

# Des images satellites aux estimations de population

Des outils d'analyse spatiale au service de la recherche en démographie

Léo Lipovac  
Lisa Astruc

Entreprise : Diginove

Encadrants : Nicolas Pech et Valérie Golaz

# Plan



Introduction : de la détection du bâti à l'estimation de population



Principales méthodes mises en œuvre jusqu'à présent



Matériel et méthodes



Conclusion et perspectives de travail



# Introduction : de la détection de bâti à l'estimation de population

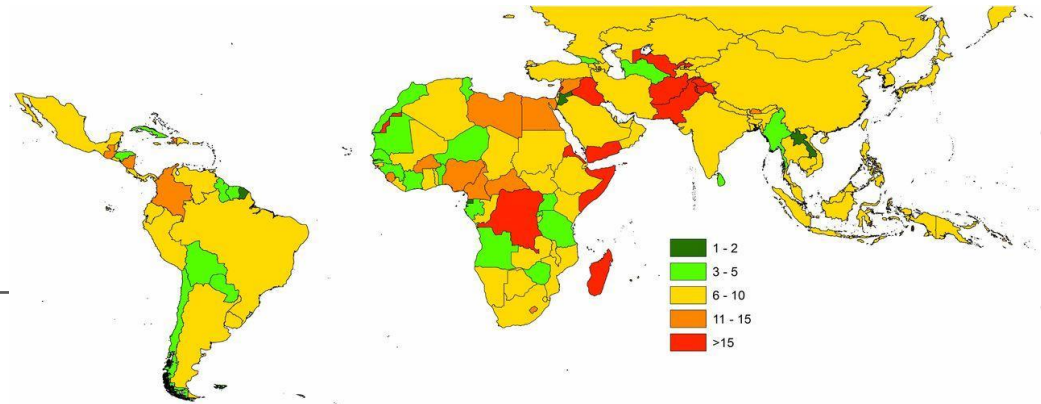
# Introduction : de la détection du bâti à l'estimation de population



- La population du continent africain :
  - 1950 : 230 millions d'habitants → 2020 : 1,3 milliard d'habitants
- Importance des cartes de population :
  - Enjeux de santé : campagnes de vaccination, pour les moustiquaires
  - Culturels
  - Économiques
  - Diminuer les inégalités



- Les recensements en Afrique :
  - Généralement anciens (moyenne ~ 10 ans)
  - Disponibles à un niveau géographique élevé - disparités spatiales



- Problématique : comment estimer la population des pays africains à l'échelle la plus fine possible ?



# Introduction : de la détection du bâti à l'estimation de population



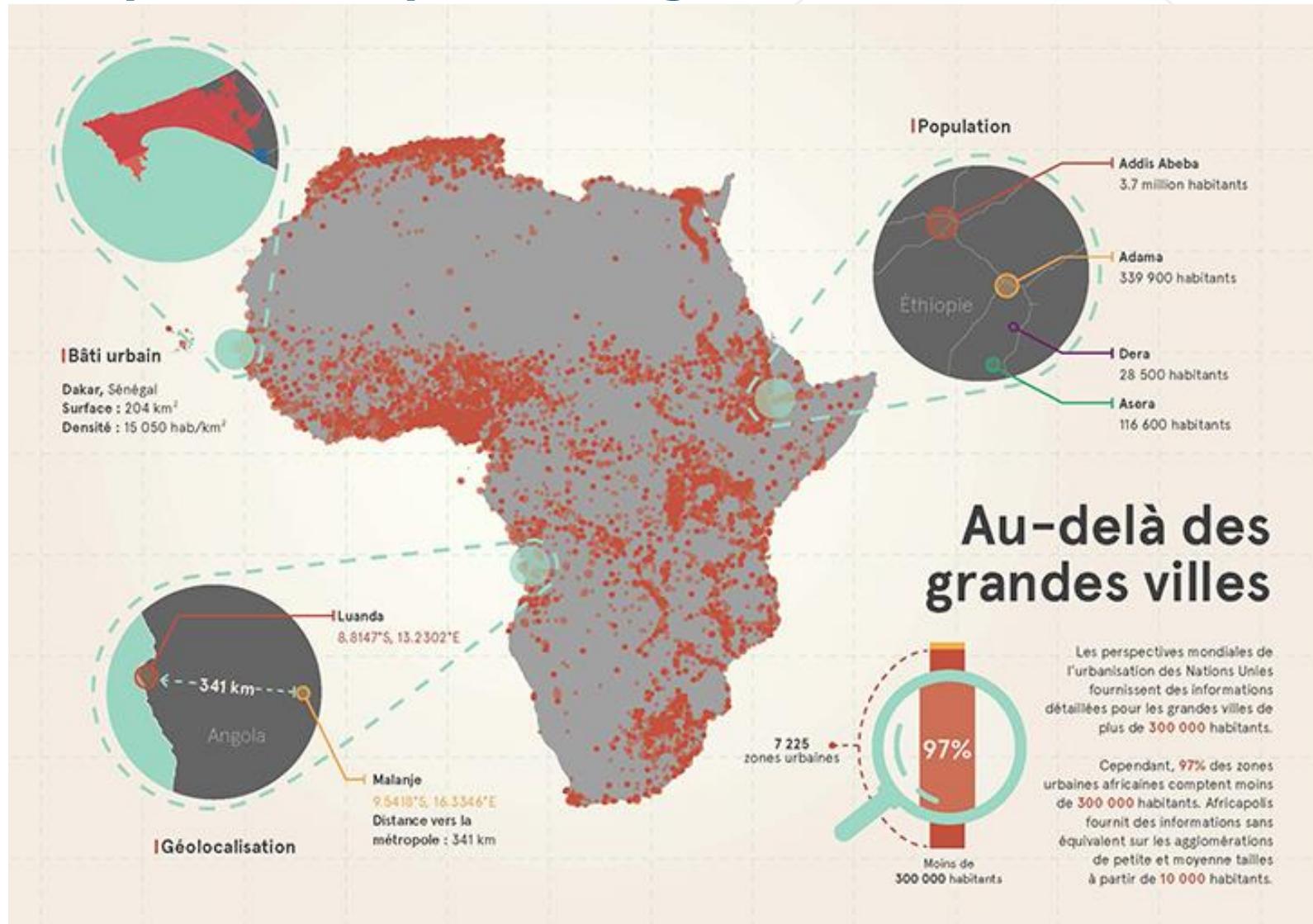
- Objectif : estimation de la population à l'échelle la plus fine possible
- Principales techniques d'estimations de population :
  - Africapolis
  - WorldPop
- Proposition :
  - Cartographier les zones bâties
  - Créer des modèles démographiques capables d'estimer la population urbaine et rurale de zones géographiques en s'appuyant sur des données de recensements et d'enquêtes
- Service Web Telecense : caractérisation de l'habitat



