

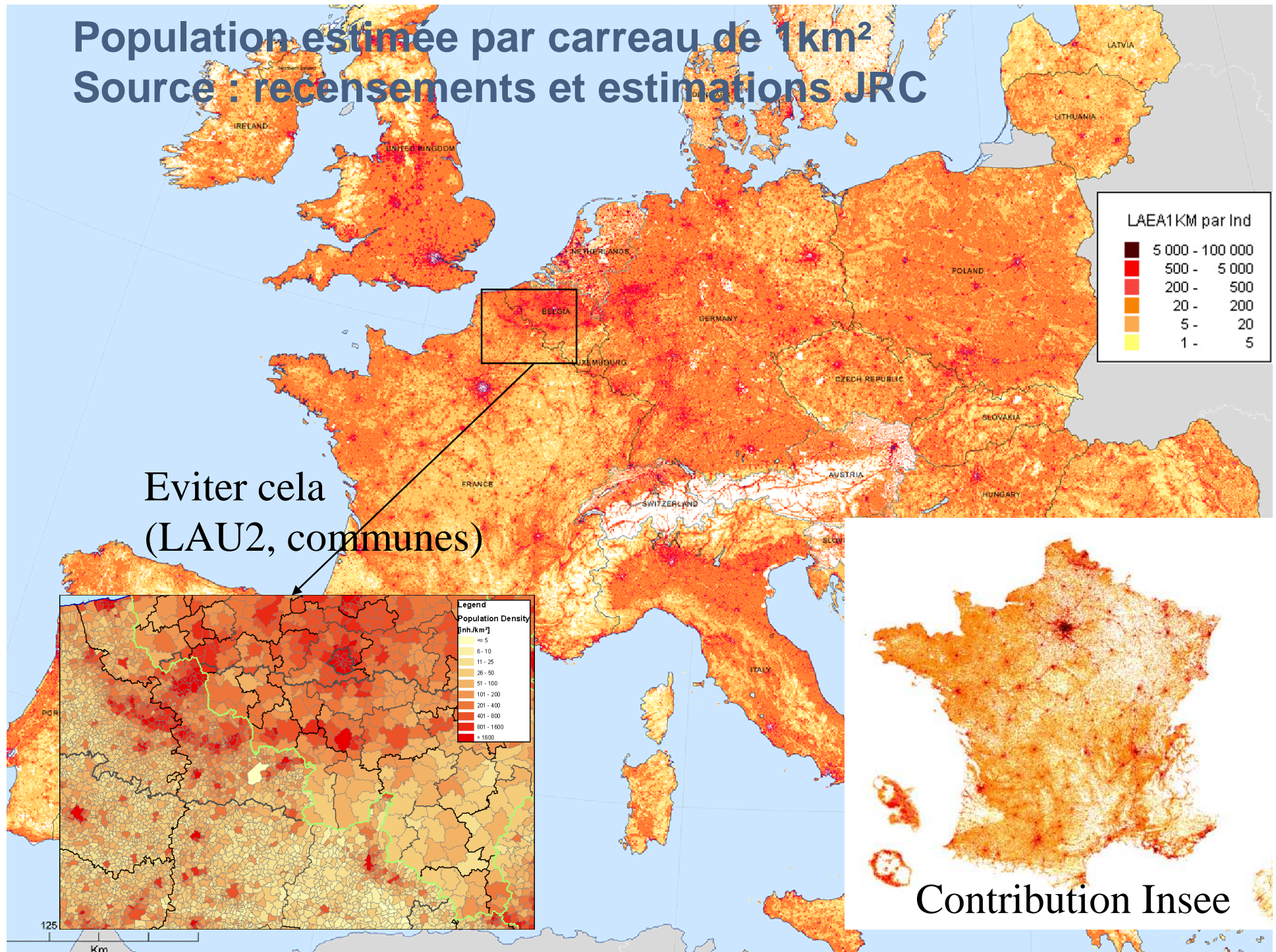


Des données carroyées à tout prix Le recensement aussi

Jean-Luc LIPATZ
INSEE/DG/DET

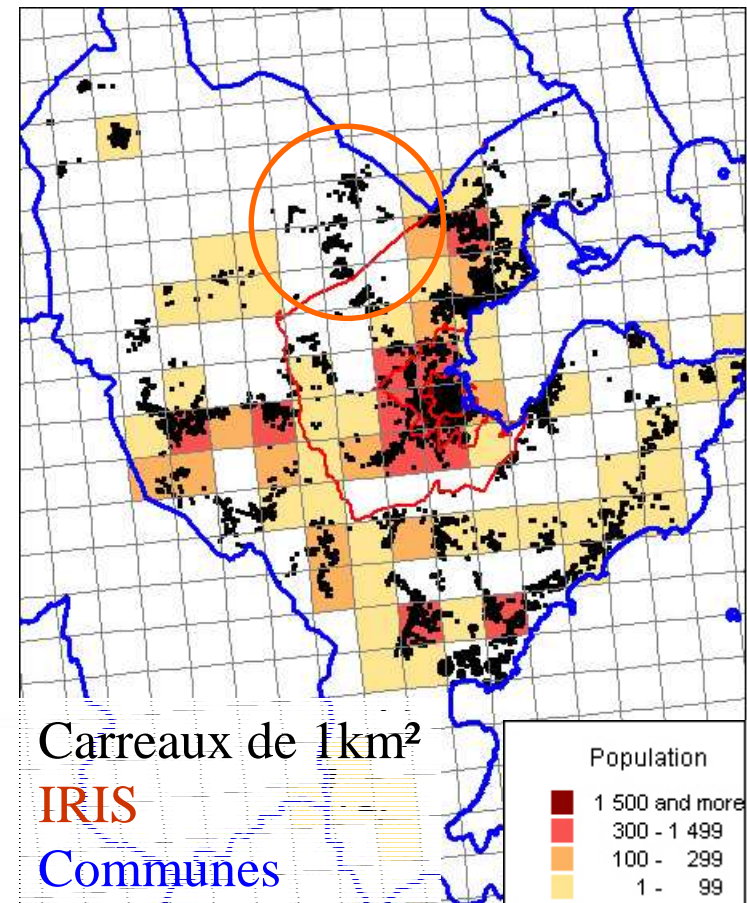


Population estimée par carreau de 1km² Source : recensements et estimations JRC



Acte I – les plus de 10 000 habitants

