

Données de recensement à l' infracommunal

Jean-Michel Floch
Département de l'action régionale



Une transition

Jean-Luc Lipatz vous a parlé des méthodes de statistique spatiale, des représentations carroyées utilisant des méthodes de statistique spatiale

Objectif = fournir un ensemble de données et de méthodes permettant de faire des analyses au niveau infra-communal

Le RP, mais pas seulement...

Le RP reste un éléments incontournable du dispositif d'information

A côté du RP des sources administratives

Riches

Parfois plus facilement géolocalisables que le RP

Exemple = Revenus fiscaux, Cnam,

Les IRIS, mais pas seulement

Les IRIS constituent le support d'information privilégié sur le site insee.fr, pour différentes sources (RP, RFL, BPE)

Mais il y a aussi

Les carroyages

DIAF, pour des exploitations spécifiques visant à pallier la disparition de l'îlot

Combiner les sources

Estimations mixtes

RP + Logements + Sources administratives

Des travaux en ce sens pour des objectifs bien particuliers (politique de la ville)

La plan

La précision dans le recensement (généralités)

Comment est pris en compte l'espace dans l'échantillon

Diaf-RP

Recensement « rénové » et précision (d'après la présentation de G.Brilhault à la SFDS)

Introduction

Nombreux facteurs influant sur la qualité des résultats de tout recensement traditionnel :

non-réponse totale, redressements des variables collectées pour non-réponse ou incohérence, erreurs de codifications, etc.

Introduction du sondage dans le recensement rénové en France : « choc culturel »

nombreux travaux sur la précision liée à l'introduction du sondage

notamment présentation au séminaire SFdS consacré au recensement rénové du 25 juin 2002

Objectif de la présentation :

point sur les différents travaux portant sur la précision du RP rénové et données de précision du recensement disponibles aujourd'hui

Plan de la présentation

- I. Rappel succinct du plan de sondage du RP rénové
- II. Evolution des méthodes de calculs de précision du RP
- III. Résultats obtenus et diffusion de ces résultats

I. Rappel succinct sur le plan de sondage du RP

On se restreint ici à la métropole et au champ des ménages

Exhaustif dans les communes de moins de 10 000 habitants (PC) :
réparties en 5 groupes équilibrés
enquête exhaustive de chaque groupe tous les 5 ans

Échantillonnage dans les communes de plus de 10 000 habitants (GC) :

répartition des adresses en 5 groupes équilibrés
grandes et nouvelles adresses enquêtées exhaustivement
chaque année, enquête par sondage des logements des « autres adresses » d'un groupe

→ au total, 40% des logements des grandes communes sont enquêtés en 5 ans

Méthode d'estimation dans ce contexte

Des enquêtes de recensement annuelles depuis 2004

Chaque année depuis 2008 (pour le RP2006) :

utilisation des 5 dernières années d'enquêtes consécutives pour :

connaître la population légale de toutes les communes

pour l'année de référence = l'année médiane

diffuser des résultats détaillés sur les variables collectées

Comment ?

en grandes communes : nombre moyen de personnes par logement sur 5 ans * nombre de logements au 1^{er} janvier de l'année médiane

(calage au niveau iris)

en petites communes : interpolation ou extrapolation pour ramener toutes les petites communes à une même date de référence

Conséquence en termes de précision

Le nouveau facteur en termes de précision provient
du sondage concernant la population des ménages
vivant dans les « autres adresses » (ni grandes, ni nouvelles)
dans les communes de plus de 10 000 habitants

Le reste de la population continue à être enquêté exhaustivement
les petites communes enquêtées exhaustivement sur 5 ans
les communautés enquêtées exhaustivement sur 5 ans
les grandes et les nouvelles adresses des grandes communes
enquêtées exhaustivement sur 5 ans
les autres populations (habitations mobiles et sans abris,
mariniers) collectées exhaustivement une fois tous les 5 ans

II. Les estimations de précision du RP avant 2009

Reposaient sur les données du RP1999

dans le cadre des réflexions sur le RP rénové
et en l'absence de données d'un cycle du RP rénové complet

Etaient basées sur des simulations

Plusieurs expériences en ce sens, variant notamment selon :

- la méthodologie retenue (strate des grandes adresses, calage, etc.)
pour mettre au point la définition de la strate des grandes adresses
- les variables d'intérêt (population seule ou jeu de plusieurs variables collectées au RP) pour anticiper la précision de la diffusion
- le nombre d'échantillons simulés
- le niveau géographique retenu (plusieurs communes, quelques iris) pour anticiper la précision de la diffusion

II. Les estimations de précision du RP avant 2009

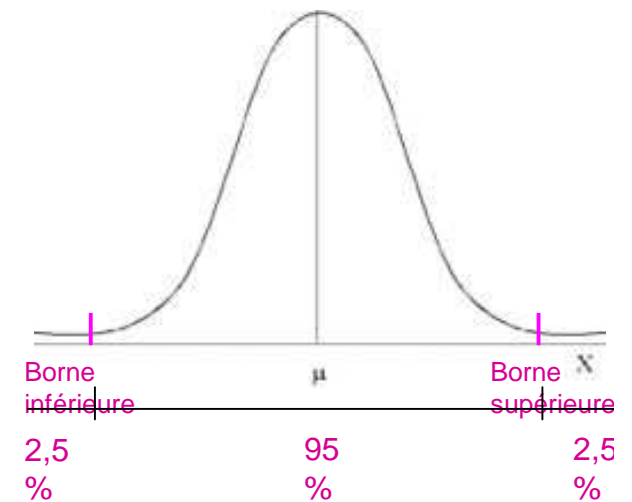
Si l'on appelle pop la population des ménages, la « vraie » valeur de pop a un peu plus de 95 % de chance de se trouver dans l'intervalle de confiance [pop - 2*pop*CV(pop) ; pop + 2*pop*CV(pop)].

Exemple : pop (18-24ans) = 250

CV = 10% → 95% de chance d'avoir

$$200 < \text{pop (18-24ans)} < 300$$

:



II. Les estimations de précision du RP avant 2009

Parmi les résultats issus de ces simulations :

les simulations de référence = rapport CNIS décembre 2005 :

La précision des résultats pour les communes de 10 000 habitants ou plus

Tranches d'effectif	Précision (CV)
50 000 ou plus	< 1,0 %
20 000 - 49 999	1,5%
10 000 - 19 999	2,0%
6 000 - 9 999	2,5%
3 000 - 5 999	3,0%
2 000 - 2 999	3,5%
1 000 - 1 999	4,5%
500 - 999	6,0%
250 - 499	8,0%
Moins de 250	> 8,0 %

Source : Rapport du CNIS « Utilisation des données produites par le recensement rénové de la population et leur diffusion » - Décembre 2005

Un contexte différent à partir de 2009

En 2009, le contexte évolue :

données du premier cycle complet du RP rénové disponibles
analyse des données des données à l'infra-communal dans la
perspective de leur diffusion fin 2009 (données du RP2006)

calcul de précision au niveau iris pour la métropole

basée sur les formules statistiques sur les échantillons équilibrés issues
des travaux de J.C. Deville et Y. Tillé (résidus)

tenant compte de l'équilibrage du tirage et du calage à l'iris

sans prise en compte de l'erreur de mesure temporelle (étalement sur 5
ans des collectes utilisées pour calculer les populations légales d'un
millésime RP)

Calculs avec ce nouvel outil depuis 2009

Ce nouvel outil permet :

des calculs plus systématiques : à différents niveaux (commune, iris), sur toutes les variables
une confirmation / un affinement des résultats de précision donnés précédemment par les simulations

Outil utilisé au département de la Démographie :

fin 2009 : pour toutes les variables du RP2006 au niveau iris
début 2010 : pour la variable population du RP2006 au niveau région, département, communes

Outil complété en 2010 pour le projet DIAF (diffusion du RP sur des zones à façon)

inclus un calage supplémentaire au niveau de la ZAF

III. Les résultats obtenus sur données RP2006

Calculs (début 2010) des CV de la variable population (des ménages) aux niveaux national, régional et départemental :

la population française est connue à + ou – 15 800 personnes
($15800 \cdot 2 / \text{pop française} = 0,05\%$)

	Distribution du CV de la variable population au niveau région	Distribution du CV de la variable population au niveau département
75% Q3	0,20	0,53
50% Médiane	0,16	0,35
25% Q1	0,12	0,25

III. Les résultats obtenus sur données RP2006 (3)

Calcul au niveau des communes de métropole :

pour 876 GC sur 892 au RP2006, la précision obtenue est meilleure que celle estimée pour le rapport CNIS de décembre 2005 (98% des GC)

distribution :

	Communes de 10000 à 19999 habitants	Communes de 20000 à 49999 habitants	Communes de 50000 à 100000 habitants	Communes de 100000 habitants et plus
	(452 communes)	(318 communes)	(79 communes)	(43 communes)
75% Q3	1,16	0,87	0,71	0,43
50% Médiane	1,02	0,78	0,56	0,39
25% Q1	0,90	0,68	0,50	0,34

Diffusion actuelle de ces données de précision

CV de la variable population du RP2006 de niveaux régional, départemental, communal :

interne à l'Insee pour l'instant, diffusion plus large prévue
information communiquée à la CNERP récemment

CV de niveau iris calculés pour toutes les variables du RP2006 :

CV de la variable population utilisés pour donner un label aux iris
en diffusion, pour qualifier l'utilisation possible des données de
niveau iris

CV pas diffusés tels quels mais sous forme de CV « résumé »
pour une meilleure robustesse des résultats

sur insee.fr : <http://www.recensement.insee.fr/basesInfracommunales.action>

Diffusion actuelle de ces données de précision

Exemple des données de niveau iris diffusées :