# Principales méthodes mises en œuvre jusqu'à présent

# Principales études effectuées

## L'exemple d'Africapolis et la gestion de l'urbain



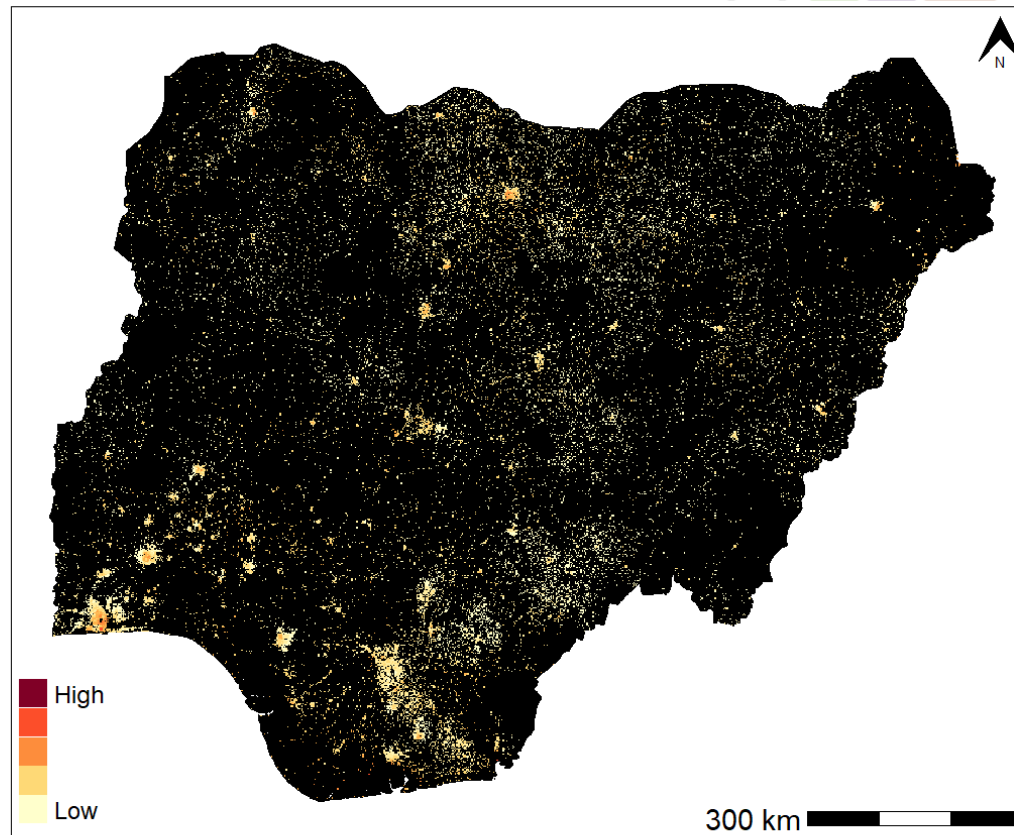
# Principales études effectuées

## L'exemple de WorldPop



### Nigeria population 2020

Estimated total number of people per grid-cell at a resolution of 3 arc seconds (approximately 100m at the equator)



WorldPop (worldpop.org - School of Geography and Environmental Science, University of Southampton)  
©2020 This work is licensed under a Creative Commons Attribution 4.0 International License





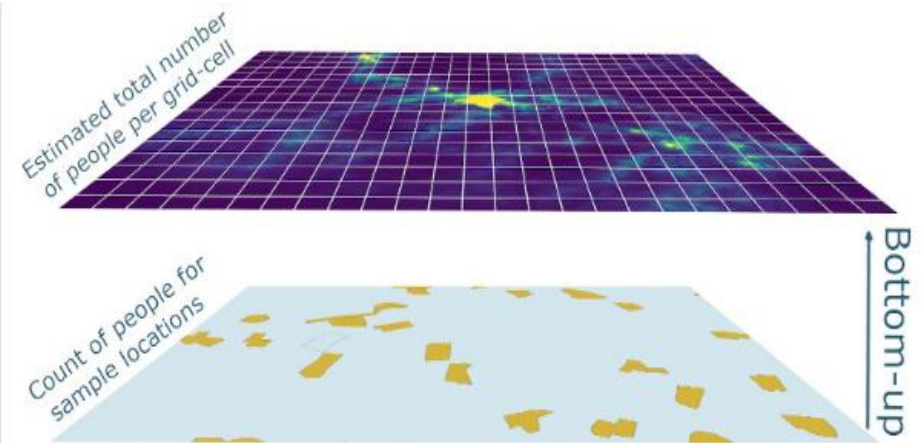
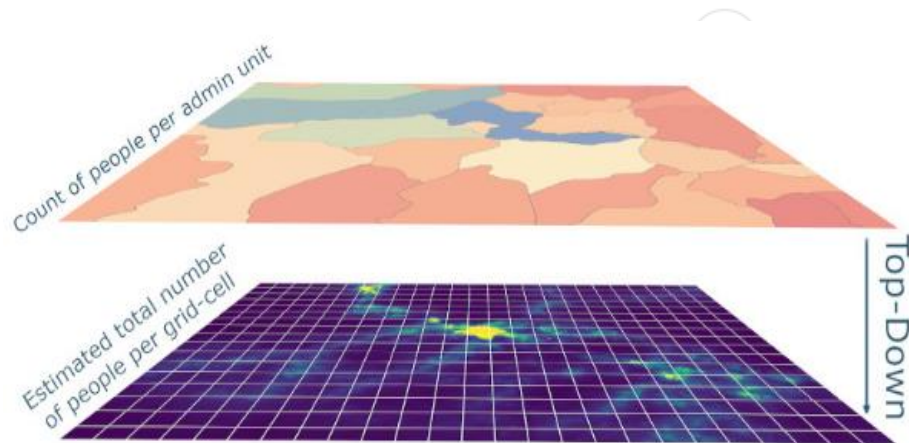
# Principales études effectuées

