peut être utilisée dans le cadre de mesures de prévention et de contrôle de la maladie.

Walz Y, Wegmann M, Dech S, Vounatsou P, Poda JN, et al. (2015) Modeling and Validation of Environmental Suitability for Schistosomiasis Transmission Using Remote Sensing. *PLOS Neglected Tropical Diseases* 9(11): e0004217.

<http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004217>

## 14. INTERACTIONS ENTRE COUCHES DE DONNÉES : RASTER – VECTEUR

L'interaction entre couches de données géoréférencées est liée au concept de superposition spatiale (*spatial overlay*). Ce concept permet le transfert de données entre différents types d'objets stockés sur des couches différentes selon leurs relations spatiales. Dans cette leçon, les types d'interactions entre couches de type raster et couches de types vecteurs seront présentés. À la fin de cette leçon, il sera possible de citer deux types d'interaction raster-vecteur et de les mettre en œuvre avec les outils de QGIS.

### INTERACTIONS ENTRE COUCHES DE DONNÉES

La superposition spatiale est la fonction qui fait la particularité et la force des SIG. En effet, l'analyse d'un phénomène dans l'espace géographique requiert l'utilisation simultanée d'informations provenant de couches distinctes. L'interaction rendue possible par le principe de superposition spatiale permet le transfert de données depuis une couche source vers une couche cible. Les contraintes à la mise en œuvre de l'interaction spatiale sont l'utilisation d'un système de projection de coordonnées identique et une couverture spatiale commune.

### INTERACTIONS ENTRE OBJETS VECTORIELS PONCTUELS ET UN RASTER

L'extraction d'une variable  $z$  stockée dans les pixels d'un raster et sa copie dans une série d'objets vectoriels ponctuels spatialement distribués et stockés dans une couche vectorielle superposée est un exemple de type d'interaction.

Voici un exemple pratique dans QGIS avec un fichier raster d'un modèle numérique de terrain et un fichier vectoriel contenant les points pour lesquelles l'altitude sera extraite. Avant tout travail de combinaison, il est nécessaire de vérifier que les deux couches respectent les conditions nécessaires à la mise en œuvre de l'interaction. C'est-à-dire que les deux couches sont définies dans un même système de projection et que leur couverture spatiale est commune. Pour cela, cliquer droit sur les couches puis cliquer sur **définir le SCR d'une couche** et choisir le système de projection approprié (ici CH1903/LV03). La couche vectorielle de point ne contient, ici, que l'identifiant des points. Il est donc nécessaire d'ajouter un attribut altitude. Pour cela installer et utiliser l'extension *Point Sampling Tool* dans QGIS (**Extension >> installer/gérer les extensions**). Pour accéder à l'extension installée, aller sous **Extension >> Analyse >> Point Sampling Tool**. Une fenêtre s'ouvre. Sous l'onglet **General**, définir la couche contenant les points, les couches de provenance des données (couche de points et couches MNT) ainsi que le nom et l'emplacement de la couche résultat. Sous l'onglet **Field** un aperçu des colonnes du fichier résultat est disponible. Il est aussi possible de renommer les colonnes. Pour finir, en cliquant sur OK le logiciel fait le transfert de données et le fichier résultats est un fichier de points vectoriel avec un nouvel attribut d'altitude provenant du fichier raster.

### INTERACTIONS ENTRE OBJETS VECTORIELS SURFACIQUES ET UN RASTER

L'extraction d'une statistique des valeurs contenues dans les pixels d'un raster et sa copie dans une série de polygones spatialement distribués et stocké dans une couche vectorielle superposée est un autre exemple de type d'interaction.

Voici un exemple pratique dans QGIS avec un fichier raster des précipitations au Kenya et un fichier vectoriel contenant les zones administratives du Kenya pour lesquelles la moyenne et la somme des précipitations seront extraites. Le fichier vectoriel ne contient pour l'instant que deux attributs (identifiant et nom). Installer l'extension *Statistiques de zones* en suivant la même démarche que dans l'exemple précédent. Pour accéder à l'extension installée, aller sous **Raster >> Statistiques de zones >> Statistiques de zones**. Une fenêtre s'ouvre. Sous **couche raster**, définir la couche raster sur laquelle les statistiques seront calculées (ici, fichier de précipitation). Sous **Couche de polygones contenant les zones**, définir la couche vectorielle contenant les zones (ici, couche des zones administratives). Cliquer sur OK. Les statistiques calculées sur le fichier raster sont directement ajoutées en tant qu'attribut dans la couche vectorielle des régions administratives.



## 15. INTERACTIONS ENTRE COUCHES DE DONNÉES : RASTER – RASTER

Les concepts des interactions entre couches de données Raster-Raster ont permis de développer l'algèbre de carte. La théorie de l'algèbre de carte a été introduite par Dana Tomlin dès 1983. Dans cette leçon, les types d'interactions entre plusieurs couches de type raster et les principes de l'algèbre de carte seront présentés. À la fin de cette leçon, il sera possible de combiner et calculer de manière pratique les informations entre fichiers raster dans QGIS.

### ALGÈBRE DE CARTE

Au même titre que les interactions entre couches raster-vecteur, les contraintes à la mise en œuvre de l'interaction spatiale entre couches raster-raster sont l'utilisation d'un système de projection de coordonnées identique et une couverture spatiale commune. Toutefois, à ces conditions préalables s'ajoute une condition sur la résolution spatiale. La taille des pixels doit être identique. D'après Tomlin, les fichiers de type raster peuvent être soumis à des opérations algébriques dont les résultats sont stockés dans de nouveaux fichiers raster. Ces fichiers sont traités comme des grilles de points et l'algèbre de carte décrit les opérations arithmétiques sur des cellules, sur des groupes de cellules ou des classes d'objets dans l'ensemble des cellules.

### ADDITION

L'opérateur d'addition entre couches raster permet d'additionner les valeurs des cellules entre ces couches. Voici un exemple pratique dans QGIS dans lequel les précipitations totales du Burkina Faso pendant la saison des pluies (juin-septembre) à partir des rasters de précipitations mensuelles des mois de juin à septembre sont calculées en chaque pixel. Pour obtenir le résultat désiré, il est nécessaire d'additionner les différents rasters entre eux.

Dans QGIS, cliquer sur **Raster >> calculatrice raster**. La calculatrice s'ouvre (figure 1, voir page suivante). Cet outil permet de sélectionner des couches raster (ici, les couches de précipitation) et de procéder à une opération algébrique entre eux (ici l'addition). Il permet aussi de définir l'emprise sur laquelle l'opération doit avoir lieu (ici l'emprise de la carte). Cliquer sur OK.

Le résultat est obtenu sous forme d'un nouveau fichier raster.

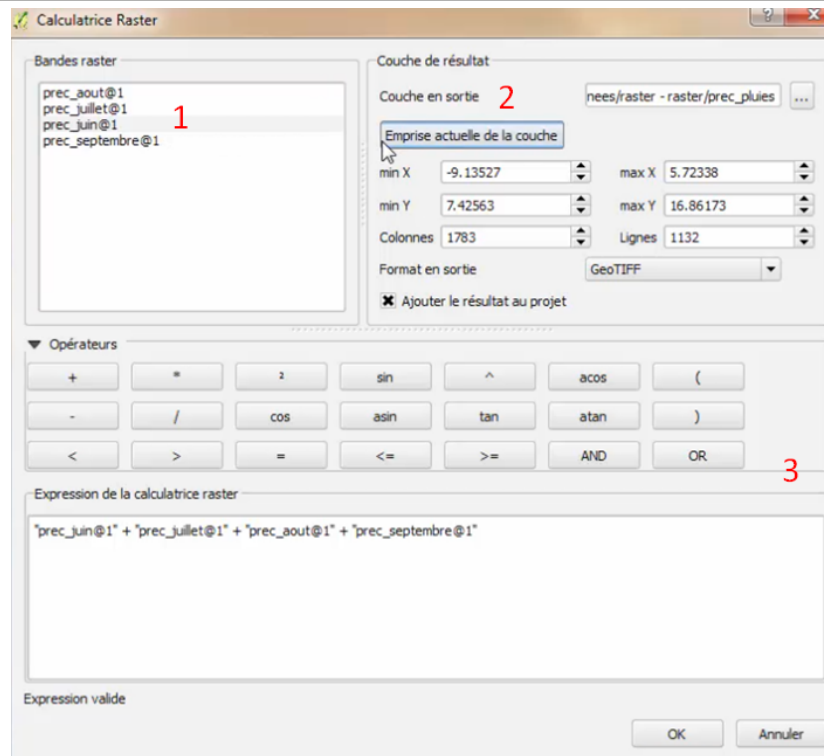


FIGURE 1

3:44

12:00

Calculatrice raster : 1) liste des couches raster disponibles pour les calculs ; 2) paramètres à définir pour le fichier de résultat ; 3) calculatrice avec ses différents opérateurs.

## SOUSTRACTION ET DIVISION

Voici un autre exemple sur QGIS dans lequel un indice de végétation par différence normalisé (NDVI) pour une région du nord de la Tunisie est calculé. Le NDVI est un indice sensible à la quantité et à la vitalité de la végétation. Il est basé sur une opération arithmétique entre deux bandes spectrales (rouge (*R*) et proche infrarouge (*NIR*)). Le NDVI est égal à la différence entre la couche *NIR* et la couche *R* divisée par la somme des deux.

$$NDVI = \frac{NIR - R}{R + NIR}$$

La normalisation de la différence permet à la fois de borner l'indice entre  $-1$  et  $1$ , mais aussi de limiter les effets de la réflectance. Les valeurs négatives sont représentatives des surfaces non végétatives telles que la neige, l'eau et les nuages. Les valeurs positives entre  $0.1$  et  $0.7$  représentent les zones de végétation et les valeurs proches de zéro sont les sols nus. Dans cet exemple, les deux bandes spectrales sont représentées sous forme raster. Il est donc possible d'utiliser la calculatrice raster (**Raster >> calculatrice raster**) pour effectuer le calcul du NDVI. Le résultat de cet exemple est obtenu sous forme d'un nouveau fichier raster. Les valeurs de ce fichier peuvent être classées en utilisant les propriétés de style de la couche.

## OPÉRATEURS RELATIONNELS

Parmi l'algèbre de carte, les opérateurs logiques et relationnels sont des outils précieux pour déterminer les pixels correspondant à des critères définis préalablement.

Voici un exemple dans QGIS dans lequel des terrains dont la pente est supérieure à  $45$  degrés avec une exposition comprise entre l'ouest et le nord sont recherchés. Deux rasters sont à disposition. Le premier contient les valeurs de pentes et le second les valeurs d'orientations. La carte indiquant les pixels satisfaisant aux deux conditions s'obtient en utilisant l'opérateur AND dans la calculatrice raster de QGIS. Le résultat de cet exemple est obtenu sous forme d'un nouveau fichier raster dont les valeurs des pixels sont soit  $1$  ou  $0$ . La valeur  $1$  indique que le pixel respecte les conditions et la valeur  $0$  indique que non.

## 16. INTERACTIONS ENTRE COUCHES DE DONNÉES : VECTEUR – VECTEUR

Dans cette leçon, les types d'interactions entre plusieurs couches de type vecteur et les outils à disposition dans QGIS seront présentés. À la fin de cette leçon, il sera possible de combiner et calculer de manière pratique les informations entre fichiers vecteurs dans QGIS.

### INTERACTION ENTRE VECTEURS

Il existe plusieurs types d'interaction possibles entre couches vectorielles. QGIS est muni d'outils qui permettent d'exploiter ces types d'interactions. Les opérations (jonctions, dénombrement, etc.) peuvent être effectuées aussi bien avec les outils disponibles dans QGIS qu'avec des requêtes SQL (voir module 2 sur les requêtes spatiales topologiques). Les opérations de bases indispensables pour exploiter les interactions entre couches vectorielles sont présentées ci-dessous.

### ÉQUIVALENT SQL

Il arrive souvent qu'un jeu de données sans identifiant unique et constitué de points contienne des attributs dans deux couches différentes. Dans l'exemple du MOOC, le jeu de donnée est constitué de deux couches représentant des hôtels sous forme de points. La couche numéro 1 possède les attributs du nom de l'hôtel et du nombre de chambres et la seconde couche contient le nombre de lits. La réunification de tous les attributs dans une seule couche se fait par jonction des deux fichiers sur la base des coordonnées des points. Deux méthodes existent pour réaliser cette jonction. La première utilise le langage SQL et la seconde utilise la fonction disponible dans QGIS. En SQL, la jonction se fait comme suit :

```
SELECT hotels1.geometry, hotels1.nom, hotels1.chambre, hotels2.lits  
FROM hotels1, hotels2  
WHERE hotels1.geometry = hotels2.geometry
```

Dans QGIS, aller sous **Vecteur >> Outils de gestion de données >> joindre les attributs par localisation**. Une fenêtre s'ouvre. Spécifier la couche vecteur de base (ici, **hotel1**) et la deuxième couche vecteur à joindre (ici, **hotel2**). Dans **Résumé de l'attribut**, il est possible d'indiquer l'action à entreprendre en cas d'attributs commun aux deux couches. Indiquer le nom et l'emplacement d'enregistrement du fichier résultats. Dans **Table en sortie**, il est possible d'indiquer si le fichier de sortie doit conserver ou non les points qui n'ont pas de correspondance. Cliquer sur OK. Le fichier final est une nouvelle couche de type vecteur point dans laquelle tous les attributs de la couche 1 et de la couche 2 sont présents.

En pratique, il est aussi utile de pouvoir compter le nombre de points d'une couche de type vecteur contenu dans les polygones d'une autre couche vecteur. Dans le deuxième exemple du MOOC, le nombre d'hôtels provenant de la couche 1 pour chaque district (polygones) de la couche 2 est compté. En SQL, l'opération s'écrit :

```
SELECT district.nom, count(hotels1.nom) as nhotels  
FROM district, hotel1  
WHERE CONTAINS(district.geometry, hotels1.geometry)  
GROUP BY district.nom
```

Dans QGIS, aller sous **Vecteur >> Outils d'analyse >> Points dans un polygone**. Une fenêtre s'ouvre. Indiquer la couche de polygones en entrée (ici, **district**) et la couche de points en entrée (ici, **hotel1**). Donner un nom au nouvel attribut de nombre d'hôtels par districts. Indiquer le nom et l'emplacement d'enregistrement du fichier résultats et cliquer sur OK. Le fichier résultats est une nouvelle couche de type vecteur contenant les polygones des districts et le nombre d'hôtels associés à chaque polygone.

