

QUINZE ANS D'ENQUETES AUPRES DES MENAGES

DANS LES DEPARTEMENTS D'OUTRE-MER

Aurore Fleuret, Jérémie Torterat ()*

*Centre de Ressources Interrégional pour les Enquêtes – Ménages (CRIEM)
Insee, Direction Régionale de La Réunion – Mayotte*

Résumé

Les *statistiques sociales par enquête dans les départements d'Outre-mer (DOM)* ont connu bien plus qu'une révolution entre 2001 et 2015. D'erratiques et à la pratique artisanale, les enquêtes de l'Insee basculent vers la modernité. Cet article revient sur ces quinze années-charnière, et sa trame suit quasiment le rythme chronologique du cheminement méthodologique.

Après une description du cadre institutionnel et un panorama général des enquêtes réalisées de 2001 à 2015 dans une première partie, la deuxième partie de l'article est consacrée aux enquêtes réalisées sur la période 2004– 2013. Le dispositif d'échantillonnage « historique » standard (méthode dite « cartographique ») sera exposé, ses limites et les entorses qui ont permis d'acquérir une première expérience de bases alternatives. Les travaux de méthode ayant conduit au dispositif actuel de tirage dans le recensement seront détaillés dans les deux parties suivantes – tant ceux traitant de la constitution de la base de sondage que ceux relatifs à la méthode d'échantillonnage. Un bilan des premières années de ce nouveau dispositif et des pistes d'amélioration seront abordées ensuite. L'article se conclura par la présentation du dispositif en place à Mayotte, nouveau département d'Outre-mer depuis 2011.

Abstract

In the past 15 years, household surveys in French overseas territories of Guadeloupe, Martinique, Guyane, La Réunion and Mayotte achieved a revolution, jumping from an *antic* stage to a *modern* one. This paper aim to provide a comprehensive overview of this period, in terms of surveys and methods, and to present the new sampling system used since 2013.

Mots clefs

Sondage, Outre-mer, Enquêtes-ménage

* Jérémie Torterat était responsable du Criem de septembre 2008 à septembre 2013. Adjointe au chef de centre de septembre 2009 à septembre 2013, Aurore Fleuret a pris la direction du Criem depuis cette date.

Introduction	4
1. Cadre institutionnel et panorama des enquêtes réalisées entre 2001 et 2015.....	7
1.1. Le cadre institutionnel	7
1.2. Panorama des enquêtes menées de 2001 à 2015	8
2. La période 2001-2013 et le dispositif « historique » à partir de 2005	11
2.1. Les premières années : 2001 – 2004.....	11
2.2. Le dispositif standard à partir de 2005 : la base cartographique	11
2.2.1. Introduction	11
2.2.2. La base de sondage dans les grandes communes	12
2.2.3. La base de sondage dans les petites communes	13
2.2.4. Le système d'échantillonnage.....	14
2.2.5. Forces et faiblesses du dispositif standard cartographique	15
2.3. Quelques entorses au dispositif standard	16
2.3.1. Panorama	16
2.3.2. Le cas de l'utilisation du recensement	17
2.3.3. Le cas de l'enquête « Loyers et Charges » en Guadeloupe	18
3. Vers une utilisation du recensement comme base de sondage standard pour les DOM « historiques »	21
3.1. Changer, certes, mais pourquoi/pour quoi ?	21
3.2. La question du nombre d'EAR en grande commune : les volumes.	22
3.2.1. Le cas de deux EAR	22
3.2.2. Le cas de trois EAR.....	24
3.3. La question du nombre d'EAR en grande commune : la précision.	33
3.3.1. Stratégie d'ensemble	33
3.3.2. Application.....	34
3.3.3. Résultats de la simulation.....	35
3.4. Les déséquilibres des groupes de rotation du recensement en grande commune	39
3.4.1. Des déséquilibres en volumes.....	40
3.4.2. Des déséquilibres en structures	43
3.4.3. La piste de l'empilement pour lisser les déséquilibres.....	47
4. Le nouveau dispositif en septembre 2013 pour les DOM « historiques ».....	50
4.1. Présentation générale.....	50
4.2. Le paradoxe de l'EECDOM : disperser le tirage, regrouper la collecte	50
4.3. En réponse au basculement vers l'EECDOM et au NCEE : un double découpage en Secteurs d'Activité des Enquêteurs (SAE) et en strates de tirage.....	51
4.4. Illustration à La Réunion : « du battant des lames au sommet des montagnes ».....	53
4.5. Associer SAE, vague et semaine de référence	55
4.6. Présentation de la base de sondage.....	57
4.7. Présentation de la procédure d'échantillonnage.....	59
4.8. Les logements neufs	60
4.8.1. La question des logements neufs : le tirage	61
4.8.2. La question des logements neufs : le calage	61

5.	Les premières années du nouveau dispositif et pistes d'amélioration	62
5.1.	Deux ans plus tard, quel bilan des SAE et du tirage dans le recensement ?	62
5.1.1.	Bilan sur l'EEC.....	62
5.1.2.	Quelle utilisation des SAE pour la répartition – enquêteurs des autres enquêtes ?.....	64
5.1.3.	Quelle utilisation des SAE pour la planification des enquêtes ?	65
5.1.4.	Un repérage facilité	66
5.1.5.	Tirer dans 3 EAR plutôt que 2 : si c'était à refaire sur Cadre de Vie et Sécurité 2015 en Guyane.	68
5.2.	Assouplir les contraintes sur les SAE ?	68
5.3.	Les tirages dans les sources fiscales : disjonction de bases	69
5.3.1.	Quid d'une enquête entièrement échantillonnée à partir des fichiers fiscaux ?.....	69
5.3.2.	Point de vigilance sur les logements neufs	71
5.4.	Quid des « petites » enquêtes ?	71
5.5.	Lorsque la stratification géographique ne peut être utilisée	72
5.6.	Possibilité d'une repondération <i>a priori</i> de la base de sondage	72
5.6.1.	Distorsion sur un facteur	73
5.6.2.	Distorsion sur plusieurs facteurs	75
6.	Les enquêtes auprès des ménages à Mayotte.....	77
6.1.	Introduction	77
6.2.	Quelle source privilégier ? Recensement versus cartographie	77
6.2.1.	Quelle utilisation possible d'un recensement à Mayotte ?.....	78
6.2.2.	Quelle utilisation possible d'une cartographie à Mayotte ?	78
6.2.3.	Inconvénients et avantages de ces options	79
6.2.4.	Zoom sur les problèmes de repérage	81
6.2.5.	Bilan, première proposition de dispositif et versions alternatives éventuelles	82
6.2.6.	Le découpage du territoire mahorais en cinq groupes de rotation équilibrés.	83
6.3.	Le dispositif d'enquêtes à Mayotte depuis 2014	84
	Bibliographie	86

Introduction

Il serait faible de dire que les *statistiques sociales par enquête dans les départements d'Outre-mer (DOM)* ont connu une révolution entre 2001 et 2015. En quinze ans, on est passé de l'*antiquité* (on trouve des traces anciennes, toujours primordiales – les grandes enquêtes structurantes du Système Statistique Public –, mais généralement erratiques et dont l'étude des fondations tient de l'archéologie) à l'*âge moderne*, avec un programme d'enquêtes proche de celui de la métropole et des méthodes de plus en plus performantes.

Avant 1993, l'âge est pour ainsi dire *préhistorique*. Quelques grandes enquêtes sont réalisées, comme celle dédiée au « budget des familles » (BdF), et posent des bases de la statistique publique. Ce sont cependant à chaque fois des événements extraordinaires, auxquels tout le personnel participait, et que les « anciens » évoquent parfois avec émotion. A partir de 1993, on passe à un âge *antique*. La pratique se répand, avec la mise en place d'une enquête sur l'emploi rudimentaire mais régulière et des enquêtes BdF en 1994/1995 et 2000/2001. Les méthodes sont cependant encore très artisanales et les dispositifs *ad hoc*. Tout n'a pas résisté au temps. Des données et métadonnées sont maintenant manquantes.

La première étincelle vers l'*âge moderne* est l'introduction d'une Enquête sur l'Emploi Annuelle dans les DOM (EEADOM) en 2001, sur un modèle proche de l'enquête-emploi annuelle métropolitaine réalisée de 1982 à 2002, mais adaptée sur certains points aux spécificités de l'Outre-mer. Elle est au départ appuyé sur un dispositif type échantillon-maître, mais celui-ci se révèle très rapidement inadéquat. Pour pérenniser l'EEADOM, un premier dispositif systématique d'échantillonnage est mis en place à partir de 2004. L'ingénieux dispositif, conçu pour être robuste et gérer des enquêtes de grande taille, mobilisait les enquêtes cartographiques réalisées pour les Enquêtes Annuelles de Recensement (EAR), tout juste introduites.

Grâce à ce dispositif novateur, les enquêtes auprès des ménages se sont systématisées progressivement dans les DOM. Entre 2005 et 2010, une grande enquête sociale structurante a pu avoir lieu tous les ans au second semestre, le premier étant systématiquement consacré à l'enquête sur l'emploi. C'était l'objectif initial de l'Insee ; il a toujours été dépassé.

Cette nouvelle « offre » d'enquêtes Outre-mer s'est en effet parallèlement accompagnée d'une « demande » toujours plus forte pour en réaliser de nouvelles, plus nombreuses, plus complexes. Un accord-cadre en 2007 entre l'Insee et la Délégation Générale à l'Outre-mer (DéGéOM) pour le financement d'extensions d'enquêtes dans les DOM a encore accentué cette dynamique. Le dispositif de 2004 est progressivement apparu obsolète pour soutenir l'harmonisation des politiques d'enquêtes nationale et ultra-marine. Adaptée aux enquêtes à grande échelle – son but initial somme toute –, la méthode « cartographique » ne permettait par exemple ni de condenser sur les territoires les fiches-adresses des enquêtes de petit volume, ni de cibler des ménages sur des critères sociaux et/ou économique.

Un projet de refonte du système d'échantillonnage des enquêtes-ménages dans les DOM s'ouvrait, et, comme un clin d'œil de l'histoire de la statistique des DOM, pouvait accompagner l'extension aux DOM de l'enquête-emploi en continu portée par le projet REFLEE (la nouvelle Enquête sur l'Emploi en Continu dans les DOM (EECDOM)). L'utilisation du recensement comme base pour tous les DOM a été l'option choisie, mais les spécificités des territoires ainsi que de leur méthode de recensement interdisaient de rejoindre le système Octopusse. Il allait falloir *inventer*.

La première partie de l'article, très rapide, va planter le décor institutionnel des enquêtes auprès des ménages de l'Insee dans les DOM, et donner un panorama général des enquêtes réalisées dans les DOM sur la période 2005 – 2015.

La deuxième partie de cet article va s'attacher à décrire la réalisation des enquêtes de l'Insee sur la période [2005-2013]. On verra ainsi comment le dispositif « historique » d'échantillonnage dans les DOM mobilise ingénieusement les enquêtes cartographiques réalisées pour le recensement et quelques enquêtes cartographiques complémentaires afin de construire un échantillonnage stratifié et équilibré d'adresses. Ce dispositif est également important car il est le « modèle » du système d'échantillonnage pérenne introduit en 2014 à Mayotte (cinquième DOM depuis 2011). La description de cette méthode standard permet enfin de visiter un certain nombre de spécificités du recensement des DOM qui conditionnent les réflexions sur les méthodes d'enquête à mettre en œuvre dans ces départements. On abordera également les limites de ce

système, qui ont conduit à faire quelques entorses au standard. L'utilisation du recensement a permis de mettre en place en Guadeloupe une enquête en continue sur les « Loyers et Charges » des locataires, et d'enregistrer une expérience précieuse pour la suite.

Le choix du Recensement de la Population (RP) comme fondement de la nouvelle méthode était assez évident. En effet, sauf à gérer des bases de sondage différenciées selon les départements, les autres fichiers classiques (fichiers fiscaux) avaient une qualité statistique trop irrégulière pour être mobilisés de façon *systématique*. **Par contre, la question de savoir avec combien d'EAR il fallait composer la base de sondage était nettement moins évidente.** Il est en effet rapidement apparu intenable de n'utiliser que la dernière EAR disponible pour constituer celle-ci. Mais fallait-il alors en utiliser deux ? trois ? voire fallait-il utiliser le cycle entier ?

La question peut s'envisager sous l'angle du volume. De quel volume dispose-t-on in fine quand on mobilise n EAR ? On ne dispose en effet de n EAR vierges qu'à l'amorce. Un raisonnement analytique a été mis en œuvre pour le cas de deux EAR. Le cas de trois EAR est plus technique ; on a dû avoir recours à une approche par les systèmes dynamiques non-linéaires discrets, Celle-ci est illustrée par plusieurs applications numériques et simulations.

La question peut s'envisager sous l'angle de la précision. De combien améliore-t-on la précision de l'enquête quand on mobilise n EAR plutôt que $n-1$ pour la constitution de la base de sondage (i.e. l'échantillon implicite de première phase) ? Un travail de simulation a permis d'y répondre.

La question peut s'envisager sous l'angle des structures internes des groupes de rotation du recensement et de leurs éventuels déséquilibres.

Répondre à ces questions est l'objet de la troisième grande partie de cet article. La base de sondage usuelle proposée s'appuie sur deux EAR dans les grandes communes et sur l'ensemble du cycle (cinq EAR) pour les petites communes.

Le choix de cette nouvelle base de sondage n'est qu'un élément de l'ensemble de la procédure de refonte. **L'objet de la quatrième grande partie de l'article est la présentation de la méthode d'échantillonnage standard menée sur la base de sondage « RP ».**

D'une part il s'agit bien sûr de savoir comment on va échantillonner à partir de l'empilement d'EAR. La méthode retenue témoigne de la volonté d'intégration organique de deux problématiques, structurantes pour les enquêtes-ménages de l'Outre-mer, qui ont émergé au même moment : l'EECDOM et la mise en place du nouveau cadre d'emploi des enquêteurs. D'autres éléments encore ont nourri la réflexion : rendre compatible la réalisation de la nouvelle enquête-emploi avec celle des autres enquêtes à partir d'une seule base de sondage dans des territoires qui ne disposaient pas de Zone d'Action d'Enquêteur (ZAE) et pour lesquels, à cause de leur petite taille et de leur diversité, il fallait raisonner sur une partition totale du territoire. L'EECDOM, enquête systématique qui cumule les contraintes, a constitué le matériau-type de la réflexion. Son échantillonnage devait répondre à deux logiques opposées :

1 – neutraliser l'effet de grappe au niveau du logement, c'est à dire interroger des logements dispersés sur le territoire pour gagner en précision. Rappelons qu'en métropole sont constituées des grappes de logements contiguës à partir de fichiers fiscaux.

2 – permettre une collecte dans des délais très contraints. Rappelons que la semaine de référence est la période sur laquelle porte l'essentiel des questions de l'enquête et que la collecte doit avoir lieu pendant les 2 semaines et 2 jours qui suivent cette semaine de référence. Il s'agit ici d'optimiser la répartition des logements tirés sur le territoire, répartition au sein des semaines de référence et répartition au sein du réseau d'enquêteurs. Chaque DOM a été découpé en 26 zones réputées équivalentes et réalisant une partition totale du territoire, chacune de ces zones étant associée à une semaine de référence. Ce sont les Secteurs d'Action d'Enquêteur (SAE), pendant ultra-marin des ZAE métropolitaines.

D'autre part, l'utilisation de plusieurs EAR pose la question des logements neufs construits entre temps dans les plus anciennes. Certes l'utilisation des deux EAR les plus fraîches limite le manque de logements neufs pour les grandes communes, mais dans les petites, le déficit est plus difficilement négligeable puisque l'on utilise le cycle complet d'EAR. En Guadeloupe, en Martinique et à La Réunion, les fichiers fiscaux sont de qualité suffisante pour offrir une base de tirage de logements neufs. En Guyane, un dispositif spécifique à partir de différentiels cartographiques peut être mis en place. Mais une fois *trouvés* les logements neufs échantillonnables, il reste cependant à déterminer *quel poids de sondage leur affecter*, et c'est là qu'est en fait l'enjeu du questionnement.