## Techniques « descendantes » et « ascendantes »



Désagrégation des données de population d'un recensement

Extrapolation à partir de données de **micro-recensement** ou d'enquêtes



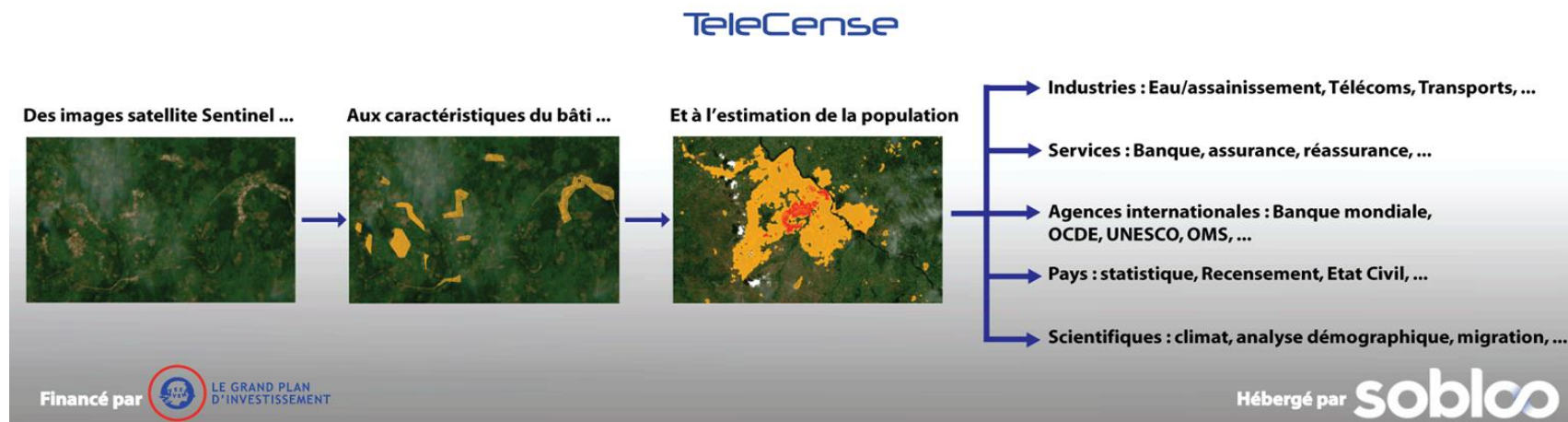
# Matériel et méthodes

# Matériel et méthodes

## Présentation de Diginove



- **Projet TeleCense :**
  - Estimations de populations localisées à échelle biannuelle à partir du traitement d'images satellites européennes en accès libre et des données démographiques venant de recensements ou d'enquêtes
  - Accompagner le développement et la planification en milieu rural et urbain



# Matériel et méthodes

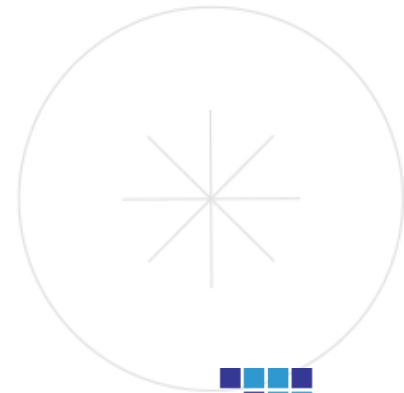
## Présentation de Diginove

Images satellites utilisées en accès libre

Saison humide



Saison sèche

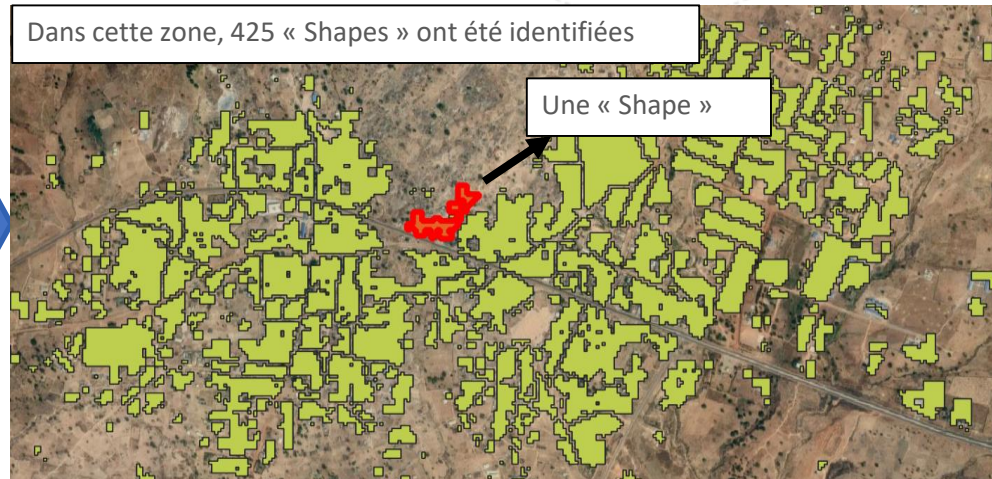


# Matériel et méthodes

## Présentation de Diginove



### Identification et caractérisation des zones de bâties



# Matériel et méthodes

## Le concept d'agglomération

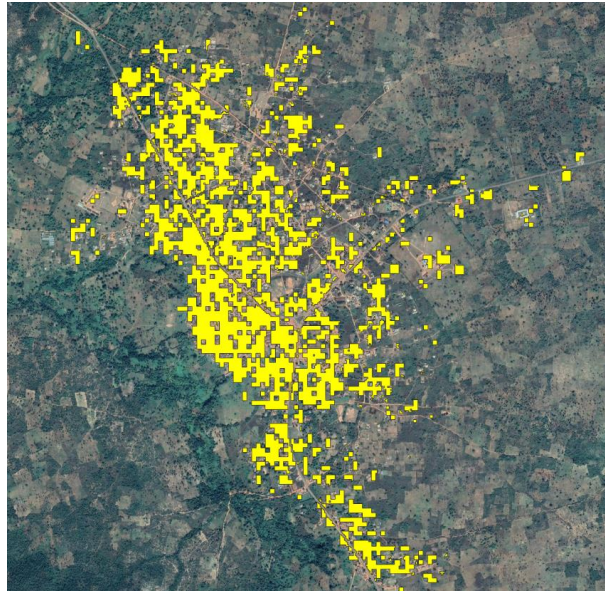


**Critère physique unique** : les zones bâties sont assemblées et considérées comme une agglomération si la distance qui sépare une zone d'une autre est inférieure à **200 mètres**

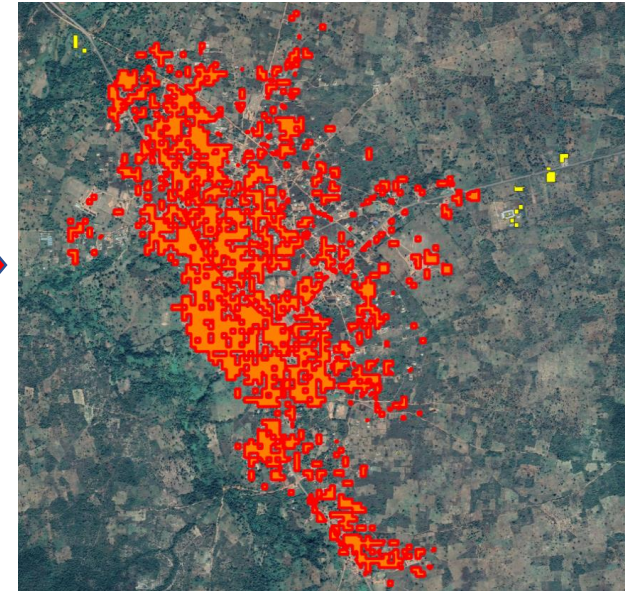
**1. Images satellites comportant des zones bâties**



**2. Détection du bâti (en jaune)**



**3. Création de l'agglomération (en rouge) en fonction du critère de distance physique entre zones**



On voit l'agglomération comme une ville vivante, elle évolue et se développe au fil du temps au-delà des limites des zones administratives