Calculer la longueur totale de lignes à l'intérieur de polygones est également une opération fréquente. Dans l'exemple du MOOC, la longueur totale d'un réseau routier pour chaque district doit être calculée. En SQL, la fonction devient compliquée et s'écrit ainsi :

```
SELECT district.nom, sum(LENGTH(INTERSECTION(district.geometry, routes.
geometry))) as lroutes
FROM district, routes
WHERE INTERSECTS(district.geometry, routes.geometry)
GROUP BY district.nom
```

Dans QGIS, aller sous **Vecteur >> Outils d'analyse >> Total des longueurs de lignes**. Une fenêtre s'ouvre. Indiquer la couche de polygones en entrée (ici, **district**) et la couche de lignes en entrée (ici, **routes**). Donner un nom au nouvel attribut. Indiquer le nom et l'emplacement d'enregistrement du fichier résultats et cliquer sur OK. Le fichier résultats est une nouvelle couche de type vecteur contenant les polygones des districts et la longueur totale du réseau routier associés à chaque polygone.

## DÉCOUPAGE

Le découpage d'une couche vectorielle par une autre peut être effectué de trois façons différentes.

- découpage entre polygones ;
- découpage de lignes par des polygones ;
- extraction de points, segment ou surface à l'intérieur d'un polygone.

Le découpage dans QGIS génère un résultat contenant uniquement les attributs de la couche découpée.

Dans le premier exemple du MOOC, le fichier du réseau routier (lignes) est découpé selon 5 districts (polygones) situés dans le sud de la région d'étude. Seuls les tronçons des routes situées dans ces 5 districts seront retenus. Dans QGIS, aller sous **Vecteur >> Outils de géotraitement >> Découper**. Une fenêtre s'ouvre. Indiquer la couche de saisie (ici, **routes**) et la couche de découpage (ici, **district**). Indiquer le nom et l'emplacement d'enregistrement du fichier résultats et cliquer sur OK. Le fichier résultats est une nouvelle couche vectorielle des routes contenue dans les 5 polygones.

Extraire des points situés dans des polygones est une opération fréquemment utilisée. Dans l'exemple du MOOC, les points de la couche vectorielle des hôtels situés dans deux districts choisis sont extraits. Dans QGIS, procéder comme pour l'exemple du réseau routier (sous **Vecteur >> Outils de géotraitement >> Découper**). La seule différence étant que la couche de saisie est maintenant la couche de points (ici, **hotel1**). Le fichier résultats est une nouvelle couche vectorielle des hôtels contenus dans les 2 polygones choisis.

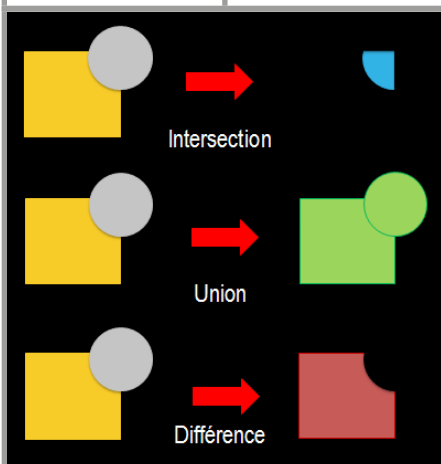


FIGURE 1

13:37 20:00

Intersection, union et différence dans le cas de deux polygones.

## OPÉRATIONS BOOLÉENNES

L'*algèbre de Boole* est une approche algébrique de la logique exprimée en termes de variables, d'opérateurs et de fonctions sur les variables logiques. Cette approche permet de traiter les expressions à deux valeurs du calcul des propositions. Dans le cas de polygones, les situations suivantes sont fréquentes (figure 1):

- l'intersection entre polygones;
- l'union entre polygones;
- la différence entre polygones.

Le premier exemple du MOOC présente l'opération booléenne Union en réunifiant les fichiers vectoriels des districts de l'île de Mahé et de l'île de Praslin. En SQL, la réunification s'écrit ainsi:

```
SELECT geometry
FROM districts_maye
UNION
SELECT geometry
FROM districts_praslin
```

Dans QGIS, aller sous **Vecteur >> Outils de géotraitement >> Union**. Une fenêtre s'ouvre. Indiquer la couche de saisie (ici, **districts\_mahe**) et la couche d'union (ici, **districts\_praslin**). Indiquer le nom et l'emplacement d'enregistrement du fichier résultats et cliquer sur OK. Le fichier résultats est une nouvelle couche vectorielle contenant les districts de Mahé et de Praslin.

Le deuxième exemple du MOOC présente l'intersection entre les districts de l'île de Mahé et deux zones d'étude choisies. Ainsi, uniquement les zones communes aux deux fichiers vectoriels sont conservées. L'utilisation de la fonction découpage permet aussi de garder les zones communes, mais le résultat contiendra uniquement les attributs de la couche découpée contrairement à la fonction intersection.

Dans QGIS, aller sous **Vecteur >> Outils de géotraitement >> Intersection**. Une fenêtre s'ouvre. Indiquer la couche de saisie (ici, **districts**) et la couche d'intersection (ici, **zones d'étude**). Indiquer le nom et l'emplacement d'enregistrement du fichier résultats et cliquer sur OK. Le fichier résultats est une nouvelle couche vectorielle des districts contenus dans les zones d'étude. Les attributs de cette couche sont les attributs contenus dans les deux couches initiales.

## 17. ÉTUDE DE CAS – LE RISQUE D'INONDATION À NIAMEY

Cette étude de cas, basée sur l'analyse des mécanismes hydrométéorologiques et de l'occupation humaine, permet de montrer de quelle manière les outils SIG sont utilisés pour cartographier les zones inondées et inondables afin d'améliorer la prise de décisions pour l'aménagement du territoire. La région urbaine de Niamey au Niger subit régulièrement, depuis 15 ans, des inondations dues à la crue du fleuve Niger et aux ruissellements après les pluies diluviennes.



### VILLE DE NIAMEY

Niamey est la capitale du Niger. Elle compte environ 1 300 000 habitants et s'étale sur plus de 240 km<sup>2</sup>. Elle est séparée en deux par le fleuve Niger et est régie par 5 arrondissements communaux (figure 1). La forte croissance démographique et socio-économique des 30 dernières années a forcé la ville à s'étendre dans des zones inondables sur la rive droite et dans les quartiers de Goudel et Saga sur la rive gauche.

FIGURE 1

2:50

19:00

Carte des arrondissements communaux de Niamey.

### PROBLÉMATIQUE DES INONDATIONS

Les inondations sont des catastrophes naturelles qui causent des dommages aux personnes, aux infrastructures et à l'environnement estimés à plusieurs milliards de francs CFA. Le Niger connaît ces dernières années une pluviométrie excédentaire. Les pluies diluviennes occasionnent des dégâts importants et les digues de protection n'arrivent plus à protéger efficacement la population provoquant des déplacements de population et des relogements.

Le régime hydrologique du fleuve Niger est composé de deux crues. L'une de décembre à janvier provenant du haut bassin et l'autre, une crue locale, en saison des pluies.

Cette étude s'intéresse aux facteurs aggravants engendrés par l'homme tel que l'imperméabilisation et la dégradation des sols dus à la déforestation et l'urbanisation croissante ainsi qu'au manque d'entretien des cours d'eau. Les mesures pouvant être mises en place se basent sur deux axes : mesures d'aménagements et mesures de prévention. Les mesures d'aménagement comportent des mesures d'interdictions de constructions dans les zones inondables, des mesures d'aménagement des berges, de construction de digues et de désensablage du lit majeur de la rivière. Les mesures de prévention comportent des mesures d'aménagements du territoire, de cartographie du danger et de modélisation des crues.





## MÉTHODOLOGIE

Les données sur les facteurs de risques d'inondation sont séparées en deux catégories. Les facteurs naturels (topographie, hydrologie et géologie) et les facteurs anthropiques (occupation du sol et tissus bâtis). Les types de données à disposition sont les images satellitaires, les cartes et les données de terrain.

La notion de risque d'inondation est définie par la probabilité que survienne une inondation et par la sévérité de ses conséquences. Il est donc important d'analyser les aléas, en se basant sur les caractéristiques morphologiques, topographiques, géologiques, pédologiques et hydrauliques contrôlant le phénomène des inondations et en évaluant la vulnérabilité des zones d'inondation potentielle en intégrant les données des activités anthropiques et socio-économiques.

La cartographie de l'aléa d'inondation se base sur l'analyse des facteurs naturels. Dans cette étude, les aléas pertinents sont l'exposition au débordement du fleuve Niger et l'exposition à l'accumulation des eaux de ruissellement suite aux pluies diluviennes. L'aléa au débordement est déterminé à partir des données historiques. En déterminant la crue centennale, il est possible d'estimer le rayon d'influence de cette crue et ainsi de déterminer l'exposition au débordement du Niger. Le rayon estimé pour cette étude est de 400 mètres. La distance linéaire de chaque unité par rapport au lit du fleuve détermine le niveau d'exposition et sa représentation graphique est une zone tampon de part et d'autre du fleuve. L'indice de Beven est un indice topographique calculé à partir d'un MNT. Il permet de caractériser le niveau d'accumulation des eaux de ruissellement en chaque point. Cet indice est utilisé pour former la carte des aléas de l'exposition à l'accumulation des eaux de ruissellement. Les deux cartes d'aléas permettent de construire une carte des zones inondables de la ville de Niamey.

La carte des enjeux permet de connaître la sévérité des conséquences d'une inondation. Elle est créée sur la base des images satellitaires Landsat de 1998 à 2013 afin de permettre la digitalisation et la classification du bâti. Pour cette étude, la vulnérabilité des différents bâtiments et infrastructure est supposée égale.

La combinaison de la carte des zones inondables et de la carte des enjeux génère une carte des risques d'inondation. Cette carte est un outil essentiel pour les décideurs de la ville dans le cadre de l'aménagement et de la prévention.

## RÉSULTATS

L'analyse des cartes de l'occupation du sol a permis de montrer l'évolution du tissu bâti de 1998 à 2013 et met en évidence l'augmentation du risque associé à cette occupation. La carte des risques d'inondation montre la situation critique de la ville de Niamey. La rive droite présente un risque élevé dans les quartiers de Lamordé et Nogaré. Les quartiers de Saga et de Goubel sont les plus exposés sur la rive gauche.

Ces résultats ont été obtenus en ne considérant qu'une partie des paramètres, mais permettent tout de même de situer les zones à risque. Ils peuvent donc servir de bases pour une prise de décision pertinente dans le cadre de l'aménagement du territoire.

## 18. LES COMPOSANTES ET FONCTIONS DE LA CARTE

La représentation de l'information géographique s'applique en suivant les règles établies en cartographie thématique. Dans cette leçon, vous apprendrez à distinguer les types de cartes thématiques et d'en expliquer les fonctions et composantes. Vous serez également capable de restituer les étapes de fabrication d'une carte.

### INTRODUCTION

La représentation de l'information géographique et la communication de l'information récoltée ou traitée sont souvent réalisées au travers de cartes thématiques statistiques. Pour permettre une bonne interprétation, ces cartes doivent suivre certaines règles. Celles-ci font intervenir la sémiologie graphique, l'échelle, la légende et les sources. Les règles et conseils appliqués dans ce cours proviennent de trois sources différentes : *De la graphique* de Jacques Bertin, de Roger Brunet et le groupement d'intérêts publics *Reclus* à Montpellier ainsi que de Micheline Cosinschi de l'Université de Lausanne.

### CARTOGRAPHIE

La *cartographie* est une science, un art et une technique. Une science, car elle se base sur des fondements mathématiques (forme de la terre, systèmes de projection, etc.). Un art, car elle est une expression graphique capable de présenter des qualités esthétiques et didactiques. Une technique, car la réalisation des cartes passe par la mobilisation d'outils et de technologies (images satellites, informatiques, etc.). La cartographie a pour but la représentation de la Terre (ou autres corps célestes) sous une forme géométrique et graphique. Elle est un système de signes synoptiques permettant de communiquer comme le langage.

La *carte* est la projection de l'espace géographique sur un plan (passage de trois à deux dimensions). La carte est donc premièrement une représentation de tout ou d'une partie de la surface de la Terre sur un plan. Sa représentation désigne l'image mentale associée aux phénomènes géoréférencés et le processus de création de la carte. Elle est aussi un outil permettant de comprendre et d'interpréter les agencements spécifiques d'objets dans l'espace géographique et elle permet la transmission et le stockage de l'information.

La *carte thématique* est une représentation graphique de la distribution spatiale d'un ou plusieurs attributs d'un phénomène quantitatif ou qualitatif. Elle représente au minimum un thème et propose une simplification de la réalité au moyen d'éléments identifiables. Elle permet donc de décrire la répartition géographique d'un phénomène et de sa variabilité spatiale sans que la localisation précise des objets soit nécessaire.

La *carte topographique* est une carte particulière représentant l'occupation du sol, mais dont la localisation précise des objets sur le territoire est primordiale. Elle n'est donc pas incluse dans la catégorie des cartes thématiques.

La *carte d'inventaire* est une carte thématique descriptive. Délicate à réaliser car elle demande de localiser avec précision une information abondante en assurant la meilleure lisibilité possible.

La *carte thématique statistique* est élaborée par la représentation simultanée d'unités spatiales (comptés, communes, etc.) et de données qualitatives ou quantitatives (taux de chômage, naissance, qualité de vie, etc.). Ces cartes requièrent un traitement de l'information faisant appel à des connaissances statistiques et méthodologiques précises et impliquent une lecture interprétative de la carte. Elles sont souvent utilisées pour représenter des phénomènes dont l'ampleur ne permet pas de les représenter efficacement sous forme de tableaux de chiffres. L'information géographique est souvent présentée sous forme de surfaces et de polygones colorés (cartes choroplèthes) ou sous forme de symboles proportionnels lorsque l'information est ponctuelle.

