

LA CONSTRUCTION DU NOUVEL ECHANTILLON DE L'ENQUETE EMPLOI EN CONTINU A PARTIR DES FICHERS DE LA TAXE D'HABITATION.

Vincent LOONIS
(Division ETSD, UMS)



JMS – 23 MARS 2009



▣ **Résultat d'un travail collectif, auquel ont participé**

- *Nicole Cadenel,*
- *Guillaume Chauvet,*
- *Marc Christine,*
- *Sébastien Durier,*
- *Sébastien Hallépée,*
- *Anne Flipo,*
- *Dominique Place.*

Plan

- ❑ Les contraintes,
- ❑ L'échantillon actuel,
- ❑ Les fichiers de la taxe d'habitation comme base de sondage,
- ❑ La construction des « grappes » et des secteurs,
- ❑ La répartition par région et le calendrier,
- ❑ La sélection de l'échantillon,
- ❑ Les procédures post sélection
- ❑ Les logements nouveaux.

Les contraintes

- ❑ l'échantillon actuel doit continuer à être mobilisé sans modifications par rapport à sa durée de vie initialement prévue,
- ❑ le schéma de rotation et d'interrogation de l'échantillon actuel doit être reconduit,
- ❑ l'objectif de 50 % doit être atteint mi-2010. Pour parvenir à cet objectif, la mise en place du nouvel échantillon doit être progressive à compter de janvier 2009. Ce calendrier implique que les Directions Régionales de l'Insee soient informées dès la fin du premier trimestre 2008 de la localisation et du nombre de zones impactées.
- ❑ le gain en précision attendu doit être, dans la mesure du possible, plus important que celui que l'on observerait lors d'une augmentation de 50 % de l'échantillon, toutes choses égales par ailleurs.



L'échantillon actuel

- ❑ Objectifs : suivi trimestriel de la situation sur le marché du travail et de son évolution,
- ❑ Solution : Echantillonnage rotatif,
- ❑ Contraintes :
 - 2 semaines et deux jours de collecte => concentration des enquêtes dans des aires, (logements contigus)
 - Charge de travail des enquêteurs : des aires d'une vingtaine de logements,



L'échantillon actuel

- ❑ Une aire est enquêtée 6 trimestres de suite
- ❑ Un secteur est un regroupement de 6 aires proches
 - Quand une aire est sortante elle est remplacée par une aire du même secteur
- ❑ Les secteurs sont répartis en 6 sous-échantillons représentatifs de l'ensemble de la population,
 - Les secteurs sont introduits progressivement afin d'assurer la rotation de l'échantillon.
- ❑ Ce schéma est reconduit pour le nouvel échantillon
 - Comment construire les « aires » et les secteurs ?
 - Une solution les fichiers de la taxe d'habitation.

Les fichiers de la TH comme base de sondage

□ Exhaustivité

– 32 millions de logements, dont 25,4 millions de résidences principales. Légèrement plus de logements qu’au recensement à cause des communautés.

□ Fraîcheur

– L’Insee reçoit à la fin de l’année N, la situation relative au premier janvier de la même année

□ Identifiabilité des logements

– Nom du dernier occupant connu, adresse, références cadastrales,

– Identifiant pérenne des logements

□ Information auxiliaire (correction de la non réponse)

– Caractéristique du logement, du ménage (âge du chef de ménage, composition, revenus...)