La cinquième partie de l'article fait le bilan des premières années du nouveau dispositif d'échantillonnage. D'abord d'un point de vue de la collecte : quels sont les résultats de l'EECDOM ? les SAE facilitent-ils effectivement l'organisation de la collecte ? Le repérage est-il simplifié avec les informations issues du RP ? **Cette quatrième partie liste par ailleurs quelques pistes d'amélioration, à l'étude ou encore en suspens, qui permettrait de rendre le nouveau système plus complet.** Des questions sur des disjonctions de bases : comment permettre des enquêtes à partir des fichiers fiscaux sur les territoires des DOM sans réinterroger des ménages déjà échantillonnés à partir des bases RP ? Que faire avec de enquêtes de petit volume ? Que faire quand la stratification géographique ne peut être utilisée ? Ou encore : peut-on repondérer *a priori* la base de sondage pour limiter l'influence des déséquilibres des groupes de rotation du recensement ?

La sixième partie de l'article va, pour terminer, présenter le système d'échantillonnage mis en place à Mayotte à partir de 2014 pour les enquêtes auprès des ménages. Le système est différent des autres DOM car le recensement y est encore quinquénal, et ne permet donc pas d'adopter la même méthode.

La récente évolution de Mayotte en *Département* d'Outre-mer (2011) a changé la donne : il fallait non seulement intégrer progressivement le territoire dans le système statistique français, mais il fallait également répondre à la pressante nécessité d'y réaliser les opérations permettant sa reconnaissance comme Région Ultra-Périphérique (RUP) de l'Union Européenne. Pour cela, il fallait là aussi créer un système pérenne pour réaliser au moins une Enquête sur l'Emploi régulière (en l'espèce *annuelle*, l'EECDOM ne concerne pas encore Mayotte) et ouvrir la voie, comme près de quinze ans plus tôt, aux enquêtes du Système Statistique Public.

1. Cadre institutionnel et panorama des enquêtes réalisées entre 2001 et 2015

1.1. Le cadre institutionnel

2001 marque un tournant pour les enquêtes ménages dans les DOM. L'Enquête sur l'Emploi est refondue dans le cadre d'un projet piloté par la DR de La Réunion. L'enquête Budget des Familles (BDF) est réalisée sur le modèle national mais avec une implication encore balbutiante de la maîtrise d'ouvrage. En 2004, un bilan de ces deux enquêtes est réalisé, notamment pour préparer la réédition de l'enquête BDF en 2005. Le dispositif type échantillon-maître utilisé pour le tirage de ces deux enquêtes et basé sur le recensement de 1999 n'est plus adapté.

De premiers travaux méthodologiques sont entrepris dans la foulée pour réformer les procédures d'échantillonnage à la Direction Régionale Insee de La Réunion, soutenus par l'Unité de Méthodes Statistiques de la Direction Générale. Ceux-ci aboutiront à ce que l'on nommera dans la suite de l'article « dispositif historique standard ».

A la suite de cette période transitoire initiale, l'essentiel des travaux méthodologiques réalisés pour les enquêtes auprès des ménages de l'Insee dans les DOM est réalisé par une équipe dédiée, division du Service Statistique de la Direction Régionale La Réunion-Mayotte.

Le Centre de Ressource Interrégional pour les Enquêtes auprès des Ménages dans les DOM (Criem) est créé en 2005. Son champ d'intervention inclut les opérations *pré-collecte* (intervention sur les questionnaires quand des adaptations sont nécessaires, échantillonnage, spécifications et tests de DataModel (questionnaire électronique), mises en forme opérationnelles des échantillons...), le suivi de l'enquête en tant que relais « local » des équipes conceptrices, les opérations *post-collecte* (redressement de non-réponse totale, calage) et l'archivage-documentation des enquêtes menées. A partir de 2009, une convention a formalisé le soutien régulier de la Direction des Statistiques Démographiques et Sociales (DSDS) de la Direction Générale de l'Insee et son Unité des Méthodes Statistiques (UMS « ménages ») au Criem, sur des opérations spécifiques (redressement de la non-réponse partielle, échantillonnage à partir de sources fiscales...).

Programmé au départ pour gérer deux enquêtes par an, l'EEADOM et une autre différente tous les ans mais commune à tous les départements, « budget des familles », « logement », ..., le Criem a suivi la mise en place d'enquêtes supplémentaires dès les premières années (enquête sur le Chikungunya à La Réunion dès 2006). C'est cependant à partir de la mise en œuvre pratique de l'accord-cadre Insee-DéGÉOM en 2007 pour le financement de compléments d'échantillons et dans la foulée de l'ambitieuse enquête « Migrations – Famille – Vieillesse » (MFV), en partenariat avec l'Ined, que le rythme des enquêtes prend une nouvelle ampleur.

Les compétences développées au Criem sont également mises à contribution dans certaines opérations spécifiques pour les statistiques sociales publiques dans les DOM, comme pour le recensement de Mayotte en 2012.

Depuis septembre 2013, le Criem a été redimensionné pour répondre à la multiplication de son activité. Une nouvelle convention le lie depuis 2014 au Département des Méthodes Statistiques (DMS) de la DG-Insee ; celle-ci reprend une part importante de l'ancienne convention en tirant partie de l'expérience acquise depuis 2009. Son élément novateur est la mise en place d'un Comité de Pilotage semestriel pour réguler l'activité du Criem. Ce comité réunit le DMS, les deux directeurs régionaux des DOM, un membre de l'Inspection Générale, la Division Maîtrise d'œuvre des Activités d'Enquête et la DSDS, principale maîtrise d'ouvrage des enquêtes ménages.

L'activité des Unités « Enquêtes auprès des Ménages » (UEM) des Directions et Services régionaux de l'Insee dans les DOM a suivi le même mouvement, avec une complexification croissante des contenus et un rythme toujours plus intense. Ajoutée à ces contraintes de fonctionnement déjà difficiles, la perspective d'une EECDOM et du nouveau cadre d'emploi des enquêteurs a motivé la mise en place d'une démarche MAIOL (Maîtrise d'œuvre Locale en Organisation du travail), conclue en 2011, sur le fonctionnement des UEM, afin de les armer face à ces nouveaux défis. Les effectifs des équipes ont été sécurisés voire revus à la hausse, des postes – dont ceux de chefs d'unité – ont

été requalifiés. Les modes de fonctionnement ont également été revus afin de mieux s'adapter aux contraintes d'une collecte en continu sur l'année (pour l'EECDOM).

Enfin, afin d'améliorer la visibilité sur le programme d'enquêtes dans les DOM, l'Insee et la DéGéOM ont signé le 23 septembre 2013 un nouvel accord cadre. Il précise la contribution financière du Ministère des Outre-mer pour les enquêtes à venir. Cette contribution permet une extension significative de la taille de l'échantillon, nécessaire à la production d'informations pertinentes sur les territoires concernés. Dorénavant, chaque Dom bénéficie chaque année d'une extension d'enquête grâce à cette contribution.

1.2. Panorama des enquêtes menées de 2001 à 2015

Les tableaux suivants présentent un panorama des enquêtes menées de 2001 à 2015 dans les DOM, quelque soit le dispositif d'échantillonnage. Le dispositif « historique » cartographique est en application et mis en œuvre au Criem à partir des enquêtes sur le terrain en 2005.

Quand l'échantillonnage n'a pas été réalisé par le Criem (à partir des enquêtes cartographiques sur le terrain en 2005), les hachures sont claires, les spécificités sont indiquées en « remarque ». Avant 2005, l'échantillonnage n'a pas été réalisé par le Criem qui n'était pas encore créé : le dispositif initial « transitoire » était en cours ; ces enquêtes sont indiquées en pointillés dans le tableau.

Dans les premières années de la période, seules la nouvelle EEADOM et la fondamentale enquête sur le « budget des familles » (utile pour les comptes régionaux des DOM et pour de premières estimations de PIB) sont réalisées.

Le développement des enquêtes ultra-marines coïncide avec la création du Criem et la nouvelle politique enquêtes de l'Insee pour les DOM.

Le principe *a priori* des deux enquêtes par an dans les DOM n'a été réalisé qu'en 2005. Dès 2006 ont lieu les premières entorses. L'urgence sanitaire de la crise du Chikungunya pousse l'Insee à dégager des moyens pour réaliser une enquête de prévalence avec l'Agence Régionale de la Santé (ARS). Les enquêtes « Information et Vie Quotidienne » (IVQ) ont quant à elles été réalisées progressivement dans les différents départements. En effet, ces enquêtes nécessitent des adaptations très importantes du questionnaire, et différentes entre les départements : l'effort à fournir par les concepteurs, même secondés par le Criem, était donc trop important pour que les quatre départements réalisent cette enquête sur une même année.

Même si quelques enquêtes prévoient auparavant des dispositions d'échantillonnage particulières (« VQS » en 2007, « IVQ » en Guadeloupe en 2008 : voir §2.), c'est cependant à partir de 2008/2009 et l'application progressive de l'accord-cadre Insee-DéGéOM de 2007 que le rythme et la complexité méthodologique des enquêtes prend une nouvelle dimension : la préparation à partir de 2008 et la mise sur le terrain en 2009 des enquêtes « Patrimoine » et « Migrations – Familles – Vieillesse » (MFV) marquent un nouveau tournant. Ces deux enquêtes, en effet, ont nécessité des investissements méthodologiques et des protocoles particuliers. Le sous-échantillon principal de « MFV » est certes tiré à partir du dispositif cartographique, mais le sous-échantillon spécifique correspondant aux « migrants de retour » est tiré à partir de l'EEADOM. Le tirage de cette enquête était ainsi nettement plus complexe que pour les enquêtes précédentes, mais la réconciliation de ces deux sous-échantillons par partage des poids et les opérations post-collecte correspondantes ont également été des procédures novatrices. L'enquête « Patrimoine » a été échantillonnée à partir des sources fiscales et gérée par l'UMS de la DG pour La Réunion, mais échantillonnée à partir du recensement pour la Guadeloupe et la Martinique par le Criem en tentant de reconstituer au mieux les modalités de stratification utilisées dans les fichiers fiscaux.

Depuis ce tournant, pas moins de six enquêtes différentes sont gérées annuellement dans les DOM et jusqu'à neuf en 2014, sans inclure dans le compte les tests « terrain » réalisés pour prévoir les enquêtes des années suivantes. Beaucoup de ces nouvelles enquêtes ont par ailleurs nécessité de mettre en place des tirages et/ou des calages plus sophistiqués, « Emploi du Temps », « Loyers et Charges », « Participation électorale », « Chlordéone »...

Tableau 0 : Enquêtes réalisées dans les Dom de 2001 à 2015

année	nom de l'enquête	département					remarque	
		Guadeloupe	Martinique	Guyane	La Réunion	Mayotte		
2001	Budget des Familles						enquête réalisée sur 2000/2001	
	EEADOM							
2002	EEADOM							
2003	EEADOM							
2004	EEADOM							
2005	EEADOM						à Mayotte : recensement, réalisée par l'Antenne	
	Budget des Familles							
2006	EEADOM						avec les autorités sanitaires, lors de l'épidémie	
	Logement							
	Information et Vie Quotidienne							
	Chikungunya							
2007	EEADOM						mise en forme électronique et marges de calage au Criem	
	Vie Quotidienne et Santé							
	Information et Vie Quotidienne							
2008	Technologie de l'Info. et de la Com.						mise en forme électronique et marges de calage au Criem	
	EEADOM							
	Handicap-Santé, volet ménages							
	Information et Vie Quotidienne							
2009	Technologie de l'Info. et de la Com.						974 : tirage UMS, marges de calage Criem - 972 : 84 FAs	
	EEADOM							
	EE - Mayotte							recensement 2007
	test EECDOM							
	Migrations - Famille - Vieillessement							
	Patrimoine							
2010	Technologie de l'Info. et de la Com.						mise en forme électronique et marges de calage au Criem	
	EEADOM							
	Budget des Familles							à Mayotte, voir §6 : mix RP07/carto RP agricole/carto
	Emploi Du Temps							
	"Energie"							pour l'Agence Régionale de l'Energie - La Réunion
	Loyers et Charges							

Note : enquête TIC : Depuis 2013, l'échantillonnage est réalisé par le DMS de la DG-Insee. Entre 2009 et 2012, le tirage est effectué à la DG à partir de l'annuaire France-Telecom pour la Guadeloupe et la Guyane, à partir de la Taxe d'Habitation pour la Martinique et La Réunion. En 2007 et 2008, un tirage est effectué à partir des numéros de téléphone collectés pour les réinterrogations (téléphoniques) de l'EEADOM.

année	nom de l'enquête	département					remarque
		Guadeloupe	Martinique	Guyane	La Réunion	Mayotte	
2011	EEADOM	■	■	■	■		
	Cadre de Vie et Sécurité				■		
	Loyers et Charges	■					
	Information et Vie Quotidienne			■	■		
	"Gisement déchets" - 974				■		expérimentation d'enquête "entreprises" / CER BTP
	Technologie de l'Info. et de la Com.	■	■	■	■		mise en forme électronique et marges de calage au Criem
2012	EEADOM	■	■	■	■		
	Conditions de Travail	■	■	■	■		
	Information et Vie Quotidienne					■	
	Loyers et Charges	■					
	Participation électorale	■	■	■	■		calé au Criem
	Technologie de l'Info. et de la Com.	■	■	■	■		mise en forme électronique et marges de calage au Criem
2013	EEADOM	■	■	■	■		
	EECDOM	■	■	■	■		
	EE - Mayotte					■	
	Logement	■	■	■	■	■	
	"Chlordécone"	■	■				avec l'ARS - Antilles
	Loyers et Charges	■					
	Technologie de l'Info. et de la Com.	■	■	■	■		mise en forme électronique et marges de calage au Criem
2014	EECDOM	■	■	■	■		
	EE - Mayotte					■	
	Enq. Nat. sur les Ressources des Jeunes adultes	■			■		
	Formation - Qualification - Profession	■			■		tiré dans la source fiscale par le DMS
	Vie Quotidienne et Santé	■	■	■	■		tiré dans la source fiscale par le DMS
	Information et Vie Quotidienne		■				
	Patrimoine	■					tiré dans la source fiscale par le DMS
	Loyers et Charges	■					
Technologie de l'Info. et de la Com.	■	■	■	■		mise en forme électronique et marges de calage au Criem	
2015	EECDOM	■	■	■	■		
	EE - Mayotte					■	
	Cadre de Vie et Sécurité	■	■	■	■		
	Patrimoine				■		tiré dans la source fiscale par le DMS
	Loyers et Charges	■					
	Entrée dans la Vie Active	■	■	■	■		
	Technologie de l'Info. et de la Com.	■	■	■	■		mise en forme électronique et marges de calage au Criem
Risques Psycho-Sociaux	■	■	■	■		panel CdT avec tirage d'entrants	

2. La période 2001-2013 et le dispositif « historique » à partir de 2005

2.1. Les premières années : 2001 – 2004

Des Unités Primaires ont été construites par regroupement d'îlots à partir du recensement de la population de 1999, afin de diminuer les déplacements des enquêteurs en limitant la dispersion géographique des échantillons. Ces réserves de logements devaient être mobilisées pour l'EEADOM, à partir de 2001, et pour d'autres enquêtes sociales éventuelles. Les UP à mobiliser dans l'année étaient échantillonnées annuellement pour le territoire réunionnais ; un tirage de dix rotations avait été effectué à la Dirag.

Ces UP contenaient entre 200 et 250 logements (recensés) et les fichiers de permis de construire devaient permettre d'ajouter les constructions neuves chaque année. Ces fichiers administratifs n'ont cependant pas été suffisants (construction très rapide et mal couverte les autorisations administratives, ce qui posait des problèmes majeurs pour le calcul des poids initiaux de l'échantillon) et un travail cartographique spécifique et important pour localiser les logements issus du recensement de 1999 et repérer les logements neufs.