La *carte thématique d'aménagement* est produite dans le cadre d'étude d'impact ou de planification du territoire. Elle associe l'information des infrastructures urbaines, des milieux naturels et des informations socio-économiques. Elle est utilisée comme support de communication pour les spécialistes et le grand public moyennant certaines adaptations. Les plans d'aménagements, les cartes de risque et les cartes de danger sont tous des exemples de cartes d'aménagement.

Les cartes permettent de représenter et communiquer une information géoréférencée. Elles peuvent représenter et explorer de gros volumes de données complexes en les synthétisant et en les structurant de manière à mettre en évidence les différences dans l'espace géographique. L'esthétique plaisante des cartes leur permet de transmettre facilement l'information. La carte est donc un moyen intégré permettant d'informer, de rechercher, de réfléchir, d'enseigner et de prendre des décisions. Ses avantages sont une expression simultanée de l'information permettant la mise en évidence de propriétés spatiale imperceptible par d'autres moyens, la localisation relative des objets les uns par rapport aux autres permettant de révéler les structures spatiales et la réduction de l'espace géographique (projection sur plan et échelle) et de l'information à traiter (légende). Néanmoins, la carte est aussi une simplification de l'information et son esthétique peut parfois dominer et fausser l'interprétation de l'information véhiculée. Les cartes en papier sont également statiques, excluant de ce fait la représentation temporelle de l'information.

### COMPOSANTES DE LA CARTE

La carte est constituée de plusieurs éléments. Elle est un système de signes régis par les règles de la sémiologie graphique. L'image constituée par la carte est l'équivalent d'un langage destiné aux fonctions cognitives de l'être humain. Elle exploite le sens de la vue et obéit à des règles définies de sémiologie graphique. Cette image est plane et déformée en raison de la projection du géoïde sur le plan. Elle est aussi réduite par l'échelle de représentation appliquée. Cette image est aussi un assemblage de signes conventionnels et de symboles utilisés pour représenter des éléments spatiaux.

Un signe est défini en sémiologie par la composition d'un *signifié* désignant la représentation mentale d'un concept et d'un *signifiant* désignant l'image acoustique d'un mot. L'ensemble est appelé *sème* ou *unité de sens*.

Les signes laissent donc beaucoup de place à l'interprétation. Pour éviter les confusions dans l'analyse des cartes, des conventions admises ou contrôlées par les règles de sémiologies graphiques sont utilisées (Nord situé en haut de la carte, légendes, etc.). La légende permet de lister les signes utilisés et donne une justification aux écarts envers les conventions.

Les composantes principales de la carte sont listées dans le [tableau ci-dessous](#). Les composantes systématiques doivent toujours apparaître sur la carte car elles en assurent une interprétation correcte. Les autres composantes dépendent du contexte.

SYSTÉMATIQUES	SELON LE CONTEXTE
Légende	Titre
Échelle	Auteur
Orientation	Date de création
Source de données	Information sur les unités spatiales
Date de données	Commentaires
	Cadre
	Système de coordonnées
	Projection
	Labels
	Zooms sur zones
	Graticule
	Nomenclature
	Index

Les cartes autonomes doivent faire apparaître, en plus des éléments systématiques, un titre, le nom des auteurs, la date de création et une information sur les unités spatiales. Si la carte autonome est sortie de son contexte, il est obligatoire de faire apparaître un commentaire. Les labels, les zooms, l'affichage du graticule, une nomenclature ou un index sont généralement utilisés dans le cas de cartes d'aménagement. La figure 1 montre l'exemple d'une carte et ses composantes associées.

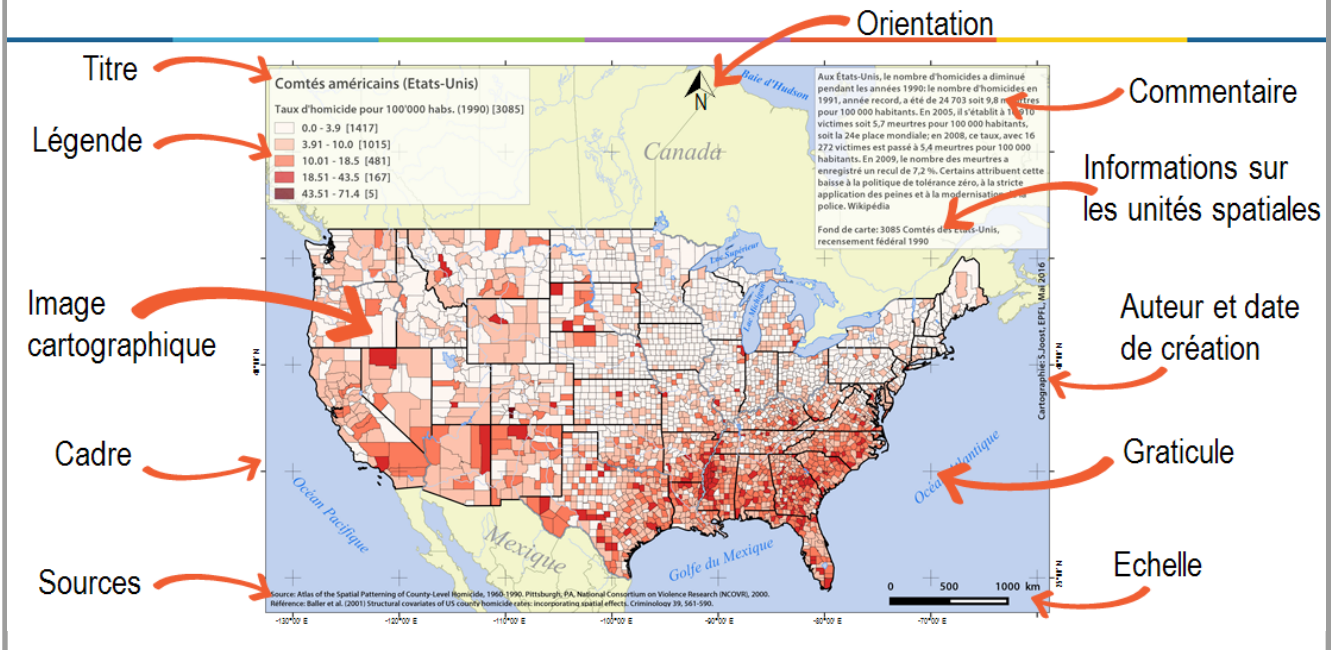


FIGURE 1

17:46

22:00

Exemple d'une carte et ses composantes associées.

## FONCTIONS DE LA CARTE

Les trois fonctions majeures de la représentation graphique identifiées par Bertin sont l'enregistrement, le traitement et la communication de l'information.

La **fonction d'enregistrement** fournit un inventaire commun et exhaustif de l'information. Cette fonction aboutit à des cartes à lire, créant une mémoire artificielle qui évite l'effort de mémorisation. C'est le cas des plans de métro.

La **fonction de traitement** est un processus de réduction de l'information (la classification, la catégorisation, etc.) qui permet d'obtenir un schéma synthétique plus facile à mémoriser. Les cartes thématiques statistiques utilisent cette fonction.

La **fonction de communication** consiste à donner les moyens de retenir une information à l'aide de la mémoire visuelle. Elle produit une image simple qui favorise la mémorisation de l'information sans qu'elle soit exhaustive.

Le processus permettant l'élaboration d'une carte thématique statistique peut être expliqué en 4 étapes. Une première étape d'*acquisition* des données formées d'attributs et d'unité spatiales. Une seconde étape de *traitement* des données dans laquelle une simplification et une classification des données peuvent être faites. Une troisième étape de *représentation* en utilisant un système de signes et de symboles conventionnels. Une dernière étape d'*édition* pour affiner et améliorer le rendu graphique.



# 19. LES VARIABLES VISUELLES

La représentation cartographique utilise un système de signes pour coder de l'information et la transmettre visuellement. L'encodage de cette information est composé de structures visuelles définies par des points, des lignes, des surfaces et des volumes. Les variations des propriétés de ces structures sont les variables visuelles et leur utilisation est régie par les règles de la technologie graphique.

Cette leçon a pour but d'expliquer le système graphique, de décrire les structures visuelles et d'expliquer les implications du bon usage du système graphique en matière de perception de l'information.

## LA GRAPHIQUE

Les cartes sont le fruit d'une réalisation ou d'une mise en scène réalisée par des moyens techniques (crayons, ordinateurs, etc.), reflétant un certain point de vue et désirant atteindre un certain but. La graphique est une discipline développée par Jacques Bertin dans les années 1950 décrivant les principes de l'information cartographique. Elle est un système de signes permettant de transcrire des relations de différences d'ordre ou de proportionnalité entre données quantitatives et qualitatives dans les deux dimensions du plan.

## LES STRUCTURES VISUELLES

Il existe quatre familles de structures visuelles :

- les points ;
- les lignes ;
- les surfaces ;
- les volumes.

L'information à cartographier est codée sur la base de ces structures visuelles. Ces dernières ne sont pas indépendantes entre elles et c'est le contexte qui permet de définir leur signification. En effet, c'est la perception visuelle qui permet la reconnaissance de modèles dans un ensemble de structures.

Les lois correspondant à l'interaction entre les structures visuelles sont définies par la *Gestaltpsychologie*, une psychologie de la forme. Elles sont listées dans le [tableau ci-dessous](#). Ces éléments sont facilement applicables à l'élaboration d'une représentation cartographique.

LOIS	EFFETS
Prégnance	Une image est facile à comprendre si sa structure est simple.
Proximité	Deux composants qui sont proches ont tendance à être perçus comme un seul composant.
Similarité	Les composants similaires sont perçus comme s'ils étaient regroupés.
Fermeture	Les contours proches sont perçus comme unifiés.
Continuité	Des éléments voisins sont perçus groupés lorsqu'ils possèdent potentiellement un trait qui les relie.
Symétrie	Des éléments sont perçus comme un élément global lorsqu'ils forment une symétrie.
Trajectoire	Des éléments qui se déplacent avec la même trajectoire semblent groupés.
Familiarité	Des éléments sont plus facilement groupables si le groupe est familier.

Lois d'interaction entre structures visuelles.

## LES VARIABLES VISUELLES

Les **variables visuelles** sont les variations des propriétés des structures visuelles (couleurs, épaisseur, etc.). Selon Bertin, il existe 8 types de variables visuelles (voir aussi figure 1) :

- La *position* est composée des deux variables X, Y.
- La *taille* traduit les variations quantitatives.
- La *valeur* ou la variation de l'intensité lumineuse du clair au foncé. Elle traduit une relation d'ordre et de différence relative.
- La *forme* exprime l'identité d'un objet.
- L'*orientation* permet de positionner un signe par rapport aux axes du graphique. Elle exprime les différences et gagne en efficacité lorsqu'elle est utilisée avec les variables de grain et de valeur.
- La *couleur* permet de traduire des différences, mais ne peut pas les ordonner entre elles.
- Le *grain* ou la résolution combine la forme et la taille. Il traduit une relation d'ordre et de différence relative.

Les variables de position, de taille et de valeur sont les *variables de l'image*. Elles permettent de construire des plages de visibilité variables (propriété dissociative), font apparaître des formes sur un fond et suggèrent des reliefs. Les autres variables sont les *variables de séparation*. Elles construisent des plages de visibilité homogènes (propriété associative) et sans relief. Elles ne font pas apparaître de forme sur le fond, mais permettent de séparer les éléments entre eux.

Dans le cas des surfaces d'une carte choroplèthe, la taille et la valeur permettent une forte différenciation. Elles créent un ordre visuel et construisent des hiérarchies. La forme ne permet pas de transcrire des différences. Elle est associative et sa sélectivité est nulle. L'orientation est associative. Elle permet de construire des plages de visibilité homogènes et sa sélectivité est très faible. La couleur associée à la valeur est fortement sélective et ordonnée. Mais une gamme de couleurs d'égales valeurs construit une image sans relief.

Le système graphique de Bertin considère 8 variables visuelles applicables à 3 types d'implémentations (points, lignes et surfaces) et possédant 5 propriétés différentes (quantitative, ordonnée, sélective, dissociative et associative) permettant de transcrire les relations de proportionnalité, d'ordre et de ressemblance entre les données.

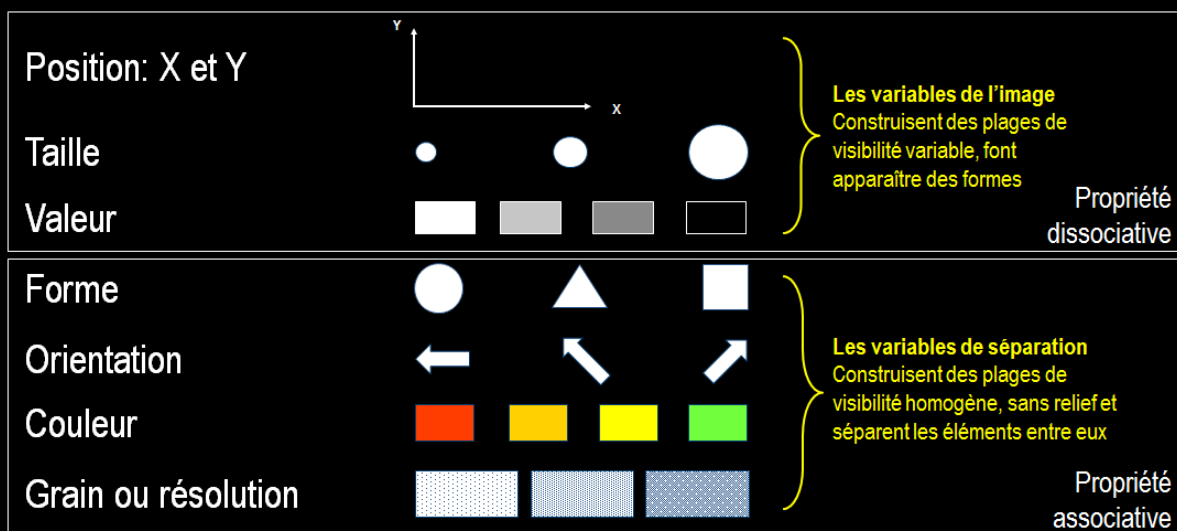


FIGURE 1

6:09

22:00



## GRAPHIQUE ET PERCEPTION

Pour maximiser l'efficacité de la représentation graphique, il est important de comprendre et de connaître les caractéristiques des variables visuelles. Chaque variable visuelle comporte un certain nombre de degrés discrets. La *longueur* d'une variable visuelle est le nombre de paliers sensible qu'une variable visuelle peut supporter et l'*étendue* d'une variable visuelle est l'écart qui sépare les valeurs extrêmes du paramètre. La *longueur* de la *valeur* est très présente dans les cartes choroplèthes car elle est utilisée pour fixer le nombre de classes dans le processus de discrétisation. L'utilisation de la longueur permet de trouver le nombre de classes dont les seuils sont perceptibles par l'œil de la majorité des utilisateurs. La valeur est liée à la notion de luminosité en théorie des couleurs et ne dépend pas de la chromaticité. Le nombre de seuils perceptifs dépend donc de la valeur du fond de carte et de la taille des éléments graphiques. Un maximum de 6 seuils perceptifs (longueur = 7) est communément admis comme idéal. La variation de la valeur étant ordonnée, il est important d'utiliser la valeur lorsque l'ordre entre variables fait sens. Par exemple, utiliser un dégradé de gris pour représenter des classes qualitatives est une erreur et compromet l'information. Il est aussi important de noter qu'il est d'usage de réserver le blanc aux données manquantes. Il ne doit pas être intégré comme classe initiale dans un dégradé de valeurs.