La constitution des grappes

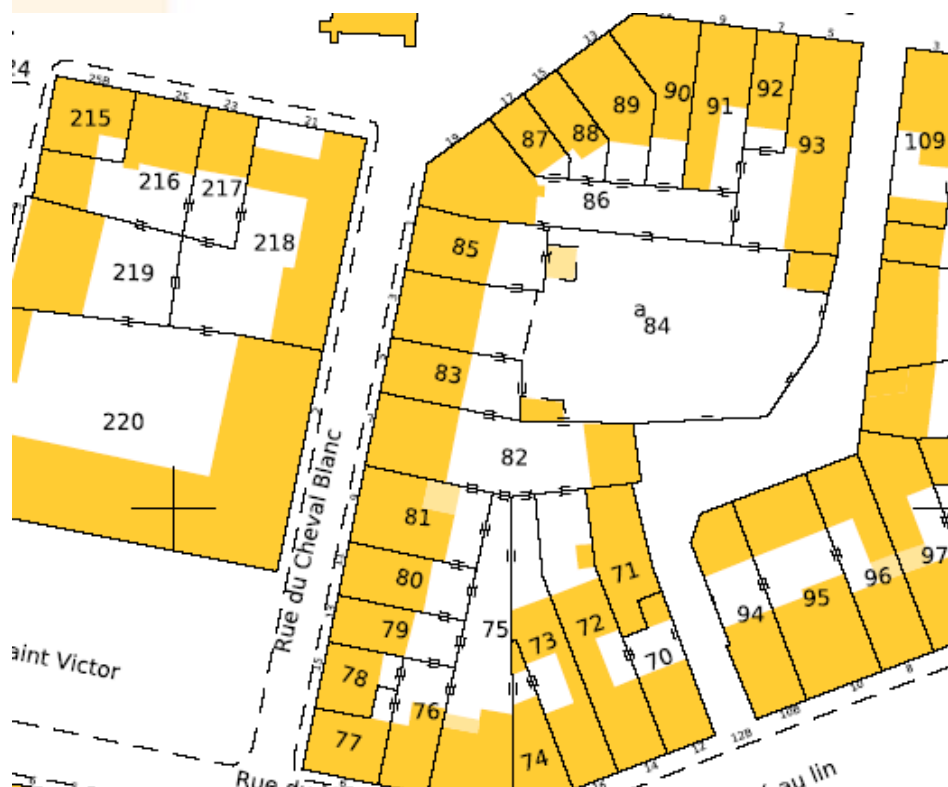
- ❑ Une grappe est regroupement de logements proches,
 - ❑ Tous les logements, y compris les logements non principaux, sont affectés à une grappe,
 - ⑩ les grappes ont un nombre de résidences principales le plus proche possible de 20.
 - ❑ Dans les immeubles collectifs, les logements d'un même étage sont inclus dans la même grappe.
-
- ❑ *On utilise le numéro de parcelle cadastrale.*

On lit le fichier de logement dans l'ordre des numéros de parcelle jusqu'à obtenir 20 résidences principales.



A l'intérieur des sections cadastrales, les parcelles sont numérotées de manière a priori contigüe.

Parfois cela ne fonctionne pas



On corrige, dans la mesure du possible, le numéro de parcelle de manière à recréer de la contigüité à partir du nom de la voie et du numéro.

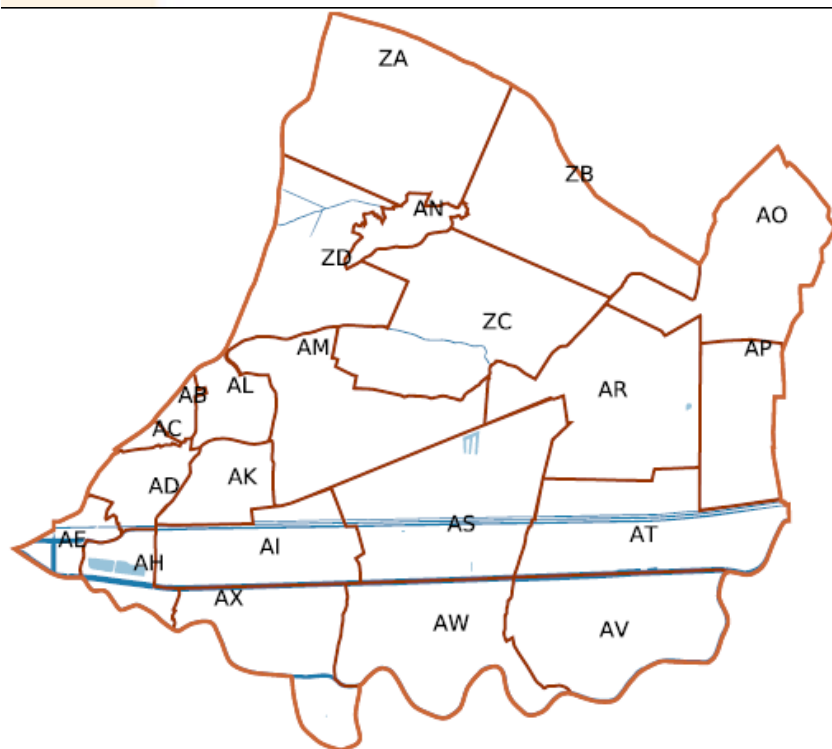
On ne peut pas tout corriger automatiquement, ce qui explique que l'on passe à de la contigüité à la proximité.