IRIS	Région	Département	Unité urbaine	Commune ou ARM	Libellé commune ou ARM	TRIRIS	Grand quartier	Libellé de l'IRIS	Type d'IRIS	Modification de l'IRIS	Label de l'IRIS en 2008	Population en 2008 (princ)
IRIS	REG	DEP	UU2010	COM	LIBCOM	TRIRIS	GRD_QUART	LIBIRIS	TYP_IRIS	MODIF_IRIS	LAB_IRIS	P08_POP
010010000	82	01	01000	01001	L'Abergement-Clémenciat	ZZZZZZ	0100100	L'Abergement-Clémenciat	Z	00	5	791
010020000	82	01	01000	01002	L'Abergement-de-Varey	ZZZZZZ	0100200	L'Abergement-de-Varey	Z	00	5	194
010040101	82	01	01302	01004	Ambérieu-en-Bugey	ZZZZZZ	0100401	Les Perouses-Triangle d'Activite	H	00	1	1774
010040102	82	01	01302	01004	Ambérieu-en-Bugey	ZZZZZZ	0100401	Longeray-Gare	H	00	1	3333
010040201	82	01	01302	01004	Ambérieu-en-Bugey	ZZZZZZ	0100402	Centre-St Germain-Vareilles	H	00	1	3396
010040202	82	01	01302	01004	Ambérieu-en-Bugey	ZZZZZZ	0100402	Tiret-Les Allymes	H	00	1	4288
010050000	82	01	01000	01005	Ambérieux-en-Dombes	ZZZZZZ	0100500	Ambérieux-en-Dombes	Z	00	5	1567
010060000	82	01	01000	01006	Ambléon	ZZZZZZ	0100600	Ambléon	Z	00	5	123
010070000	82	01	01000	01007	Ambronay	ZZZZZZ	0100700	Ambronay	Z	00	5	2296
010080000	82	01	01302	01008	Ambutrix	ZZZZZZ	0100800	Ambutrix	Z	00	5	657
010090000	82	01	01000	01009	Andert-et-Condon	ZZZZZZ	0100900	Andert-et-Condon	Z	00	5	331
010100000	82	01	01000	01010	Anglefort	ZZZZZZ	0101000	Anglefort	Z	00	5	926
010110000	82	01	01000	01011	Apremont	ZZZZZZ	0101100	Apremont	Z	00	5	339
010120000	82	01	01000	01012	Aranc	ZZZZZZ	0101200	Aranc	Z	00	5	285
010130000	82	01	01000	01013	Arandas	ZZZZZZ	0101300	Arandas	Z	00	5	164
010140000	82	01	01401	01014	Arbent	ZZZZZZ	0101400	Arbent	Z	00	5	3433
010150000	82	01	01000	01015	Arbignieu	ZZZZZZ	0101500	Arbignieu	Z	00	5	473
010160000	82	01	01000	01016	Arbigny	ZZZZZZ	0101600	Arbigny	Z	00	5	368
010170000	82	01	01000	01017	Argis	ZZZZZZ	0101700	Argis	Z	00	5	421
010190000	82	01	01000	01019	Armix	ZZZZZZ	0101900	Armix	Z	00	5	23

Précision (coefficient de variation en %) selon la tranche d'effectif de la variable

VAR00_49	VAR50_99	VAR100_199	VAR200_299	VAR300_499	VAR500_699	VAR700_999	VAR1000_1999	VAR2000_2999	VAR3000P
> 3	> 3	> 3	> 3	> 3	> 3	> 3	3	3	2

Diffusion actuelle de ces données de précision

Exemple des données de niveau iris diffusées :

Données infra-communales - Diplômes - Formation							
France - IRIS							
Découpage géographique au 01/01/2009							
© Insee		Source : Insee, Recensement de la population 2007 exploitation principale.					
IRIS	Région	Libellé commune ou ARM	Libellé de l'IRIS	Type d'IRIS	Label de l'IRIS en 2007	Pop scolarisée 2-5 ans en 2007 (princ)	Pop scolarisée 6-14 ans en 2007 (princ)
IRIS	REG	LIBCOM	LIBIRIS	TYP_IRIS	LAB_IRIS	P07_SCOL0205	P07_SCOL0614
010040101	82	Ambérieu-en-Bugey	Les Perouses-Triangle d'Activite	H	1	56	110
010040102	82	Ambérieu-en-Bugey	Longeray-Gare	H	1	117	393
010040201	82	Ambérieu-en-Bugey	Centre-St Germain-Vareilles	H	1	114	345
010040202	82	Ambérieu-en-Bugey	Tiret-Les Allymes	H	1	152	559

Données infra-communales - Diplômes - Formation							
Précision des variables							
Pour les IRIS appartenant à une commune de France métropolitaine de 10 000 habitants ou plus							
© Insee		Source : Insee, Recensement de la population 2007 exploitation principale.					
Variable	Libellé	Précision (coefficient de variation en %) selon la tranche d'effectif de la variable					
VAR_ID	VAR_LIB	VAR00_49	VAR50_99	VAR100_199	VAR200_299	VAR300_499	VAR500_699
P07_SCOL0205	Pop scolarisée 2-5 ans	> 18	18	13	8	< 8	so
P07_SCOL0614	Pop scolarisée 6-14 ans	> 18	18	15	11	9	7

Quelques compléments : tirage d'échantillon et contraintes spatiales

La méthode du Cube : une idée simple

La macro CUBE est un algorithme d'échantillonnage qui réalise des tirages équilibrés : elle assure le choix aléatoire d'un échantillon apte à restituer les vraies structures de la base de sondage pour des informations auxiliaires données. En outre, cette méthode améliore également, parfois de manière notable, la précision des résultats obtenus par l'enquête.

Le concept d'échantillon équilibré est simple et intuitif : un échantillon est dit équilibré sur une ou plusieurs variables disponibles dans la base de sondage, lorsque pour chacune d'entre elles, l'estimateur de Horvitz-Thompson³ du total coïncide exactement avec le vrai total issu de la base de sondage.

Un échantillon S équilibré sur la variable de contrôle X respecte donc la contrainte suivante :

$$\sum_{k \in S} \frac{X_k}{\pi_k} = \sum_{k=1}^N X_k$$

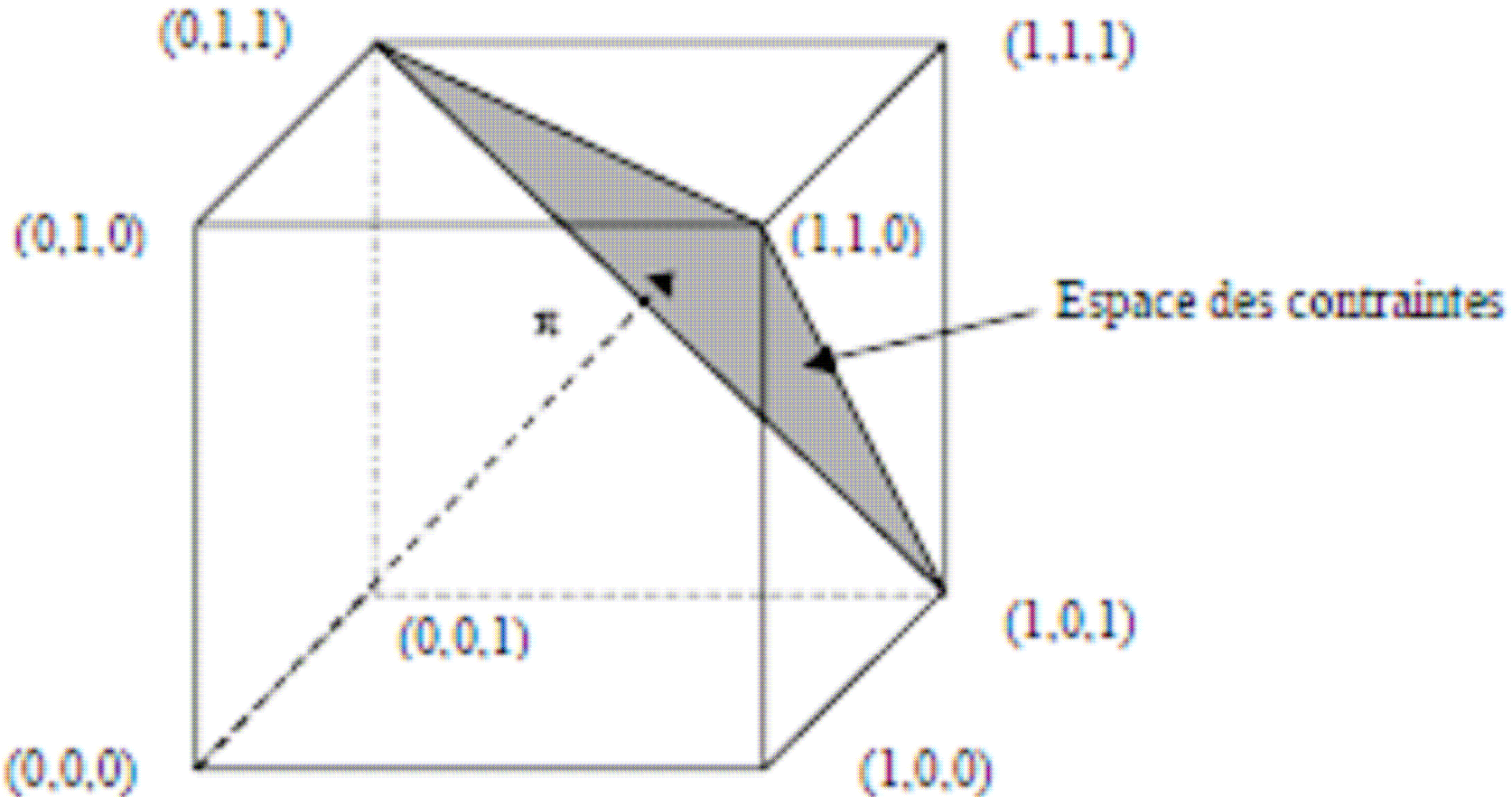
où pour tout individu k de la base de sondage ($k = 1$ à N), π_k désigne sa probabilité d'être sélectionné dans S et X_k la valeur qui lui est associée pour la variable X .

Une formalisation mathématique complexe

La méthode du Cube tire son nom d'une représentation géométrique des différents échantillons possibles d'une base de sondage de taille N dans un hypercube de dimension N . Chacun des sommets de l'hypercube désigne l'un des 2^N échantillons sans remise possibles, y compris l'ensemble vide. Les N coordonnées d'un sommet sont définies par les valeurs des indicatrices d'appartenance à l'échantillon de chacun des N individus. Par exemple, le sommet de coordonnées $(1,0,\dots,0,1)$ désigne l'échantillon de taille 2 composé du premier et du dernier individu de la base de sondage.