Les cartes thématiques statistiques sont des supports pour la pensée et permettent d'améliorer le processus cognitif. L'efficacité de ces cartes, en plus de la qualité de leurs représentations, dépend également de la qualité de l'image à être une forme visualisable et perceptible dans l'instant minimum de vision. Toute représentation cartographique présentant plus de trois composantes ne peut donc pas être construite en une seule image et nécessite l'élaboration de plusieurs cartes.

## COULEUR

La couleur peut être décrite par trois composantes :

- La teinte est la forme pure de la couleur.
- La saturation est l'intensité de la coloration.
- La luminosité est l'impression de clarté associée à la couleur.

Les couleurs sont codées informatiquement afin de permettre la combinaison efficace des niveaux de teintes, de saturations et de luminosités lors de leur application aux structures visuelles. Informatiquement, les couleurs sont représentées par un triplet de valeurs se rapportant à chacune des couleurs primaires de la synthèse additive RVB (rouge, vert et bleu). Les codes des couleurs sont des nombres entiers bornés entre 0 et 255, soit 256 valeurs sur un octet. Néanmoins, l'utilisation des couleurs primaires pour représenter les couleurs n'est pas évidente à utiliser. Pour cela, les logiciels d'édition de cartes offrent la possibilité d'ajuster la teinte, la saturation, la valeur et la transparence et traduisent directement ces paramètres en RVB. Néanmoins, ce système reste toujours compliqué et ne permet pas de choisir efficacement une palette de couleur adéquate pour une représentation cartographique.

Un choix judicieux peut être fait soit en utilisant les palettes de couleurs prédéfinies dans le logiciel QGIS par exemple, ou en utilisant l'outil développé par Cynthia Brewer du GeoVISTA Center. Cet outil permet de visualiser directement l'effet des choix des paramètres sélectionnés sur la base d'une série de palettes prédéfinies garantissant une efficacité optimale.

La couleur est fortement sélective et permet une bonne discrimination dans les teintes rouges et violettes, mais la discrimination est moins bonne dans les teintes jaunes. La couleur est ordonnée si sa relation d'ordre est transcrite par la variable visuelle *valeur*. Dans le cas de l'utilisation de couleurs différentes, la valeur transcrivant la relation d'ordre est perçue avant la teinte qui, elle, permet la sélection. L'efficacité de la couleur diminue également avec la surface des structures visuelles. La combinaison de la variation de la teinte et de la variation de la valeur offre plus de seuils perceptifs.

La couleur à l'avantage d'être associée à une forte attractivité psychologique et rends l'information plus facilement mémorisable. Son désavantage est le coût élevé de la diffusion des documents couleur non électroniques et la difficulté à tenir compte des utilisateurs souffrant d'anomalie de la perception visuelle.

## 20. ATTRIBUTS THÉMATIQUES ET MISE EN CLASSES

Cette leçon a pour objectif d'inculquer les notions de cartographie thématique statistique, de présenter les types de cartes thématiques et d'expliquer les méthodes de discrétisation.

### NOTIONS DE BASE

Le contenu de l'information géographique forme la composante thématique de la carte. Il est représenté par les attributs des objets géographiques. Dans la phase de traitement de l'information, il est nécessaire de généraliser l'information au moyen de la classification et de la catégorisation. Cette généralisation se base sur plusieurs notions importantes :

- L'unité spatiale. Une unité d'observation indivisible et géoréférencée (polygones, lignes, points).
- La distribution spatiale est la notion de fréquence à laquelle un phénomène apparaît dans l'espace géographique.
- La structure spatiale est la notion d'organisation de l'information.
- Le processus spatial est le mécanisme qui engendre les structures spatiales des distributions.

L'information est généralement stockée sous forme de tableau à double entrée. Les lignes horizontales représentent les unités spatiales étudiées et les colonnes verticales stockent les attributs thématiques. Les unités spatiales sont toujours marquées par un identifiant unique contenu dans une variable spéciale. Cet identifiant permet d'assurer l'unicité des unités spatiales. Il existe 2 principaux types de variables thématiques. Les variables quantitatives et les variables qualitatives.

Une *variable est quantitative* lorsque sa valeur prend des nombres reflétant une notion de grandeur (comptage, mesures, etc.). Les variables quantitatives sont discrètes ou continues. Elles sont discrètes lorsque l'information qu'elles représentent prend un nombre fini de valeurs entre deux valeurs quelconques et continues lorsque cette information prend un nombre infini de valeurs entre deux valeurs quelconques. Les variables quantitatives discrètes dénombrement et les variables quantitatives continues approximent.

Une *variable est qualitative* lorsque l'information caractérisant chaque unité spatiale n'est pas représentative d'une quantité. Les valeurs prises par cette variable sont appelées *modalités*. Une variable qualitative est dite *nominale* si ses modalités ne peuvent être ordonnées. Elle est *ordinaire* dans le cas contraire.

Les variables peuvent donc être quantitative discrète, quantitative continue, qualitative nominale ou qualitative ordinaire.

Certaines règles s'imposent dans le cas de variables quantitatives. Une variable thématique brute issue d'un comptage ou d'un dénombrement doit obligatoirement être représentée en utilisant la variable visuelle de la taille. En effet, la taille peut directement représenter les variations quantitatives par une variation de la surface d'un symbole (cercles proportionnels par exemple). Si la variable statistique est relative (moyenne, taux, rapport, etc.), elle est valable en tout point de la surface des unités spatiales. Il est possible de la représenter en utilisant la variable graphique de la valeur et de la couleur.

## CARTES EN SYMBOLES PROPORTIONNELS

La représentation cartographique d'une valeur absolue en symboles proportionnels s'applique par une relation mathématique entre la distribution statistique des valeurs et la surface des symboles à représenter sur la carte. La surface des symboles est donc le paramètre qui varie proportionnellement.

L'application de cette relation mathématique demande la connaissance préalable de la distribution statistique de la variable choisie. Dans QGIS, il est possible de visualiser les statistiques descriptives à l'aide de l'outil d'[analyse vecteur statistique](#) ou via le plugin [Statist](#).

Le choix d'un ratio entre la valeur de la variable et la surface du symbole proportionnel se base sur plusieurs contraintes. L'étendue (plage de valeur entre min et max), qui est la base sur laquelle ce choix est défini. Les caractéristiques de la dispersion, qui ont un impact sur la construction de la carte. En effet, une dispersion faible des valeurs risque de mener à la construction d'une carte dont les symboles ont tous une taille similaire et une dispersion forte risque de faire apparaître uniquement des petits et des gros symboles. La distribution spatiale et la taille des unités spatiale, qui influence la répartition des points supportant les symboles.

De manière concrète, il est primordial de dimensionner les plus gros symboles pour qu'ils ne masquent pas le fond de carte et que leur emprise n'affecte pas les plus petits symboles. Lorsque le dimensionnement ne parvient pas à satisfaire l'exigence de lisibilité, il peut être nécessaire de décomposer la carte en deux vues, par un zoom par exemple, en prenant soin d'ajuster la règle de proportionnalité pour les deux vues. La représentation d'une variable dont la dispersion n'est pas trop forte commence par l'identification de la valeur maximum à représenter pour déterminer la taille du plus gros symbole. Le ratio de représentation de toutes les autres valeurs est basé sur la valeur maximum. Ainsi pour une valeur maximum et un rayon maximal à représenter, il est possible de calculer la surface ( $S$ ) en pixel<sup>2</sup> ( $S = \pi r^2$ , pour un cercle) du rayon maximum et ensuite, à l'aide d'une règle de trois, en déduire la surface et les rayons des autres valeurs. Lorsque la variable à représenter est trop dispersée, il est nécessaire de transformer le ratio en passant d'un ratio linéaire ( $valeur * ratio = Surface$ ) à un ratio logarithmique ( $\log(valeur) * ratio = surface$ ) ou racine ( $\sqrt{valeur} * ratio = surface$ ). Pour éviter de confondre l'utilisateur, il est obligatoire d'indiquer dans la légende le type de transformation opérée.

La version 2.8 de QGIS permet de fabriquer des cartes en cercle proportionnel en utilisant l'extension [analyse en ronds et secteurs proportionnels](#). Cette extension est celle utilisée dans le cadre de l'exemple qui suit. La carte de cet exemple est celle des comtés américains et l'exercice consiste à représenter la population totale par comté exprimée en nombre d'habitants.

Dans QGIS, aller sous [Extension >> installez gérer les extensions](#) et installer l'extension [Analyse en ronds et secteurs proportionnels](#). Pour permettre la représentation, il est nécessaire de connaître la distribution des valeurs. Pour cela, sélectionner les comtés désirés puis aller sous le menu [Vecteur >> Outil d'analyse >> Statistiques basiques](#). Une fenêtre s'ouvre. Indiquer la couche vectorielle d'entrée (ici, comtés), spécifier s'il faut utiliser uniquement les valeurs sélectionnées (ici, oui) et indiquer pour quelle variable les statistiques simples doivent être élaborées (ici, [population](#)). Cliquer sur Ok. La statistique s'affiche.

Il est maintenant possible de lancer l'extension en allant sous le menu [Vecteur >> analyse en ronds proportionnels](#). Une fenêtre s'ouvre. Dans l'onglet [Analyse et légende](#), sélectionner la couche de fond (ici, comtés) et indiquer la variable étudiée. Définir ensuite l'échelle en utilisant la valeur maximum obtenue par la statistique simple et préciser le rayon maximum du cercle. Des valeurs personnalisées pour la légende peuvent être précisées. Par défaut, QGIS indique le maximum, le maximum divisé par trois et le maximum divisé par 9. En bas de la fenêtre, choisir le moyen de stockage de la légende et des cercles proportionnels. Cliquer sur OK.

La couche des cercles proportionnels et la couche contenant la légende apparaissent. Le style peut être modifié sous les propriétés de la couche. Une transparence de 40%, la couleur noire pour le remplissage et une largeur de bordure de 0.3 permet de bien mettre évidence l'information.

## CARTES EN PLAGES DE COULEURS

Les cartes thématiques en plages de couleur sont réparties en deux groupes. Le groupe concernant les données qualitatives nominales et ordinales et les cartes chloroplèthes concernant les données quantitatives relatives.

La carte en plages de couleurs se construit en plusieurs étapes. La **discrétisation** est une opération permettant de classer la distribution statistique des valeurs d'une variable. Elle a pour but de simplifier l'information en regroupant les objets géographiques de même caractéristiques en classes homogènes et distinctes. La discrétisation est l'étape finale de la réduction, de l'organisation et de la hiérarchisation de l'information. Elle doit satisfaire aux règles de la sémiologie graphique et de la statistique afin de perdre le moins d'information possible tout en optimisant la perception visuelle et la transmission de l'information. La discrétisation est applicable aux variables quantitatives et qualitatives. Dans le cas de la discrétisation de variables qualitatives nominales, l'information est généralisée en fonction de la hiérarchie. Par exemple, les comtés des États-Unis peuvent être classés selon leur appartenance aux États ou selon des classes comprenant les États du Nord et les États du Sud par exemple. Dans de telles cartes, il est important de maximiser la discrimination entre unités spatiales. Pour les variables qualitatives ordinales, la discrétisation se base sur le rang.

Le prochain exemple présente la création d'une carte qualitative nominale dans QGIS. L'exemple porte sur la distinction des comtés américains selon leur appartenance aux États.

Double-cliquer sur la couche et accéder aux propriétés et à l'onglet style. Dans le menu déroulant, choisir *catégoriser* puis sélectionner la colonne désirée (ici, le nom des États). Pour maximiser la distinction entre les États, il est nécessaire de catégoriser selon des couleurs différentes et non selon un dégradé. Pour cela, sélectionner **Couleurs au hasard** dans le menu déroulant **Palette de couleur** puis cliquer sur **classer**. Le résultat est visualisable en cliquant sur **Appliquer**. Lorsque les couleurs attribuées aux États sont jugées adéquates, il est judicieux d'enregistrer le style (menu déroulant **style >> enregistrer le style**) pour une utilisation future. Pour valider les couleurs, cliquer sur OK.