A l'intérieur des communes on agrège les grappes par 6 selon l'ordre de création et selon l'ordre de la section cadastrale

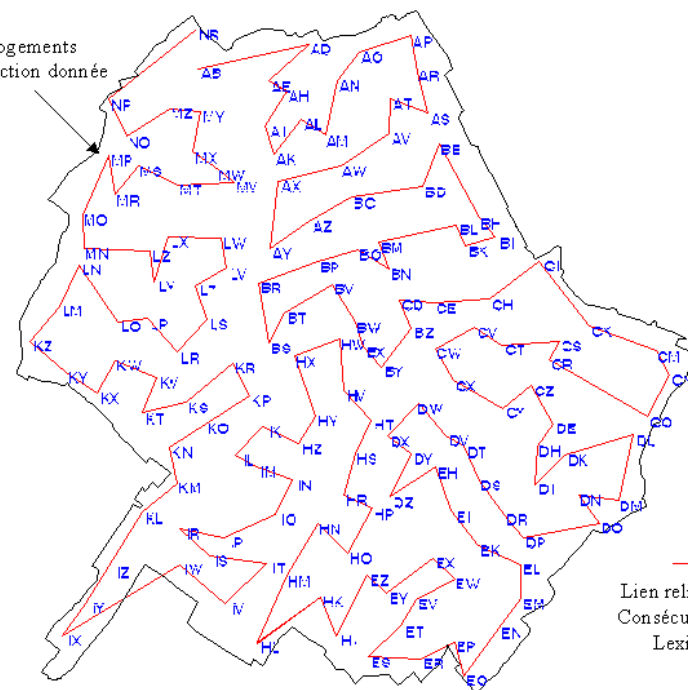


Toutes les communes françaises sont découpées en sections cadastrales

Deux sections d'une même commune ayant des codes qui se suivent selon l'ordre lexicographique seront limitrophes