Ce travail de mise à jour des UP était cependant contraint par des questions budgétaires et ne pouvait pas être réalisé tous les ans. La mise en place à partir de 2003 des enquêtes cartographiques nécessaires pour la nouvelle méthode de recensement allait contraindre de manière encore accrue la réalisation de ces mises à jour dont dépendait la qualité des enquêtes. Enfin, parallèlement à cette mise à jour inflationniste, on assistait à une distorsion progressive des UP : à La Réunion, pour la campagne 2004, un tiers des UP était inutilisable (par exemple à cause de la construction du boulevard Sud de Saint-Denis qui a nécessité de raser plusieurs quartiers).

Dès ce dispositif initial, les échantillons sont ventilés dans l'ensemble des communes. Par comparaison, il faut se souvenir que pour une enquête métropolitaine, on trouve des agglomérations de taille appréciable (jusqu'à 100 000 habitants) qui ne sont jamais concernées par les enquêtes-ménages. C'est à l'inverse une caractéristique des DOM que de devoir enquêter *a priori* dans l'ensemble des communes pour garantir une précision suffisante et ne pas générer de biais. Les dispositifs d'échantillonnage qui suivront (basés sur la base cartographique puis sur les enquêtes annuelles de recensement) continueront à être conçus pour couvrir toutes les communes du territoire.

Néanmoins, le cas de la Guyane est différent car compte tenu de la difficulté liée au terrain, les enquêtes se concentrent en Guyane dite "côtière" (les zones intérieures sont d'accès réglementé), en excluant en sus certaines parties difficilement accessibles de cette zone côtière - ce qui fait qu'*in fine* 12% environ de la population de Guyane n'est pas couverte par les enquêtes -ménages¹.

2.2. Le dispositif standard à partir de 2005 : la base cartographique

2.2.1. Introduction

Le dispositif de sondage appuyé sur un échantillon maître issu du recensement de 1999 a eu cours au début des années 2000 est donc rapidement devenu obsolète.

Il fallait pourtant achever la pérennisation de la toute récente EEADOM – lancée en 2001, en rythme normal d'exploitation en 2003 – qui souffrait déjà du dispositif initial. Cependant, la modification du dispositif n'avait pas seulement une incidence méthodologique sur l'EEADOM ; en effet, le caractère *annuel* de l'EEADOM obligeait de plus à utiliser un dispositif permettant d'*articuler* plusieurs enquêtes sur l'année. Or un dispositif type « échantillon maître » venait justement de montrer ses limites pour les DOM. Il fallait donc inventer un dispositif original, et ce rapidement puisqu'il était attendu de réaliser en 2005 l'enquête sur le « Budget des Familles ».

¹ Les petites îles guadeloupéennes des Saintes et de la Désirade ainsi que des zones isolées de la Réunion (cirque de Mafate par exemple) sont aussi exclues du champ des enquêtes ménages, mais elles représentent une part négligeable de la population.

L'introduction, dans les Départements d'Outre-mer comme en métropole, du Recensement Rénové de la Population allait ouvrir des pistes. Les DOM n'ont pas rejoint le dispositif OCTOPUSSE, mais l'option développée a permis de profiter par contre de la logistique du nouveau RP : on n'allait pas utiliser les Enquêtes Annuelles de Recensement (EAR) mais les *bases cartographiques* mobilisées pour réaliser l'échantillon annuel à recenser. Ces bases cartographiques sont réalisées dans l'ensemble des communes des DOM, alors qu'en métropole elles ne concernent que les communes de plus de 10 000 habitants (dites « grandes communes » (GC) par opposition aux « petites communes » (PC) de moins de 10 000 habitants) ; les Systèmes d'Information Géographiques (SIG) Insee, en cycle quinquennal, contiennent ainsi la quasi intégralité des bâtis des territoires des DOM², aux logements neufs des anciennes cartographies près... Ces bases, bien que grossières, étaient une opportunité intéressante, alliant fraîcheur et robustesse.

Différents travaux méthodologiques ont été menés à l'époque, en particulier par Marc Christine, Véronique Daudin et Delphine Artaud qui y consacrent un mémoire. On ne va s'attacher dans la suite qu'au dispositif finalement mis en place.

Le RP gère de façon différente les petites communes (PC : moins de 10 000 habitants) et les grandes communes (GC : plus de 10 000), ce qui va impliquer un processus de sondage différencié selon le type de communes.

2.2.2. La base de sondage dans les grandes communes

Le territoire de ces communes est divisé en cinq groupes de rotation équilibrés (par rapport au RP99) ; ces groupes sont constitués d'îlots. A la différence de la métropole, le recensement des départements d'Outre-mer ne mobilise pas une strate de logements neufs (qui sont ensuite répartis entre les cinq groupes). Les groupes de rotation des GC des DOM pour le recensement sont donc purement géographiques.

Chaque année, une EAR est menée sur 40% *des adresses* d'un groupe de rotation. Préalablement à cette EAR, une enquête cartographique a lieu répertoriant l'ensemble des bâtis de ce groupe. Seuls les DOM bénéficient d'une telle enquête cartographique, la constitution du RIL en métropole s'appuyant uniquement sur des fichiers administratifs et une expertise des communes. Cette enquête cartographique permet d'établir la Base de Sondage des Adresses (BSA) du recensement.

Ces mêmes enquêtes cartographiques sont mobilisées pour constituer une base de sondage pour les enquêtes auprès des ménages à partir des 60% d'adresses non recensées (les grandes adresses de plus de 21 logements, rares, sont cependant conservées dans la base de sondage des enquêtes auprès des ménages même lorsqu'elles sont recensées).

Tableau 1 : enquêtes cartographiques, RP et enquêtes auprès des ménages en grandes communes

	<i>partie de la base carto mobilisée pour le RP</i>	<i>partie de la base carto disponible pour le tirage des enquêtes auprès des ménages</i>
GR 1		
GR 2		
GR 3		
GR 4		
GR 5		

² Quelques zones particulièrement difficiles d'accès sont exclues : l'intérieur guyanais, certains « Hauts » de La Réunion

2.2.3. La base de sondage dans les petites communes

Dans le RP, les petites communes sont recensées exhaustivement tous les cinq ans, et sont également réparties dans cinq groupes. Ces communes sont trop peu nombreuses (un groupe contient 3 ou 4 PC en Guadeloupe, 4 ou 5 en Martinique, 3,4 ou 5 en Guyane, 1 ou 2 à La Réunion) et trop hétérogènes pour que ces groupes puissent être équilibrés comme en métropole, et leur constitution a plutôt répondu à une logique pragmatique. Il n'était ainsi pas satisfaisant d'utiliser les groupes de rotation du recensement comme tels pour la conduite d'enquêtes auprès des ménages.

Les territoires des petites communes ont donc ainsi été divisés en cinq groupes de rotation interne, sur le même modèle que les grandes communes (regroupements d'îlots).

Une petite commune recensée l'année n a été cartographiée exhaustivement en $n-1$. Les bases de sondage correspondant à cette commune sont réalisées, année après année, selon le schéma suivant :

- En n : on utilise la cartographie constituée en $n-1$ sur le groupe de rotation de l'année (GR m)
- En $n+1$, on utilise la cartographie constituée en $n-1$ sur le groupe de rotation de l'année (GR $m+1$)
- En $n+2$, on utilise une « extension » de cartographie sur le GR $m+2$ réalisée en $n+1$
- En $n+3$, on utilise une « extension » de cartographie sur le GR $m+3$ réalisée en $n+2$
- En $n+4$, on utilise une « extension » de cartographie sur le GR $m+4$ réalisée en $n+3$

[en $n+4$, on réalise une cartographie complète de la commune pour le recensement qui a lieu en $n+5$, le cycle se ré-enclenche].

L'« extension » était une opération de cartographie complémentaire à la cartographie nécessaire au recensement (on « étendait » les opérations normales de cartographie aux petites communes non concernées par le recensement, sur un cinquième déterminé de leurs territoires), qui avait pour seule fin de permettre une réalisation correcte des enquêtes auprès des ménages

Ce système des extensions n'a été mis en place à La Réunion que depuis les tirages de 2006 et depuis 2007 pour la DIRAG. Auparavant, la base des petites communes était constituée de l'intégralité des enquêtes cartographiques. Le sous-échantillon de 2005 a donc été tiré sur une base comportant l'intégralité des îlots des petites communes cartographiés en 2003 et en 2004 et le sous-échantillon de 2006 pour la DIRAG sur l'intégralité des îlots des petites communes cartographiés en 2003, 2004 et 2005.

Tableau 2 : schéma d'articulation des enquêtes cartographiques du recensement et des extensions pour constitution des bases de sondage dans les petites communes

Année enquête cartographique ↓	PC groupe de rotation 1 (EAR 2004 et 2009)					PC groupe de rotation 2 (EAR 2005 et 2010)					PC groupe de rotation 3 (EAR 2006 et 2011)					PC groupe de rotation 4 (EAR 2007 et 2012)					PC groupe de rotation 5 (EAR 2008 et 2013)				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
2003																									
2004																									
2005																									
2006																									
2007																									
2008																									
2009																									
2010																									

Légende :

	Enquêtes cartographiques RP
	Extension des enquêtes cartographiques pour constitution des bases de sondage
	Bases de sondages 2005 à 2010 pour échantillons 2006 à 2011

2.2.4. Le système d'échantillonnage

La base de sondage « cartographique » de l'année $n-1$ utilisée pour les tirages des enquêtes ménages de l'année n (ou du début de l'année $n+1$) est donc issue de quatre types d'enquêtes cartographiques empilées :

- « GC-RP n » : enquêtes cartographiques réalisées en $n-1$ sur les îlots des grandes communes recensés en n (pour la base cartographique 2009, ce sont les îlots du groupe de rotation 2) : ces enquêtes contiennent un cinquième des adresses de chaque grande commune, seules les adresses non recensées l'année n (soit 60 % des adresses) alimentent la base cartographique ;
- « PC-RP n » : enquêtes carto réalisées en $n-1$ sur les petites communes recensées en n (pour la base cartographique 2009, ce sont les îlots du groupe de rotation 2) : ces enquêtes contiennent l'ensemble des adresses de ces communes, seules 1/5^{ème} de ces adresses (celles du groupe d'îlots 5 pour la base cartographique 2009) alimentent la base cartographique ;
- « PC-RP $n-1$ » : enquêtes cartographiques réalisées en $n-2$ sur les petites communes recensées en $n-1$ (pour la base cartographique 2009, ce sont les îlots du groupe de rotation 1) : ces enquêtes contiennent l'ensemble des adresses de ces communes, seules 1/5^{ème} de ces adresses (celles du groupe d'îlots 5 pour la base cartographique 2009) alimentent la base cartographique ;
- « PC-extension » : enquêtes cartographiques « extension » réalisées en $n-1$ sur 1/5^{ème} de chacune des petites communes recensées en $n-4$, en $n-3$ et en $n-2$ (pour la base cartographique 2009, ce sont les adresses du groupe d'îlots 5 des PC des groupes de rotation 5, 4 et 3) : elles alimentent intégralement la base cartographique.³

Cinq strates de tirage sont ensuite définies, trois dans les grandes communes et deux dans les petites, sur la base de sondage cartographique constituée de l'empilement de ces quatre types d'enquêtes. Ces cinq strates correspondent à des *types d'adresse*.

Dans les grandes communes, on mobilise :

- une strate de monologement (bâtis ne contenant qu'un seul logement)
- une strate de « petites adresses » (bâtis de logements collectifs de moins de 21 logements) : ces adresses sont enquêtées intégralement
- une strate de « grandes adresses » (bâtis de logements collectifs de 21 logements au moins) : 21 logements seulement sont enquêtés dans ces adresses

Dans les petites communes, on mobilise :

- une strate de monologement (bâtis ne contenant qu'un seul logement)
- une strate de « petites adresses » (bâtis de logements collectifs de moins de 21 logements)⁴

Des allocations par type d'adresses spécifiques sont calculées pour aboutir, *in fine*, à une équipondération des *logements* (quel que soit d'ailleurs le type de commune auquel appartient le logement) dans l'échantillon de *logements* finalement mobilisé.

Des variables d'équilibrage étaient introduites dans le tirage afin d'en améliorer la précision. Certaines étaient communes à tous les DOM :

- des zones géographiques : regroupements de communes, souvent proches des zones d'emploi ;
- l'aspect du bâti : variable fournie avec la base adresses (classement des habitations lors des enquêtes cartographiques en quatre modalités : habitation de fortune, case traditionnelle, habitation ou immeuble en bois, habitation ou immeuble en dur).

³ Ce système est en place à La Réunion depuis les tirages de 2006 et depuis 2007 pour la DIRAG. Auparavant, la base des petites communes était constituée de l'intégralité des enquêtes cartographiques.

⁴ Les grandes adresses en petites communes sont extrêmement rares et ne permettent pas de constituer une strate. Les quelques grandes adresses sont alors « fragmentées » en autant de (petites) adresses qu'il est nécessaire pour constituer plusieurs petites adresses fictives à la place de la grande adresse de la petite commune. A l'identifiant géographique des adresses (département/commune/ilot/rang d'adresse) est adjoint alors un numéro de fragment.

Une variable complémentaire de nature socio-économique était différente entre les Antilles-Guyane et La Réunion :

- Pour la DIRAG, l'équilibrage se fait sur l'évolution de la population entre 1990 et 1999 ;
- Pour La Réunion, l'équilibrage se fait sur une typologie de quartiers créée par la DR à partir du RP99.

2.2.5. Forces et faiblesses du dispositif standard cartographique

La base de sondage cartographique est une base d'adresses (et non de logements), qui contient (sous l'hypothèse d'une construction/destruction équivalente sur les groupes de rotation) environ 20% (1/5^{ème}) des adresses des PC et 12 % des adresses des GC (60 % de 1/5^{ème}), réparties sur l'ensemble des communes. Chaque année, la couverture de la base cartographique change, l'ensemble du territoire étant couvert au bout de 5 années.

Les principaux points positifs sont :

- Ce dispositif⁵ est robuste et adaptée aux enquêtes volumineuses, et fournit pour celles-ci une précision acceptable (+/- 1,5 points sur le taux de chômage dans l'EEADOM par exemple).
- Il est permis également de coordonner proprement le tirage de plusieurs enquêtes dans l'année et sur plusieurs années.

Pour ce qui concerne spécifiquement la base, on peut apprécier :

- sa représentativité : elle est représentative de l'ensemble des adresses du DOM ;
- sa fraîcheur : elle ne contient que des adresses cartographiées 1 voire 2 ans avant le tirage (et donc la collecte) ;
- les plans associés au SIG sont une aide précieuse pour le repérage, en plus des adresses postales contenues dans la base ;
- elle limite les sollicitations auprès des enquêtés :
 - o disjonction par construction des bases de sondage sur 5 ans : un logement tiré en n pour une enquête ménage ne peut être retiré avant l'année n+5, à l'exception des logements des grandes adresses des GC qui peuvent être tirés dans plusieurs enquêtes ménages d'une même année (ces adresses peu nombreuses ne sont pas retirées de la base de sondage après tirage pour ne pas déséquilibrer la base, un contrôle par le CRIEM est toutefois fait après tirage pour limiter ces réinterrogations) ;
 - o seules certaines adresses de la base cartographique peuvent être recensées et enquêtées la même année : les adresses des PC recensées dans l'année, c'est-à-dire 1/5^{ème} des adresses des PC figurant dans la base de sondage (adresses appartenant au groupe d'ilôts de l'année) et uniquement les grandes adresses (de plus de 21 logements) des GC (qui ne sont pas exclues de la base de tirage).

Les principaux points faibles sont :

- Tout échantillon est éclaté sur l'ensemble des territoires régionaux, ce qui est un gain pour la précision des enquêtes à échantillon volumineux, mais cela est peu performant pour de petits échantillons (contribution à l'échelle régionale pour la statistique nationale) qui devraient être concentrés afin de réduire les coûts importants de déplacement. Le dispositif était rigide pour moduler des plans de sondage en fonction d'arbitrages coûts/précision.
- Le dispositif ciblant des adresses sans aucune information sur les habitants ne permet pas de cibler des échantillons sur de petites sous-population (personnes âgées ou jeunes, individus actifs occupés ou non etc.). On devait alors estimer la part de la population cible dans la population totale et calibrer ensuite la taille d'un échantillon en population entière qui permette d'obtenir (en moyenne) la taille de la population cible.