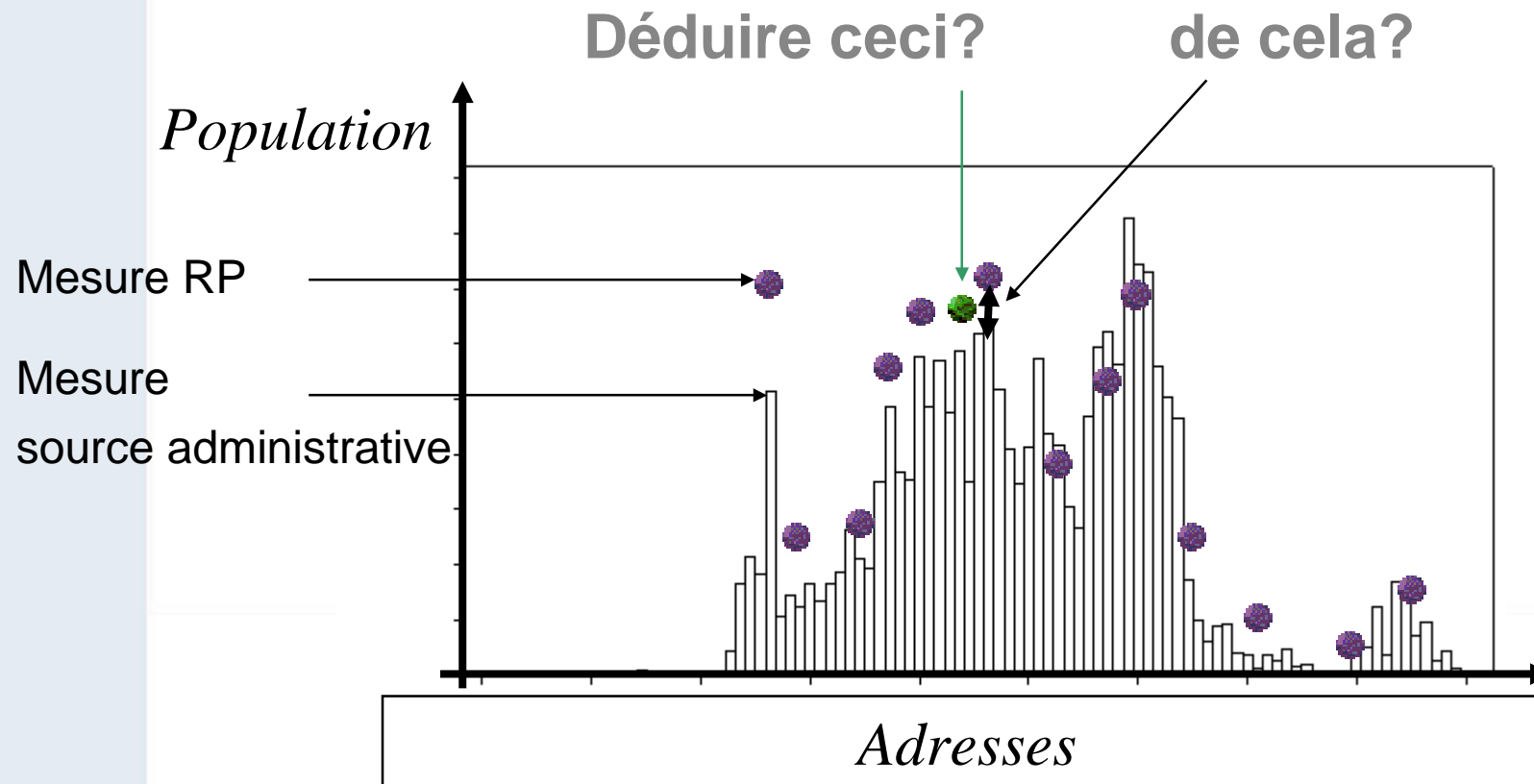
- › L'échantillon ne garantit pas une couverture spatiale
 - ➔ IRIS et zones DIAF-RP (pas toujours)
- › Mais on dispose de beaucoup d'information auxiliaire géolocalisée
 - RIL
 - Données administratives



Estimation mode DIAF-RP