Barycentre des logements appartenant à une section donnée

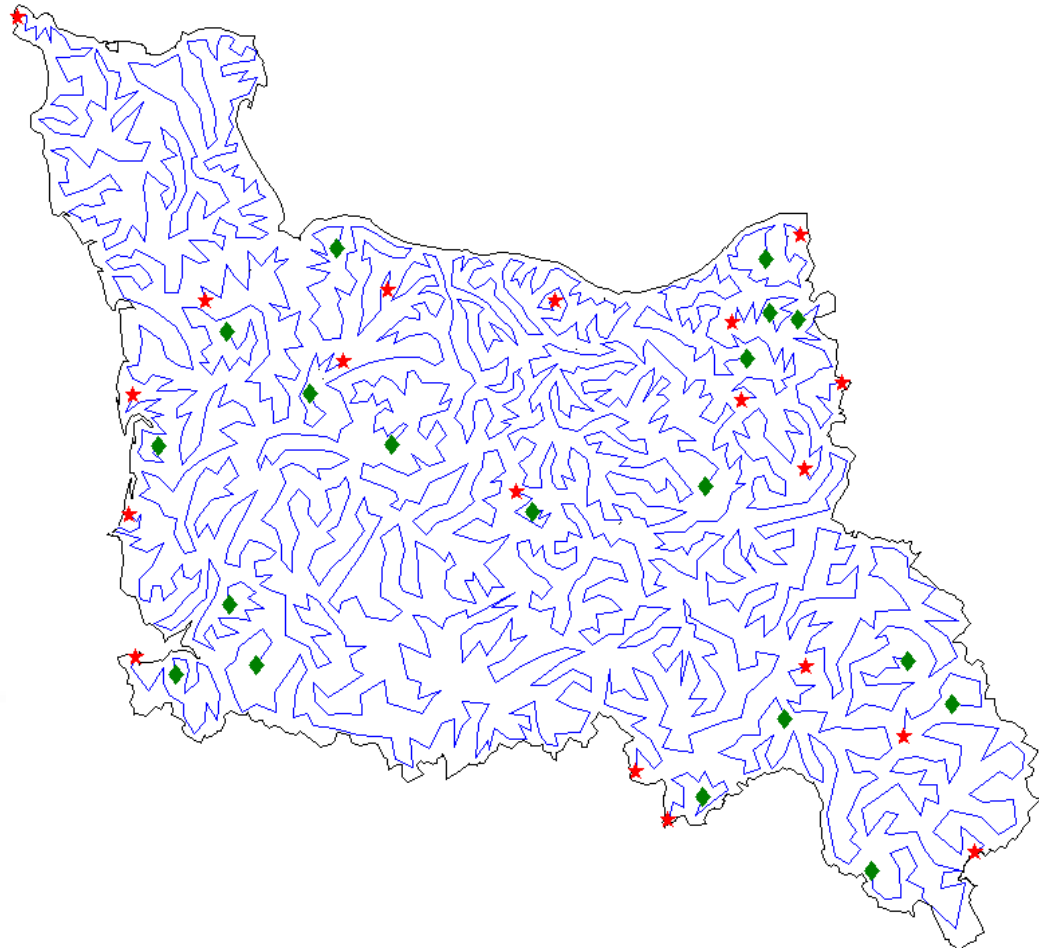




Comment obtenir un numéro de commune tel que la commune i soit limitrophe de la commune $i+1$?

Pour les petites communes, on agrège les grappes par ordre de création à l'intérieur de la section, par section, et par numéro de communes, tel que la commune i est à côté de la commune $i+1$.

Les secteurs sont soit dans une commune, soit dans des communes limitrophes.





Bilan...

- ❑ 32 millions de logements regroupés en 1,2 millions de grappes dont la taille moyenne est de 21,1 résidences principales
- ❑ 200 000 secteurs constitués dont 96 % de 6 grappes et 4 % de 7 grappes.
- ❑ Un nouvel échantillon de 67 800 résidences principales à repérer (+50 %) => **3 211 secteurs à sélectionner.**