Pour ce qui concerne la base de sondage, on peut déplorer les points suivants :

- elle contient peu d'information. Ainsi, habituellement, les variables intervenant dans le tirage pour stratifier et/ou équilibrer l'échantillon sont les types d'adresse, de bâti, l'appartenance à une zone géographique / zone d'emploi, la typologie des quartiers. Pour le tirage de l'enquête « Information et Vie Quotidienne » (IVQ) en Martinique et à La Réunion, le taux de non

⁵ Pour une description fine de la précision et de l'efficacité du dispositif, voir le rapport IG Ardilly-Domergue.

diplômés (issu du RP99 et défini à l'IRIS) a été introduit, mais une telle expérience n'a pas été reconduite.

- Un autre défaut, plus difficilement quantifiable, provient des erreurs d'appréciation de l'enquêteur cartographique, par exemple sur le nombre de logements d'une petite ou grande adresse qui peut être sous évalué ou sur évalué.
- Elle contient mécaniquement des logements « hors champ » : résidences secondaires, logements occasionnels, logements vacants.

2.3. Quelques entorses au dispositif standard

2.3.1. Panorama

A partir de 2007, progressivement, quelques sources de tirage alternatives ont été mobilisées pour résoudre certains problèmes spécifiques d'échantillonnage (cf. tableau 0).

En analysant ces dispositifs *a priori* comme *a posteriori*, des différences majeures sont apparues entre La Réunion d'une part et les départements de la Caraïbe et de Guyane d'autre part. Les sources administratives classiques utilisées en métropole sont souvent (mais pas toujours) mobilisables de façon pertinentes à La Réunion. L'enquête « Patrimoine » de 2009 par exemple a pu être échantillonnée à partir de la taxe d'habitation (TH), tandis que, côté DFA, elle n'a pu qu'être échantillonnée à partir du recensement qui a permis, à défaut de mieux, un certain profilage à partir des catégories socio-professionnelles déclarées dans le recensement.

Tableau 3 : enquêtes réalisées dans les DOM de 2005 à 2011 (hors EEDOM) et les bases de sondage associées :

	Carto	EAR	Autre	Observations
EEDOM dps 2005	X			Tirage d'adresses
Budget des familles 2005	X			Tirage d'adresses
Logement 2006	X			Tirage d'adresses
Enquête Chik Réunion - 2006	X			Tirage d'adresses
Vie Quotidienne et Santé 2007		X	EEDOM	EAR pour échantillon national (2006 pour les GC, 2004-2006 pour les PC) + enquête accolée à EEDOM 2007 pour échantillon extension
Information et Vie Quotidienne 2006 Martinique - 2007 Réunion - 2008 Guadeloupe	X	X 971		Pour Guadeloupe : tirage EAR 2007 pour les GC ; tirage EAR 2004-2007 et carto 2007 pour les PC Pour autres DOM : Tirage d'adresses dans la carto
Tech. Info. et Com 2007-2008			EEDOM	
Tech. Info. et Com. 2009			annuaire	annuaire France Télécom (tirage DG)
Handicap-Santé volet Ménages 2008			VQS	
Migrations-Famille-Vieillessement 2009	X		EEDOM	Tirage d'adresses dans la carto + EEDOM 2006-2008 pour le ssech des migrants retour
Patrimoine 2009		X Antilles (1)	TH Réunion	EAR 2004-2008 pr Antilles (éch proportionnel à pop en Martinique, soit 84 FA), tirage DG dans la TH pour Réunion
Emploi du Temps 2010	X			Tirage en 2 phases, IRIS puis logements
Enquête Energie ARER - 2010 La Réunion		X		EAR 2005-2009 pour les PC, EAR 2009 pour les GC ; tirage en 2 phases, IRIS puis logements
Loyers et Charges 2010-2011		X		EAR 2005-2010/11 pour les PC, EAR 2009/10/11 pour les GC
Budget des Familles 2010	X			
Cadre de Vie et Sécurité 2011	X Réunion			Tirage d'adresses
Information et Vie Quotidienne 2011 Guyane/Réunion		X		EAR 2006-2010 pour GC et PC

(1) éch de 84 FA pour 972

2.3.2. Le cas de l'utilisation du recensement

Pour la première fois en 2007, les bases des enquêtes annuelles de recensement (EAR) ont été utilisées pour le tirage d'une enquête dans les DOM, en l'occurrence pour les besoins de l'enquête VQS (Vie quotidienne et Santé). Les EAR ont été la base de sondage alternative la plus utilisée par la suite, et ont été réservées aux cas où :

- il est décidé que les DOM doivent se rapprocher du protocole de collecte national (VQS) ;
- il est jugé que l'utilisation d'informations auxiliaires au niveau du logement, du ménage ou de l'individu, absentes de la base cartographique, améliorerait significativement la qualité du tirage et/ou de la collecte : niveau de diplôme pour le tirage d'IVQ en Guadeloupe, CSP et âge pour se rapprocher, dans les DFA, de la stratification utilisée en métropole et à La Réunion pour l'enquête Patrimoine, restriction aux ménages locataires pour Loyers et Charges.

Contrairement au tirage métropolitain (utilisant l'outil Octopusse), le tirage Outre-Mer ne peut se contenter de la dernière EAR disponible, l'équilibrage des groupes de rotation se détériorant progressivement sur les grandes communes (GC), et n'existant pas dans les petites communes (PC). Les tirages dans le recensement n'ont donc pu s'effectuer qu'après plusieurs années de mise en œuvre du nouveau dispositif. Les tirages dans le recensement ont systématiquement été faits dans plusieurs EAR. Compte-tenu de la disponibilité des EAR, l'enquête VQS en 2007 n'a fait appel qu'aux 3 EAR 2004 à 2006 et l'enquête IVQ Guadeloupe en 2008 aux 4 EAR 2004 à 2007 (complétées par l'enquête cartographique préalable à l'EAR 2008). Les tirages effectués en 2009 et 2010 ont été fait sur un cycle complet de recensement (5 EAR).

Ces expérimentations montrent qu'une base de sondage constituée par l'empilement de 5 EAR consécutives peut être une base correcte de logements ou d'individus. Elle contient l'ensemble des adresses des PC et 40 % de celles des GC. L'expérience acquise au court de ces enquêtes a permis d'apprécier les atouts et faiblesses de l'utilisation des EAR pour l'échantillonnage des enquêtes auprès des ménages.

Les principaux points forts sont :

- sa représentativité : elle est représentative de l'ensemble des logements / individus du DOM ;
- sa richesse en information ;
- la limitation des « hors champ » : les résidences secondaires, les logements vacants, les logements occasionnels... peuvent être exclus avant le tirage. Par exemple, la différence dans les résultats entre enquête Patrimoine et enquête MFV se mesure ainsi : sur le département de la Guadeloupe, Hors champ Patrimoine (base-RP) = 12,5 % (dont 7,9 % de vacants) / Hors champ MFV (base carto + EE) = 17,3 % (dont 10,9 % de vacants) ; sur le département de La Réunion, Hors champ Patrimoine (base TH) = 6,1 % (dont 3,5 % de vacants) // Hors champ MFV (base carto + EE) = 10 % (dont 7 % de vacants) ;

Ses principaux points faibles (en l'état) sont :

- le manque de préparation méthodologique pour une utilisation raisonnée de la base et pour des calculs de précision ;
- une impossibilité d'appariement des enquêtes cartographiques avec les logements échantillonnés dans les petites communes, ce qui interdit l'utilisation de plan. Le repérage est détérioré ;
- des données de fraîcheur variable : elle contient des logements recensés jusqu'à 5 ans avant le tirage (et donc la collecte) ;
- Petite taille et renouvellement lent de la base : des tirages répétés peuvent poser problème. Par exemple, la « ponction » des locataires de logements loués vides (enquête « Loyers et Charges en Guadeloupe) déséquilibre la base de tirage ;
- risque de lassitude des enquêtés :
 - o L'ensemble des adresses de la base ont été recensées dans les 5 ans qui précèdent l'enquête (par construction !). En particulier, les adresses tirées dans la dernière EAR ont été recensées l'année précédant l'enquête (1/5^{ème} des adresses des PC, davantage dans les GC si l'on se limite à la dernière EAR disponible comme cela a

été fait pour Loyers et Charges où toutes les adresses des GC tirées ont été recensées l'année précédente).⁶

- Par ailleurs, une procédure de marquage des adresses tirées dans les EAR pour les besoins des enquêtes ménages doit être systématisée pour éviter l'interrogation de mêmes logements à des périodes rapprochées. Elle doit être construite dans les petites communes à cause du problème d'appariement précédemment cité.

2.3.3. Le cas de l'enquête « Loyers et Charges » en Guadeloupe

Cette enquête a pour but de construire un indice des loyers en Guadeloupe ; elle doit notamment fournir des éléments utiles à la confection des indices de prix. Elle s'adresse aux ménages louant leur logement vide, et est organisée en continu sur un rythme trimestriel (en gros, collectée le premier mois de chaque trimestre).

Son échantillonnage était soumis à la nécessité de ciblage des logements (« ménages locataires »), à la continuité de collecte (ménages à interroger cinq fois, tous les trois mois), mais également à deux autres contraintes, opposées entre elles : suffisamment disperser les fiches-adresse sur le territoire pour obtenir une représentation satisfaisante, mais également suffisamment les regrouper pour que les déplacements des enquêteurs ne soient pas exorbitants. Le plan de sondage a été établi dans un travail concerté entre le Criem et l'UMS « ménages ».

2.3.3.1. La méthode d'échantillonnage

En premier lieu, parmi les bases utilisables en Guadeloupe, le ciblage nécessaire de ménages louant vide leur logement a imposé l'utilisation du recensement. Pour des raisons de fraîcheur de données, il a été choisi d'utiliser la dernière EAR disponible en grande commune (le dernier « cinquième » disponible) et le dernier recensement – intégral – en petite commune.

Pour assurer une couverture acceptable du territoire, les quatre « micro-régions » guadeloupéennes (Pointe-à-Pitre, Grande-Terre, Basse-Terre Nord/Est, Basse-Terre-Ville) ont servi à stratifier géographiquement le tirage. Pour obtenir l'échantillon entrant de logements, on sélectionne au premier degré un certain nombre d'Unités Primaires (UP) par strate à probabilités égales, puis un nombre fixé de logements par UP retenue au premier degré. Les poids d'un échantillon ainsi tiré sont variables, mais la méthode a été jugée plus robuste et praticable. Le nombre d'UP tirées par micro-région suit le nombre de logements loués vide de la zone. La zone de Pointe-à-Pitre contient 25 UP, la zone de la Grande-Terre 17, la zone de la Basse-Terre-Nord/Est 16, celle de la Basse-Terre-Ville 21.

Concrètement, l'échantillon entrant comporte 132 logements, organisé en 11 UP de 12. La zone de Pointe-à-Pitre verra entrer 5 UP (et donc en aura 25 actives simultanément pour l'enquête), la zone de la Grande-Terre 2 UP (donc 10 actives simultanément), tout comme les deux zones de la Basse-Terre. Les 25 UP de la zone de Pointe-à-Pitre seront donc mobilisées en permanence (on aura une « vraie » stratification et pas un premier degré de tirage) ; cinq « groupes de rotation » sont donc définis une fois pour toute, avec un équilibrage sur le nombre de logements loués vides.

⁶ Le problème de lassitude des enquêtés a été en particulier constaté à La Réunion pour l'enquête « Emploi du temps » en 2010 : suite à quelques erreurs de repérage, soit au cours de la dernière campagne de recensement, soit au cours de l'enquête, les logements enquêtés et recensés au cours de la même année ont été plus nombreux qu'habituellement et les enquêteurs ont rapporté plus de plaintes quant aux sollicitations de l'INSEE.

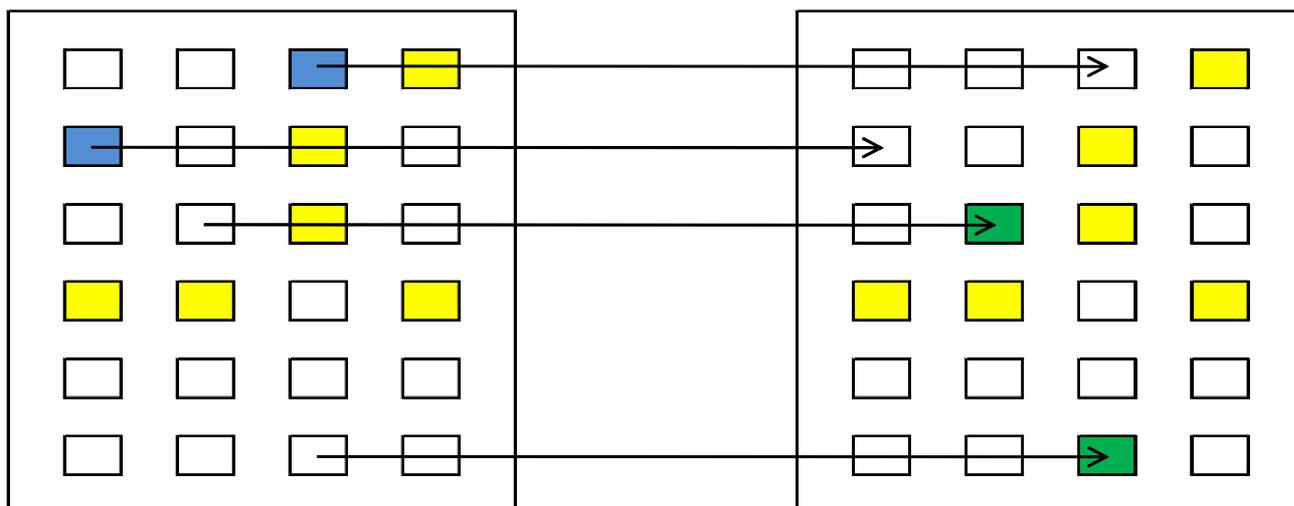


Schéma de rotation sur une micro-région sur deux trimestres

Note de lecture :

Les deux grands cadres représentent une même micro-région lors de deux trimestres consécutifs. A l'intérieur les UP sont représentées par les petites cases.

La micro-région fictive représentée comprend 24 UP. Chaque trimestre, 8 de ces 24 UP sont actives : ce sont les UP colorées.

A la fin du premier des trimestres, situation de gauche, les UP bleues ont été mobilisées pour construire l'échantillon cinq trimestres auparavant, les UP jaunes ont été introduites quatre, trois, deux et un trimestre auparavant. Les UP non actives sont non-colorées.

Au trimestre suivant, on aura la situation de droite. Les UP bleues ont été désactivées, mais des UP vertes ont été échantillonnées parmi les UP précédemment inactives ; les UP bleues actives au trimestre précédent ont été inactivées.

2.3.3.2. Constitution des Unités Primaires du tirage

L'ensemble des IRIS fournit *a priori* un modèle d'Unités Primaires plaisant. Déjà constitué, il permet d'obtenir facilement de nombreuses données agrégées mobilisables si nécessaire pour équilibrer les tirages. Il faut cependant s'assurer alors que chaque IRIS contienne suffisamment de logements loués vides, non seulement au moins le nombre à prélever pour un échantillon entrant, mais au nombre correspondant à tous les prélèvements *possibles* jusqu'au renouvellement de l'EAR. Puisque le rythme de renouvellement diffère entre petites et grandes communes, deux seuils minimaux pour une UP acceptable doivent être calculés. Cette question de volumétrie est importante car la part de locataires est assez faible en Guadeloupe, voire très faible dans certaines zones ; par ailleurs, la méthode du recensement contraint fortement le nombre de logements mobilisables dans les grandes communes.

Dans les petites communes, le recensement est exhaustif. L'ensemble des logements de chaque IRIS est présent dans la base pour cinq ans sans modification. Comme l'échantillon de « Loyers et Charges » est renouvelable par 1/5^e tous les trimestres, *une même UP est potentiellement échantillonnée au premier degré quatre fois* au cours d'un cycle quinquennal de recensement.