Remplir les trous de collecte corriger les biais administratifs et temporels





Des régressions locales

L'espace considéré comme une variable explicative

$$Pop_{RP} = \alpha(s).Pop_{Adm} + \beta(s).Logements + \varepsilon$$

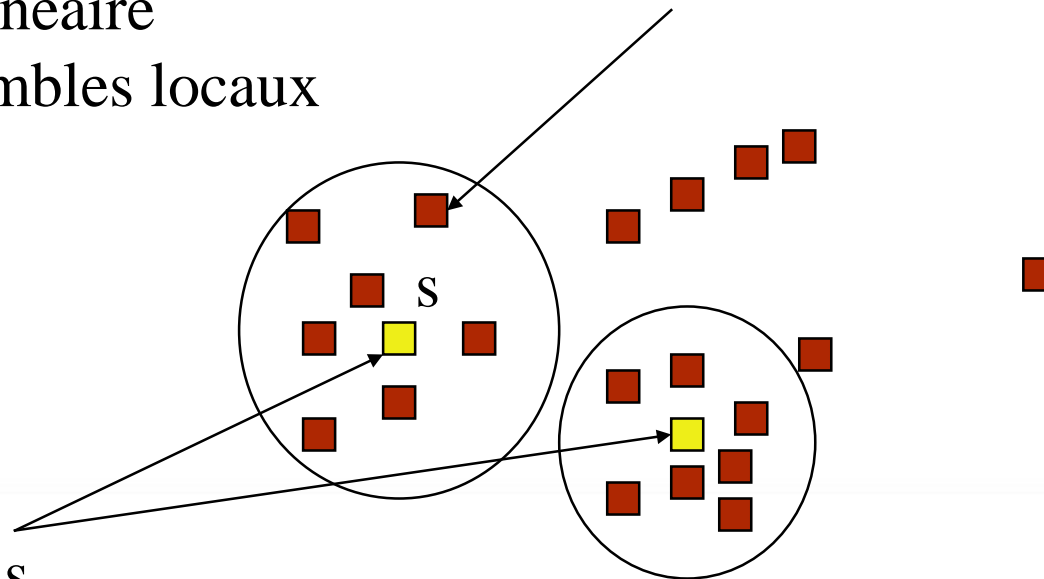
Régressions linéaire
sur sous-ensembles locaux