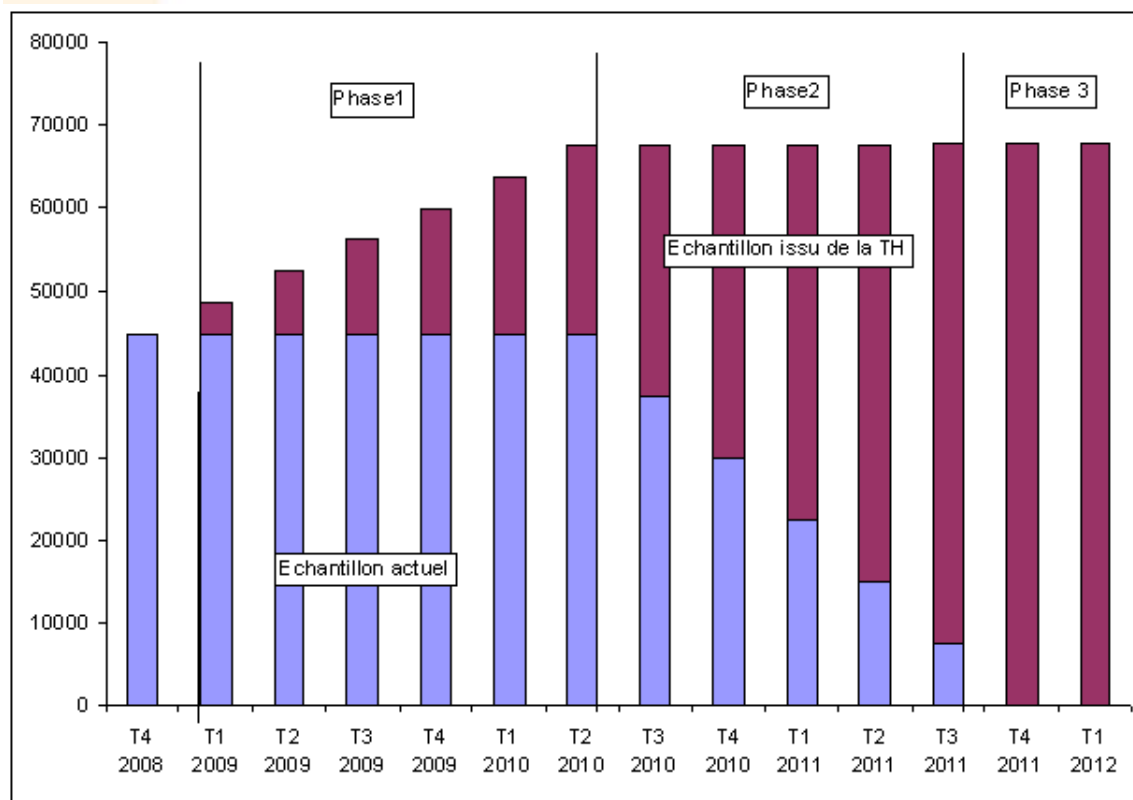
La répartition de l'échantillon par région

- ❑ Actuellement allocation de Neyman avec surreprésentation des petites régions (contraintes Eurostat).
- ❑ Dans le nouvel échantillon la répartition est proportionnelle au nombre de résidences principales avec correction pour les plus petites régions.



Calendrier d'introduction : 3 phases

Nombre total de résidences principales
repérées selon la date et le type d'échantillon



**Du T1 2009 au T2 2010,
1/18ème de l'échantillon est introduit
chaque trimestre**

**Du T3 2010 au T4 T2011 inclus, 1/9ème
de l'échantillon est introduit.**

**A partir du T4 2011, l'échantillon
est exclusivement TH.**

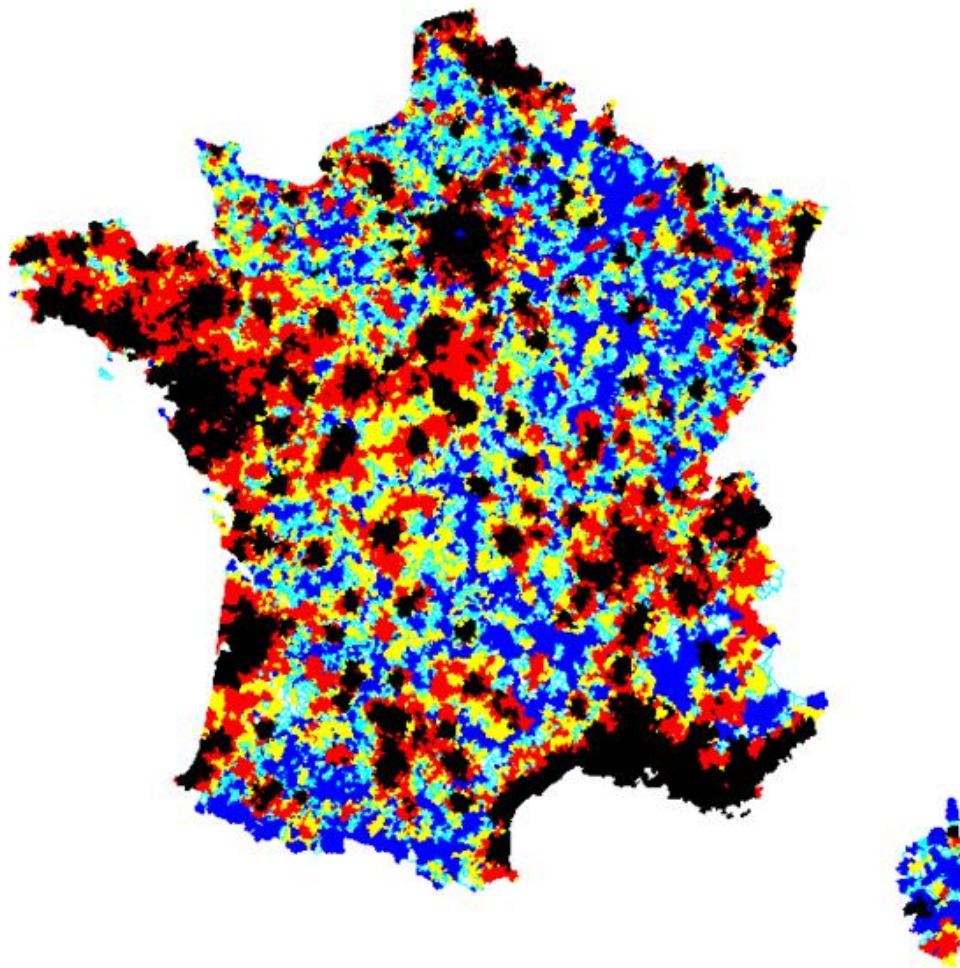
Une sélection descendante de l'échantillon