Les cartes chloroplèthe sont particulièrement adaptées à la représentation des variables quantitatives relatives. Elles présentent des mesures transformées par calcul pour produire des rapports, des pourcentages et des distributions continues. L'élaboration de ce type de carte se base sur une mise en classe pertinente et un choix adéquat des teintes et des valeurs de teinte. L'indice de Huntsberger définit le nombre de classes idéales par la formule  $N_{cl} = 1 + 3.3 \log_{10}(N_{obs})$  où  $N_{cl}$  indique le nombre de classe et  $N_{obs}$  le nombre d'observations. Toutefois, cet indice n'est pas satisfaisant en présence d'une grande quantité d'objets à représenter car le nombre de seuils perceptifs du système visuel humain est de maximum 6. Un nombre impair de classes est souvent choisi pour représenter un comportement moyen.

CL'augmentation du nombre de seuils exploitable peut être réalisée en utilisant un gradient de valeurs pour deux teintes. Une des teintes est utilisée pour représenter les valeurs les plus basses et l'autre pour les valeurs élevées. Les deux teintes s'articulent autour d'une même teinte neutre. L'utilisation d'un nombre impair de classes est donc généralement recommandée sauf dans le cas où la valeur traitée représente deux comportements opposés, par exemple dans un système de votation. Cette méthode permet d'augmenter le nombre maximum théorique de classes à 12, mais reste délicate car il est nécessaire d'expliquer le système de la double gradation aux utilisateurs et le choix des couleurs doit être cohérent avec le thème présenté.

Lorsque le nombre de classes est connu, il est nécessaire de répartir les données dans les classes en procédant à la discrétisation de l'information. La nature de l'information définit la méthode utilisée pour la discrétisation. Les variables qualitatives nominales ne demandent pas de méthode spécifique, mais sont regroupées selon un critère commun. Les variables qualitatives ordinales conservent la hiérarchie de la distribution et les classes sont constituées sur la base des rangs. Les variables quantitatives demandent de considérer 3 types d'information pour déterminer les méthodes de discrétisation à utiliser.

L'ordre de grandeur révélé par les valeurs centrales de la distribution (mode, moyenne ou médiane). Ces valeurs peuvent être utilisées comme limite des classes. La dispersion est la mesure de l'inégalité des valeurs (écarts-types, intervalle interquantiles). Ces valeurs peuvent être choisies comme amplitude des classes. La forme de la distribution (normale, asymétrique, exponentielle, etc.) exprime le comportement d'un phénomène et permet de déterminer la méthode de mise en classe la plus adaptée. Le tableau ci-contre résume les 5 méthodes de discrétisation les plus utilisées.

MÉTHODE	DÉFINITION	CALCUL	REMARQUES	TYPE DE DISTRIBUTION STATISTIQUE
Écart à la moyenne	Toutes les classes ont une même étendue égale à l'écart-type, sauf les classes extrêmes	Moyenne et écart-type	Si le nombre de classes est impair, la classe centrale est à cheval sur la moyenne. Intérêt : se repérer par rapport à la moyenne, mettre en évidence les extrêmes, comparer plusieurs cartes	Normale (Gauss) avec concentration des données autour de la moyenne, éventuellement peu asymétrique
Égale amplitude	Les intervalles de classes sont égaux (intervalles constants)	(Valeur maximale - valeur minimale) / nombre de classes	Méthode simple, facile à interpréter, mais peu utilisée car ne convient pas si la distribution des valeurs est asymétrique : les classes peuvent être très inégales et certaines vides. Pas de comparaison possible	Distribution uniforme (données réparties uniformément sur toute l'amplitude de la série), normale (Gauss) avec concentration des données autour de la moyenne
Seuils naturels	Seuils observés	Observation de l'histogramme de fréquences cumulées triées dans l'ordre croissant	Prend en compte les discontinuités de la série Cartes non comparables	Toute distribution qui présente des discontinuités Distribution plurimodale
Quantiles	Chaque classe a le même nombre d'individus	Effectif total / nombre de classes Limites de classes = nb d'individus défini par le calcul ci-dessus, statistique d'ordre	Ne tiens pas compte des valeurs exceptionnelles. Des limites de classes peuvent être discutables (valeurs très proches réparties dans des classes différentes) Permet les comparaisons	Distribution uniforme ou autre distribution sans discontinuités À éviter si présence de valeurs extrêmes ou si trop grand nombre de valeurs proches
Box map	Classes basées sur l'analyse en boxplot ou boîte à moustaches	6 classes : quartile 1 (Q1), Q2, Q3, Q4, outliers inférieurs (Q1 - [1.5*(Q3-Q1)], outliers sups. (Q1 + [1.5*(Q3-Q1)])	Statistique d'ordre, permet les comparaisons, implémentée dans Geoda	Toute distribution, adéquat pour traiter des valeurs exceptionnelles

Cinq méthodes de discrétisation et leurs caractéristiques.

La mise en classe à partir de ces méthodes de discrétisation est également régie par les règles suivantes :

- Les classes doivent couvrir l'ensemble de la distribution statistique et doivent être contiguës.
- Une valeur ne doit appartenir qu'à une seule classe.
- Les classes ne peuvent pas être vides.
- Le choix des valeurs limites doit reposer sur des critères robustes.
- Des valeurs non significativement différentes ne doivent pas être placées dans des classes distinctes.

Le prochain exemple décrit la création d'une carte choroplèthe dans QGIS. L'exemple analyse la distribution spatiale du coefficient de Gini dans les comtés américains en 1989.

Le coefficient de Gini mesure l'inégalité des revenus au sein d'une population. Sa valeur varie entre 0 (égalité parfaite) et 1 (inégalité totale). Plus de détails sur ce coefficient sont disponibles dans l'article original de Corrado Gini.

Pour déterminer la méthode de discrimination adéquate, il est nécessaire de connaître la distribution statistique de la variable du coefficient de Gini. Pour cela, dans l'extension **Statist**, sélectionner la couche **comté** et indiquer la variable désirée (**GI 89**, ici) puis cliquer sur OK. Un graphique de la distribution et des statistiques apparaissent. Dans l'exemple, la statistique est presque normale, légèrement asymétrique et la moyenne est un petit peu supérieure à la médiane. La méthode des écarts à la moyenne est utilisée. L'étendue d'une classe est égale à l'écart-type. Pour utiliser cette méthode, double-cliquer sur la couche **comté** pour atteindre le menu des **propriétés**. Choisir **gradué** dans le menu déroulant et sélectionner la colonne **GI 89**. Choisir 5 classes et la palette de couleur prédéfinies : red, yellow, green. Cette palette est divergente et sa classe centrale est de couleur jaune clair. Inverser la palette afin que les inégalités élevées soient indiquées en rouge. Indiquer le mode de discrétisation (ici, **écart-type**). Cliquer sur **classer**. Les couleurs associées à chaque classe peuvent également être redéfinies à partir des couleurs obtenues sur l'outil **ColorBrewer** (<http://http://colorbrewer2.org/>). Lorsque toutes les couleurs ont été attribuées, cliquer sur OK.

Dans QGIS, la discrétisation de l'écart à la moyenne n'autorise que 5 ou 9 classes et la classe [0, 2\* écart-type] prend la valeur neutre jaune tandis que sa classe symétrique [-2\*écart-type, 0] prend une teinte vert clair. Dans ce cas-là, il serait nécessaire de faire intervenir un nombre pair de classes et une palette en deux teintes afin de représenter l'opposition entre ces deux classes.

Le résultat permet de mettre en évidence les inégalités dans les comtés du sud du Texas et le long du Mississippi.

## 21. HABILLAGE ET PRODUCTION DE CARTES THÉMATIQUES

Cette leçon a pour but d'expliquer le rôle de l'habillage des cartes et de présenter le composeur d'impression de QGIS.

### HABILLAGE CARTOGRAPHIQUE

Comme vu lors de la leçon 18, un signal constitué par une représentation cartographique doit être accompagné d'un certain nombre de composantes d'habillage de la carte. Certaines de ces composantes sont obligatoires lors de l'élaboration d'une carte tandis que d'autres seront utilisées selon le contexte de la carte. Une composante est jugée indispensable lorsque le but poursuivi par la carte et le contexte de sa publication implique son utilisation. Les composantes communément admises comme obligatoires sont :

- un titre et sous-titre ;
- l'année de production des données statistiques et spatiales ;
- la légende ;
- l'échelle graphique ;
- une indication de l'orientation (par défaut le nord est en haut de la carte) ;
- source des données ;
- nom des auteurs et copyright ;
- l'année de fabrication de la carte.

L'habillage complète la carte et permet à l'utilisateur de mieux la comprendre. Une carte thématique ne devrait donc pas être présentée sans son contexte géographique. En effet, tout territoire représenté est contigu à d'autres régions. L'information contenue dans ces régions contiguës comme les cours d'eau, les lacs, les chaînes de montagnes, les grandes villes ou les axes routiers et ferroviaires majeurs améliorent le repérage spatial et permettent une meilleure compréhension des réalités des situations contextuelles du phénomène étudié comme la situation socio-économique par exemple.

L'esthétique des documents cartographiques repose sur trois principes fondamentaux.

- L'*harmonie* est la relation entre les éléments de la carte. Elle permet de produire un signal homogène. Elle est l'équilibre visuel entre tous les éléments.
- La *composition* est l'arrangement des éléments sur la carte. Elle est l'organisation et la hiérarchisation des composantes de la carte en fonction de leur valeur informative.
- La *clarté* est la cohérence des choix des couleurs, la sobriété dans le choix des polices de caractères et l'optimisation de l'agencement des éléments de l'habillage dans l'espace disponible. Une clarté optimale aidera le lecteur dans le décodage du signal.

La phase de finalisation des aspects esthétiques est souvent faite par une retouche dans un logiciel d'édition graphique.

### LE COMPOSEUR D'IMPRESSION DE QGIS

L'exemple qui suit décrit étape par étape la création de la carte de la distribution spatiale du taux d'homicide dans les comtés américains. Cette création et habillage de la carte se fait dans le composeur d'impression de QGIS. Pour ouvrir un nouveau composeur d'impression, aller sous le menu déroulant **Projet >> Nouveau composeur d'impression**. Les différents composeurs d'impression peuvent être gérés en allant sur **Projet >> Gestionnaire de composition**. Après avoir donné un nom au nouveau composeur, l'interface graphique du composeur s'ouvre dans une nouvelle fenêtre. Pour faciliter l'utilisation du composeur, aller sous **Vue >> Barres d'outils** et cocher toutes les cases. Les fonctionnalités de la barre d'outils et les emplacements importants du composeur d'impression de QGIS sont présentés dans la figure 1.



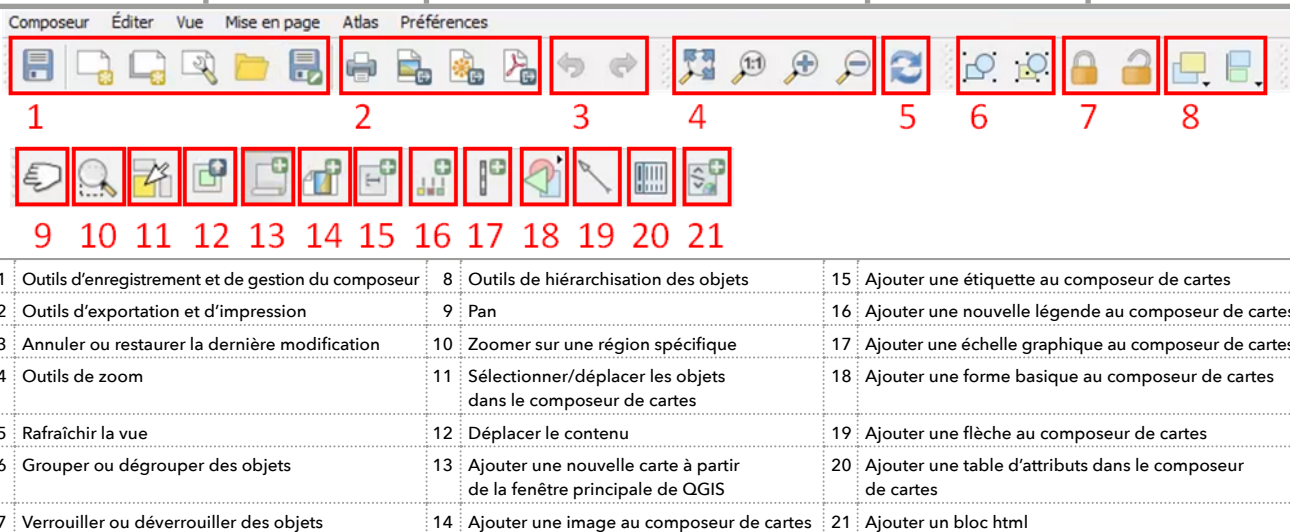


FIGURE 1

6:40

16:00

Barres d'outils du composeur d'impression.

À l'aide de l'outil **Ajouter une nouvelle carte**, ajouter une nouvelle carte. La mise en page de la carte se fait au travers du **panneau de propriétés de l'objet**. Les propriétés de l'objet permettent de définir l'emprise et l'échelle de la carte afin de définir l'étendue géographique représentée. L'étendue peut également être modifiée en utilisant l'outil **Déplacer le contenu de l'objet**. Un graticule peut être ajouté en choisissant, sous l'onglet **Graticules** du panneau de propriétés, le type de grille, l'unité, la taille des intervalles et le type de lignes. Toujours sous l'onglet **graticule**, il est possible d'ajouter un cadre et d'afficher les coordonnées sur les côtés désirés de la carte. L'onglet **Position et taille** permet de spécifier la position et la taille de la carte à l'intérieur du composeur d'impression. Pour cet exemple, la carte est centrée au milieu de la page et une marge de 4 mm de chaque côté est ajoutée. L'onglet **Cadre** permet d'ajouter un cadre autour de la carte, l'onglet **Fond** permet de définir la couleur du fond de carte et l'onglet **Rendu** décrit le mode de fusion et la transparence de la carte.