$\alpha(s), \beta(s)$

Estimations

$$\hat{Pop}_{RP}(s) = \alpha(s).Pop_{Adm}(s) + \beta(s).Logements(s)$$

Pondérations décroissantes
avec la distance

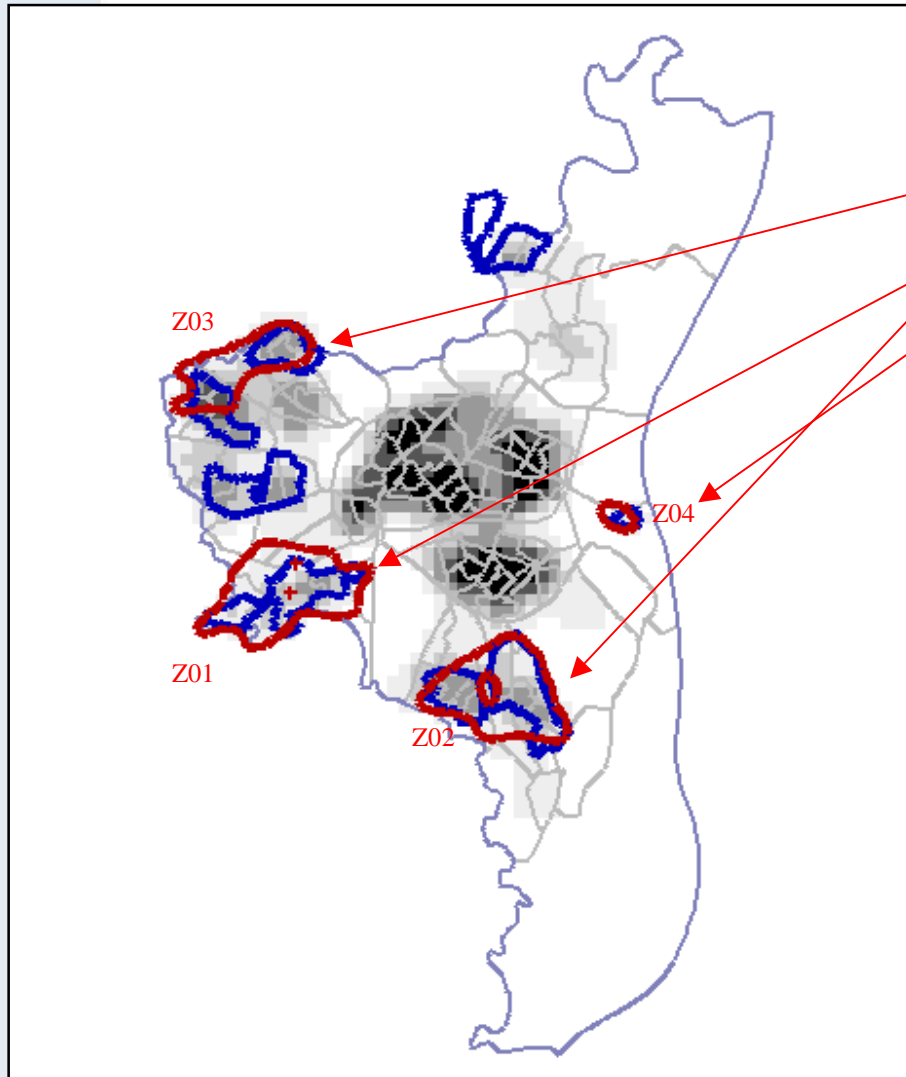




- › Taille des voisinages : nombre d'observations fixe
 - Indicateurs politique de la ville : nombre variable déterminé après analyse des AICC; travail sur des intersections carreaux - zonages
 - Population carroyées pour Eurostat : 50 logements