- ❑ 3 211 secteurs stratifiés par région et équilibré, sur des critères d'âge, revenus, type d'espace, type de logements... Sélection proportionnelle au nombre de résidences principales.
- ❑ Sélection emboîtée de 6 groupes de même taille, probabilité constante et équilibrage (stratification par région avec mise en commun des phases d'atterrissage)(G. Chauvet 2008),
- ❑ Dans chaque groupe partition en 2 sous-échantillons de taille respective $1/3$ et $2/3$ (pour retrouver $1/18^{\text{ème}}$ et $1/9^{\text{ème}}$)
- ❑ Dans chaque secteur attribution aléatoire et équilibrée d'un numéro de 1 à 6 aux grappes pour le rang d'interrogation.

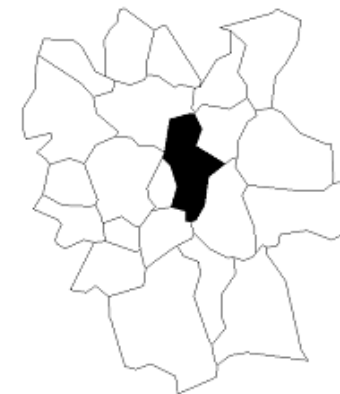
Prise en compte de la charge de travail des enquêteurs



Degré d'isolement des communes : sondage stratifié par région, répartition proportionnelle au nombre de résidences principales, probabilité d'inclusion proportionnelle à la taille, équilibré.



Pour une commune donnée, on calcule, par simulation, la probabilité de trouver dans un environnement proche au moins 3 secteurs pour un plan de sondage donné.