L'algorithme génère une marche aléatoire dans cet hypercube qui part des probabilités d'inclusion individuelles fixées a priori, et qui arrive à un sommet symbolisant l'échantillon final. Il se décompose en deux phases successives dites de vol et d'atterrissage. La deuxième étape ne se déclenche que dans le cas où le vol n'aurait pas abouti à un sommet respectant exactement les contraintes et permet alors de désigner l'échantillon final. En pratique, cette seconde phase se déclenche dans la plupart des cas, pour résoudre des problèmes d'arrondis notamment.

Le schéma



LA VARIANCE DANS LE CADRE DU CUBE

La méthode du cube permet de faire des estimations de variance (par régression de la variable d'intérêt sur les variables de calage)

$$V(\hat{t}_{y\pi}) = \sum_h \sum_{k \in Sh} \frac{b_k}{\pi_k} (y_k - \beta_k' z_k)^2 \quad b_k = \pi_k(1 - \pi_k)$$

La variance d'échantillonnage, sous de bonnes hypothèses (réalistes) est égale à la somme des carrés des résidus de la régression de la variable d'intérêt sur les variables d'équilibrage

D'autres résultats sont obtenus par des méthodes de rééchantillonnage (cf. pages suivantes)

Deux phases dans l'algorithme

Phase de vol

Un sommet du cube

Phase d'atterrissage

Relâchement des contraintes

L'équilibrage initial

Les variables d'équilibrage

Objectif = les résultats communaux, les populations légales

De 2004 à 2008, les critères d'équilibrage sont les mêmes que ceux qui ont servi à constituer les groupes de rotation : le nombre de logements total, le nombre de logements collectifs, le sexe, la répartition par groupes d'âges au recensement de 1999, auxquels s'ajoute le poids actuel des IRIS en nombre de logements qui, lui, n'a pas été utilisé dans la constitution des groupes de rotation. Ce dernier critère est là pour s'assurer que l'échantillon est bien réparti sur tout le territoire de la commune. Toutefois, il intervient en dernier et donc n'est pas toujours très opérant. Le tirage ne garantit pas, au niveau de l'IRIS, un taux de sondage annuel des logements de 8%.

L'équilibrage actuel

Les variables d'équilibrage

Socio-démographiques

A partir de l'enquête de 2009, le plan de sondage est modifié : l'équilibrage n'est plus effectué que sur le nombre de logements, le nombre de logements en collectif et le nombre de logements à l'IRIS issus du dernier recensement, car les variables du recensement de 1999 ne sont plus utilisées. En effet, les petites adresses nouvelles ayant été enquêtées lors du premier cycle sont devenues des petites adresses connues et à ce titre participent désormais au tirage de l'échantillon. Or, pour elles, on ne dispose pas de données démographiques de 1999 car, n'existant pas à l'époque, elles n'ont pas été enquêtées lors du RP1999. Cela contribue à un meilleur équilibre du nombre de logements entre les IRIS.

Spatiales

IRIS, ce qui veut dire des contraintes nombreuses, et qui donc seront les plus susceptibles d'être relâchées en cas de difficultés

=> quid des zones qui ne sont pas des Iris (ZUS, ZAF, carreaux)

La réalisation des contraintes spatiales

Les IRIS, ce qui veut dire des contraintes nombreuses, et qui donc seront les plus susceptibles d'être relâchées en cas de difficultés

=> quid des zones qui ne sont pas des Iris (ZUS, ZAF, carreaux)

Position du problème

La représentativité spatiale

Equilibrage à l'Iris

Qu'en est-il des autres zones

On va construire une petite expérience de pensée

Un exemple simplifié

Une ville fictive à 10 IRIS ne comprenant que des adresses ordinaires (constitué à partir de vraies observations)

Données spatialisées (IRIS, xy, carroyages de tailles diverses)

3 types de logement

Individuel

Collectif

Social

Pas de groupes de rotation

Un échantillon de 40% équilibré sur les 3 types et les IRIS => 13 contraintes d'équilibrage, tiré à l'aide de la macro CUBE

Plusieurs maillages : Iris, carreaux de 50 à 1000 m

Simplifié mais pas trop simpliste

La vérification des contraintes

Le résultat du tirage est le suivant :

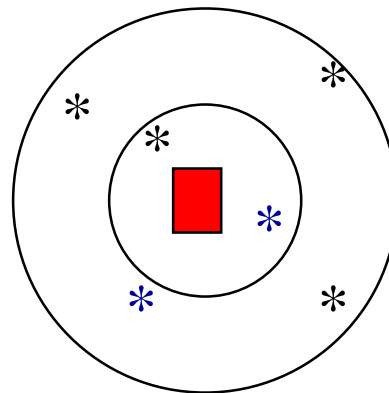
IA Individuel
IB Collectif
IC Social
I04xx Iris 04xx

Valeur réelle		Valeur estimée		Ecart relatif (en %)	
IA	3394	IA	3392.5	IA	-0.04
IB	13452	IB	13455	IB	0.02
IC	2826	IC	2857.5	IC	1.11
I0401	2245	I0401	2222.5	I0401	-1.00
I0402	3250	I0402	3240	I0402	-0.31
I0403	1376	I0403	1372.5	I0403	-0.25
I0404	2442	I0404	2447.5	I0404	0.23
I0405	1203	I0405	1200	I0405	-0.25
I0406	2065	I0406	2112.5	I0406	2.30
I0407	1652	I0407	1627.5	I0407	-1.48
I0408	1958	I0408	1952.5	I0408	-0.28
I0409	2159	I0409	2170	I0409	0.51
I0410	1322	I0410	1360	I0410	2.87

Adéquation échantillon/population

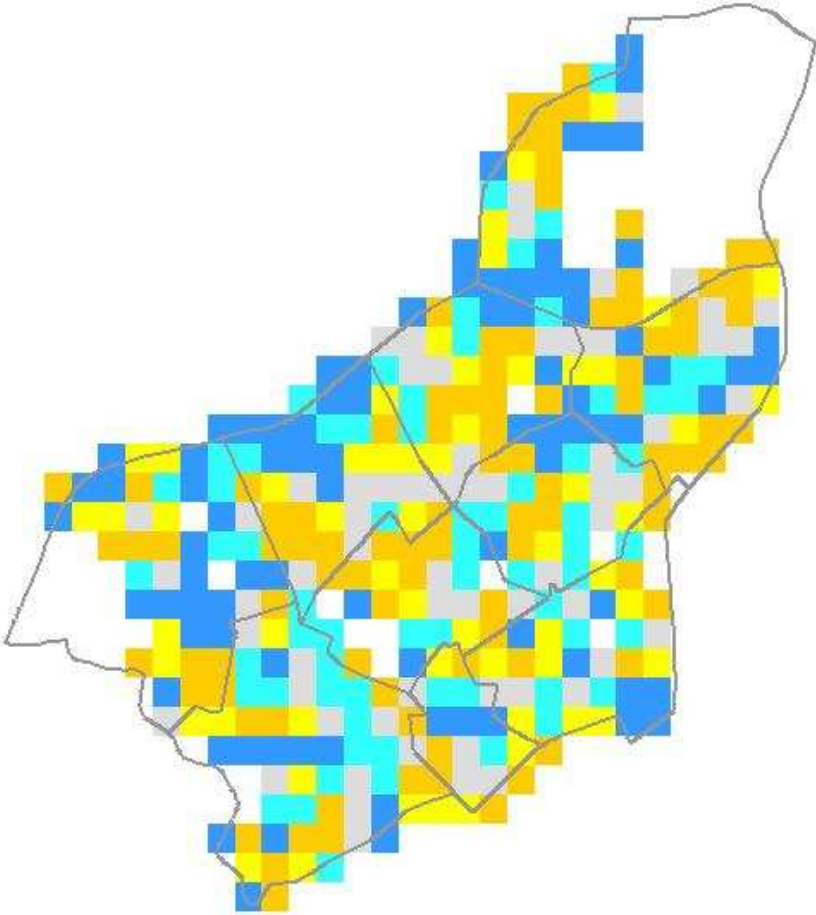
On calcule pour chaque point d'une grille le rapport entre le nombre de logements dans un rayon de r m et le nombre de logement de l'échantillon

On fait varier r pour voir à partir de quel moment l'échantillon est spatialement proche de la base de logement