Maintenant que la carte principale a été mise en page, les éléments d'habillage doivent être mis en place. L'échelle peut être ajoutée en allant sous l'onglet **Mise en page >> Ajouter une échelle graphique** ou en cliquant directement sur l'icône de l'outil correspondant. Une échelle par défaut apparaît. Dans **Propriétés de l'objet**, il est possible de modifier le type d'échelle, les unités, la longueur affichée, la police de caractères de l'objet et sa taille. La légende peut être ajoutée en allant sous l'onglet **Mise en page >> Ajouter une légende** ou en cliquant directement sur l'icône de l'outil correspondant. Dans **Propriétés de l'objet**, il est possible de définir le titre. Sous **Objets de légende**, les différentes couches à afficher peuvent être sélectionnées en décochant la mise à jour automatique et en supprimant les couches non désirées. Il est également possible de redéfinir le nom d'une couche en cliquant sur le bouton **Modifier** (crayon). La police, la couleur du fond et la transparence sont également modifiables sous les autres onglets de propriétés. L'orientation, peut être affichée en utilisant l'outil **Ajouter une image** disponible sous l'onglet **Mise en page** et en traçant un rectangle sur la carte. Dans les propriétés de l'objet sous l'onglet **Recherche dans le répertoire**, sélectionner un symbole de flèche indiquant le nord. L'outil **Ajouter une étiquette**, permet d'ajouter des zones de textes (pour définir les noms de pays, d'océan, etc.). La police, la transparence, l'orientation et le fond peuvent être définis dans les **propriétés de l'objet**. Les étiquettes sont aussi utilisées pour ajouter un commentaire, le nom des auteurs, les sources et les références.

Pour créer un encadré représentant un zoom sur une région, il est nécessaire de verrouiller la carte de base. Pour cela dans les propriétés de la carte principale, sélectionner la case **verrouiller les couches pour cette carte**. Cette option permet que les modifications faites sur le projet QGIS ne soient pas répercutées sur la carte principale. Dans le projet QGIS, zoomer sur la zone d'intérêts. Dans le composeur d'impression, ajouter une nouvelle carte et placer là à l'endroit désiré.

Maintenant que la carte est terminée, il est utile de l'exporter. Pour cela, plusieurs options sont possibles. La carte peut être enregistrée en format image (jpeg, png, etc.) ou en format vectoriel (svg, pdf, etc.) retouchable dans des logiciels d'édition graphique. Avant l'enregistrement, il est important de spécifier, dans l'onglet composition, la taille du papier et la résolution de l'image lors de l'export.

## 22. CARTOGRAPHIE DYNAMIQUE ET INTERACTIVE SUR INTERNET

La cartographie sur internet est un thème d'actualité. En effet, de plus en plus d'applications informatiques du monde des systèmes d'information géographique s'appuient sur l'internet pour proposer des solutions multiplateformes (ordinateurs de bureau, tablettes, smartphones, etc.) sur différents systèmes d'exploitation (Windows, macOS, Linux, etc.). L'objectif de cette leçon est d'examiner l'influence d'internet sur l'organisation et le fonctionnement des SIG et d'explorer le thème du Web SIG. À partir de ces explications, il sera possible d'expliquer le principe des architectures client-serveur et de créer une page Web SIG simple.

### ARCHITECTURE CLIENT-SERVEUR ET SIG

Un *serveur* est un logiciel ou une machine capable d'écouter le réseau, de recevoir des requêtes émanant de différents clients et de répondre à ces requêtes. Son objectif est de partager des données ou des ressources matérielles et/ou logicielles entre clients dans une architecture dite client-serveur. Les serveurs peuvent être de type différent. Il existe des serveurs d'impression, de courriel, de base de données, des serveurs Web, etc.

Un *client* est un logiciel ou une machine capable de formuler des requêtes et de les envoyer à un serveur. Il récupère ensuite leurs réponses et les interprète pour affiner l'information transmise.

Un *interlogiciel* ou *Middleware* permet de gérer les communications entre clients et serveurs.

Les serveurs écoutent le trafic sur le réseau internet et identifient les requêtes qui leur sont adressées. Le *port logiciel* est le portail permettant cette écoute sur le réseau. Il est le point d'accès au système d'exploitation d'un ordinateur et est identifié par un numéro exprimé sur 16 bits. Il existe donc un maximum de 65536 ports par ordinateur. Plusieurs serveurs peuvent fonctionner simultanément sur une même machine en utilisant chacun un port spécifique. Les numéros de ports fréquemment utilisés sont :

- 20/21 pour l'échange de fichier via FTP ;
- 80 pour la consultation d'un serveur http ;
- 443 pour la consultation d'un serveur HTTPS ;
- 5432 pour un serveur de base de données PostgreSQL.

Un logiciel serveur peut également être hébergé sur une machine locale en même temps que les logiciels clients pour autant que la communication client-serveur passe par les mêmes ports logiciels qu'utiliserait ce serveur depuis une machine distante. Le terme *localhost* est le nom habituel désignant une interface logique sur un ordinateur local. Il est également associé à l'adresse IPv6::1 et à la plage d'adresses IPv4 127.0.0.0/8 (adresses comprises entre 127.0.01 et 127.255.255.255). Pour accéder un port particulier par une adresse URL il faut ajouter deux points et le numéro du port. Ainsi, pour accéder le port 3000 sur une machine locale, l'adresse URL est `http://localhost:3000`.

Une architecture client-serveur est donc généralement constituée d'un serveur, d'un client et d'une communication client-serveur (figure 1).



FIGURE 1

Dans le contexte des SIG (figure 2), le serveur est souvent constitué d'un serveur de données ou de deux serveurs en série. L'un pour la gestion des requêtes et des réponses et l'autre pour la fourniture de données. Le client peut être composé d'un logiciel SIG (QGIS, Manifold, ArcGIS, etc.) ou d'un Web SIG (navigateur Web comprenant un composant cartographique). Les communications client-serveur sont gérées par des protocoles de communication. Le Middleware le plus souvent utilisé est Open Data Base Connectivity (ODBC). Ce logiciel est capable de transmettre au serveur les requêtes SQL formulées par le client. La plupart des logiciels de gestion de base de données possèdent des pilotes ODBC pour permettre la connexion aux bases de données utilisant ce protocole. C'est le cas de QGIS.

OLEDB (Object Linking and Embedding Database) et ADO.NET sont deux autres Middleware fréquemment utilisés dans le monde Windows.

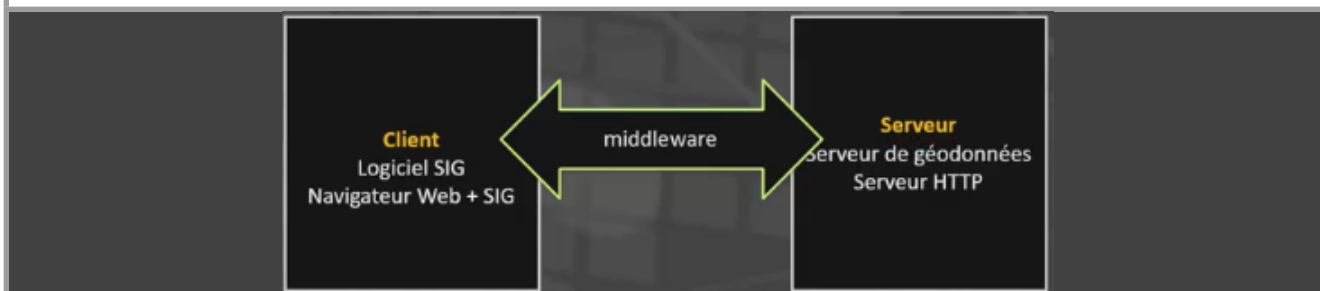


FIGURE 2

6:34

42:00

Architecture client-serveur dans le contexte des SIG.

Le protocole http est un cas particulier d'architecture client-serveur (figure 3). Le côté serveur est constitué d'un serveur http ou de géoservices de type WMS (Web Map Service) et WFS (Web Feature Service). Le côté client est constitué d'un Web SIG ou d'un logiciel SIG. La communication client-serveur s'appuie sur le protocole http.



FIGURE 3

9:22

42:00

Architecture client-serveur dans le cas d'un protocole http.

## PAGE WEB – HTML ET CSS

La compréhension des fondements de l'organisation et de la construction d'une page web est un prérequis pour l'intégration de composants géographiques dans une page web.

Une page Web est un document HTML (HyperText Markup Language) soit un simple fichier texte annoté afin de définir divers types de contenus (en-têtes, paragraphes, titres, liens, images, etc.). Les annotations (Markup) sont formées par des balises (tag) c'est-à-dire des mots-clés placés entre chevrons et signalant le début (< >) et la fin (</>) d'un élément HTML.

Les documents HTML (figure 4) débutent et se terminent par des balises de début et de fin (<html></html>) suivi par une section d'en-tête (<head></head>) et une section de contenu (<body></body>). L'html n'est pas sensible à la casse, mais l'utilisation de la minuscule est le standard recommandé.

L'en-tête contient des métadonnées, des références et le titre de la page (<title></title>). Les éléments html apparaissant dans la zone de contenu sont placés entre les balises *body*.

Voici une liste des balises les plus courantes :

- <h1>...<h6>, titres de niveau 1 à 6
- <p></p>, paragraphes de texte
- <em></em>, marquer un élément de texte (en italique par défaut)
- <strong></strong>, marquer fortement un élément de texte (en gras par défaut)
- <br>, saut de ligne
- <hr>, ligne horizontale
- <a>, lien hypertexte vers un site web ou vers un élément de la page défini par son id (voir figure 4).
- <ol></ol>, listes ordonnées (listes numérotées)
- <ul></ul>, listes non ordonnées (listes à puces)
- <li></li>, éléments de listes
- <img>, pour inclure une image. Cette balise est particulière car elle ne nécessite pas de balise de fin. Elle possède deux attributs obligatoires et leurs valeurs doivent être écrites entre guillemets simples ou doubles. Les deux attributs obligatoires sont :
  - src, le chemin d'accès à la source, sous forme d'URL ou de chemin local (relatif)
  - alt, un texte alternatif affiché lorsque la source n'est pas disponible.
 Exemple : <img src = "sources.jpg" alt="Photo des sources" width = 220 />  
 La définition du chemin d'accès suit la règle standard.
  - Si la source est placée dans le même dossier que le fichier HTML alors le chemin sera du type src = "sources.jpg" ou src = "./sources.jpg".
  - Si la source est placée dans un dossier images au même niveau que le fichier HTML alors le chemin sera du type src = "./images/sources.jpg".
  - Si la source est placée dans un dossier images au même niveau que le fichier HTML contenant le fichier html alors le chemin sera du type src = "../images/sources.jpg".
  - Si la source est hébergée sur un site web alors le chemin sera du type src = "http://www.mysite.ch/OrontesImg/sources.jpg".
- <table></table>, définir une table. La table est composée de :
  - <tr></tr>, lignes (table row). Ces lignes contiennent des cellules (table data) définie par <td></td>.

La liste complète de toutes les balises est disponible sous le site web : <http://www.w3schools.com/html>.

```
<a href="http://www.epfl.ch"> Site Web de l'EPFL </a><br>
<br>
```

[Site Web de l'EPFL](http://www.epfl.ch)

```
<a id="top">This is the top</a><br>
<a href="#top" > Back to top </a>
```

This is the top  
[Back to top](#)

FIGURE 4

Les fichiers html contiennent également des éléments neutres (figure 5) définis par les balises `<div></div>` ou `<span></span>`. Ces éléments sont des conteneurs pour d'autres éléments html et servent à structurer et styliser le document. L'élément `div` est encadré par des sauts de lignes (éléments blocs) et l'élément `span` ne possède pas de formatage par défaut (éléments en lignes).

```
<h1>My <span style="color:red">Important</span>Heading</h1>

<p>This is some text.</p>
<div style="color:#0000FF;background:lightgrey">
  <h3>This is a heading in a div element</h3>
  <p>This is some text in a div element</p>
</div>
<p>This is some text.</p>
```

My Important Heading

This is some text.

This is some text in a div element

This is some text in a div element

This is some text.

FIGURE 5

16:40

42:00

Exemples d'utilisation des balises `<div>` et `<span>`.

Les éléments html possèdent un attribut `style` dont la forme est définie comme *nom:valeur* et séparé par des points-virgules. Bien qu'il puisse être intéressant de définir les éléments de styles directement comme attributs des éléments html, cette habitude comporte des inconvénients. En effet, la mise à jour des éléments stylistiques est difficile, la lisibilité de la page html se dégrade et l'augmentation de la taille des pages engendre un temps de chargement plus long.

Pour éviter ces désagréments, les éléments de style sont regroupés dans une section particulière définie par la balise `<style></style>`. Les styles sont ensuite assignés à des objets, par l'utilisation du sigle `#`, ou à une classe d'objets en utilisant le point. Il est aussi possible d'assigner un style à tous les objets html d'un certain type sans distinction. Cette section particulière de style peut aussi être sortie du fichier html principal pour former un fichier de style en `.css` (cascading style sheet) qui pourra être appliqué à l'ensemble des pages html du site web. Cette page de style est référencée dans la page html en utilisant la balise `<link></link>` (voir figure 6).

Les éléments de style sont donc organisés en référençant les objets (html, «id» ou «class») et en regroupant les éléments de style dans une section de la page html ou dans un fichier de style séparé. La hiérarchie d'application des styles est importante. Un style défini localement (comme attribut d'un objet) aura la priorité sur les styles plus généraux.