# Matériel et méthodes

## Croisement de plusieurs sources de données

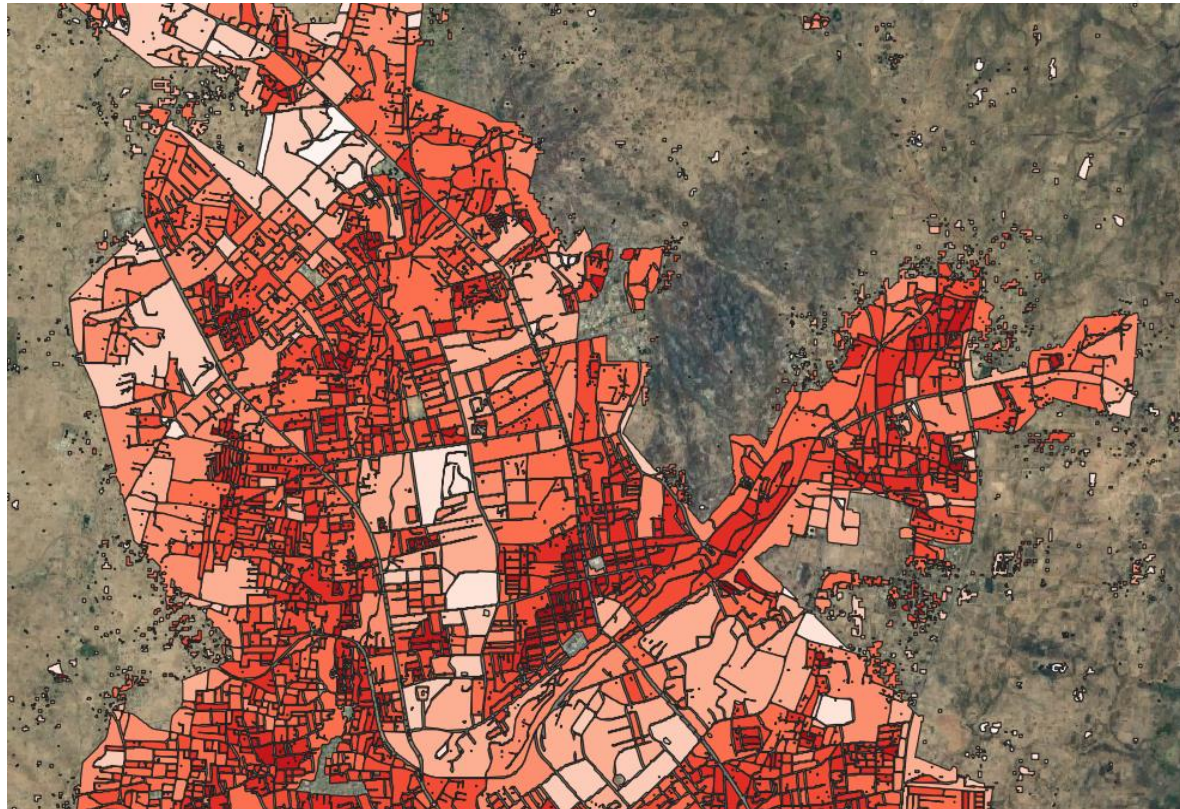


- Recensements et enquêtes auprès des ménages
  - Population, structure en âge et en sexe, composition du ménage, migration...
- Télédétection
  - Surface, densité de bâti, altitude, nombre de « shapes », indices de convexité, distance à la route, aux rivières et fleuves, aux hôpitaux et centres de santé
- Autres sources comme e-geopolis, ESA
  - Couverture du sol (GlobCover), température, précipitations, écoles, intensité de la lumière la nuit, données mobiles etc...



# Matériel et méthodes

## Exemple des variables pour une agglomération urbaine



Résultats de l'identification

| Entité   | Valeur                        |
|--|-------------------------------|
| ▼ 20200407-T32PKQ-T32PKR-T32PLQ-T32PLR-cities copier |                               |
| ▼ fid  | 10032                         |
| ▶ (Dérivé)   |                               |
| ▶ (Actions)  |                               |
| fid  | 10032                         |
| area   | 62149000                      |
| elevation  | 254                           |
| elevationMin   | 197                           |
| elevationMax   | 348                           |
| slope  | 0                             |
| width  | 14060                         |
| height   | 12430                         |
| perimeter  | 2629320                       |
| externalPoints                                       | 14                            |
| solidity   | 0,43                          |
| concavity  | 58,49                         |
| nbShapes   | 4747                          |
| BigCityDist  | 0                             |
| BigCityName  | Minna                         |
| BigCityInd   | 799                           |
| Preci  | 1138,2056118920564            |
| EcoRegion  | Guinean forest-savanna mosaic |
| Temp   | 27,119116250874953            |
| startelec  | 1999                          |
| EL2016MxT  | 0,9999961853027344            |
| WPop2020   | 1442650                       |
| distanceToRoad                                       | 0                             |
| distanceToWater                                      | 0                             |
| distanceToHealth                                     | 0                             |
| administrative-bounda...                             | (2;5,7)                       |



# Matériel et méthodes

## Estimation de la population au niveau de l'agglomération

Typologie et modélisation basée sur les caractéristiques des agglomérations

- Typologie des zones de bâties
  - Centres urbains
  - Périphéries urbaines
  - Zones rurales



Densité de population diffère selon la zone étudiée

Urbain ou rural  
Aride ou humide  
Couverture du sol

- Modélisation de la population par régression linéaire multiple

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k x_{ki} + \varepsilon_i$$

- Prise en compte de la structure spatiale

$$Y = \rho \cdot WY + X \cdot \beta + WX \cdot \theta + u \quad \text{Où } u = \lambda \cdot Wu + \epsilon$$



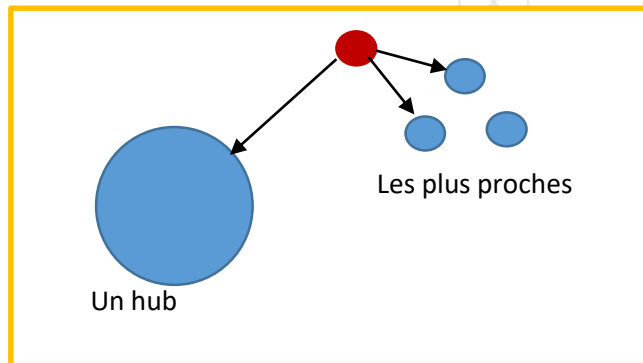
# Matériel et méthodes

## Estimation de la population au niveau de l'agglomération

Pourquoi faire de la statistique spatiale ?