100 m

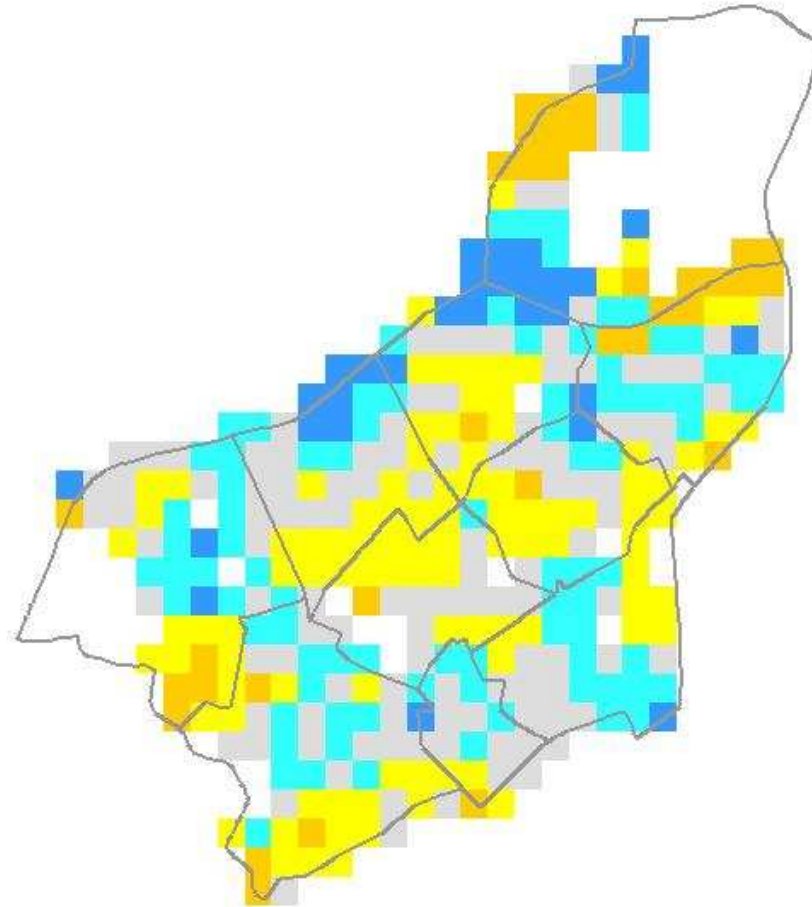
Sur/sous échantillonnage à 100 m



rap >> > = < <<
INSEE-RP 2007

200 m

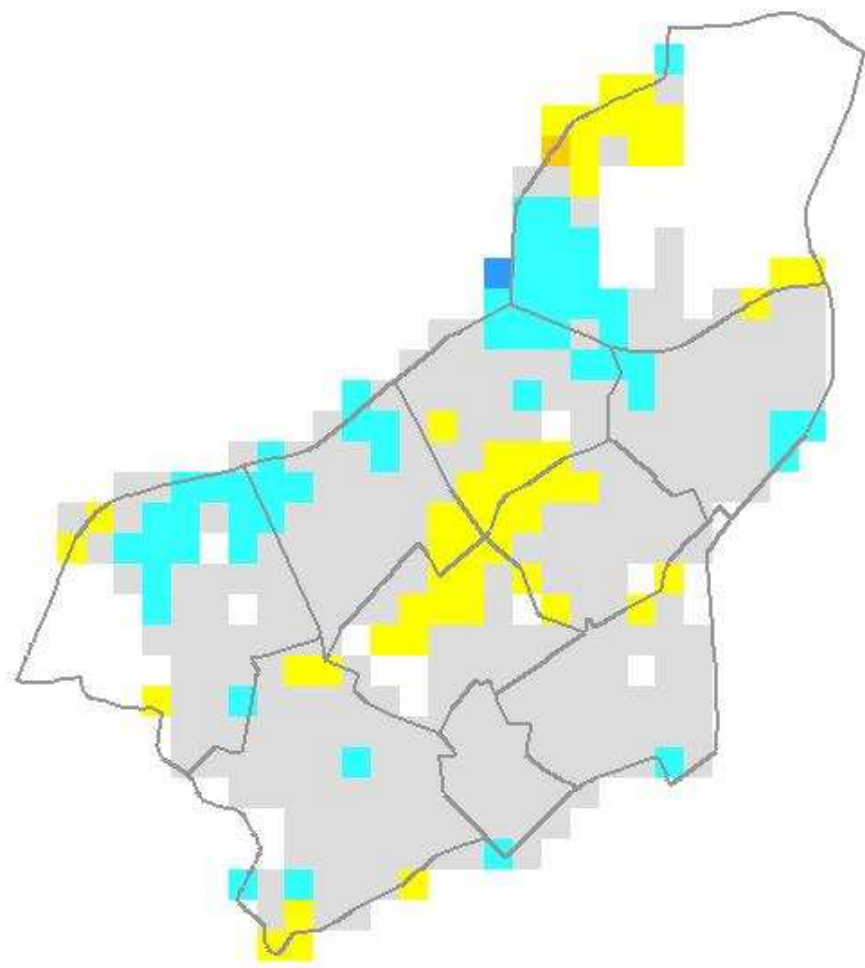
Sur/sous échantillonnage à 200 m



rap >> > = < <<
INSEE-RP 2007

400 m

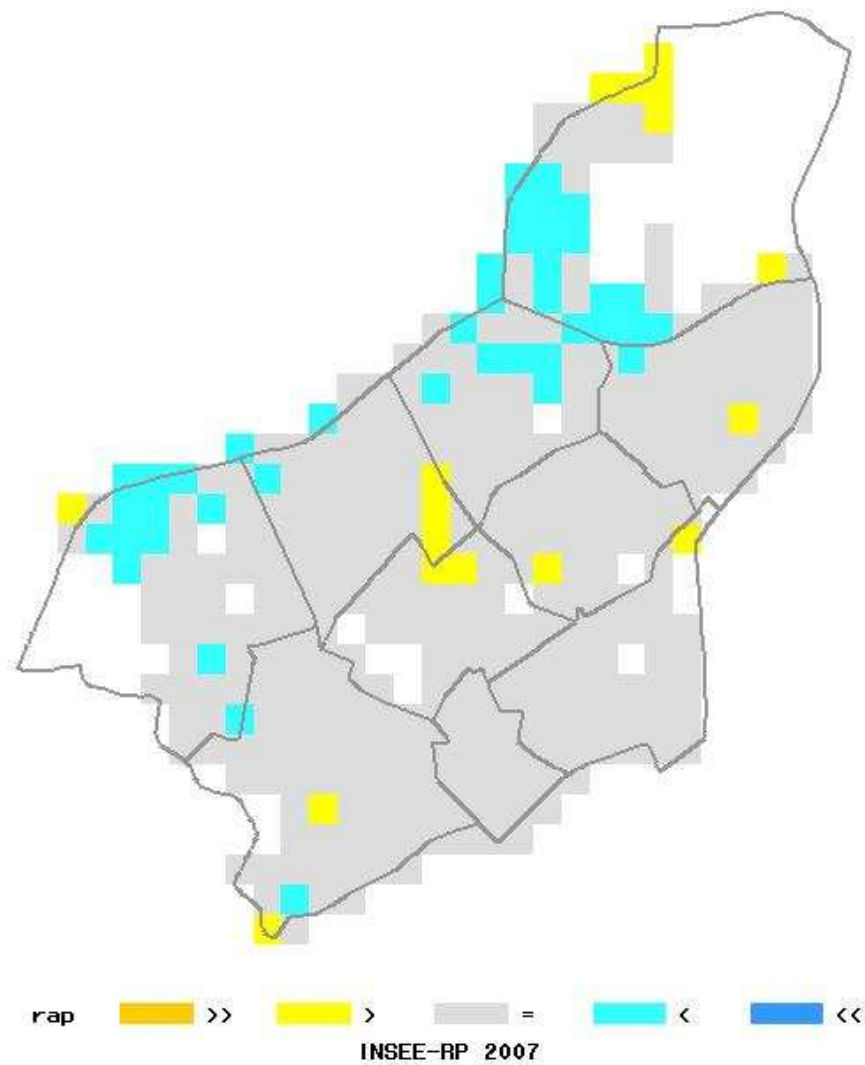
Sur/sous échantillonnage à 400 m



rap ■ >> ■ > ■ = ■ < ■ <<
INSEE-RP 2007

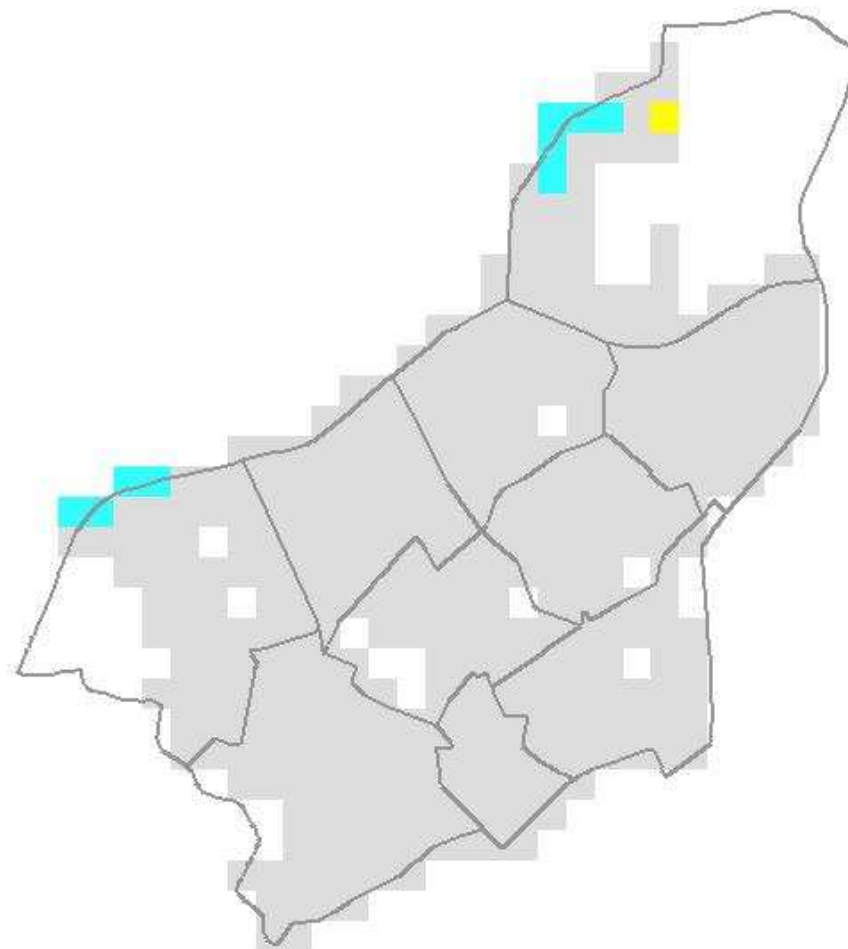
500 m

Sur/sous échantillonnage à 500 m



750 m

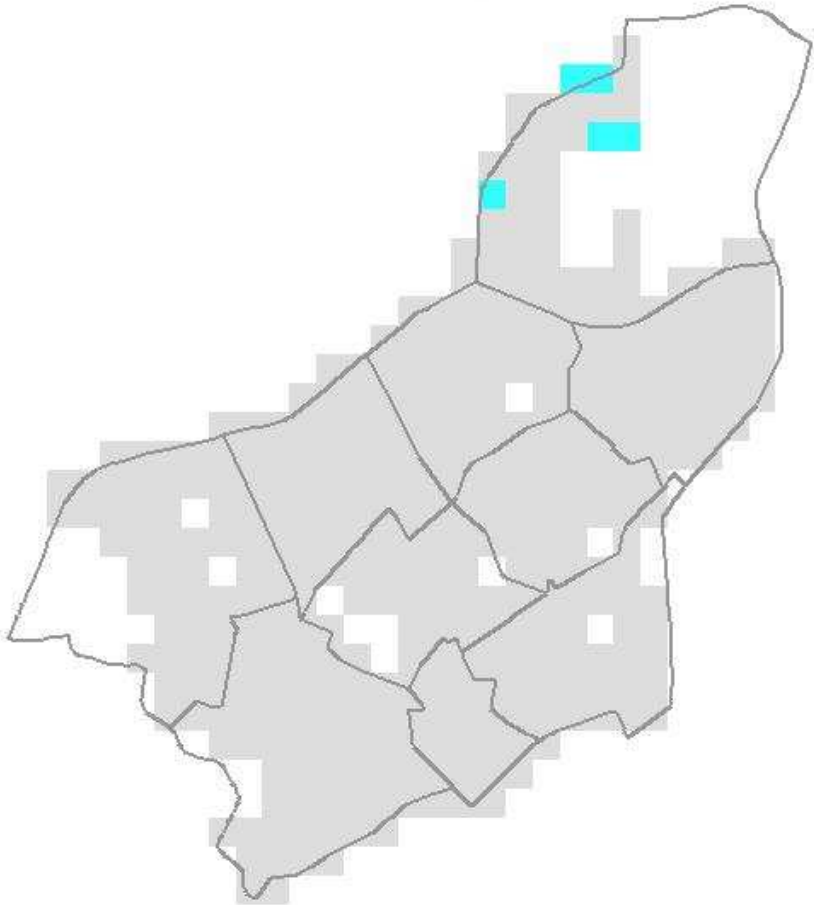
Sur/sous échantillonnage à 750 m



rap >> > = < <<
INSEE-RP 2007

1000 m

Sur/sous échantillonnage à 1000 m



rap >> > = < <<
INSEE-RP 2007

Les distributions des surreprésentations locales

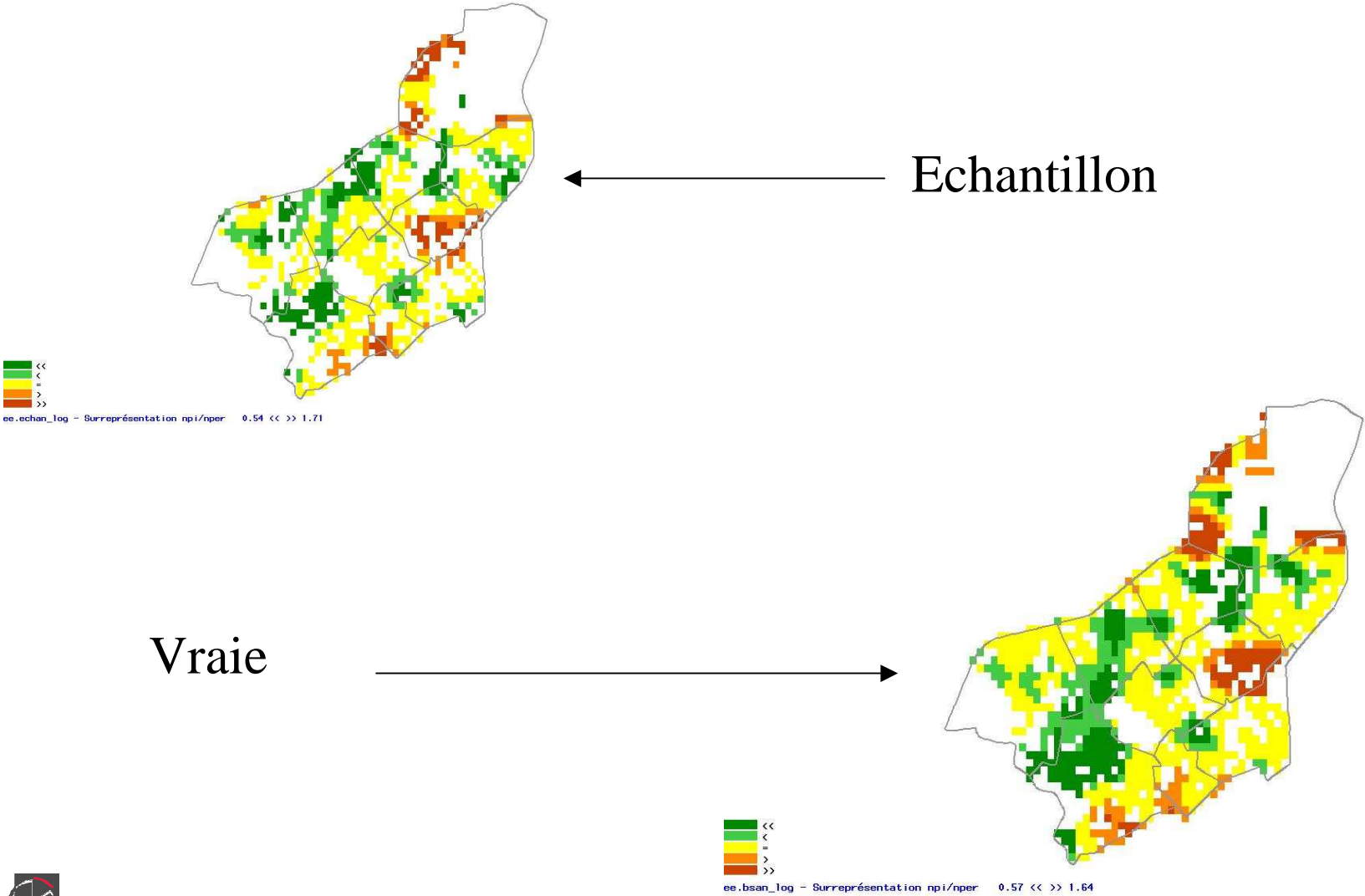
	100 m	200 m	400 m	500 m	750 m	1000 m
Max	2,25	1,57	1,23	1,10	1,08	1,03
P90	1,31	1,16	1,07	1,04	1,02	1,02
Q3	1,17	1,08	1,03	1,02	1,01	1,01
Med	1	1	1	1	1	1
Q1	0,81	0,91	0,97	0,97	0,99	0,99
P10	0,63	0,83	0,93	0,95	0,98	0,98
0,19	0,19	0,43	0,77	0,88	0,89	0,93

Des limites pour l'analyse locale

id200	Popula tio		Pop.immig rée	
	Es tima tio n	CV	Es tima tio n	CV
178	984	3,4	154	13,1
158	854	3,2	30	24,0
146	784	3,6	53	18,5
140	723	3,2	67	19,5
105	708	4,1	60	21,2
110	697	3,9	32	21,9
192	689	2,8	34	21,4
195	641	4,7	84	18,3
56	637	4,0	20	33,3
157	635	4,4	72	21,0
142	628	4,0	22	28,5
141	619	2,9	17	32,6
161	617	5,2	84	26,5
196	612	3,8	19	28,8
213	602	4,6	56	21,3
127	576	5,2	44	37,6
193	576	3,2	58	20,7
165	540	3,9	103	12,0
233	535	5,4	60	23,0
126	534	5,0	42	27,9
214	532	3,8	14	30,7
124	518	4,3	39	19,9

id200	Popula tio		Pop.immig rée	
	Es tima tio n	CV	Es tima tio n	CV
77	497	4,1	11	33,2
91	495	4,2	20	24,4
59	476	4,2	8	55,9
111	463	5,5	16	34,1
129	458	5,2	8	38,8
88	452	5,7	51	23,8
73	451	4,4	16	38,7
72	445	3,8	0	
144	440	5,6	22	34,0
41	427	5,1	27	23,6
109	422	4,9	23	40,1
76	421	5,0	22	25,3
176	419	5,5	59	17,6
93	401	5,8	18	34,2
106	393	4,9	17	26,5
198	391	4,6	30	19,2
42	391	5,9	16	35,5
159	388	4,8	7	38,6
145	386	5,7	25	31,5
74	378	4,5	18	29,3
107	373	5,8	13	47,3
90	372	5,5	5	52,2

Une vision globale satisfaisante





DIAF-RP

Qu'est ce qu'une zone à façon (ZAF) ?

Une zone à façon (ZAF) est un ensemble d'adresses récupérées à partir d'un contour tracé par l'utilisateur sur un site dédié.

A la différence d'un IRIS ou d'un quartier de la politique de la ville, cette zone n'est pas prédéfinie.