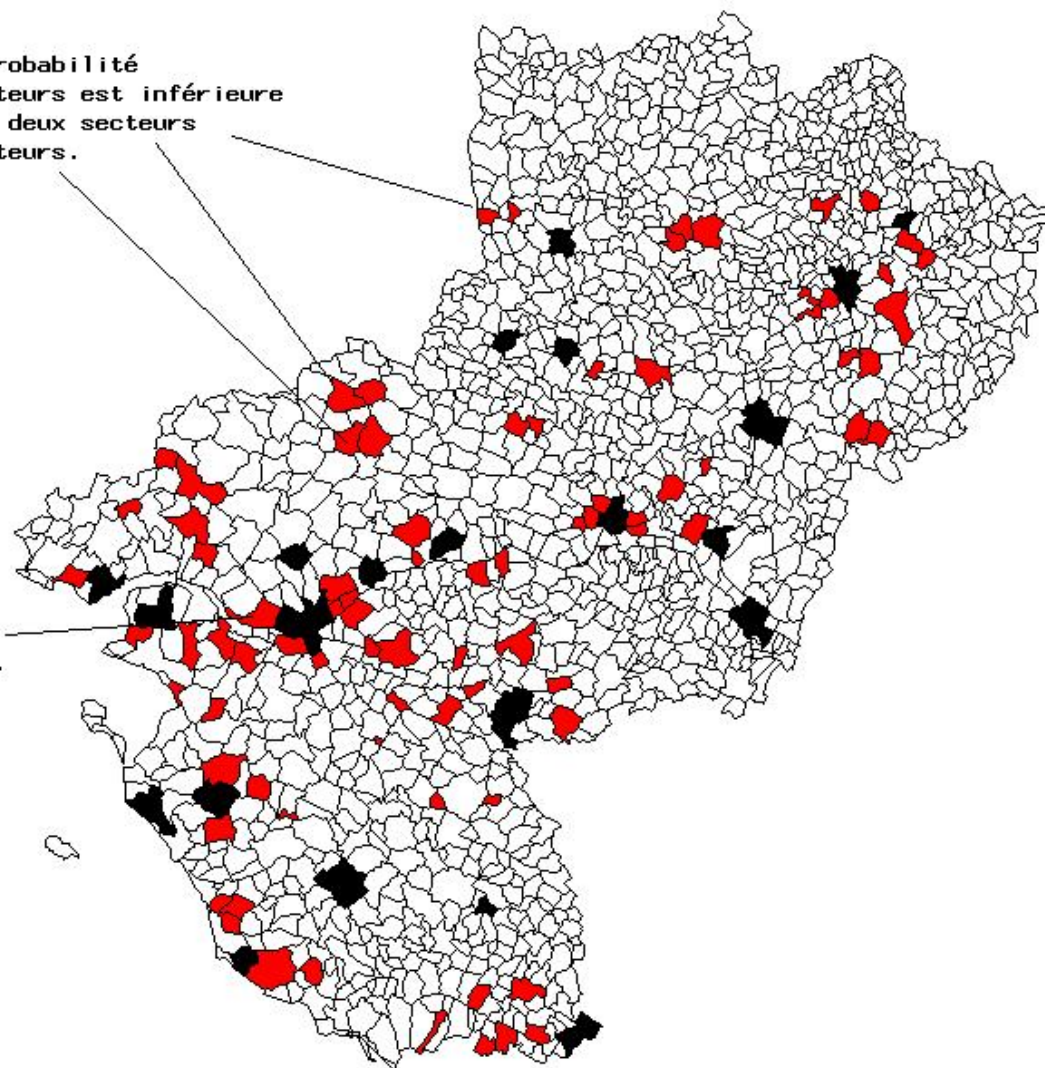


Probabilité
Croissante



Dans les zones où la probabilité d'avoir au moins 3 secteurs est inférieure à 0.95, on sélectionne deux secteurs par canton : les bisecteurs.

Dans les zones où la probabilité d'avoir au moins 3 secteurs est supérieure à 0.95 on sélectionne des secteurs.