En effet, sur les vingt trimestres (5 ans) de vie de l'EAR, considérons une UP entrant en période 1 ; elle reste en périodes 2, 3, 4, 5, puis est inactivée pendant la période 6. Elle *peut* (pire cas) entrer à nouveau en 7, et alors rester jusqu'en 11 ; elle est à nouveau inactivée en 12. En suivant la logique du pire, elle *peut* entrer à nouveau dans l'échantillon en 13 jusqu'en 17, avoir une inactivation en 18, et être mobilisée une nouvelle fois pour la période 19. Pour ne pas réinterroger deux fois un

même logement dans le délai raisonnable de cinq ans, une UP devrait contenir au moins $4 \times 12 = 48$ logements loués vides.

Ces considérations de taille ont conduit à la définition de 26 UP sur les petites communes (deux fusions ont été nécessaires) contenant de 62 à 330 logements loués vides.

Dans les grandes communes, les EAR sont renouvelées chaque année. Une UP ne peut donc être mobilisée qu'une seule fois par EAR. La limite de logements est donc dans ce cas très basse : une UP semble devoir contenir au moins 12 logements seulement. La limite semble moins forte que dans le cas des petites communes, mais la méthode de recensement induit des difficultés spécifiques. En effet, une fraction seulement de chacun des IRIS est enquêtée au titre du recensement. Les groupes de rotation du recensement des grandes communes des DOM « historiques » sont constitués géographiquement à partir d'îlots, répartis par équilibrage entre les GR à partir du recensement de 1999. Une année sur cinq, chaque *îlot* sera présent dans la base de sondage de l'EAR. En moyenne (mais seulement en moyenne), 1/5^e seulement des îlots d'un IRIS sera enquêté dans une EAR donnée. Ceci implique que la base utile dans un IRIS donné (en termes d'îlots mobilisables) sera modifiée chaque année ; rien n'interdit qu'elle ne soit en fait nulle si un IRIS n'est pas représenté sur un ou plusieurs GR. La brique de base pour étudier la conformité d'un découpage en UP est ainsi le croisement UP x GR. La contrainte des 12 logements s'applique à *cette* brique. Cette contrainte est plus forte que celle en petite commune, non seulement l'UP doit contenir au moins 60 logements au total, chaque année 12 au moins doivent être mobilisable dans l'EAR correspondante. Enfin, la limite des 12 logements doit nécessairement s'interpréter à la hausse : seulement 40% des *adresses* sont enquêtées (on ne dispose *dans la base de sondage* que d'environ 40% des nombres estimés *par le recensement*), et les grandes adresses ne sont pas incluses *de facto* dans l'échantillon de recensement. La brique de base du raisonnement est le croisement IRIS x GR, qui est très fin : l'estimation du nombre de logements loués vides d'une telle zone est donc instable et doit être manipulé avec précaution.

Peu d'IRIS respectent le critère des 12 logements par croisement EAR x GR. Une application stricte du principe est astronomique et conduit à un tel nombre de regroupements, que cette opération remet finalement en cause le principe du plan de sondage : constituer des UP pour limiter les déplacements des enquêteurs. Le découpage finalement proposé a desserré un peu la contrainte en tolérant quelques croisements inférieurs à la limite. On accepte ainsi, sous contrainte du terrain, de risquer d'avoir quelques sous-échantillons entrants légèrement amputés. 37 IRIS ont été regroupés, on construit *in fine* 53 UP sur les grandes communes.

Ce phénomène de « traces » géographiques des EAR sur des zones de collecte et/ou de tirage spécifiques induira régulièrement des difficultés. Ici, la constitution des UP est largement contrainte ; dans le nouveau système (cf. §4.), la variabilité des « traces » induira une variabilité dans les charges *a priori* affectées aux Secteurs d'Activité des Enquêteurs (SAE).

3. Vers une utilisation du recensement comme base de sondage standard pour les DOM « historiques »

3.1. Changer, certes, mais pourquoi/pour quoi ?

Régulièrement, l'idée d'une refonte du dispositif de sondage avec l'utilisation, comme en métropole, du recensement a ainsi été avancée ; on l'a déjà vu, le système « historique » manquait singulièrement de flexibilité, même si le recensement était parfois mobilisé dans des cas d'échantillonnage plus complexe. Ce recours était cependant « systématiquement *ad hoc* » : il manquait un travail approfondi pour en explorer les possibilités et limites.

Au delà, utiliser le recensement permet d'obtenir l'impression au Centre National Informatique de Nantes (CNIN) de fiches-adresses contenant les scans des adresses (déclarées) par les individus, comme c'est le cas pour les enquêtes réalisées en France métropolitaine dans le dispositif OCTOPUSSE, qui permet d'échantillonner à partir du recensement. Or la question du repérage des logements est centrale pour les unités Enquêtes-Ménages (UEM) de la Dirag ; tout à la fois les outils des Systèmes d'Information Géographiques (SIG) y sont moins performants qu'à La Réunion et les adresses non normalisées encore plus fréquentes : l'adresse déclarée dans le recensement, combinée aux outils SIG, semblait une piste d'amélioration importante pour le travail des UEM.

Le rapport de l'Inspection générale sur la « stratégie d'enquête auprès des ménages dans les DOM » réalisé par Philippe Domergue et Pascal Ardilly a également abordé cette question. La sixième recommandation du rapport est ainsi la suivante : « *l'échantillonnage des logements à partir d'une base d'adresses issues de l'enquête cartographique ne va pas nécessairement dans le sens de l'efficacité, compte tenu du recul dont on dispose désormais sur les enquêtes de recensement dans les DOM. Il conviendrait de reprendre cette question et de s'interroger sur l'intérêt du passage à une base de logements recensés.* »

Les (lourds) travaux entrepris dans le cadre du projet REFLEE de refonte de l'enquête sur l'emploi et de son extension dans les départements ultra-marins a été l'occasion d'entamer des réflexions sur une refonte parallèle des dispositifs d'échantillonnage des DOM « historiques » (Mayotte réalisant quant à elle une Enquête sur l'Emploi Annuelle depuis 2013). Compte tenu des spécificités de gestion de l'enquête (petit échantillon entrant, petit nombre d'enquêtes en face à face, gestion des déplacements des enquêteurs sous ces contraintes en particulier), le maintien (même avec amélioration éventuelle) d'un échantillonnage cartographique posait plus de problèmes qu'il n'en résolvait.

Plusieurs sources pour l'échantillonnage étaient virtuellement ouvertes : recensement, sources fiscales en particulier. Si les sources fiscales semblaient utilisables à La Réunion et « probablement » utilisable en Martinique (« Patrimoine » en 2009 a été échantillonné avec succès à La Réunion à partir de la taxe d'habitation ; une rapide analyse sur les volumes menée par V. Loonis sur les DFA avait conclu à la possible utilisation de cette même source pour la Martinique seulement), ce n'était pas le cas en Guadeloupe et en Guyane. L'utilisation des sources fiscales laissait par ailleurs entier le problème du repérage dans les Antilles-Guyane. La volonté de conserver un dispositif commun entre les quatre DOM « historiques » a ainsi conduit à ne pas (encore) utiliser ces sources fiscales comme dispositif standard dans aucun département d'Outre-mer.

La piste « recensement » a ainsi été privilégiée.

Plusieurs études à caractère méthodologique ont été réalisées au Criem pour orienter la mise en place du nouveau dispositif d'échantillonnage.

La fraîcheur des données est un facteur de qualité important, auquel répond le choix dans le dispositif OCTOPUSSE de ne mobiliser que la dernière EAR, dans les petites communes comme dans les grandes. Le faible nombre de petites communes (PC) ainsi que leurs caractères très spécifiques

obligeaient le dispositif, pour les petites communes, à utiliser l'ensemble du cycle PC. Pour les grandes communes, cependant, la question n'était pas naturellement tranchée.

3.2. La question du nombre d'EAR en grande commune : les volumes.

L'utilisation d'une seule EAR conduit à une base de sondage de volume trop faible pour une politique d'enquête ambitieuse pour l'Outre-mer. Une EECDOM *doit* pouvoir être réalisée partout ; on *devrait* pouvoir espérer réaliser une ou deux enquêtes supplémentaires exploitables localement, donc avec un volume conséquent. Le volume dont on ampute la base de sondage peut être particulièrement conséquent (et supérieur au volume de fiches-adresse finalement enquêtées) quand il faut surreprésenter des sous-populations. En effet, dans ce cas, il faut amputer la base de sondage à la mesure de la surreprésentation de la sous-population, pour ne pas « déséquilibrer » le reste de la base de sondage où seront échantillonnées des enquêtes ultérieures.

Si on appuie le système standard d'échantillonnage sur une seule EAR, la situation est à la suivante. Une EECDOM peut être réalisée dans tous les DOM, mais la Guyane est très contrainte pour la mise en œuvre d'une enquête de volumétrie conséquente ou avec surreprésentation. La Guadeloupe et la Martinique ne peuvent organiser qu'une enquête de volumétrie importante ou avec surreprésentation. Ce seul argument de volumétrie conduit à penser qu'il est délicat de passer à un système entièrement appuyé sur une seule EAR, sans qu'il y ait régulièrement à inventer des solutions *ad hoc*, en parallèle d'un nouveau système standard.

Il faut donc *empiler* des EAR pour dégager un volume suffisant important de logements à échantillonner pour soutenir la politique d'enquêtes. Mais l'estimation du volume finalement disponible par empilement nécessite un peu d'attention. En effet, quand on tire dans plusieurs EAR, les EAR « anciennes » ont déjà subi des ponctions lors des échantillonnages précédemment réalisés ; ainsi, la taille d'une base annuelle de tirage appuyée sur n EAR sera inférieure à la taille de l'ensemble des EAR mobilisées. Si on considère un empilement (glissant) de deux EAR, le volume de logements disponibles dans la base n'est doublé qu'à l'initialisation ; dès la deuxième année, l'EAR la plus ancienne aura déjà été ponctionnée pendant la première année : le volume disponible pendant la deuxième année n'est que l'addition des volumes de la nouvelle EAR et de l'ancienne, *mis à part les logements échantillonnés préalablement*.

Pour défricher la situation, on va *raisonner en moyenne* dans la situation simplifiée où les EAR ont une taille constante X et où on doit mobiliser tous les ans Y ⁷ fiches-adresses.

3.2.1. Le cas de deux EAR

Notons P_n le nombre de FAs qui seront tirées sur l'EAR de l'année n ⁸. P_{n-1} FAs ont été tirées l'année précédente dans l'EAR de l'année $n-1$.

La base de l'année n est ainsi constituée des X observations de l'EAR de l'année n et de $X - P_{n-1}$ observations de l'EAR de l'année $n-1$. Autrement dit, Y FAs sont à tirer dans $X + (X - P_{n-1})$.

⁷ On montre par une récurrence simple qu'on peut, dans ce cadre, assimiler le tirage successif de plusieurs enquêtes pendant l'année (qui est la situation standard) dans deux EAR au tirage d'une seule enquête dont le volume est la somme des volumes des enquêtes successives, si les spécifications des tirages sont les mêmes (équilibre, surreprésentation). Pour simplifier les expressions numériques, on raisonnera donc sur une unique ponction globale de Y FAs sur l'année.

⁸ L'EAR de l'année n étant disponible pendant l'année $n+1$, il faudrait préciser en toute rigueur que $P(n)$ est le nombre de FAs tirées dans l'EAR de l'année n pendant l'année $n+1$, afin de ne pas introduire d'ambiguïté. On tirera aussi dans l'EAR de l'année n pendant l'année $n+2$, dans le cas de figure décrit : ce volume n'est pas « noté » mais vaudra $X - P(n+1)$

En moyenne, on va tirer pendant l'année n :

- $Y \times \frac{X - P_{n-1}}{2.X - P_{n-1}}$ dans l'EAR de l'année n-1
- $Y \times \frac{X}{2.X - P_{n-1}}$ dans l'EAR de l'année n, soit P(n)

La fonction à analyser est ainsi : $P_n = Y.X / (2.X - P_{n-1})$

Pour tout n, on a la relation $0 < P_n < Y < X$.

La fonction de récurrence est contractante, on en déduit que la valeur limite des P est :

$$X - \sqrt{X.(X - Y)} = X - X.\sqrt{1 - \varepsilon} = X.(1 - \sqrt{1 - \varepsilon})$$

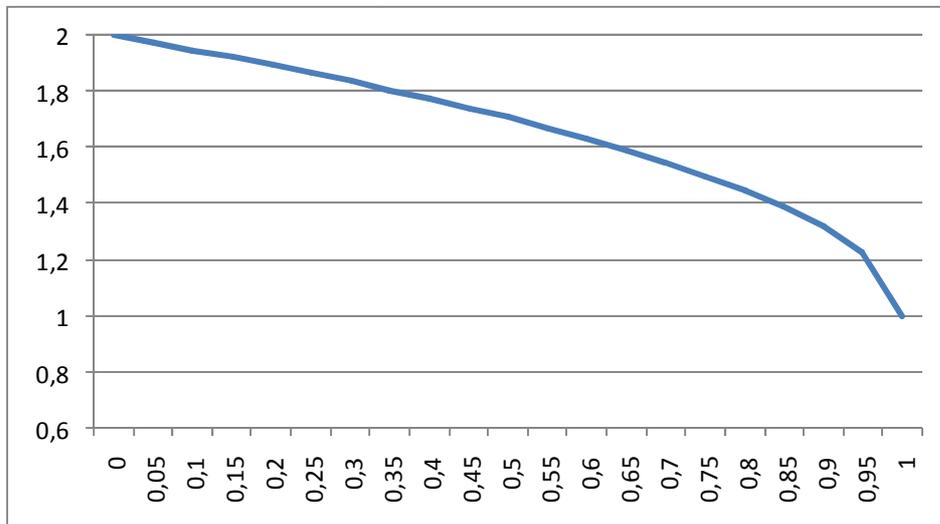
En notant $\varepsilon = Y/X$

En régime constant, la taille de la base sera donc : $X.(1 + \sqrt{1 - \varepsilon})$.

Elle varie donc de près de deux fois le volume d'une EAR (si le volume d'enquête est petit devant le volume de la base) à un peu plus du volume d'une EAR seule (si le volume d'enquêtes est proche du volume d'une EAR).

Ainsi, plus Y est proche de X (plus le volume de FAs annuel à tirer est proche de la taille d'une EAR), plus la part tirée dans la nouvelle EAR sera grande.

A la limite, si X est très proche de Y, on tirera « toutes » les FAs dans la nouvelle EAR.



En régime constant, taille de la base utile (base 1) en fonction de ε, la part que représente le volume de fiches-adresse tirées annuellement dans le volume d'une EAR

Une application numérique proche de la situation guyanaise :

Une EAR guyanaise comporte environ 8 000 logements. Imaginons que le tirage annuel y est de 6 000 logements sur deux EAR.

En n, on tire (en moyenne) 3 000 logements dans les EAR n-1 et n

En n+1, la base est amputée (8 000 - 3 000 = 5 000 observations disponibles pour l'EAR n). On va tirer 3 700 observations dans l'EAR n+1.

En n+2, la base est plus amputée encore (8 000 - 3 700 = 4 300 observations disponibles pour l'EAR n+1). On va tirer 3 900 observations dans l'EAR n+2

En n+3, on va tirer 3 970 observations dans l'EAR n+3.

En n+4, on va tirer 3 990 observations dans l'EAR n+4.

La « croissance » des parts à tirer dans la nouvelle EAR est en racine carrée, et converge vers :

$$8000 - \sqrt{8000 \times (8000 - 6000)} = 4000$$

Dans cette configuration, en régime courant, les 2/3 des observations de l'échantillon seront tirés dans la nouvelle EAR ; autre point de vue, le volume effectivement disponible en enquêtant en régime constant sur deux EAR est celui d'une EAR et demi.

En réalité, on constate que les tirages annuels sont plutôt autour de 4 000 logements en Guyane. Chaque année environ 1400 logements sont tirés pour l'enquête emploi en continu, auxquels s'ajoute 1500 à 3000 logements d'une 2^{ème} grosse enquête, Cadre de Vie et Sécurité en 2015, Logement en 2013-2014, Conditions de Travail en 2012. Dans cette configuration, en régime courant, 2340 observations de l'échantillon seront tirés dans la nouvelle EAR, 1660 dans l'EAR n-1.