Pour rappel, 2 types de zones (séquence 1) :

- Zone isolée d'un seul tenant au sein d'une commune
- Partition d'une commune

Zones isolées ou partitions

Données infracommunales à façon du recensement de la population

INSEE Informations générales Aide Documentation

Chercher une rue : Chercher **Lyon 6e Arrondissement**

Affichage

- Contour
- Iris
- Nom de rues
- Immeubles

Vérifier si votre zone atteint 1000 logements Enregistrer votre travail sans validation

diatrp - v1.1.0 Mentions légales et crédits | © INSEE

Données infracommunales à façon du recensement de la population

INSEE Informations générales Aide Documentation

Message d'information

Le polygone a été importé au sein du contour de travail. Tout vide compris entre le contour de travail et des polygones intérieur a été comblé par des polygones.

Chercher une rue : Chercher **Gennevilliers**

Affichage

- Contour
- Iris
- Nom de rues
- Immeubles

Vérifier un élément de la partition Vérifier l'intégralité de la partition Enregistrer votre travail sans validation

Pour des zones de plus de 1000 logements

Le champ des résultats fournis pour les ZAF

- Des données issues du RP :
 - Pour l'ensemble des logements ordinaires et des individus vivant en communautés
 - Exclus : les sans-abris, les personnes vivant dans des habitations mobiles, les mariniers
 - Issues de l'exploitation principale et complémentaire
 - Données diffusées au lieu de résidence

Calcul des résultats des ménages ordinaires (1/2)

- Population des ménages ordinaires issu du « RIL Habitation » enquêtée par sondage
- Affectation d'un poids pour que les personnes recensées comptent pour celles non recensées
- Poids de diffusion des résultats standards du RP calés au niveau IRIS

Calcul des résultats des ménages ordinaires (2/2)

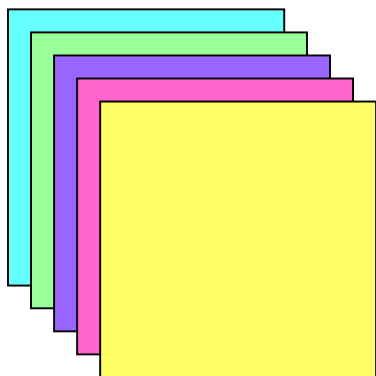
- Chaque logement et individu du logement a un poids de diffusion qui est le produit de deux poids :
 - poids de sondage (type d'adresses)
 - taux de calage sur le RIL médian au niveau IRIS (ou regroupement d'IRIS)
- Le taux de calage sert à ramener toutes les données collectées des 5 années à une année centrale

Calage sur le RIL médian des ZAF

Pour chaque ZAF, le nombre de logements et d'individus est calculé en sélectionnant les adresses de la zone et en appliquant le poids de diffusion standard.