- Objectif : comprendre les **liens** et **relations** entre agglomérations proches

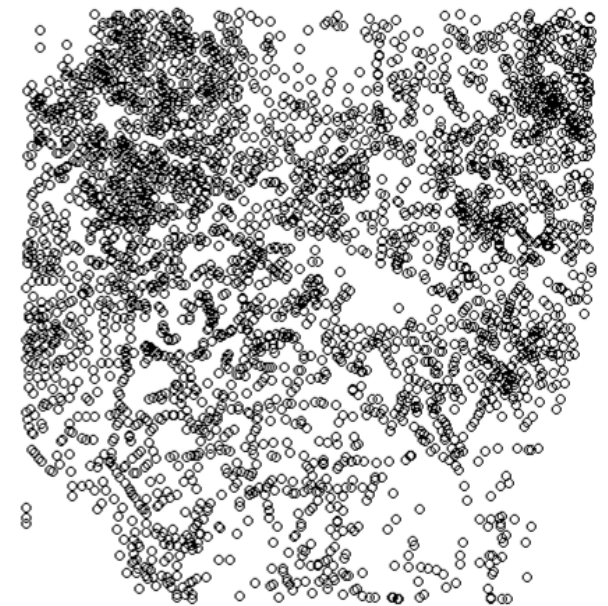
En études des foules un nœud se **connecte** préférentiellement avec les nœuds les plus **visibles** : les plus **proches** ou les plus **gros**



Mécanisme similaire dans notre **réseau d'agglomérations**

- Permet de surpasser les insuffisances de la modélisation linéaire classique (Autocorrélation des résidus et non-respect de l'hypothèse d'homogénéité des résidus)

Centroïdes des agglomérations de la région d'Abuja

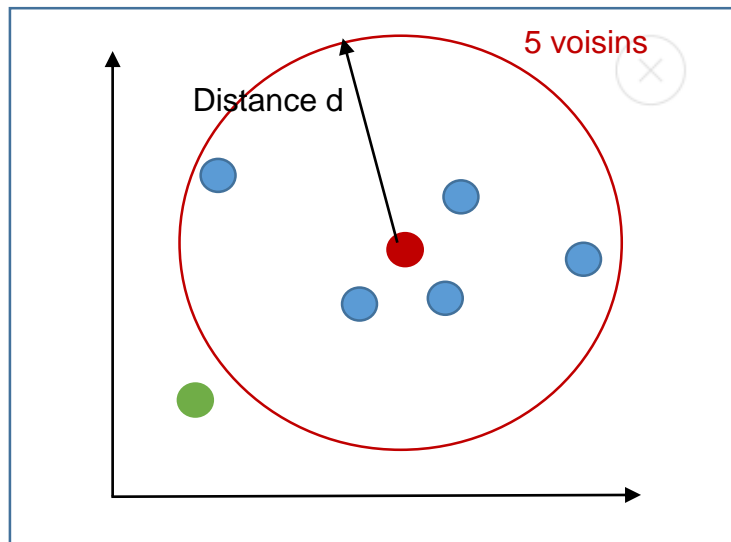


# Matériel et méthodes

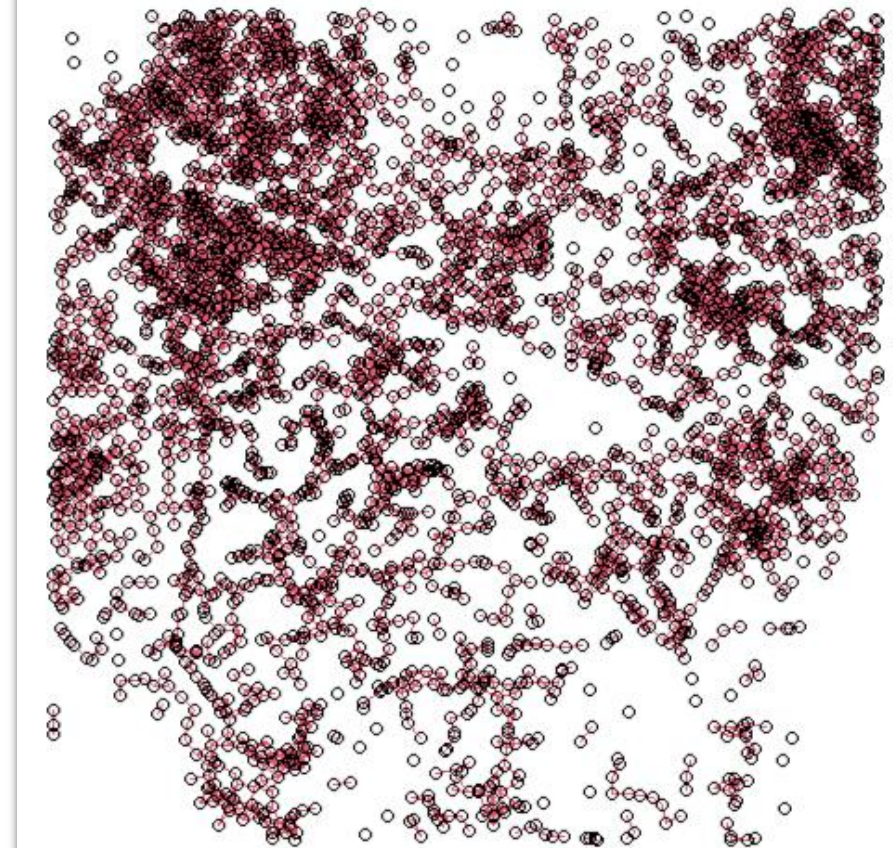
## Estimation de la population au niveau de l'agglomération

Codifier la structure du voisinage

Exemple du voisinage à distance radiale



Connexions entre les agglomérations



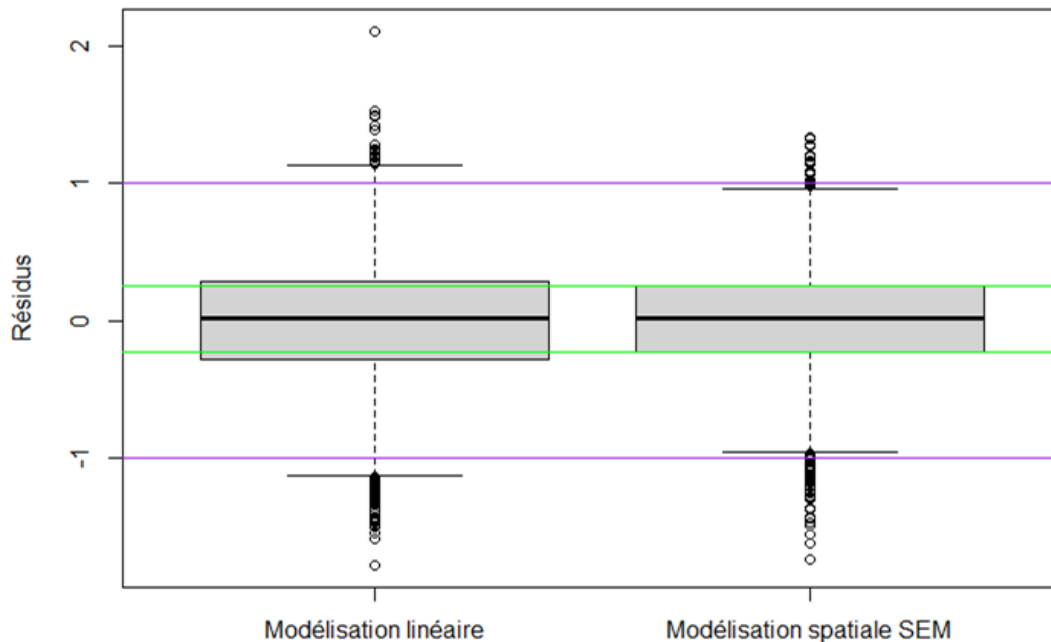
# Matériel et méthodes

## Estimation de la population au niveau de l'agglomération

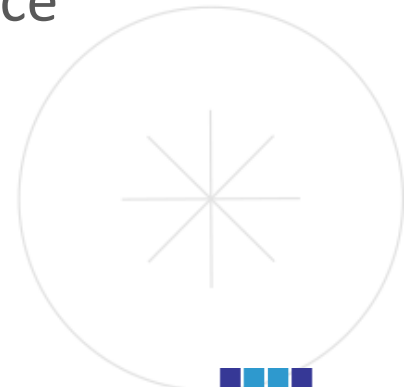
### Effets et améliorations du modèle spatial