3.2.2. Le cas de trois EAR

La résolution analytique de l'équation de récurrence est autrement plus difficile dans le cas de trois EAR.

Pour commencer à analyser le problème, on peut faire une simulation numérique pour étudier le déroulé de l'utilisation de trois EAR, avec les volumes de l'exemple précédent. Le tableau suivant réalise une telle simulation, avec une initialisation au fil de l'eau du dispositif.

		reste dans l'EAR n	reste dans l'EAR n-1	reste dans l'EAR n-2	échantillonné dans l'EAR n	échantillonné dans l'EAR n-1	échantillonné dans l'EAR n-2	échantillonné dans l'EAR n (%)	échantillonné dans l'EAR n-1 (%)	échantillonné dans l'EAR n-2 (%)	taille de la base de l'année n
année d'enquête 1	6000	8000	0	0	6000	0	0	100,00	0,00	0,00	8000
année d'enquête 2	6000	8000	2000	0	4800	1200	0	80,00	20,00	0,00	10000
année d'enquête 3	6000	8000	3200	800	4000	1600	400	66,67	26,67	6,67	12000
année d'enquête 4	6000	8000	4000	1600	3529	1765	706	58,82	29,41	11,76	13600
année d'enquête 5	6000	8000	4471	2235	3264	1824	912	54,40	30,40	15,20	14706
année d'enquête 6	6000	8000	4736	2647	3120	1847	1032	52,01	30,79	17,21	15383
année d'enquête 7	6000	8000	4880	2889	3044	1857	1099	50,73	30,95	18,32	15768
année d'enquête 8	6000	8000	4956	3023	3004	1861	1135	50,07	31,02	18,92	15979
année d'enquête 9	6000	8000	4996	3095	2983	1863	1154	49,72	31,05	19,23	16091
année d'enquête 10	6000	8000	5017	3133	2972	1864	1164	49,54	31,06	19,40	16150
année d'enquête 11	6000	8000	5028	3153	2966	1864	1169	49,44	31,07	19,49	16181
année d'enquête 12	6000	8000	5034	3164	2964	1865	1172	49,39	31,08	19,53	16197
année d'enquête 13	6000	8000	5036	3169	2962	1865	1173	49,37	31,08	19,55	16205
année d'enquête 14	6000	8000	5038	3172	2961	1865	1174	49,35	31,08	19,57	16210
année d'enquête 15	6000	8000	5039	3173	2961	1865	1174	49,35	31,08	19,57	16212
année d'enquête 16	6000	8000	5039	3174	2961	1865	1175	49,34	31,08	19,58	16213
année d'enquête 17	6000	8000	5039	3174	2960	1865	1175	49,34	31,08	19,58	16214
année d'enquête 18	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16214

Un système sur trois EAR possède beaucoup plus d'inertie qu'un système sur deux : il arrive en régime quasi-constant beaucoup plus lentement (réciproquement quatre ans et une douzaine d'années).

En régime constant, on tire cette fois seulement la moitié des FAs dans l'EAR la plus fraîche (et un tiers dans la précédente, et un cinquième dans la plus ancienne). La taille de la base est doublée dans ce cas de figure.

Si on initialise le tirage sur trois EAR par un tirage sur trois EAR (et non au fil de l'eau comme pour la simulation précédente), on atteint un peu plus rapidement un régime quasi constant, et ce par valeurs supérieures pour le volume disponible de la base de l'année n (cf. tableau suivant).

		reste dans l'EAR n	reste dans l'EAR $n-1$	reste dans l'EAR $n-2$	échantillonné dans l'EAR n	échantillonné dans l'EAR $n-1$	échantillonné dans l'EAR $n-2$	échantillonné dans l'EAR n (%)	échantillonné dans l'EAR $n-1$ (%)	échantillonné dans l'EAR $n-2$ (%)	taille de la base de l'année n
année d'enquête 1	6000	8000	8000	8000	2000	2000	2000	33,33	33,33	33,33	24000
année d'enquête 2	6000	8000	6000	6000	2400	1800	1800	40,00	30,00	30,00	20000
année d'enquête 3	6000	8000	5600	4200	2697	1888	1416	44,94	31,46	23,60	17800
année d'enquête 4	6000	8000	5303	3712	2821	1870	1309	47,02	31,17	21,82	17016
année d'enquête 5	6000	8000	5179	3433	2889	1871	1240	48,16	31,18	20,67	16612
année d'enquête 6	6000	8000	5111	3309	2923	1868	1209	48,72	31,13	20,15	16419
année d'enquête 7	6000	8000	5077	3243	2941	1866	1192	49,02	31,11	19,87	16320
année d'enquête 8	6000	8000	5059	3210	2950	1866	1184	49,17	31,09	19,73	16269
année d'enquête 9	6000	8000	5050	3193	2955	1865	1180	49,25	31,09	19,66	16243
année d'enquête 10	6000	8000	5045	3184	2958	1865	1177	49,29	31,09	19,62	16229
année d'enquête 11	6000	8000	5042	3180	2959	1865	1176	49,32	31,08	19,60	16222
année d'enquête 12	6000	8000	5041	3177	2960	1865	1175	49,33	31,08	19,59	16218
année d'enquête 13	6000	8000	5040	3176	2960	1865	1175	49,33	31,08	19,59	16217
année d'enquête 14	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16216
année d'enquête 15	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16215
année d'enquête 16	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16215
année d'enquête 17	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16215
année d'enquête 18	6000	8000	5040	3175	2960	1865	1175	49,34	31,08	19,58	16215

Quand on veut échantillonner pour l'année n , la base de sondage est ainsi constituée :

EAR : année n	EAR : année $n-1$	EAR : année $n-2$
	ponctionné en $n-1$	ponctionné en $n-1$ / ponctionné en $n-2$

La base de tirage « effective » est ainsi constituée de l'intégralité de l'EAR n , de la partie non ponctionnée (en $n-1$) de l'EAR $n-1$, et de la partie non ponctionnée (ni en $n-1$ ni en $n-2$) de l'EAR $n-2$.

On va se placer dans la même configuration que pour le cas de l'utilisation de deux EAR : l'EAR annuelle est de taille fixe égale à X et on tire tous les ans un échantillon globale de Y fiches-adresse pour les besoins d'enquêtes.

Adoptons les notations suivantes :

$$\begin{cases} P_n: \text{nombre de FAs ponctionnées en } n \text{ dans l'EAR } n \\ Q_n: \text{nombre de FAs ponctionnées en } n \text{ dans l'EAR } n-1 \\ R_n: \text{nombre de FAs ponctionnées en } n \text{ dans l'EAR } n-2 \end{cases}$$

On a la relation $P_n + Q_n + R_n = Y$

On déduit immédiatement le système de suites numériques suivant :

$$\begin{cases} P_n = Y \times \frac{X}{(X) + (X - P_{n-1}) + (X - P_{n-2} - Q_{n-1})} \\ Q_n = Y \times \frac{X - P_{n-1}}{(X) + (X - P_{n-1}) + (X - P_{n-2} - Q_{n-1})} \\ R_n = Y \times \frac{X - P_{n-2} - Q_{n-1}}{(X) + (X - P_{n-1}) + (X - P_{n-2} - Q_{n-1})} \end{cases}$$

$$\frac{Q_n}{P_n} = \frac{X - P_{n-1}}{X}$$

Il suit que :

$$Q_n = P_n \times \left(1 - \frac{P_{n-1}}{X}\right) \text{ puis } Q_{n-1} = P_{n-1} \times \left(1 - \frac{P_{n-2}}{X}\right)$$

En injectant cette expression dans la première équation du système, on obtient la relation de récurrence suivante :

$$X \times Y = P_n \times \left(3 \times X - 2 \times P_{n-1} - P_{n-2} + \frac{P_{n-1} \times P_{n-2}}{X}\right)$$

La relation est non-linéaire, on n'a pas de solution standard sous la main.

Supposons dans un premier temps qu'il existe bien une limite à cette suite, et tentons de l'exhiber. Cette suite étant définie sur un compact (à valeur dans $[0 ; Y]$), elle possède d'ailleurs au moins une valeur d'adhérence (Bolzano-Weierstrass). Cette limite éventuelle est nécessairement finie, comprise entre 0 et Y.

La fonction liant P_n à P_{n-1} et P_{n-2} est continue, passons à la limite à la hussarde dans l'expression précédente. Une limite P de $(P_n)_n$ sera solution de l'équation du troisième degré suivante :

$$X \times Y = P \times \left(3 \times X - 2 \times P - P + \frac{P \times P}{X}\right) \quad (1)$$

Soit donc :

$$P^3 - 3X.P^2 + 3X^2.P - X^2.Y = 0$$

Une étude simple de fonction montre que cette fonction polynomiale est strictement croissante sur \mathbb{R} , à part au point d'inflexion d'abscisse X . Elle n'admet donc qu'un unique zéro (réel).

Aucune solution « évidente » n'apparaît, on peut avoir recours à la méthode de Tartaglia-Cardan pour calculer analytiquement les zéros de ce polynôme.

La méthode revient à résoudre le polynôme en Z dit « réduit » en suivant ensuite un développement spécifique:

$$Z^3 + (0) \times Z + (X^2 \times (X - Y)) = Z^3 + (X^2 \times (X - Y)) = 0 \text{ en posant } P = Z + X$$

Sans avoir besoin de recourir finalement plus avant à la méthode de Tartaglia-Cardan, on fait apparaître ici une première (et on l'a vu, unique) solution réelle :

$$Z_0 = -\sqrt[3]{X^2 \times (X - Y)}$$

Il suit que la valeur du zéro du polynôme P correspondant est :

$$P_0 = X - \sqrt[3]{X^2 \times (X - Y)}$$

La valeur du discriminant de l'équation « réduite » peut nous informer sur la forme de l'ensemble des solutions de l'équation, sans avoir à faire la réduction directe du polynôme.

$$\Delta = -\left[4 \cdot (0)^3 + 27 \cdot (X^2 \times (X - Y))^2\right] < 0$$

Le discriminant est négatif ; on en conclut que le polynôme admet une racine réelle et deux racines complexes conjuguées.

« La » candidate pour la limite de la suite $(P_n)_n$ est ainsi $X - \sqrt[3]{X^2 \times (X - Y)}$.

Si on note ε la part que représente le volume annuel de Y par rapport au volume X une EAR ($Y = \varepsilon \cdot X$), on peut écrire cette candidate sous la forme :

$$P_0 = X - \sqrt[3]{X^2 \times (X - Y)} = X - \sqrt[3]{X^2 \times (X - \varepsilon \cdot X)} = X - X \cdot \sqrt[3]{1 - \varepsilon} = X \cdot (1 - \sqrt[3]{1 - \varepsilon})$$

Sous cette hypothèse, la taille « utile » de la base de sondage serait alors en régime constant :

$$\frac{X \cdot Y}{X - \sqrt[3]{X^2 \times (X - Y)}} = \frac{Y}{1 - \sqrt[3]{1 - \varepsilon}} = \frac{\varepsilon \cdot X}{1 - \sqrt[3]{1 - \varepsilon}}$$

Retour sur l'application numérique précédente :

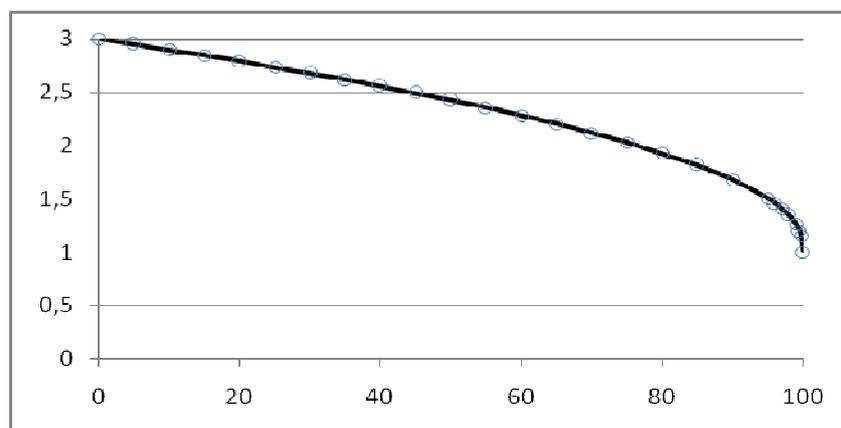
La simulation introductive utilisait une EAR de taille fixe $X = 8\,000$ pour des tirages annuels de $Y = 6\,000$ fiches-adresse.

La résolution analytique donne la solution suivante :

$$P = 8\,000 - \sqrt[3]{8\,000^2 \times (8\,000 - 6\,000)} \cong 2\,960$$

C'est bien la valeur « limite » proposée par la simulation.

Le graphe suivant présente la taille de la base utile (l'unité est ici X , la taille de l'EAR annuelle) en fonction du volume Y de fiches-adresse échantillonnées rapporté au volume X d'une EAR.



De la même manière que dans le cas de deux EAR, quand le volume de l'ensemble des fiches-adresse échantillonnées dans l'année est très faible devant la taille de l'EAR, la taille de la base « utile » est maximal (ici, de près de trois fois le volume de l'EAR) et inversement, si le volume échantillonné est de l'ordre de la taille d'une EAR, le volume « utile » est minimal (de l'ordre d'une EAR).

La décroissance est lente : quand le volume total des fiches-adresse est de l'ordre de la moitié du volume d'une EAR, le volume de la base utile est encore de 2,5 fois le volume d'une EAR ; il faut que le volume des fiches-adresse soit de l'ordre de 80% d'une EAR pour que le volume « utile » passe sous la barre de deux EAR.

Pour aborder sérieusement la question de la convergence de la suite $(P_n)_n$, il faut recourir à une mise en forme un peu plus sophistiquée. On peut tenter de résoudre le problème en le considérant comme un système dynamique discret sur \mathbb{R}^2 . Ce qui tiendra lieu de *conditions initiales* seront les volumes offerts dans les « pseudo bases ponctionnées antérieures » à la première étape.

On va considérer la suite $(P_n)_n$ de couples réels (ou, selon le point de vue, de points du plan) définie par la relation de récurrence suivante :

$$\mathcal{P}_{n+1} = (P_{n+1}; P_n) = \phi(\mathcal{P}_n) = \phi((P_n; P_{n-1}))$$

$$\text{avec } \phi \left\{ \begin{array}{l} \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ \begin{pmatrix} x \\ y \end{pmatrix}' \rightarrow \begin{pmatrix} f(x, y) = \frac{X^2 \cdot Y}{3 \cdot X^2 - 2X \cdot x - X \cdot y + x \cdot y} \\ g(x, y) = x \end{pmatrix}' \end{array} \right.$$

$(P_n)_n$ converge si et seulement si $(P_n)_n$ converge ; on peut donc travailler sur la suite $(P_n)_n$ en la considérant analytiquement comme une suite d'ordre 1, à valeurs complexes et définie par une fonction continue (et même holomorphe), ou bien comme un système dynamique discret du plan.

Dans le système qui nous intéresse, ϕ est restreinte au compact $[0; Y]^2$. Le théorème du point fixe nous garantit alors l'existence d'un point fixe pour la suite récurrente complexe définie par ϕ .

Soit donc (x^*, y^*) un point fixe de ϕ . Cela se traduit par le système suivant :

$$\begin{cases} x^* = \frac{X^2 \cdot Y}{3 \cdot X^2 - 2X \cdot x^* - X \cdot y^* + x^* \cdot y^*} \\ y^* = x^* \end{cases}$$

On est ramené à la résolution du même polynôme que précédemment. Le point fixe de l'application ϕ est donc :

$$(x^*, y^*) = (P; P) \quad \text{avec } P = X - \sqrt[3]{X^2 \times (X - Y)} = X \cdot (1 - \sqrt[3]{1 - \varepsilon}) \quad \text{avec } \varepsilon = Y/X$$

Il faut maintenant montrer que ce point fixe est attractif pour qu'il soit valeur d'adhérence de $(P_n)_n$ (et donc de $(P_n)_n$). Si on montre que le rayon spectral de la matrice de Jacobi de l'application ϕ en son point fixe $(P; P)$ est strictement inférieur à 1, on sera assuré d'une convergence locale, c'est-à-dire qu'avec de « bonnes » conditions initiales, $(P_n)_n$ convergera vers la valeur-limite qu'on vient d'exhiber.