 - › Données administratives :
 - Indicateurs politique de la ville : CNAM, appariée au bâtiment RIL sur coordonnées géographiques
 - Population carroyées pour Eurostat : logements seuls
- Simulation (commune de Saintes, 59 carreaux, 25 000 habitants, 100 échantillons à partir de la TH et de RFL)
- somme des écarts type par carreau :
- | | |
|-----------------|------------|
| logements seuls | 1057/25000 |
| ménages | 966/25000 |
- plus imprécis en zone éparses limite de commune

Politique de la ville (Strasbourg)



Effectif total	cluster	Part Sous-population	Carreaux	x	y
255227	total	0.082	.	.	.
18380	Z02	0.141	66	1000421	2408177
14966	Z03	0.127	45	996056	2413764
14612	Z01	0.132	73	996675	2409871
1405	Z04	0.139	8	1002696	2411206

A mettre en rapport avec l'estimation RP des ménages: 259 399

Estimation
(3 années)

Chiffres finaux
(5 années)

dc	zus	est_pop
67482	4201030	4303
67482	4201090	7689
67482	4201100	9357
67482	4201120	13412
67482	4201130	11188
67482	4201140	7515
67482	4201180	11436
67482	4201190	1346
67482	HORSZONE	188980

7437
9665
13476
11425
7821
11740
1391

En rouge: Zones de concentration de chômeurs
En gris: Distribution spatiale de la population (d'autant plus sombre qu'elle est concentrée)
En bleu: Contours des ZUS

Source: Estimations à partir des enquêtes annuelles de recensement 2004 à 2006

Jean-Luc LIPATZ



Acte II – les moins de 10 000 habitants

Des données fiscales aux données RP

Deux sources parlant de la même réalité

Avec des interprétations / descriptions différentes :

- Qualité (RP : valeurs manquantes, fisc : défauts d'enregistrement)
- Etudiants (RP : au lieu de résidence, fisc : dans leur famille)
- Enfants (RP : données complètes, fisc : seulement l'âge)
- Résidence secondaires
- Notion d'immeuble
- Personnes hors ménages ordinaires

Mais 95 % des enregistrements parlent des mêmes personnes

→ Importer la localisation depuis les données fiscales



Le processus

- 1) Appariement flou des deux fichiers individuels en utilisant :
 - La commune de résidence
 - Le sexe
 - La date de naissance
 - La commune (éventuellement seulement le département) de naissance

➔ Appariements multiples
- 2) Calcul d'un score individuel
 - Qualité de la reconnaissance du lieu de naissance?
 - Enregistrement administratif correspondant à une résidence secondaire, à autre chose qu'une maison ou un appartement?
- 3) Calcul par agrégation d'un score au niveau immeuble RP
- 4) Choix de l'appariement le plus probable, invalidation des appariements improbables
 - Le recensement est collecté par grappes
- 5) Interpolation des appariements invalidés et des échecs d'appariement
 - Entre deux enregistrements RP voisins dans le fichier, lorsqu'ils ont été identifiés dans la même rue



Les résultats

Rien n'est parfait

- Le lieu de naissance n'est parfaitement connu d'aucune des deux sources : intégré au calcul du score.
- La date de naissance des personnes âgées est parfois mal saisie
- Les données manquantes du recensement sont comblées par clonage : impossible à corriger, les ménages correspondants sont placés au barycentre du district de collecte.

➔ Mais le résultat est plus que satisfaisant

Département de la Vienne, communes de moins de 10 000 habitants,
population des ménages ordinaires

Positionné au barycentre	0.6 %
Résultant du repérage de l'immeuble	97.4 %
Interpolé	2.0 %

Des carreaux pour réconcilier les approches (RP 2007, nouveaux arrivants) carreaux LAEA 200m

Plus de 10 000 habitants
Uniquement des données carroyées