- Diminution de l'AIC et de la variance → meilleures prédictions

Comparaison du modèle classique et du modèle spatial



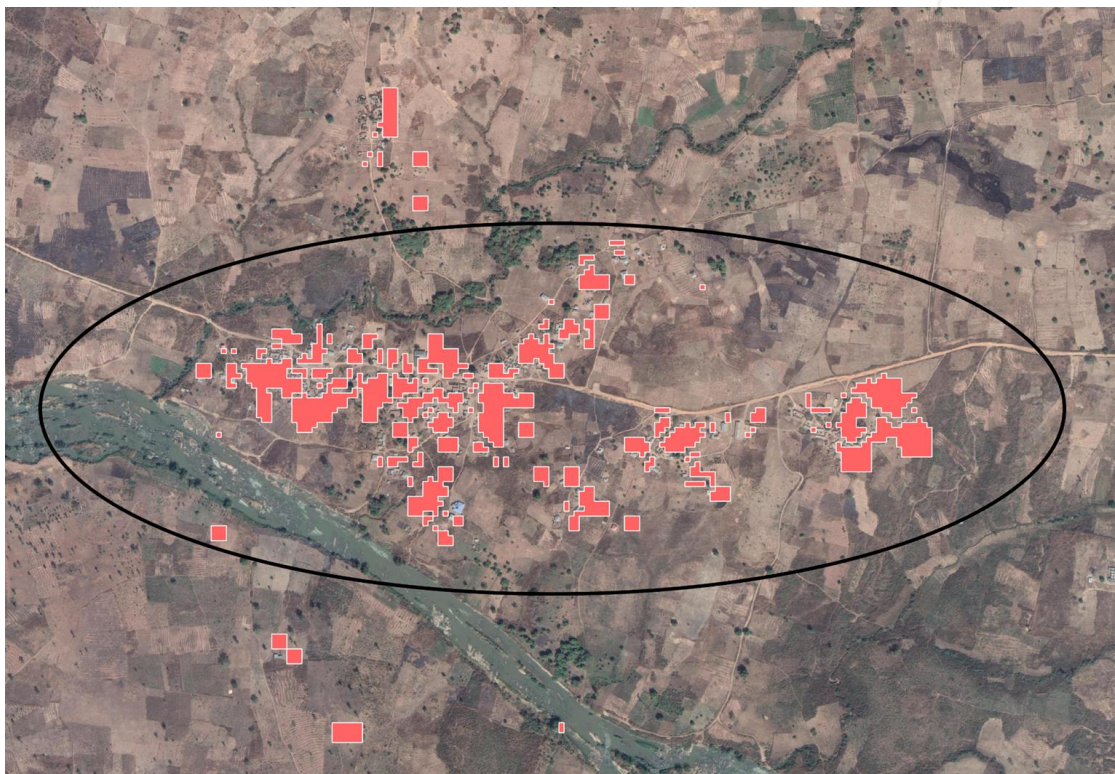
Effet indirect de la surface



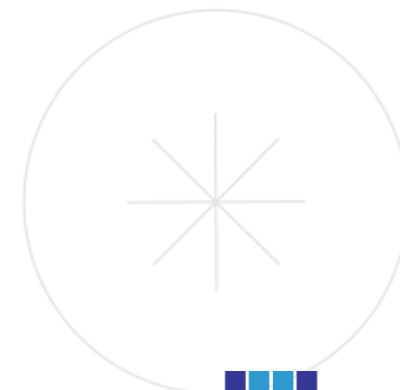
# Matériel et méthodes

## Estimation de la population au niveau de l'agglomération

Résultat pour une agglomération et suite du travail



Agglomération avec  
1144 habitants à  
répartir dans ses 94  
shapes



# Matériel et méthodes

## Passage de l'agglomération à la shape

### Le concept de désagrégation

- Désagrégation = passage d'une zone géographique fixée à une zone plus petite
- Notre utilisation : répartir la population d'une agglomération dans plusieurs shapes qui la constituent



Nombre de maisons  
au niveau du pays  
(connu)



Nombre de maisons  
au niveau du  
département (à  
trouver)

zones sources = pays



zones cibles = départements



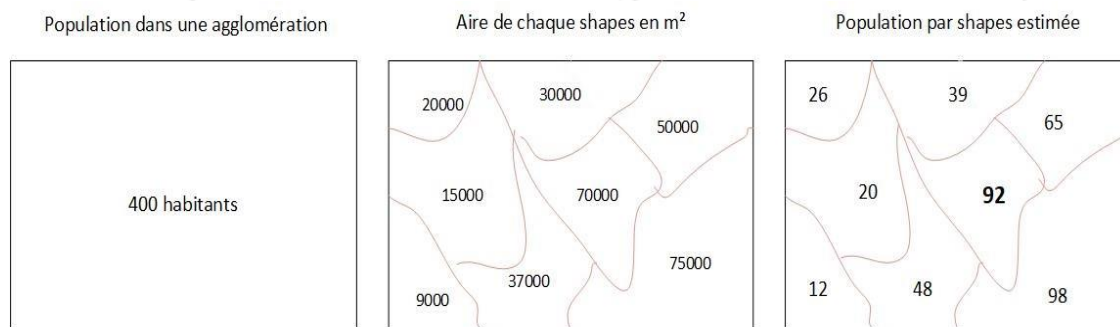
# Matériel et méthodes

## Passage de l'agglomération à la shape

Méthode d'allocation proportionnelle à l'aire

**Principe** : on répartit la population de l'agglomération proportionnellement à l'aire de chaque shape

**Problème** : deux shapes (quartiers) de même aire peuvent avoir des structures d'habitats différentes



Aire totale = 306 000 m<sup>2</sup>

$$400 * 70\,000 \div 306\,000 = 92$$



# Matériel et méthodes

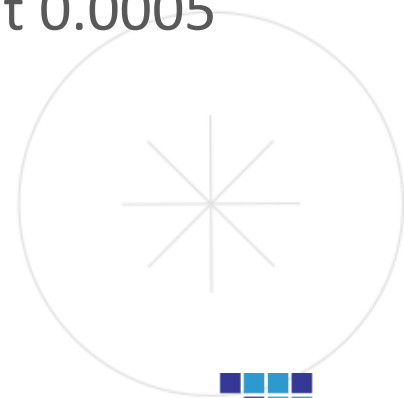
## Passage de l'agglomération à la shape

Méthode dasymétrique avec zones de contrôles

**Principe** : prend en compte l'aire et une autre information auxiliaire catégorielle. On connaît ou estime la densité d'habitant par zone de contrôle.