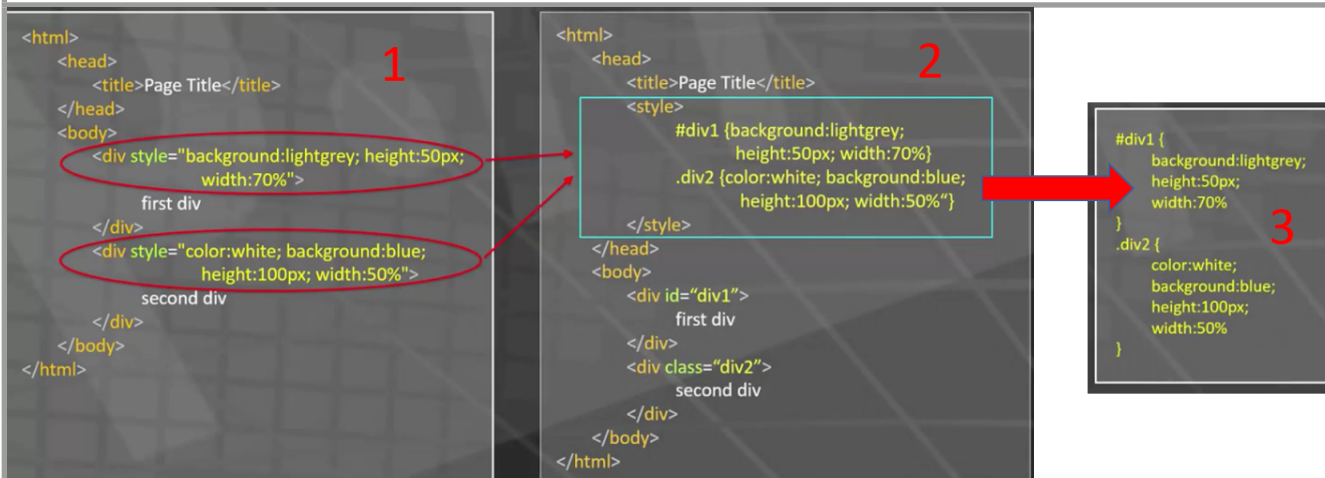


FIGURE 6

19:55

42:00

Étapes de rationalisation des styles. 1) Éléments de style définis au niveau des objets. 2) Éléments de styles définis dans une section de l'en-tête. 3) Éléments de style définis dans un fichier css.



## WEB SIG

Le langage html est limité et son manque de complexité ne lui permet pas d'intégrer correctement des éléments de cartographie. Plusieurs possibilités existent pour compléter le langage html du côté client.

- En ajoutant du code exécutable dans la page html (JavaScript, vbscript, dhtml).
- En utilisant du code exécutable à télécharger et utiliser avec une page html (scripts, ActiveX, java applet).
- En utilisant des interfaces-applications spécifiques à installer sur le client (plug-in).

Il est également possible d'améliorer l'architecture du côté serveur.

- En utilisant des .exe, .dll ou des scripts exécutés sur le serveur. Cela permet de créer dynamiquement des pages html (Common Gateway interface, CGI).
- En utilisant du code exécutable au niveau du serveur (PHP, java servlet, ASP-active serveur page, ColdFusion).

Ce cours présente uniquement le JavaScript. JavaScript est un langage fonctionnel (les fonctions sont considérées comme des objets). Il permet d'accéder aux éléments de la page html en utilisant le Document Object Model (DOM) et d'accéder les éléments du navigateur en utilisant le Browser Object Model (BOM).

Dans l'exemple simple qui suit (figure 7), l'implémentation d'un bouton dont l'actionnement fait apparaître la date et l'heure actuelle est présentée. La syntaxe du fichier html ne comporte pas d'en-tête et le corps contient un titre, le bouton et un paragraphe permettant de contenir le texte de la date et l'heure générée par la fonction. Les propriétés de l'objet bouton sont le *type* et l'événement *onclick* dont l'attribut est *myFunction()*. Cette fonction est définie entre les balises *script*. Elle utilise l'objet dont l'identifiant est *démo* et assigne la date courante au contenu html de cet objet.

Comme pour les styles, le code des fonctions JavaScript du code html peut être regroupé dans un fichier de type .js. Ce fichier est ensuite appelé dans le code html en utilisant la balise *script* et en indiquant le chemin de la source du fichier JavaScript.

Dans le cadre des SIG, de nombreuses bibliothèques open source JavaScript existent :

- Openlayers (<http://openlayers.org/>)
- Leaflet (<http://leafletjs.com/>)
- Mapstraction (<http://mapstraction.com/>)
- Polymaps (<http://polymaps.org>)
- Cesium (3D-Globes) sous <http://cesiumjs.org/>
- Modest Maps, JQuery Geo, etc.

Il existe également des bibliothèques non open source tels que Google Maps, Bing Maps, MapQuest, Esri ou encore Via Michelin.

```
<html>
  <head>
    <script>
      function myFunction() {
        document.getElementById("demo").innerHTML = Date();
      }
    </script>
  </head>
  <body>
    <h4>This is a JavaScript example</h4>
    <button type="button"
      onclick="myFunction()">
      The time is?
    </button>
    <p id="demo"></p>
  </body>
</html>
```

FIGURE 7



## EXEMPLE D'APPLICATION

Cet exemple présente l'utilisation de la librairie cartographique Openlayers 3 pour l'élaboration d'une page html contenant une carte géographique interactive de l'île de Mahé aux Seychelles. Pour débiter, créer un fichier intitulé *Openlayers\_Seychelles.html* (figure 8.1). Ajouter les balises de base c'est-à-dire le type de document (Doctype), l'en-tête (head) et le corps (body). Dans l'en-tête, ajouter un titre et deux références. La première lie la librairie Openlayers et la seconde pointe vers le fichier CSS comprenant le style par défaut d'Openlayers. Dans le corps (body) de la page, ajouter un conteneur (div) identifié par l'id « map ». Ensuite, ouvrir une section style dans l'en-tête et y insérer premièrement un style indiquant la volonté que la page html et le corps occupent 100% de la largeur et 100% de la hauteur et deuxièmement un style pour l'objet « map » auquel on attribue 100% de la hauteur et une couleur de fond gris clair. Ajouter une section « script » dans le corps pour y écrire le code JavaScript. Ce code contient la définition d'une nouvelle variable « map » d'Openlayers (ol.Map). Indiquer ensuite dans le code la cible (target) de cette variable. Ici la cible est le conteneur dont l'id est « map ». Créer une nouvelle variable indiquant la vue de la map (view) centrée sur la latitude et longitude (55.47, -4.67) et avec un zoom de niveau 12. Pour l'affichage de cette map, il est nécessaire d'ajouter une couche. Pour cela, créer une nouvelle couche de type tuile et de source OSM puis ajouter cette couche à la variable map. Il est maintenant possible de visualiser cette carte dans un navigateur.

Pour améliorer la carte et ajouter des éléments spécifiques, il convient de déplacer le code JavaScript contenu dans la page html vers un nouveau fichier nommé *Openlayers\_Seychelles.js* (figure 8.2). Le nouveau fichier doit être référencé dans l'en-tête du fichier html. Dans le nouveau fichier JavaScript, séparer la définition de la variable map et définir le reste du code comme une fonction exécutée au chargement de la page et permettant l'affichage de la carte.

L'ajout des districts des Seychelles stocké dans un fichier de type GeoJSON se fait en créant une nouvelle variable de couche vecteur définie par un style. Indiquer le chemin d'accès des données et ajouter la nouvelle variable à la couche. Dans le début du fichier, définir un style appliqué à la variable districts. Le style indique que l'objet sera rempli d'une certaine couleur avec un degré de transparence et que sa bordure aura une certaine couleur et transparence. D'autres couches peuvent être ajoutées de la même façon (hôtels, routes, etc.). Il est également possible, moyennant l'ajout d'un code plus complexe, de faire intervenir une fonction permettant d'afficher les attributs des couches dans une fenêtre « popup » à l'action d'un clic sur un district.

Toutefois, pour que ces districts soient visibles, il est nécessaire de rendre le fichier *district.json* accessible sur un serveur. Un serveur statique est suffisant et peut être créé en utilisant la librairie *Node.js*. Pour cela, créer un fichier nommé *serveur.js* (figure 8.3) dans lequel quelques lignes de codes suffisent à créer le serveur. Dans ce fichier, Node Express est utilisé, le dossier public est défini comme l'endroit où les fichiers seront accessibles et le port d'écoute du serveur est défini (port 3000). Les fichiers qui ont donc été créés et utilisés pour le fonctionnement de la page web sont le dossier GeoJSON contenant les fichiers des districts, des hôtels et des routes. Le dossier des modules Nodes, les fichiers *Openlayers\_Seychelles.html*, *Openlayers.Seychelles.js* et *serveur.js*.

Pour démarrer le serveur utiliser un invite de commande et aller jusqu'au répertoire de *Openlayers\_Sample* (dossier comprenant les fichiers précités.) et taper *node serveur.js*.

Le serveur est maintenant en route. En ouvrant un navigateur Web et en tapant l'adresse *localhost3000/Oplayers\_Seychelles.html*, le code javascript de chargement de la page va s'exécuter et charger les districts, les routes et les hôtels sur la carte.

```

1 <!doctype html>
2 <head>
3   <script src="http://openlayers.org/en/v3.10.1/build/ol.js" type="text/javascript"></script>
4   <link rel="stylesheet" href="http://openlayers.org/en/v3.10.1/css/ol.css" type="text/css">
5   <script type="text/javascript" src="/Openlayers_Seychelles.js"></script>
6   <style>
7     html,body{
8       height:100%;
9       width:100%;
10    }
11    #map {
12      height:100%;
13      background-color: lightgrey;
14    }
15  </style>
16  <title>Openlayer Seychelles</title>
17 </head>
18 <body>
19   <div id="map">
20   </div>
21 </body>

```

1

```

1 var express = require('express');
2 var app = express();
3
4 app.use(express.static(__dirname));
5 app.listen(process.env.PORT || 3000);
6 console.log('server up and running. Listening on port 3000')

```

3

```

1 var map;
2
3 //Définition de quelques styles pour les couches vecteur
4 var District_Style = new ol.style.Style({
5   fill: new ol.style.Fill({ color: 'rgba(200, 10, 10, 0.2)' }),
6   stroke: new ol.style.Stroke({ color: 'rgba(255,0,0,1)', width: 1 })
7 });
8
9 //Script executé lorsque la page est chargée
10
11 $(document).ready(function(){
12
13   map = new ol.Map({
14     target: 'map',
15     view: new ol.View({
16       center: ol.proj.fromLonLat([55.47, -4.67]),
17       zoom: 12
18     })
19   });
20
21 // create and add the tile layer
22 var osmlayer = new ol.layer.Tile({
23   source: new ol.source.OSM()
24 });
25 map.addLayer(osmlayer);
26
27 //Ajout de fichiers geojson
28 var vector = new ol.layer.Vector({
29   style: District_Style,
30   source: new ol.source.Vector({
31     url: './geojson/districts.geojson',
32     format: new ol.format.GeoJSON(),
33   })
34 });
35 map.addLayer(vector);
36
37
38 });

```

2

FIGURE 8

35:30

42:00

Syntaxes des différents fichiers nécessaires à l'élaboration de la carte interactive de l'exercice sur les Seychelles. 1) Syntaxe du code html. 2) Syntaxe du code JavaScript pour la map interactive. 3) Syntaxe JavaScript pour le serveur.



## 23. AUTRES FORMES DE REPRÉSENTATION : LA TROISIÈME DIMENSION

Cette leçon porte sur la représentation de l'information géographique en trois dimensions. Elle permettra de d'acquérir des notions de base dans la manipulation et dans le rendu des données en trois dimensions. Les buts de cette leçon sont de faire un rappel des différentes techniques permettant de visualiser le relief et d'expliquer comment construire un modèle 3D à partir d'un modèle numérique de terrain. L'utilisation et la manipulation du plug-in Qgis2threejs seront aussi présentées par des cas pratiques.

### FORME DE REPRÉSENTATION DU RELIEF

La représentation géographique incluant le rendu d'une variable d'altitude permet de représenter l'environnement de manière plus concrète et plus réaliste. La vue en trois dimensions est également plus attractive pour l'utilisateur et lui donne un avantage dans son utilisation pour les études d'impacts ou les mises à l'enquête face à la représentation 2D. Le rendu 3D du volume des objets favorise une meilleure compréhension et perception des relations entre les objets composant le paysage. La troisième dimension permet donc de visualiser le relief et un attribut thématique de manière simultanée.

Dans le cadre de la leçon dédiée aux modèles numériques de terrain, plusieurs techniques de représentation du relief ont été évoquées. Ces représentations sont appelées 2.5D, 2D et demie ou pseudo 3D. Les techniques pour la représentation du relief sont :

- faire varier la valeur d'une teinte des pixels d'un MNT ;
- les courbes de niveau ;
- les modèles ombrés ;
- une combinaison des modèles ci-dessus.

### CONSTRUCTION D'UN MODÈLE 3D

Le *drapage* est une technique permettant de caler une image sur un modèle 3D afin de représenter le relief en trois dimensions. La modélisation 3D NURBS (Non Uniform Rational Basic Spline) est construite à partir d'un MNT de type raster où chaque pixel est caractérisé par ses coordonnées X, Y et Z. Les pixels sont élevés à leurs altitudes Z puis une surface est construite à l'aide des courbes de Bézier reliant les points centraux de chaque pixel. L'information est parfois ré-échantillonnée afin d'ajuster la résolution du modèle numérique de terrain. Pour finir, un drapage est effectué. L'étape de drapage engendre souvent la déformation des pixels de l'image lorsque la différence d'altitude est grande sur la surface.

### QGIS2THREEJS : EXEMPLE DES SEYCHELLES

Le plugin QGIS2threejs utilise la librairie JavaScript threejs et permet de visualiser les éléments 3D dans QGIS. Dans l'exemple qui suit, la mise en œuvre du plugin utilise le modèle numérique de terrain de l'île de Mahé ainsi qu'une image aérienne en rouge vert et bleu. Dans QGIS, installer le plugin en allant sous le menu **Extensions >> Installer/Gérer les extensions** puis chercher et installer l'extension **Qgis2threejs**. Lorsque l'installation est terminée, l'icône correspondant au plugin apparaît (figure 1). Cliquer sur l'icône. La fenêtre des paramètres du plugin apparaît. L'onglet **World**, permet de définir l'emprise du modèle (ici toute l'île). L'onglet **Controls** donne la possibilité de définir les combinaisons de touches du clavier et de la souris pour la navigation dans l'image générée par le modèle. L'onglet **DEM** (Digital Elevation Model) permet de désigner la couche contenant le MNT à utiliser. Dans cet exemple, le MNT de l'île de Mahé est utilisé. Dans le même onglet, la résolution du modèle peut être ajustée. Une résolution trop élevée engendrera de la lenteur dans le traitement de l'information et une résolution trop faible ne permettra pas de faire ressortir les détails du MNT et du drapage. Il est donc conseillé d'utiliser la fonction avancée **quadtree** permettant d'ajuster la résolution en fonction du niveau de zoom. Le nombre de niveaux de hauteur de quadtree est défini sur 4 pour cet exemple. Toujours dans le même onglet, choisir le point sur la carte pour lequel le plus de détail est désiré et sélectionner l'image à draper (ici, l'image satellite de l'île de Mahé). Les onglets suivants permettent de gérer plus de MNT et d'ajouter des objets vectoriels. Cliquer sur Run. Une nouvelle fenêtre s'ouvre dans le navigateur Web et il est possible de naviguer dans le modèle 3D.