Avec les notations précédentes, on peut écrire :

$$f(x, y) = \frac{X^2 \cdot Y}{3 \cdot X^2 - 2X \cdot x - X \cdot y + x \cdot y} = \frac{X^2 \cdot Y}{X^2 + (x - X)(y - 2X)}$$

Cela nous permet d'obtenir facilement :

$$\frac{\partial f}{\partial x}(x, y) = \frac{X^2 \cdot Y \cdot (y - 2X)}{[X^2 + (x - X)(y - 2X)]^2} \quad \frac{\partial f}{\partial y}(x, y) = \frac{X^2 \cdot Y \cdot (x - X)}{[X^2 + (x - X)(y - 2X)]^2}$$

$$\frac{\partial g}{\partial x}(x, y) = 1 \quad \frac{\partial g}{\partial y}(x, y) = 0$$

On en déduit la matrice jacobienne de ϕ en tout point de $[0; Y]^2$:

$$J(\phi)(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x}(x, y) & \frac{\partial f}{\partial y}(x, y) \\ \frac{\partial g}{\partial x}(x, y) & \frac{\partial g}{\partial y}(x, y) \end{pmatrix} = \begin{pmatrix} \frac{X^2 \cdot Y \cdot (y - 2X)}{[X^2 + (x - X)(y - 2X)]^2} & \frac{X^2 \cdot Y \cdot (x - X)}{[X^2 + (x - X)(y - 2X)]^2} \\ 1 & 0 \end{pmatrix}$$

Exprimons maintenant $J(\phi)(P, P)$.

On sait que P vérifie la relation (1), ce qui implique, en multipliant à gauche et à droite par X qui est non nul :

$$\begin{aligned} X^2 \times Y &= P \times \left(3 \times X^2 - 2 \times P \times X - X \times P + \frac{X \times P \times P}{X} \right) = P \times (3 \times X^2 - 3 \cdot X \times P + P^2) \\ &= P \times (X^2 + (P - X) \times (P - 2 \cdot X)) \end{aligned}$$

Il suit que :

$$\begin{aligned} J(\phi)(P, P) &= \begin{pmatrix} \frac{X^2 \cdot Y \cdot (P - 2X)}{[X^2 + (P - X)(P - 2X)]^2} & \frac{X^2 \cdot Y \cdot (P - X)}{[X^2 + (P - X)(P - 2X)]^2} \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{X^2 \cdot Y \cdot (P - 2X)}{[X^2 \cdot Y / P]^2} & \frac{X^2 \cdot Y \cdot (P - X)}{[X^2 \cdot Y / P]^2} \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{P^2 \cdot (P - 2X)}{X^2 \cdot Y} & \frac{P^2 \cdot (P - X)}{X^2 \cdot Y} \\ 1 & 0 \end{pmatrix} \end{aligned}$$

En notant λ et μ les deux valeurs propres de $J(\phi)(P, P)$, on remarque que :

$$\text{tr}(J(\phi)(P, P)) = \lambda + \mu = \frac{P^2 \cdot (P - 2X)}{X^2 \cdot Y} < 0 \quad (2)$$

$$\det(J(\phi)(P, P)) = \lambda \times \mu = -\frac{P^2 \cdot (P - X)}{X^2 \cdot Y} > 0 \quad (3)$$

Si λ et μ sont réelles, alors (3) nous apprend que λ et μ sont de même signe ; dans ces conditions, on déduit de (2) que λ et μ sont toutes les deux négatives.

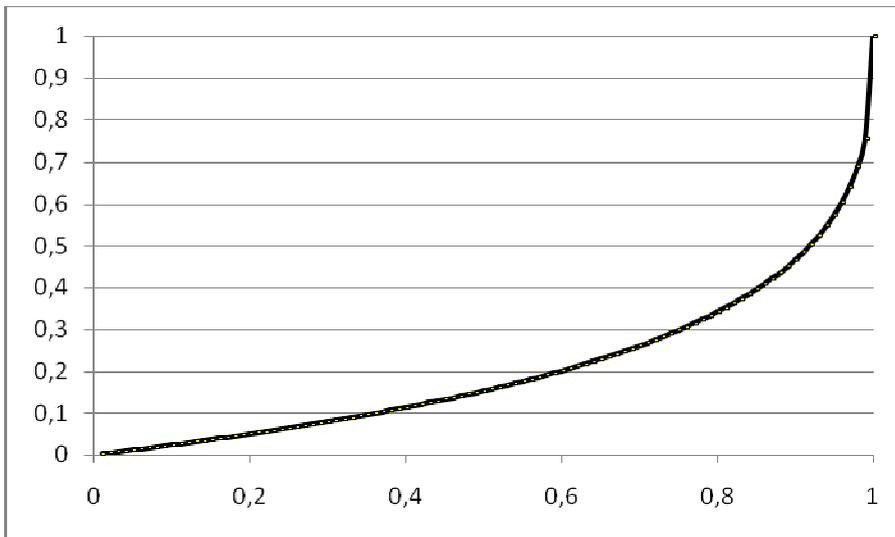
On a : $P = X \cdot (1 - \sqrt[3]{1 - \varepsilon})$ en notant $\varepsilon = Y/X$

Il suit alors :

$$\begin{aligned} |tr(J(\phi)(P, P))| &= - \frac{(X \cdot (1 - \sqrt[3]{1 - \varepsilon}))^2 \cdot (X \cdot (1 - \sqrt[3]{1 - \varepsilon}) - 2X)}{\varepsilon \cdot X^3} \\ &= - \frac{(X - X \cdot \sqrt[3]{1 - \varepsilon})^2 \cdot (X - X \cdot \sqrt[3]{1 - \varepsilon} - 2X)}{\varepsilon \cdot X^3} = + \frac{(X - X \cdot \sqrt[3]{1 - \varepsilon})^2 \cdot (X + X \cdot \sqrt[3]{1 - \varepsilon})}{\varepsilon \cdot X^3} \\ &= \frac{X^2 \cdot (1 - \sqrt[3]{1 - \varepsilon})^2 \cdot X \cdot (1 + \sqrt[3]{1 - \varepsilon})}{\varepsilon \cdot X^3} = \boxed{\frac{(1 - \sqrt[3]{1 - \varepsilon})^2 \cdot (1 + \sqrt[3]{1 - \varepsilon})}{\varepsilon}} \end{aligned}$$

Cette expression de la valeur absolue du jacobien au point fixe ne dépend des contraintes de système X et Y que dans la mesure de leur rapport ε .

L'étude de fonction est technique. Elle est strictement croissante sur $]0;1[$. Elle est prolongeable par continuité en 0, avec pour valeur 0. Elle vaut 1 en 1. Son graphe est présenté ci-dessous :



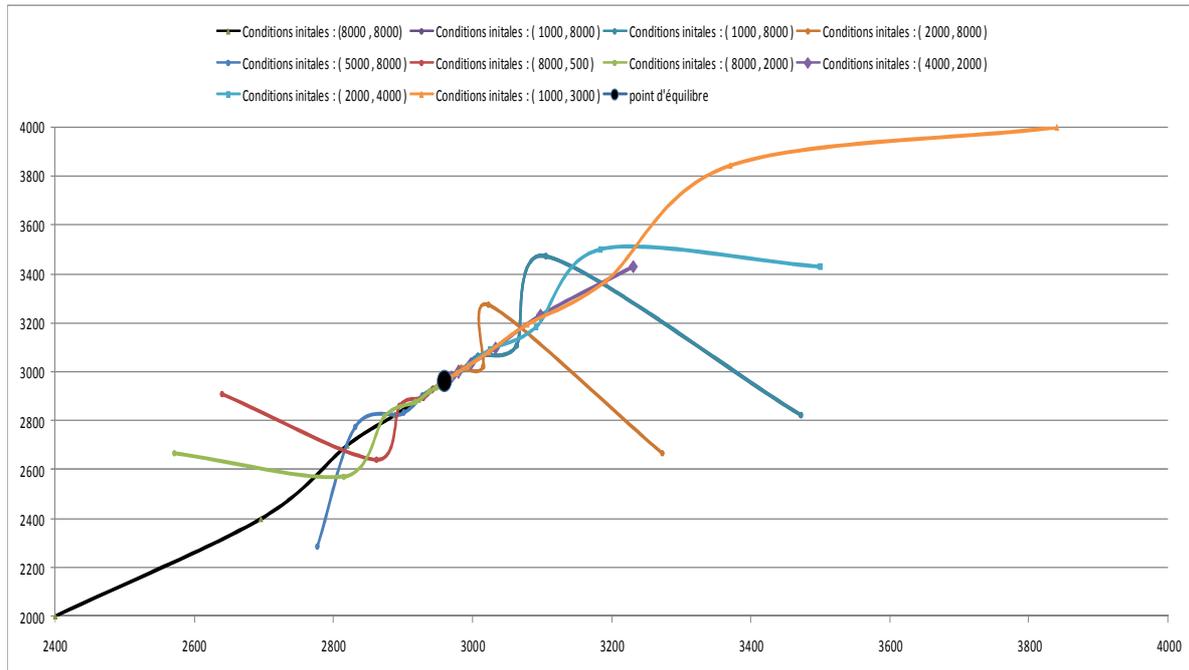
Sauf le cas où $\varepsilon=1$, on remplit donc la condition suffisante pour que le point fixe soit attractif (toujours sous hypothèse que les valeurs propres de la jacobienne en ce point fixe soient réelles).

Si $\varepsilon=1$, et donc $P=X$, on a :

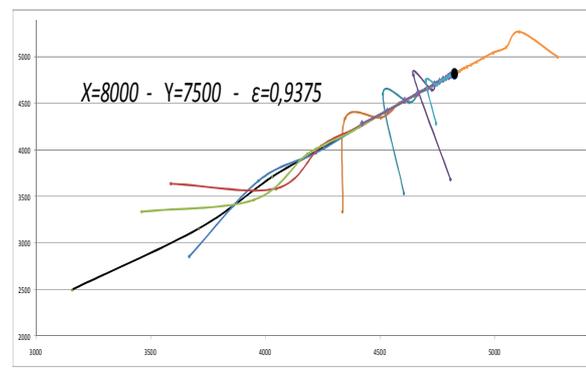
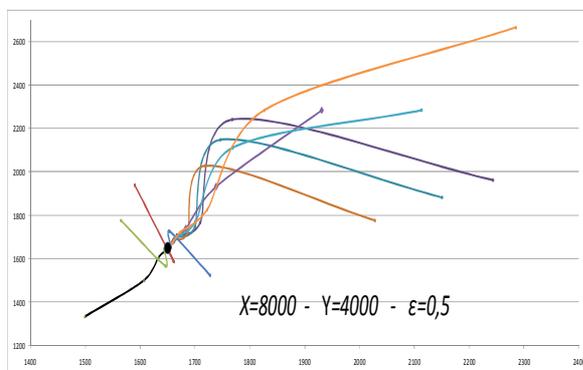
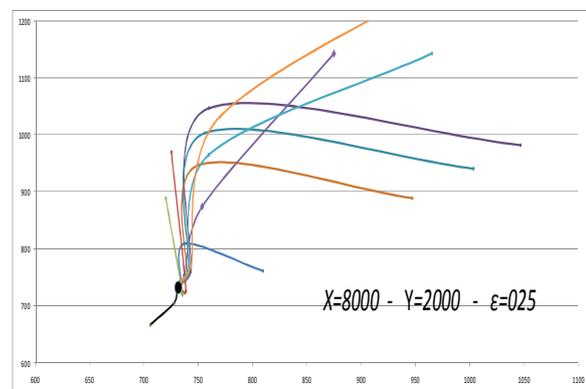
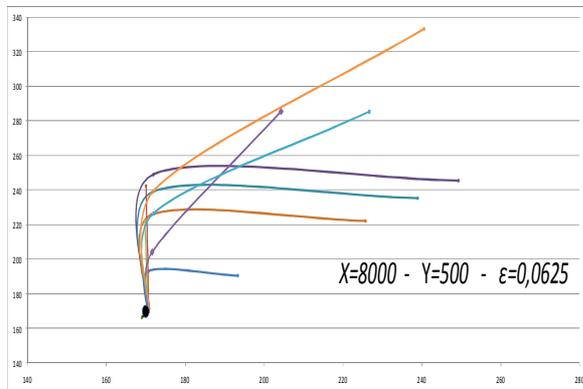
$$\det(J(\phi)(P = X, P = X)) = \lambda \times \mu = - \frac{X^2 \cdot (X - X)}{X^2 \cdot Y} = 0$$

Une des valeurs propres est nulle. La valeur de la trace est donc égale à l'autre valeur propre : -1. C'est un cas dégénéré.

Le graphe ci-dessous présente les trajets de suites $(\mathcal{P}_n)_n = (P_n, P_{n-1})_n$ avec des conditions initiales différentes, mais toujours les contraintes de système de l'exemple numérique, un tirage annuel $Y=6000$ fiches-adresse dans des EAR de tailles fixe $X=8000$ ($\varepsilon=2/3$).



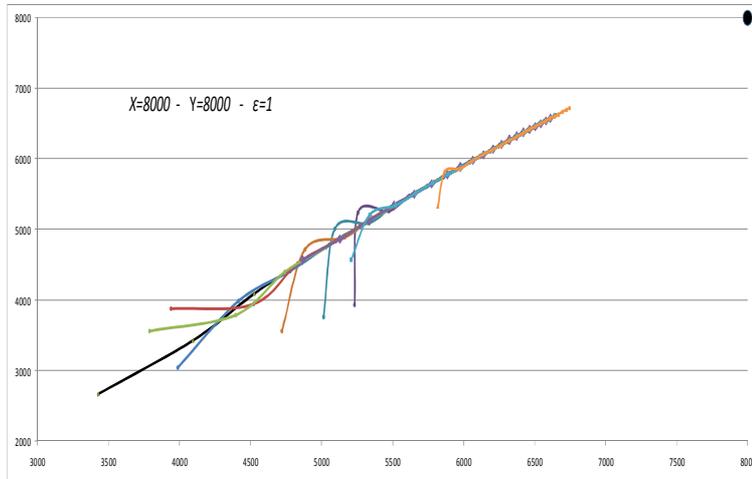
Les exemples suivants conservent une même taille d'EAR $X=8000$ et les différentes conditions initiales déjà utilisées ; on fait varier Y (et donc ε).



Quand ε augmente, les oscillations sont plus importantes et la valeur limite est atteinte beaucoup plus lentement : avec les conditions initiales de l'exemple, l'équilibre est atteint en deux itérations pour $\varepsilon=0,0625$; il faut plus de vingt itérations si $\varepsilon=0,9375$.

Que se passe-t-il avec $\varepsilon=1$?

Représentons graphiquement le cas, toujours avec le même ensemble de conditions initiales.



Il semble que même dans le cas dégénéré, la convergence se fasse vers le point d'équilibre ; le système est régulièrement oscillant sur les premières itérations, mais il est ensuite strictement croissant ; la convergence semble par contre se faire à un rythme extrêmement lent. Avec la condition initiale (8000 ; 8000), P_n atteint 7650 après 300 itérations. Les résultats standards consultés ne permettent cependant pas de conclure.

On a supposé jusqu'à ici que les valeurs propres de la matrice jacobienne étaient réelles. Supposons maintenant qu'elles soient complexes (conjuguées). Leur produit est égal au déterminant de la matrice ; la racine de leur produit est égale à leur module :

$$|\mu| = \sqrt{\mu \cdot \bar{\mu}} = \sqrt{\det(J(\phi)(P, P))} = \sqrt{\frac{P^2 \cdot (X - P)}{X^2 \cdot Y}} = \sqrt{\frac{P^2}{X \cdot Y} \times \frac{(X - P)}{X}} < 1 \quad \text{car } P < Y < X$$

On peut donc maintenant conclure généralement que le point fixe exhibé est asymptotiquement stable. Il serait maintenant intéressant de montrer qu'il y a convergence vers cet attracteur pour toute condition initiale, mais les articles consultés ne donnent pas de pistes pour chercher à conclure à une convergence générale et non plus locale.