Prenons **un exemple** avec comme zones de contrôles les zones urbaines ou rurales, on a  $0.0026 \text{ hab/m}^2$  en zone urbaine et  $0.0005 \text{ hab/m}^2$  en zone rurale

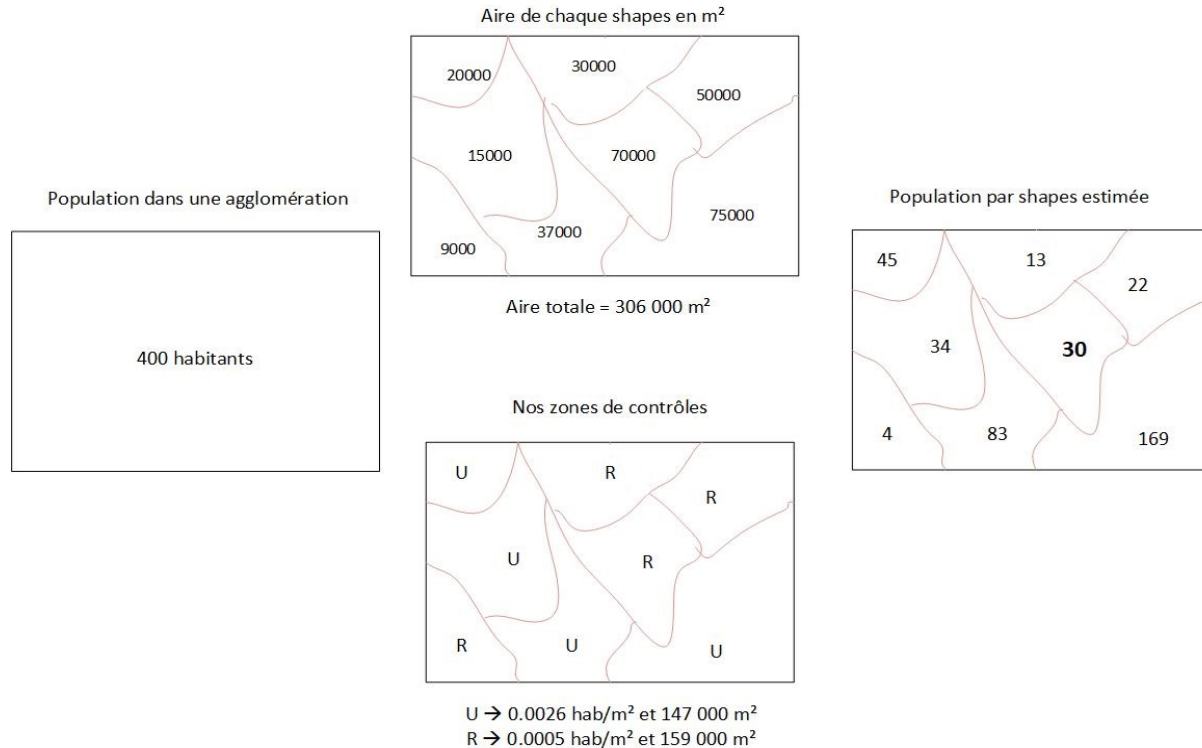




# Matériel et méthodes

## Passage de l'agglomération à la shape

### Méthode dasymétrique avec zones de contrôles



$$400 * \frac{70000 * 0.0005}{147\ 000 * 0.0026 + 159\ 000 * 0.0005} = 400 * \frac{35}{461.7} = 30$$



# Matériel et méthodes

## Passage de l'agglomération à la shape

Nos zones de contrôles

Nos zones de contrôles sont les zones de différentes densités de bâti:

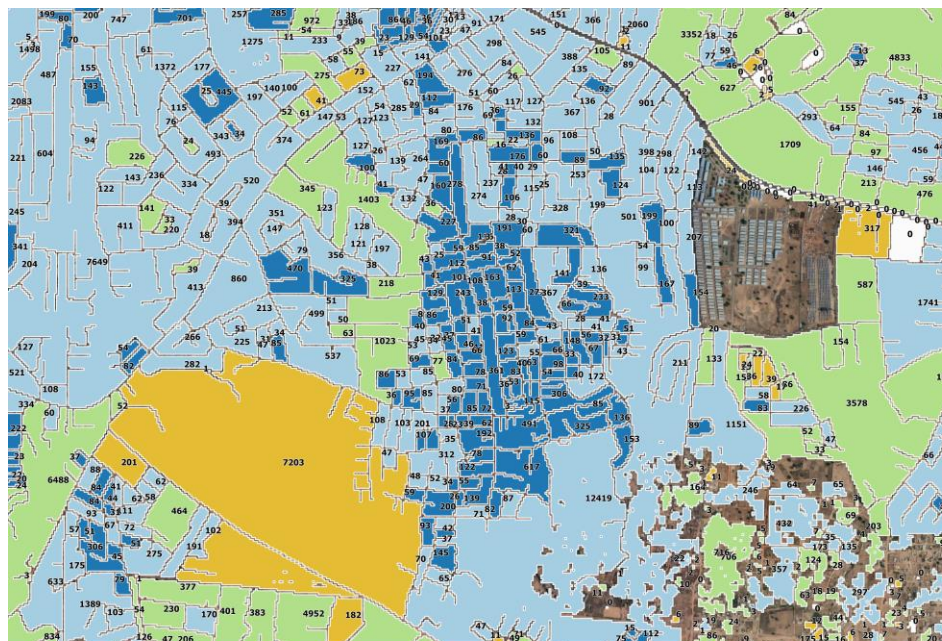
- Zone inhabitée
- Tissu urbain faible discontinu
- Tissu urbain moyen discontinu
- Tissu urbain dense discontinu
- Tissu urbain continu

Pour chacune de ces zones on estime la densité d'habitants et en combinant cette info à l'aire on obtient le nombre d'habitants