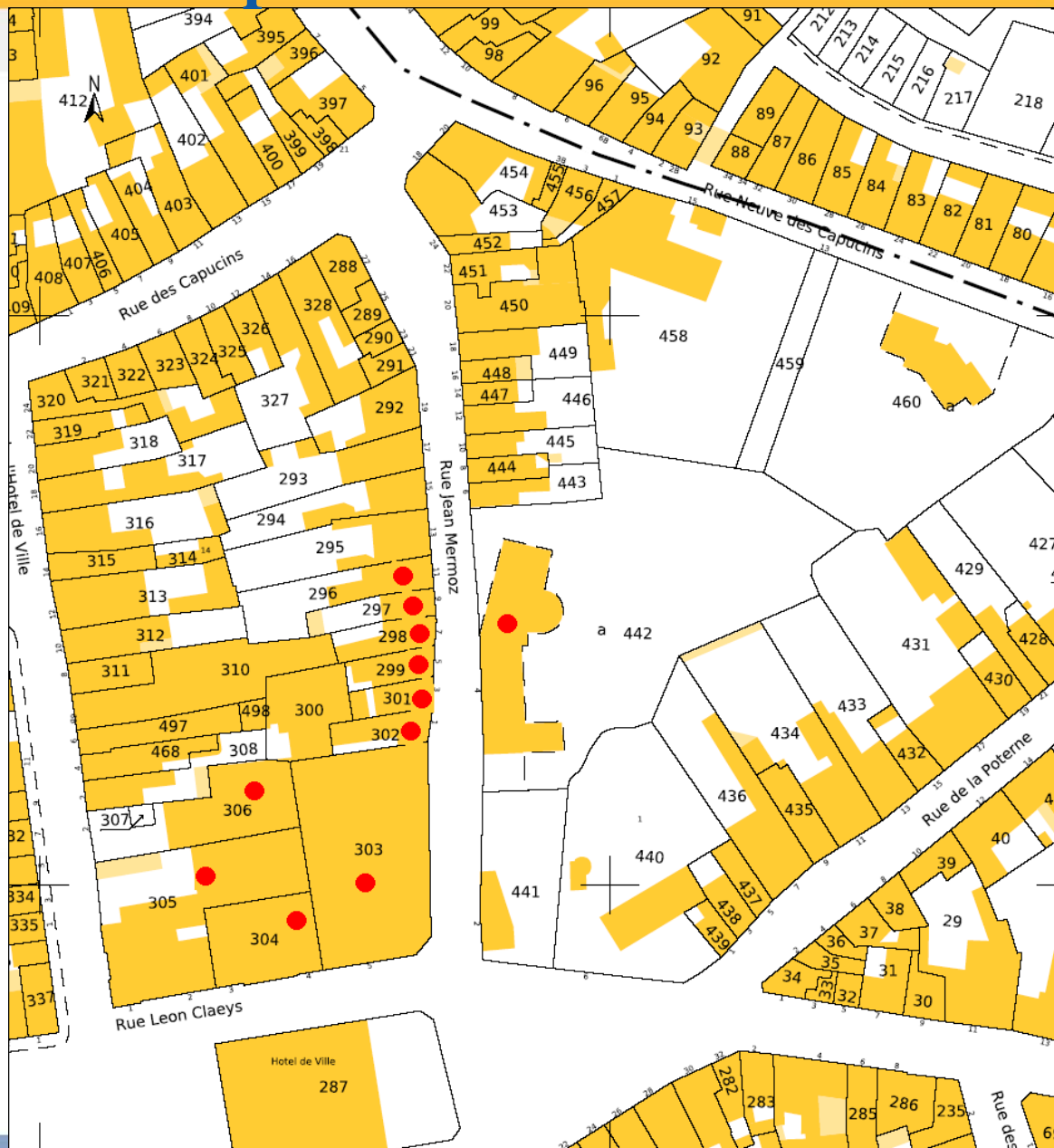
Nbre de secteurs par commune 0 1 2+



procédures post-échantillonnage

- ❑ La sélection des logements non principaux
- ❑ La réallocation
 - Corriger les anomalies géographiques au sein des grappes d'un secteur
 - Mettre en place les documents de collecte,
- ❑ La sélection des logements nouveaux.

Exemple de document de collecte



La sélection des logements nouveaux



- ❑ Par comparaison des identifiants pérennes, on repère les logements nouveaux en N+1,
- ❑ On trie le fichier N+1 selon les critères qui ont prévalu dans la construction des grappes (notion de distance),
- ❑ On associe un logement nouveau au premier logement non nouveau le précédant la commune
- ❑ Si le logement associé est un logement échantillonné, alors le logement nouveau est échantillonné.
- ❑ L'application de la formule du partage des poids montre que la pondération des logements « anciens » se transmet aux logements nouveaux.



▣ Merci pour votre attention...