Puis, on effectue un recalage sur le nombre de logements de la ZAF dans le RIL de l'année médiane pour une meilleure précision.

Exemple de calcul avec les poids pour les données au 1er janvier 2007



Calcul du nombre de logements issus des EAR en cumulant les poids de diffusion des 5 EAR (de 2005 à 2009)

Nombre de logements = 1150



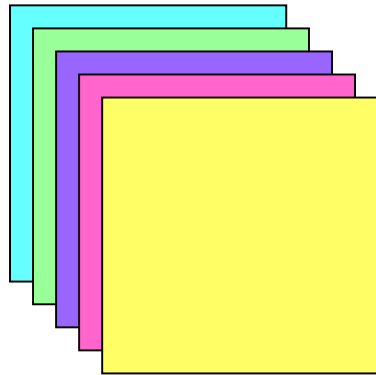
Calcul du nombre de logements du RIL médian

Moyenne du RIL de juillet 2006 corrigé de la collecte 2007 et du RIL de juillet 2007 corrigé de la collecte 2008

Nombre de logements = 1100

Différence du nombre de logement entre les 2 fichiers : choix de caler sur le nombre de logements du RIL médian

Exemple de calcul avec les poids pour les données au 1er janvier 2007



Nombre de logements = 1 150



Nombre de logements = 1 100

$$\text{Taux de calage de la ZAF} = 1\ 100 / 1\ 150 = 0,957$$

$$\begin{aligned} &\text{Poids de diffusion des logements de la ZAF} = \\ &\text{Poids de diffusion des logements} * \text{Taux de calage de la ZAF (0,957)} \end{aligned}$$

On applique le poids de diffusion à toutes les variables.

Calcul des résultats finaux

Cumul des résultats des différentes populations:

- issus des ménages ordinaires du RIL d'habitation

- issus des communautés (y compris logements de fonction)

Quelques résultats

Des résultats satisfaisants

Fractile	Zaf	Zaf_hors_rues	Zaf_rues	Iris	Iris_32
P95	6.03	5.90	7.13	4.54	5.10
P90	4.96	4.94	5.16	3.96	4.31
Q3	3.88	3.88	4.06	3.19	3.52
Médiane	3.02	2.97	3.26	2.57	2.84
Q1	2.36	2.35	2.47	2.13	2.29
P10	1.77	1.77	1.96	1.79	1.82

Page 6



JM Floch

DIAF

2 juin 2010



Qualité des résultats et diffusion

Diffusion des IRIS, rappel

Label choisit à partir d'indicateurs de :

- précision
- stabilité

Trois labels d'IRIS :

label 1 : les tables de précision sont utilisables pour les IRIS
(environ 90% des IRIS)

label 2 : IRIS à regrouper en Triris ou équivalent pour pouvoir
utiliser les tables de précision (environ 10% des IRIS)

label 3 : IRIS très fragiles, les tables de précision ne sont pas
utilisables (quelques dizaines d'IRIS)

Attribution du label de qualité des ZAF

- Volonté d'une politique de diffusion similaire au IRIS pour les ZAF avec 3 niveaux de qualité
- Choix d'attribution de l'indicateur de qualité à partir de trois critères :
 - Coefficient de variation qui permet de mesurer la précision
 - Taille de la zone
 - % de population en IRIS de label 2 et 3 qui permet de prendre en compte la particularité de la zone

À quelles conditions les résultats seront-ils diffusés ? (1/2)

Si $(CV < 4 \%) \Rightarrow \text{QUALITE}=1$

On peut garantir à l'utilisateur du service une précision comparable à celle des données à l'IRIS :

- L'utilisateur reçoit les données et doit se reporter aux tables de précision pour les diffusions des données à l'IRIS afin d'obtenir le CV pour la variable et l'effectif correspondant. Il pourra ainsi calculer des intervalles de confiance.

~~À quelles conditions les résultats seront-ils diffusés ? (2/2)~~

Si CV compris entre 4% et 6% => **QUALITE=2**

- › diffusion des résultats (si l'utilisateur le souhaite) et du CV de la population des ménages, sans toutefois que l'on puisse garantir la précision de toutes les variables.

Si CV compris entre 4% et 6% et si la part de la population en iris de label 2 ou 3 est supérieure ou égale à 70% pour une ZAF de moins de 3 500 habitants

=> **QUALITE=3,**

- › ni les résultats, ni le CV ne seront diffusés

Si CV > 6%,

=> **QUALITE=3,**

- › ni les résultats, ni le CV ne seront diffusés

Conseils et préconisations à partir de l'examen des indicateurs de qualité

Dans les cas où QUALITE vaut '3'

- › l'utilisateur peut agrandir sa zone pour améliorer la précision des résultats et les rendre ainsi diffusibles

Dans les cas où QUALITE vaut '2'

- › l'utilisateur peut agrandir sa zone pour améliorer la précision des résultats pour plus de « sécurité » dans leur utilisation.

Résultats des tests effectués pour la diffusion des ZAF

Tests qui ont donnés des résultats similaires au IRIS pour les ZAF:

- Une majorité de ZAF en qualité 1 (environ 90%)
 - Précision correcte
- Quelques ZAF en qualité 2 (environ 9%)
 - Précision médiocre
- Très peu de ZAF en qualité 3 (environ 1%)
 - Précision mauvaise

Exemples de zones à façon

ZAF n°1 : zone de qualité 1

ZAF n°2 : zone de plus de 3500 habitants de qualité 2

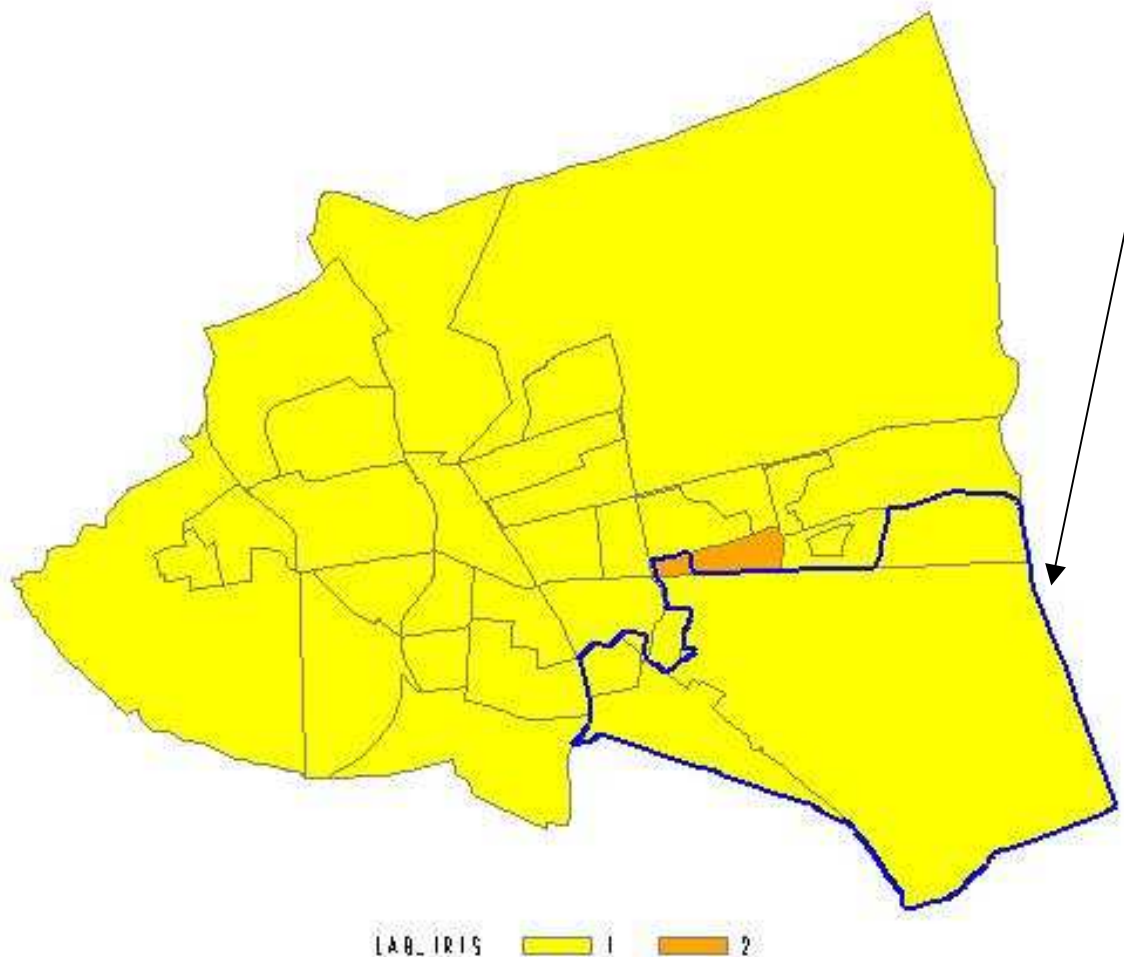
ZAF n°3 : zone de qualité 3

du fait du CV de la population des ménages

ZAF n°1

Zone de qualité 1 : cas le plus fréquent

ZAF n°1 : Calais



Zaf n°1



- 1375 logements pour 3 634 habitants
- CV pop = 2,01
- Pop en IRIS de label 2 ou 3 < 1 %



Zone de qualité 1 :
résultats diffusés à
l'utilisateur

ZAF n°1 : Calais

Zone de qualité 1 : résultats diffusés à l'utilisateur

Pour la précision des données, l'utilisateur doit s'appuyer sur les tables de précision.