# Matériel et méthodes

## Passage de l'agglomération à la shape

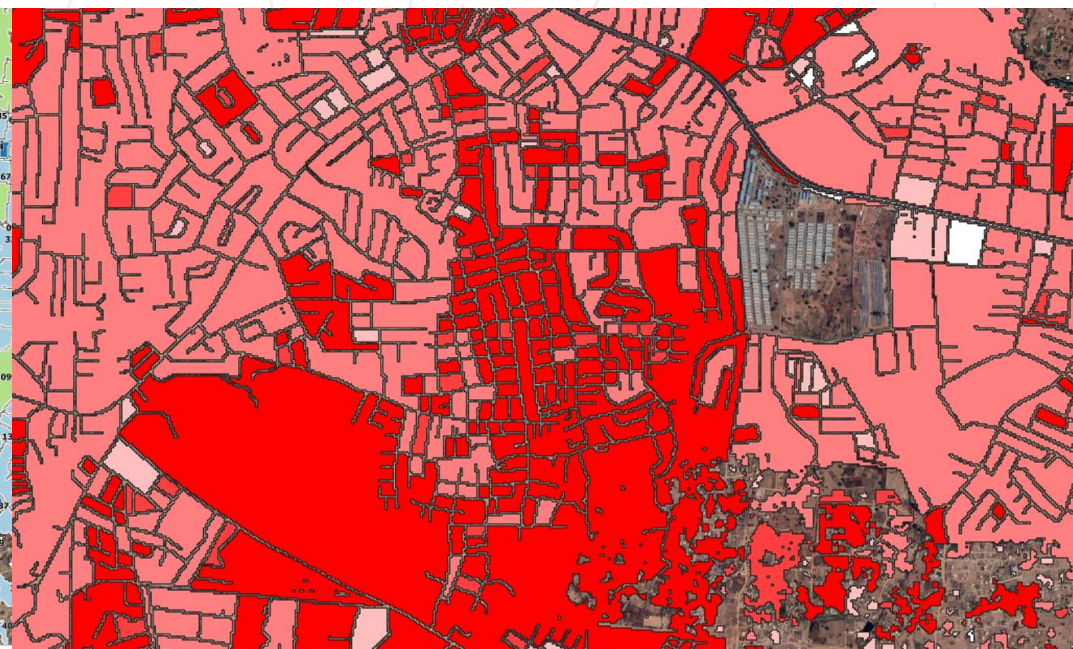
### Illustrations des résultats



Densité de population

(hab/m<sup>2</sup>)

- Zone inhabitée (D\_0=0)
- Tissu urbain faible discontinu (D\_1=0.01357)
- Tissu urbain moyen discontinu (D\_2 = 0.02191)
- Tissu urbain dense discontinu (D\_3 = 0.023012)
- Tissu urbain continu (D\_4 = 0.027863)



Densité de population

(hab/m<sup>2</sup>)

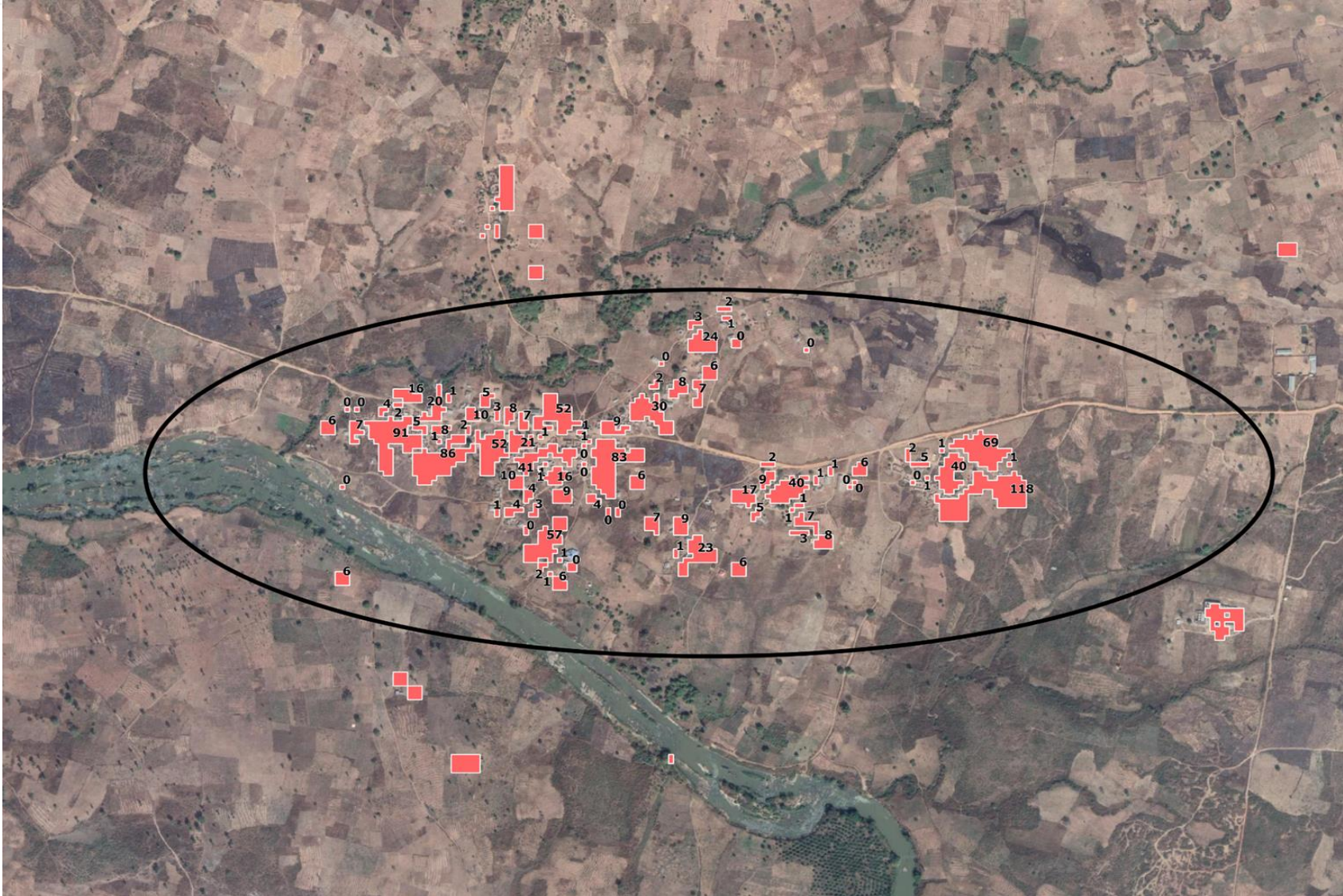
- 0,00000 - 0,00500
- 0,00500 - 0,00903
- 0,00903 - 0,00953
- 0,00953 - 0,01000
- 0,01000 - 99,91000



# Matériel et méthodes

## Passage de l'agglomération à la shape

Illustrations des résultats

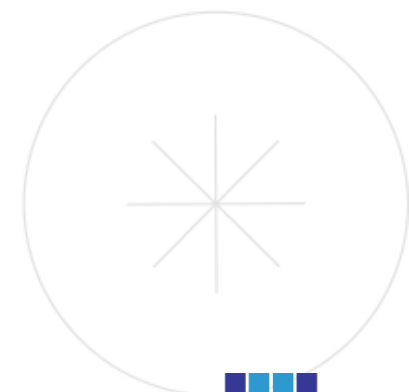


# Matériel et méthodes

## Passage de l'agglomération à la shape

### Perspectives de travail

- Possibilité d'utiliser d'autres variables que l'aire et la densité de bâti
- Utilisation des forêts aléatoires pour prendre en compte plus de variables explicatives :
  - Validation du modèle sur une région avec micro-recensement
  - Généralisation du modèle dans les autres régions



# Conclusion

# Conclusion



- Objectif : trouver la meilleure méthode pour estimer la population ainsi que la désagréger
- Perspective de travail : désagrégation à l'aide des forêts aléatoires en passant de la plus petite zone administrative pour laquelle des données de recensement sont publiées aux agglomérations.
- Intérêt pour Lisa : travail sur la structure de la population, estimation des proportions de femmes en âge de procréer et enfants de moins de 5 ans → savoir où sont les personnes à risque pour le paludisme
- Intérêt pour Léo : Thèse et extrapolation des résultats sur le Kenya





**Merci !**