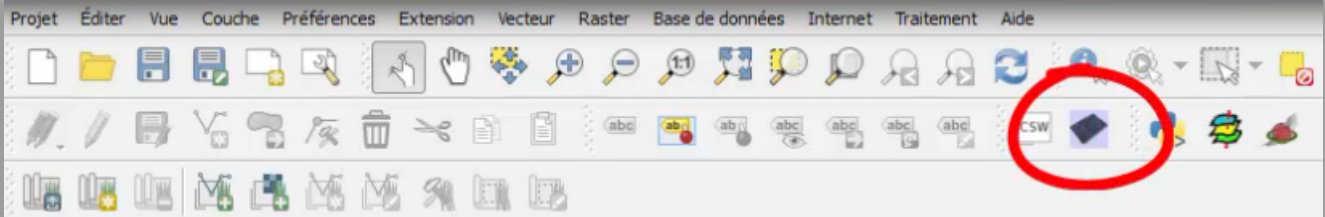


FIGURE 1

6:50

11:00

Icône de l'extension QGIS2THREEJS dans la barre d'outil de QGIS.

### QGIS2THREEJS: EXEMPLE DE GENÈVE

La modélisation 3D peut également être accompagnée d'objets vectoriels. Dans l'exemple proposé sur la modélisation se base sur les données du MNT d'un quartier de la ville de Genève, d'une image aérienne (RVB) de 40 cm de résolution spatiale et d'une couche vectorielle de l'emprise au sol des bâtiments. La couche vectorielle contient un attribut de hauteur des bâtiments. Comme pour l'exemple de l'île de Mahé, ouvrir la fenêtre des paramètres du plugin QGIS2threejs. Sélectionner le MNT désiré et définir le ré-échantillonnage sous **simple** (la zone d'étude est petite et le calcul n'est donc pas long). Dans l'onglet **Polygon**, sélectionner la couche vectorielle de l'emprise au sol des bâtiments et indiquer l'attribut correspondant à la hauteur. Cliquer sur Run. Le résultat s'ouvre dans une page Web. Il est maintenant possible de naviguer dans le modèle 3D du quartier de Genève et de ses bâtiments.



Modèle 3D d'un quartier de Genève produit dans QGIS avec l'extension QGIS2threeJS.

## 24. AUTRES FORMES DE REPRÉSENTATION : SIG ET RÉALITÉ AUGMENTÉE

La réalité augmentée est une technologie informatique permettant de superposer en temps réel un modèle virtuel (2D ou 3D) à une perception de la réalité. Combiner les SIG et la réalité augmentée offre des perspectives intéressantes. Les buts de cette leçon sont d'expliquer la technologie de la réalité augmentée (RA) et de présenter des exemples d'intégration de RA avec les SIG.

### LA RÉALITÉ AUGMENTÉE (RA)

D'après Jens Ingensand, professeur de géoinformatique à l'institut INSIT de la Haute Ecole d'Informatique et de Gestion (HEIG) du canton de Vaud, la réalité augmentée est une technologie permettant de superposer des informations virtuelles à la réalité. Cette superposition est très souvent liée à l'utilisation d'un appareil (tablette, etc.) dont la position et la zone visible sont connues.

En 1901, l'écrivain américain Lyman Frank Bauman imagine déjà, dans le livre *Le Magicien d'Oz*, des lunettes permettant à son porteur de voir apparaître sur le front des gens une lettre correspondant à leurs caractères. En 1960, Morgon Heilig invente le Sensorama permettant au spectateur d'un film de s'immerger dans un environnement non seulement visuel, mais aussi olfactif et sonore. Il s'agit donc d'une superposition d'effets sensoriels synchronisés avec une image. Dès les années 1970, avec l'arrivée des premiers microprocesseurs intégrés, les techniques de traitement d'image débutent et conduisent au développement de logiciel rendant possible l'interaction avec des objets virtuels. Un tournant à lieu lorsqu'en 1980 sort le processeur Motorola 68000 à 32 bits.

Le principe de fonctionnement de la réalité augmentée est décrit par Krueger en 1972 (figure 1).

- Capture d'une image de la réalité (1).
- Les images sont collectées en temps réel par un logiciel de réalité augmentée (2).
- Le logiciel détecte des points d'accroche prédéfinis (points GPS, par exemple) (3).
- Le logiciel sélectionne l'information associée au point d'accroche (4).
- L'image réelle est augmentée de cette information (5).
- L'image est projetée sur un écran (lunettes, tablettes, smartphone, etc.) (6).
- L'algorithme tourne en boucle, en temps réel et en interactivité.

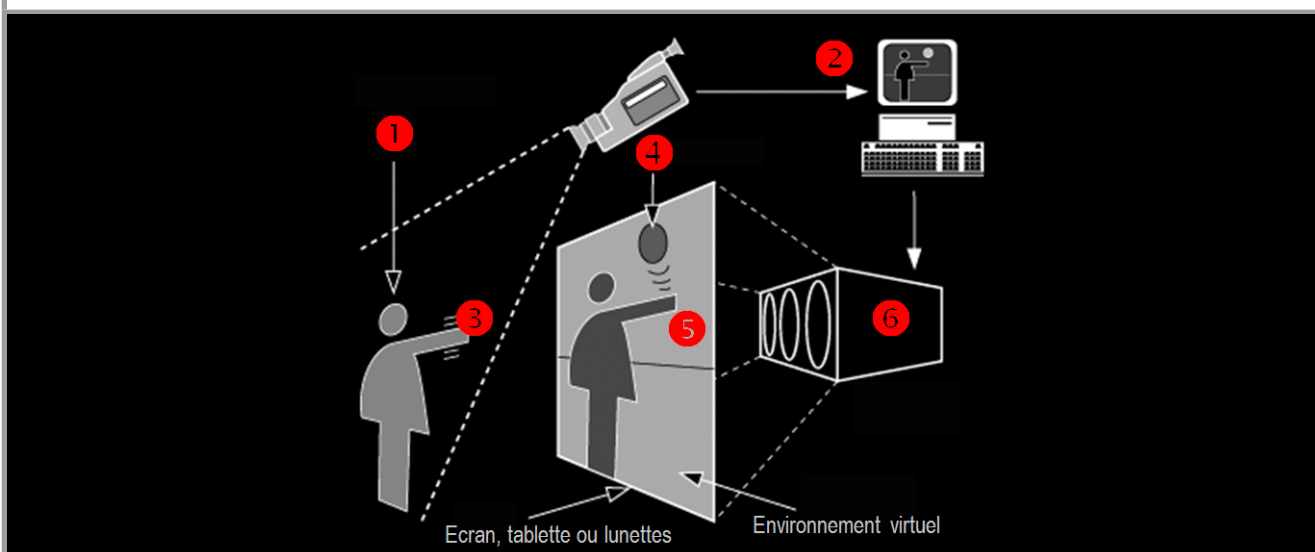


FIGURE 1

4:14

14:00



### COUPLAGE DE LA RÉALITÉ AUGMENTÉE ET DES SIG

Les SIG et la réalité augmentée sont relativement complémentaires et permettent, lorsqu'ils sont combinés, d'apporter une autre vision de l'information géographique. De manière générale, les données utilisées dans un SIG sont en deux dimensions et le lien entre la carte bidimensionnelle et la réalité est fait par l'utilisateur. Mais dans le cas de la RA, le lien est direct et ne nécessite aucune interprétation de l'utilisateur puisque les informations sont simultanément superposées à la réalité. La RA permet aussi de démocratiser les SIG à un plus large public. En effet, les tablettes, smartphones et autres appareils adaptés à l'association des SIG et de la RA rendent les données géographiques accessibles à tous.

### EXEMPLES D'APPLICATIONS INTÉGRÉES RA ET SIG

Le premier exemple d'association des SIG et de la RA est le projet Biosentiers de la HEIG-VD dont le but est de sensibiliser les utilisateurs à la biodiversité urbaine présente et de leur permettre de saisir de nouvelles informations.

Le projet Biosentier est fondé sur une application de RA sur smartphone. Sur un chemin défini entre la gare d'Yverdon-les-Bains et la réserve naturelle du centre Champ-Pittet, les utilisateurs peuvent utiliser l'application pour voir les différentes espèces (plantes, oiseau, insectes, etc.) observées par d'autres utilisateurs le long de ce sentier (figure 2). Un descriptif pour chaque espèce est disponible directement. Ce sentier permet également la prise de conscience de l'augmentation de la biodiversité entre la gare et la réserve naturelle.



FIGURE 2

8:20

14:00

Exemple d'interface de RA pour le projet Biosentier.

Le deuxième exemple présente une démarche didactique d'une interface tangible. Le bac à sable à réalité augmentée ou Augmented Reality Sandbox permet d'expliquer la notion de base de la topographie en générant à la main des cartes topographiques avec des courbes de niveau. La Sandbox permet également un apprentissage facilité des sciences de la terre en permettant la visualisation des processus se produisant habituellement sur de grandes échelles spatiales et temporelles. L'installation est composée d'un bac à sable de 100 x 75 x 20 cm remplis de 50 dm<sup>3</sup> de sables fins blancs. Une caméra *Kinect for Xbox* et un beamer sont suspendus 1 m en dessus du bac (figure 3).



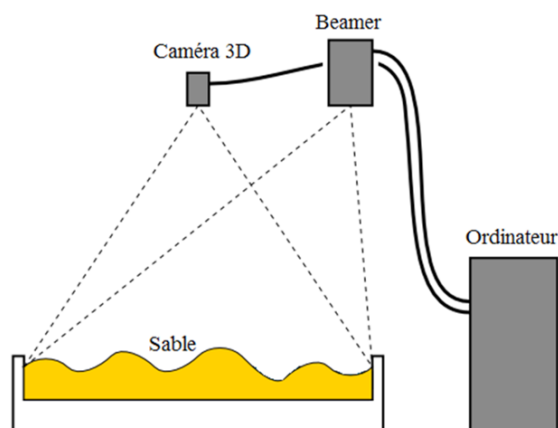


FIGURE 3

9:35

14:00

Schéma descriptif de la Sandbox.

La Kinect est composée d'une caméra couleur et d'une caméra infrarouge de profondeur. La géométrie de la surface est reconstituée grâce à la caméra infrarouge et par une approche de lumière structurée. La réalité augmentée utilisée pour la Sandbox se base sur une boucle fermée constituée de plusieurs étapes clés. Premièrement, la Kinect reçoit une matrice de distance brute avec une fréquence de 30 images par seconde. Ces données passent dans un filtre statistique dont le but est d'identifier les objets en mouvements (les mains des utilisateurs, sable qui bouge, etc.), de réduire les perturbations du signal et de compenser les données manquantes provenant des données de profondeur de la Kinect. La surface topographique est ensuite projetée sur le sable de manière à correspondre à celle façonnée dans la Sandbox. Le logiciel utilise une palette colorimétrique spectrale. Le bleu indique les zones basses et le vert et rouge pour les zones plus élevées et le blanc pour les points culminants (figure 4).

La Sandbox est utilisée pour l'enseignement des sciences de la terre. Les trois expériences disponibles pour la Sandbox sont :

- La reconstitution d'un relief existant.
- La modélisation de l'évolution historique d'un paysage par comparaison du relief.
- L'analyse de bassins-versants par une modélisation du ruissellement basé sur un modèle hydrologique et la simulation des fluides.

Elle est également utilisée dans des activités de recherche comme l'intégration de différents modes d'interaction ou dans l'étude de l'optimisation de la visualisation dans le cadre de la sémiologie graphique.

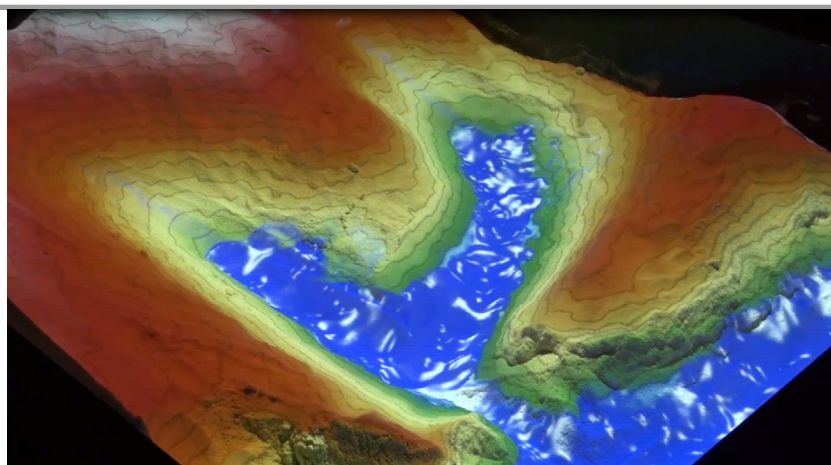


FIGURE 4

11:00

14:00

Sandbox en cours d'utilisation.



**BOOCs**  
**EPFL**

Systèmes d'Information  
Géographique – Partie 2  
Stéphane Joost, Marc Soutter,  
Fernand Kouamé et Amadou Sall



## IMPRESSUM

© EPFL Press, 2017.  
Tous droits réservés.

Graphisme :  
Emphase Sàrl, Lausanne

Résumé : Yvan Chevalley

Développés par EPFL Press, les BOOCs (Book and Open Online Courses) sont le support compagnon des MOOCs proposés par l'École polytechnique fédérale de Lausanne. Valeur ajoutée aux MOOCs, ils rassemblent l'essentiel à retenir pour l'obtention du certificat et constituent un atout pédagogique. Learn faster, learn better. Bonne révision !

ISBN 978-2-88914-424-2