Exercice :

Calculer l'intervalle de confiance pour la population des ménages de la zone:

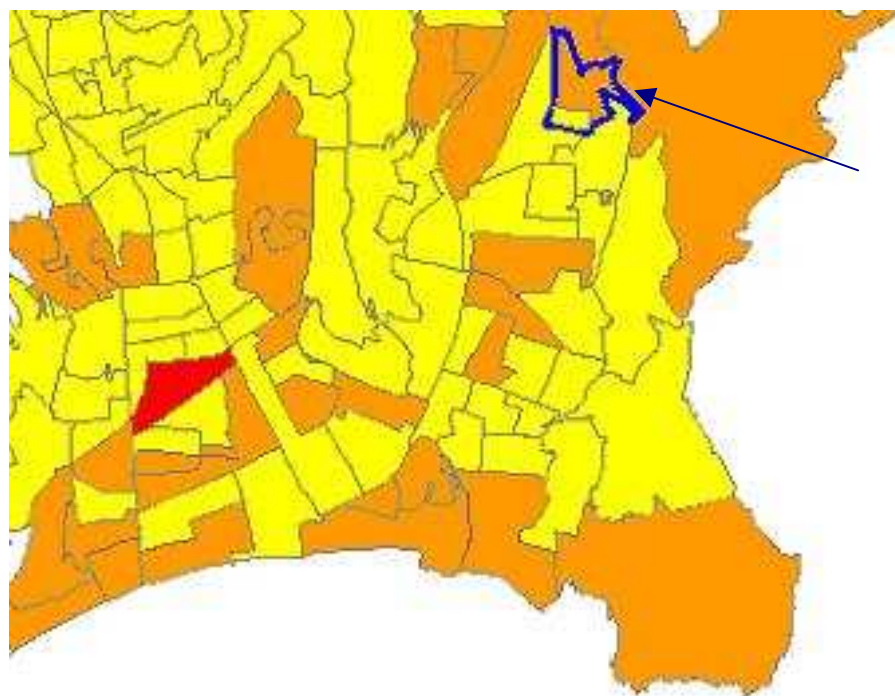
Pop ménages = 3 634

CV pop = 2%

ZAF n°2

Zone de plus de 3500 habitants de qualité 2

ZAF n°2 : Nice



Zaf n°2 {
- 1 788 logements pour 3 746 habitants
- CV pop = 4,5
- Pop en IRIS de label 2 = 69 %



Zone de qualité 2

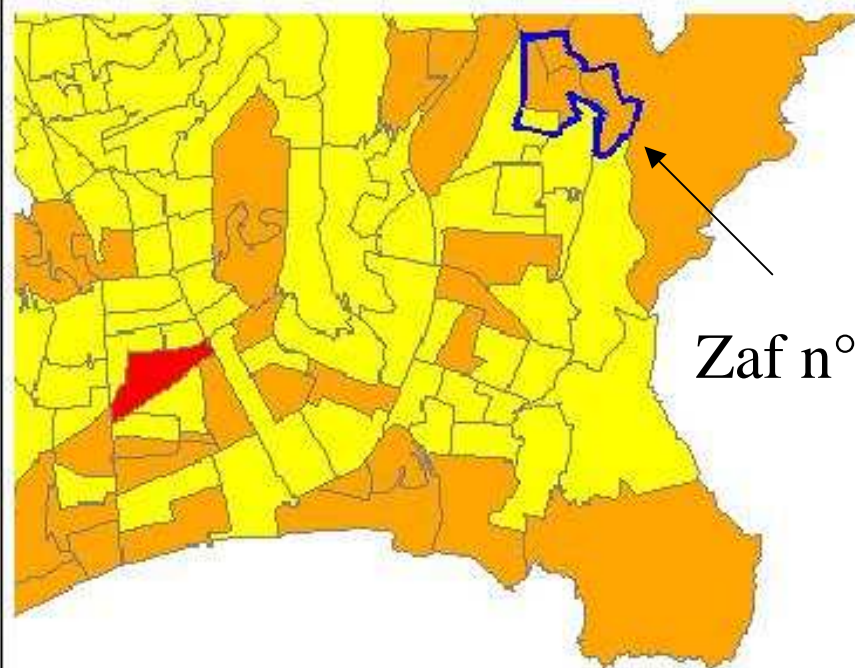


ZAF n°2 : tentative d'élargissement de la zone

Ilots ajoutés à la ZAF initiale



ZAF élargie



Zaf n°2 bis

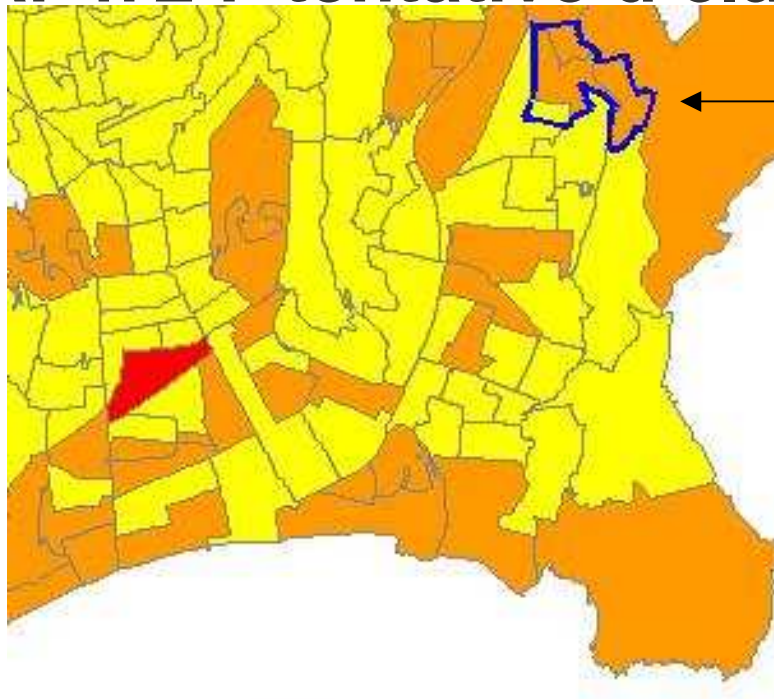
LAB_IRIS

1

2

3

ZAF n°2 : tentative d'élargissement de la zone



Zaf n°2 bis



- 2 094 logements pour 4 425 habitants
- CV pop = 3,8
- (Pop en IRIS de label 2 ou 3 = 73 %)



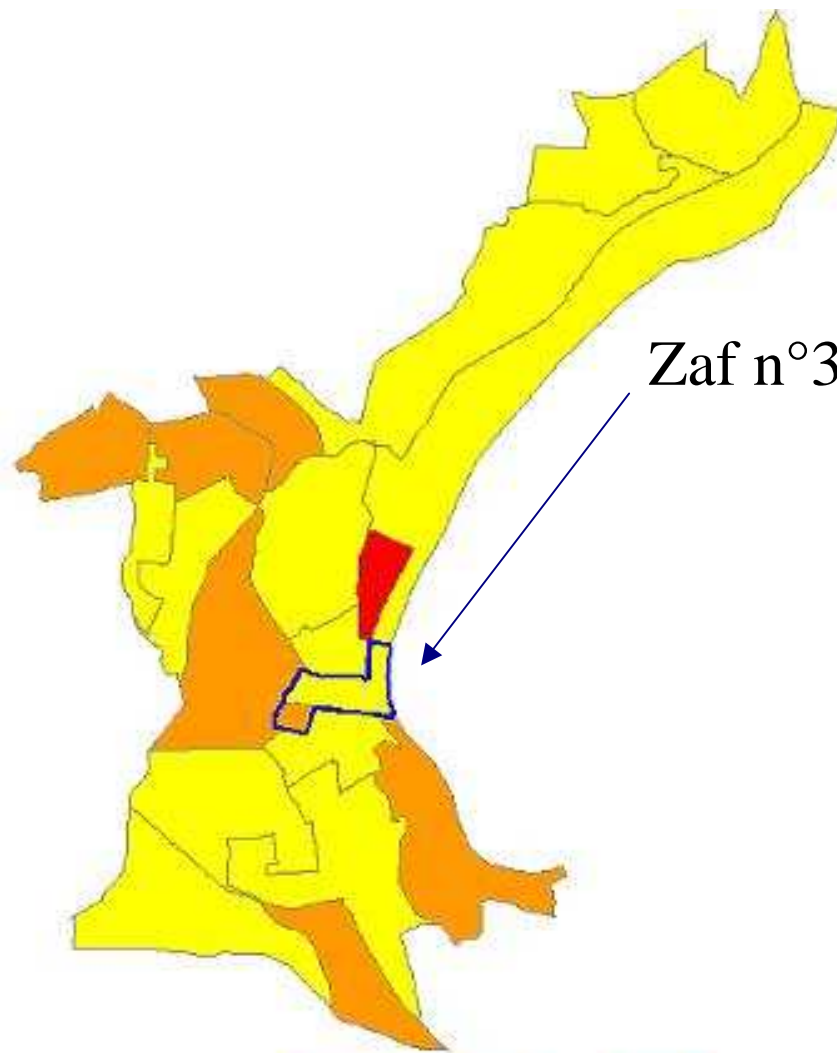
LAB_IRIS 1 2 3

Zone de qualité 1 : résultats diffusés à l'utilisateur

ZAF n°3

Zone de qualité 3 uniquement à cause du CV de la population des ménages

ZAF n°3 : Lyon 9e



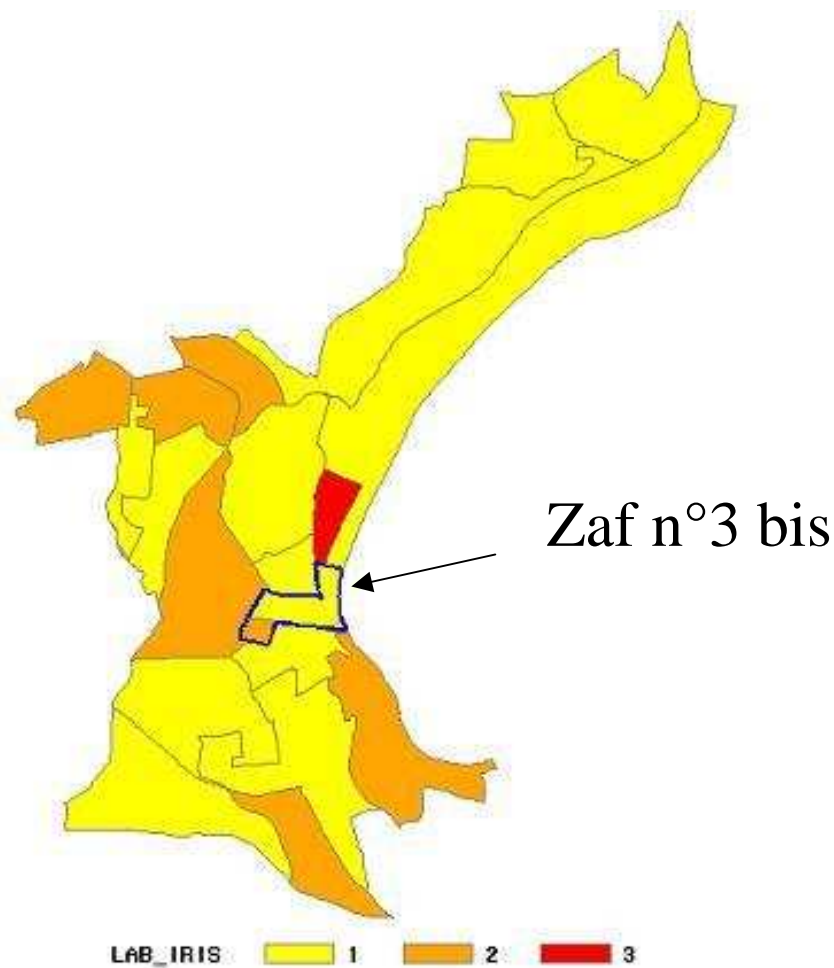
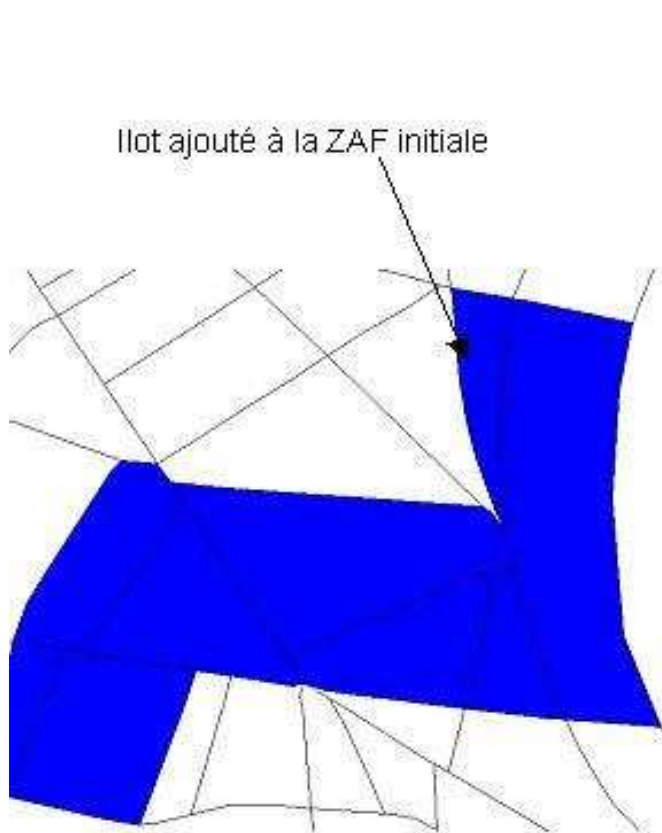
Zaf n°3

- 1 221 logements pour 1 762 habitants
- CV pop = 6,42
- Pop en IRIS de label 2 ou 3 = 35 %

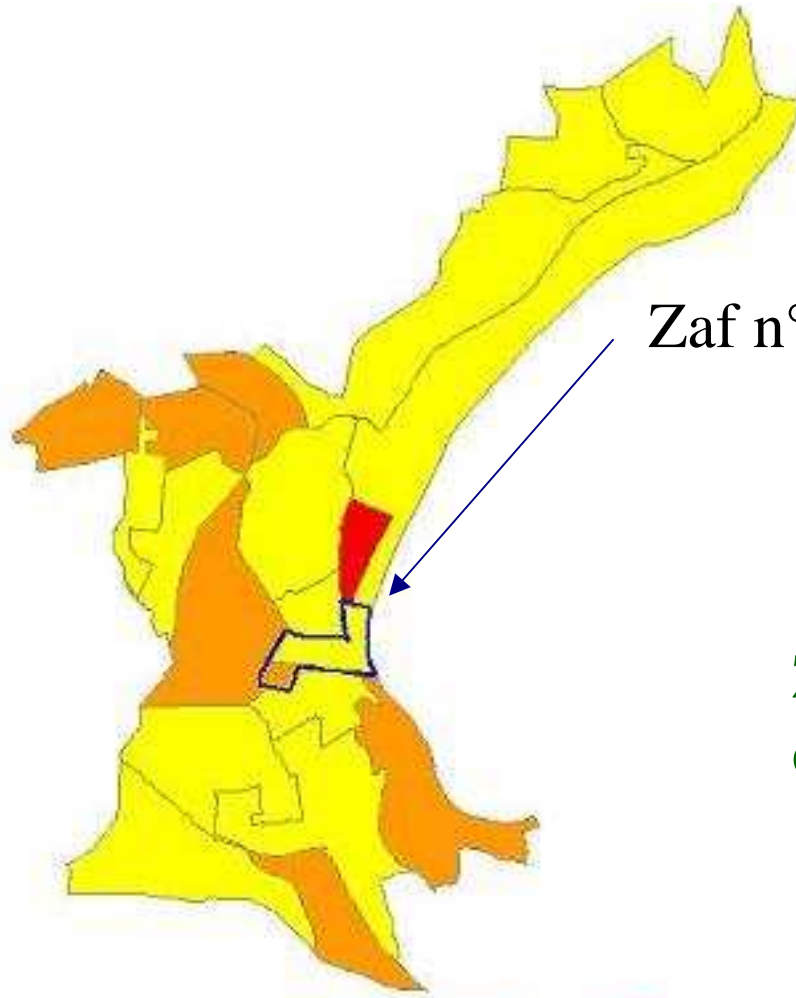


Zone de qualité 3 :
résultats non diffusés

ZAF n°3 : tentative d'élargissement de la zone



ZAF n°3 : tentative d'élargissement de la zone



Zaf n°3 bis

- 1 396 logements, 2 067 habitants
- CV pop = 5,8
- Pop en IRIS de label 2 ou 3 = 29 %



Zone après élargissement passe en qualité 2 :
les résultats seront diffusibles...
...mais toujours fragiles



Les 6 fichiers résultats

✓ 5 fichiers excel thématiques :

- Activité des résidents
- Diplômes - Formation
- Famille - Situation matrimoniale
- Logement
- Population

dans chaque fichier :

Un onglet résultats, un onglet dictionnaire des variables,
un indicateur de coefficient de variation

✓ 1 fichier excel indicateurs essentiels:

La diffusion de l'information

Les 5 fichiers thématiques peuvent être mis en commun entre les différents participants d'un projet ou d'une action de service public, **uniquement dans le cadre de ce projet ou de cette action**

Dans le cadre de l'information du public concerné par un projet, les **seules informations** qui pourront être **rediffusées** par les bénéficiaires sont celles issues du 6ème fichier **indicateurs essentiels**.

Une diffusion infracommunale riche et variée

Diaf RP, une offre ciblée

- pour certains organismes
- sur des données exclusivement issues du recensement
- sur les communes de 10 000 habitants ou plus
- sur des zones de plus de 1000 logements

L'Insee propose d'autres approches et d'autres sources à l'infracommunal...

- sur les Iris, les Zus et les quartiers Cucs
- avec aussi des données sur les revenus, le tissu économique, ...

...résultant d'importants travaux en amont.

Diap RP n'est pas forcément le bon outil si:

Le zonage proposé est très proche d'un découpage à l'Iris

- **Il est préférable de s'appuyer sur les données Iris**

Le zonage proposé correspond à une Zus ou un quartier Cucs

- **Il est préférable de s'appuyer sur les informations déjà disponibles dans l'espace « Bases de données » du site de l'Insee**

Les thématiques concernent le chômage, les revenus, l'activité dans un quartier (Iris ou regroupement d'Iris)

- **Il est préférable de s'appuyer sur les informations disponibles dans l'espace « Bases de données » pour trouver une réponse**
- **L'insee peut vous proposer un diagnostic de quartier**

Diaf RP peut être le bon outil quand:

Un EPCI s'intéresse au tracé d'une future ligne de transports en commun traversant une ou plusieurs communes de 10 000 habitants ou plus.

Une commune souhaite connaître les caractéristiques des habitants d'un quartier rénové qui ne correspond pas à un Iris ou regroupement d'Iris.

Une commune souhaite optimiser l'implantation d'une crèche en s'affranchissant des limites d'Iris.

A retenir (1/2)

Les avantages :

- Obtenir des résultats RP validés sur des zonages adaptés aux problématiques
- Disposer du nombre de logements à l'adresse en complément du service
- Pouvoir partager des informations avec les acteurs du projet
- Bénéficier d'une tarification adaptée
- Pouvoir dessiner la zone de son choix directement sur internet
- Rapidité du service

A retenir (2/2)

Les limites :

- Pas de zone sur les communes ayant moins de 10 000 habitants
- Pas de zone « à cheval » sur plusieurs communes
- Pas de résultat sur les zones de moins de 1 000 logements, sauf le nombre de logements

Données urbaines infra-communales

Les quartiers de la politique de la ville

Les Iris

Des données de recensement, des données administratives



Merci de votre attention