3.3. La question du nombre d'EAR en grande commune : la précision.

Pascal Ardilly a proposé un travail de simulation pour approfondir la question du nombre d'EAR à mobiliser en grandes communes, cette fois du point de vue de la précision de l'enquête afin d'appréhender deux effets dans un dispositif appuyé sur des EAR. D'une part, il s'agissait de vérifier – et de quantifier – l'intuition selon laquelle la précision relative (par rapport à un sondage en une passe) se dégrade à mesure que le volume de fiches-adresse augmente, à taille de base de sondage fixée (i.e. à nombre d'EAR fixé pour constituer la base de sondage) : « plus la taille de l'échantillon va être proche de la taille de la base, plus sera prépondérante l'imprécision due à la première phase de l'échantillonnage ». D'autre part, il s'agissait de quantifier également la variation de précision, à taille d'enquête constante, en fonction du nombre d'EAR mobilisées. L'intuition dans ce cas est que la précision augmentera avec plus d'EAR mobilisées : « il y aura moins d'erreur *in fine* car mobiliser une fraction de population plus importante en première phase limite l'erreur due à cette première phase ».

3.3.1. Stratégie d'ensemble

« On veut comparer 2 situations, dans une grande commune donnée :

- Situation 1 (de référence) : sondage aléatoire simple (SAS) de logements dans l'ensemble des logements des adresses recensées dans la commune
- Situation 2 : tirage d'ilots et d'adresses (selon la méthode des EAR), puis en seconde phase SAS de logements dans l'ensemble des logements recensés.

Sur le principe, on reconstruit dans chaque ilot une population complète des adresses – et donc des logements – à partir des données de l'EAR la plus récente qui le concerne (l'information est suffisamment large pour que ça ait un sens - puisque dans chaque ilot on dispose d'un recensement complet de 40% des adresses (...)), puis on tire des ilots et des adresses suivant le schéma des EAR, et ensuite on tire des logements pour reproduire le tirage EECDOM. A ce stade, l'hypothèse faite est que le tirage final EECDOM est de type aléatoire simple (ce qui ne sera probablement pas le cas *in fine* - on introduira presque sûrement un échantillonnage plus performant, au moins une stratification).

On perd au passage l'effet d'équilibrage des ilots, mais ça ne paraît pas dramatique (ce serait sensiblement plus compliqué à implanter et probablement en décalage avec l'enjeu – une borne supérieure de variance doit suffire, surtout si on conclut à une dégradation de variance limitée par rapport à la situation de référence).

On ne doit pas ignorer la phase de tirage des adresses du recensement (le fameux tirage d'adresses à 40%) : en effet, la part de variance due à cette phase doit être assez proche de la part de variance due au tirage des ilots. Cette intuition vient du fait que la taille d'échantillon de logements après tirage des adresses d'une EAR s'avère significativement plus petite que le nombre total de logements présents dans l'ensemble des ilots de cette EAR, mais sans en être très différente (cette situation intermédiaire ne permet ni de négliger l'incertitude due au tirage des adresses ni de la rendre prépondérante).

En réalité, le schéma d'ensemble est celui d'un échantillonnage en 3 phases :

- phase 1 : tirage des ilots de l'EAR,
- phase 2 : tirage des adresses de l'EAR,
- phase 3 : tirage des logements EECDOM.

Or les phases 1 et 2 sont à l'origine d'incertitudes dont on peut penser que les ordres de grandeur sont comparables, ce qui fait qu'on a deux façons de raisonner :

- soit on dit d'emblée – en s'appuyant sur la théorie mais effectivement sans "preuve numérique" – que les phases 1 et 2 ont un impact négligeable par rapport à la phase 3 - auquel cas l'affaire est entendue et il ne sert à rien de faire toutes ces simulations;
- soit on veut mesurer objectivement l'impact du tirage du recensement dans l'ensemble du processus et si on prend en compte l'incertitude de la phase 1, alors logiquement il faut aussi prendre en compte l'incertitude de la phase 2 - en tout cas, dans la simulation, il n'y a pas de justification a priori à ne faire que la moitié du chemin, c'est-à-dire à tirer des ilots sans tirer juste derrière les adresses qui complètent le dispositif de sélection de l'échantillon EAR.

Les simulations qui suivent sont conçues dans un cadre simplifié où les tirages de toutes les vagues enquêtées un trimestre donné ont été effectués dans le même ensemble d'EAR (soit toutes, soit une seule). Sur les 6 vagues concernées chaque trimestre, il y a au moins 4 vagues où c'est vrai - mais il y a au moins une vague pour laquelle ce n'est pas vrai (en pratique, par exemple si la règle est de tirer chaque vague dans la dernière EAR disponible au moment du tirage, il y a toujours 2 EAR mobilisées pour l'enquête EECDOM, rarement 3). Affecter les tirages de différentes vagues à différentes EAR compliquerait les programmes sans que les conclusions ne soient sensiblement modifiées. »⁹

3.3.2. Application

La simulation est réalisée pour plusieurs tailles d'échantillon, afin de vérifier si, comme attendu, pour un nombre fixé d'EAR au départ, plus la taille de l'échantillon est importante, plus la perte de précision relative (comparativement au sondage sans sélection préalable des îlots) est importante.

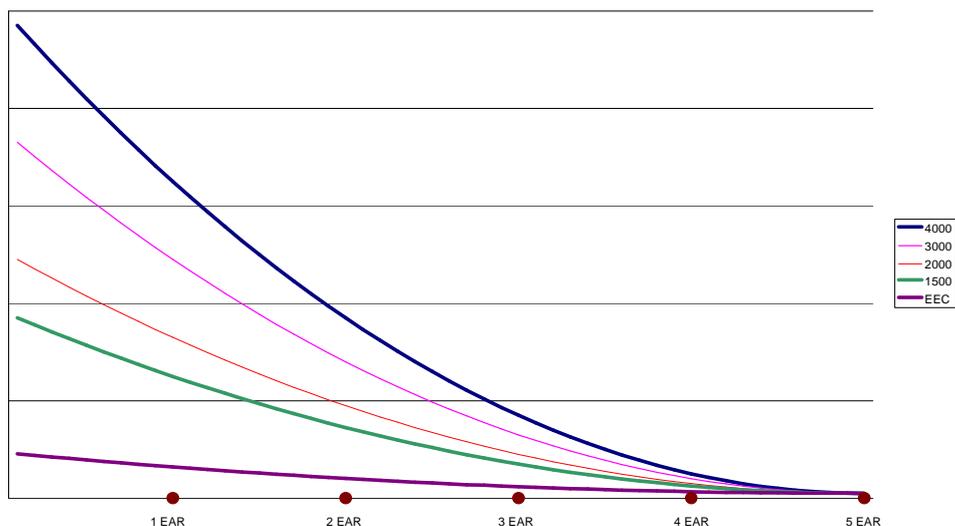
Pour chaque grande commune de chacun des départements, la simulation a donc été menée pour cinq tailles d'enquête (à l'échelle régionale, 338 FAs – taille EEC-DOM –, puis 1500 FAs – taille courante des enquêtes réalisées avec extension –, 2000, 3000, 4000 FAs – situations plus rares d'enquêtes particulièrement volumineuses comme « MFV » ou « Logement »).

Parallèlement à l'étude de l'impact de la taille de l'échantillon, on a également décliné la simulation selon le nombre d'EAR à mobiliser. Pour chaque grande commune de chaque département et chaque taille d'échantillon, on a donc réalisé quatre simulations, qui mobilisent respectivement pour la constitution de la base de sondage l'équivalent de une à quatre EAR (la situation à cinq EAR correspond à la situation de référence).

Vingt simulations sont donc menées pour chacune des 46 grandes communes d'Outre-mer. Chacune ramènera quatre estimations, deux à contrôler (estimations du nombre d'actifs occupés et du nombre de chômeurs dans la situation à contrôler, x EAR, n FAs) et deux de contrôle (estimations du nombre d'actifs occupés et du nombre de chômeurs en faisant un échantillonnage en une passe).

Dans chacune des 20 situations à contrôler, les deux fois mille simulations communales sont agrégées pour obtenir deux fois mille simulations pour chaque département, et on ajoute également à partir des estimations déjà produites, celle d'un taux de chômage, également sur l'ensemble des grandes communes de chaque département.

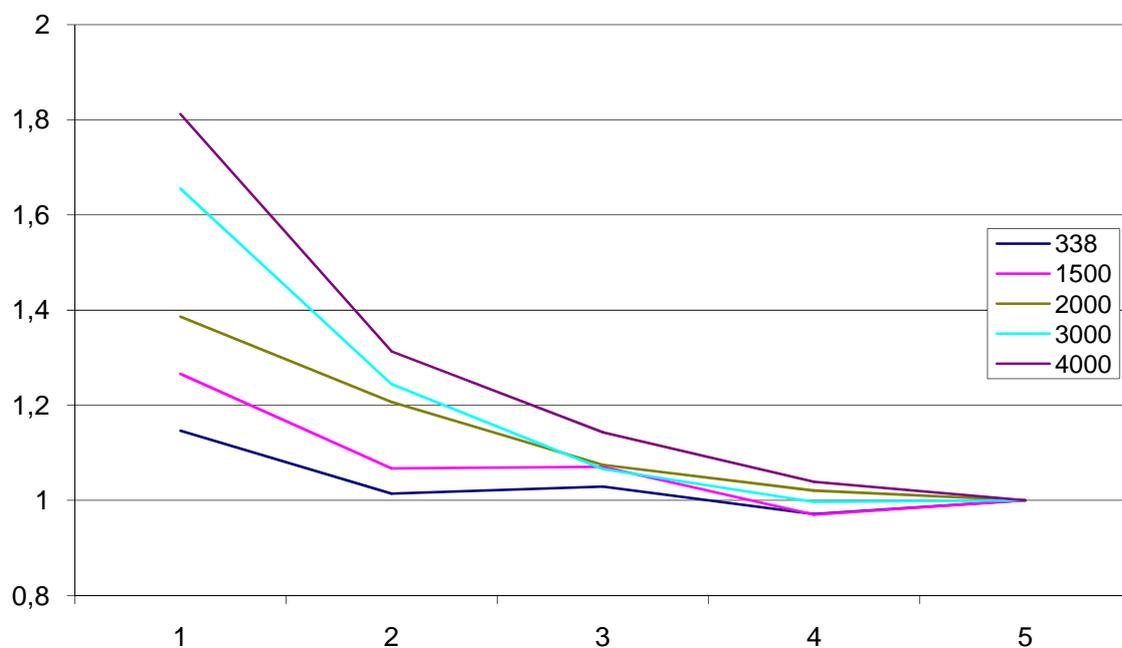
Les intuitions qu'on teste conduisent à imaginer obtenir, pour chaque département et chaque variable (sur les grandes communes, nombre d'actifs occupés, nombre de chômeurs, taux de chômage), un graphique du type suivant, avec en abscisse le nombre d'EAR mobilisées, et en ordonnée le rapport entre la précision du sondage « avec intermédiaire d'EAR » et celui en une passe (en l'espèce rapport entre les écarts-types dans les deux cas à comparer), les courbes représentant une taille d'échantillon particulière, à l'échelle régionale :



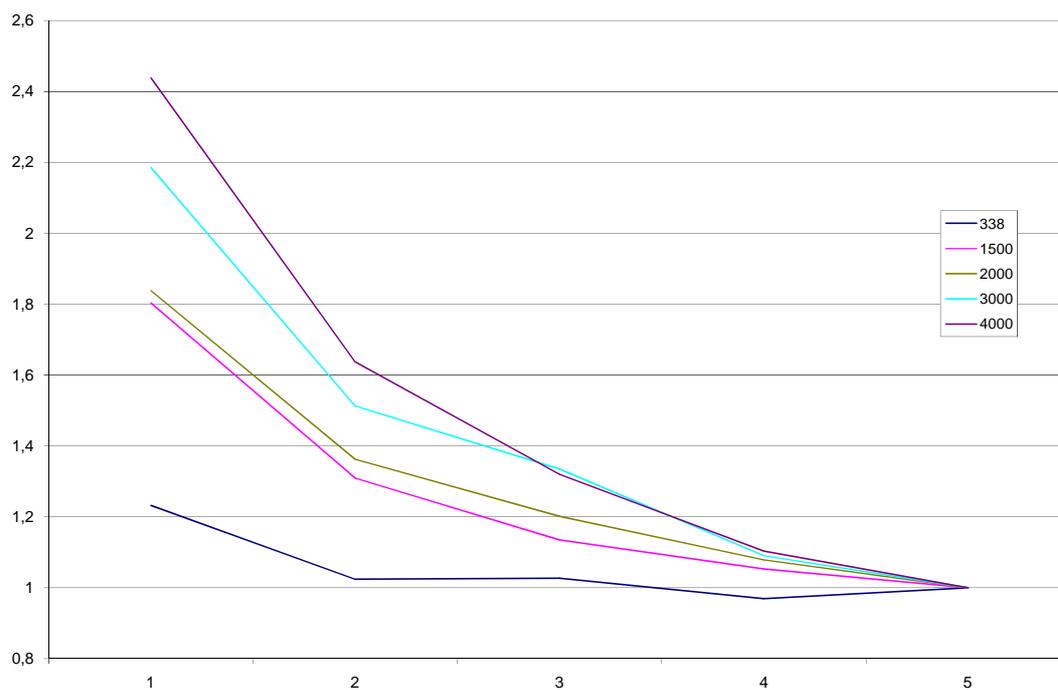
⁹ Extrait de « Opération d'estimation de l'impact du tirage des îlots EAR sur l'EECDOM », Pascal Ardilly, note Insee N°2579/DG75-F401/

3.3.3. Résultats de la simulation

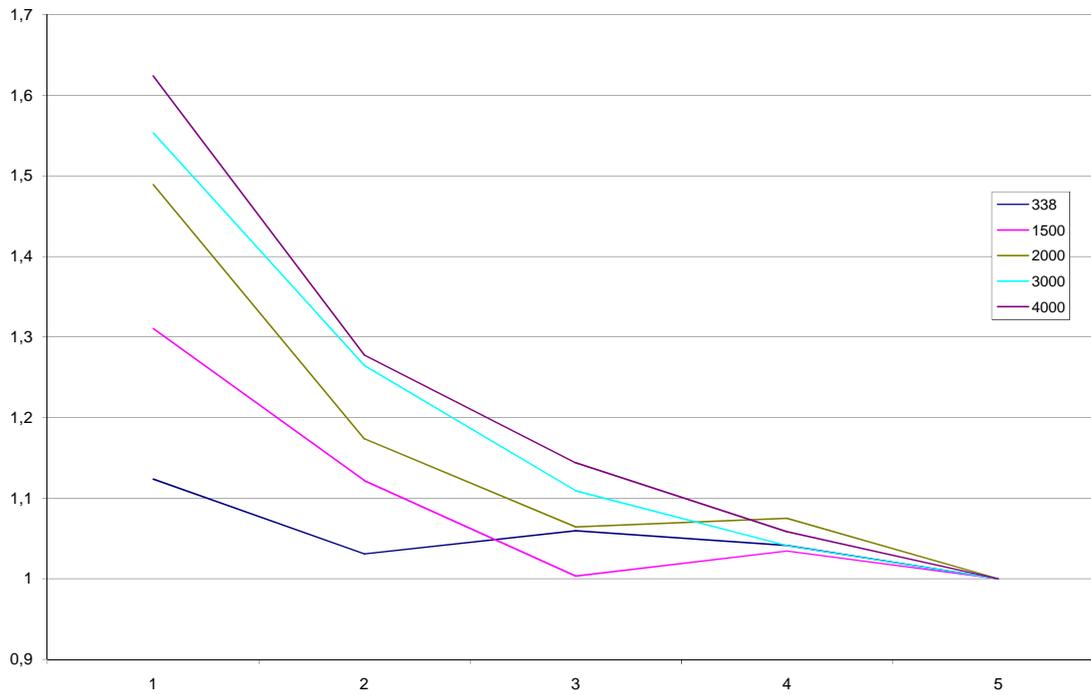
Rapports des écarts-types de l'estimation du nombre de chômeurs sur les grandes communes de Guadeloupe, en fonction du nombre d'EAR mobilisées



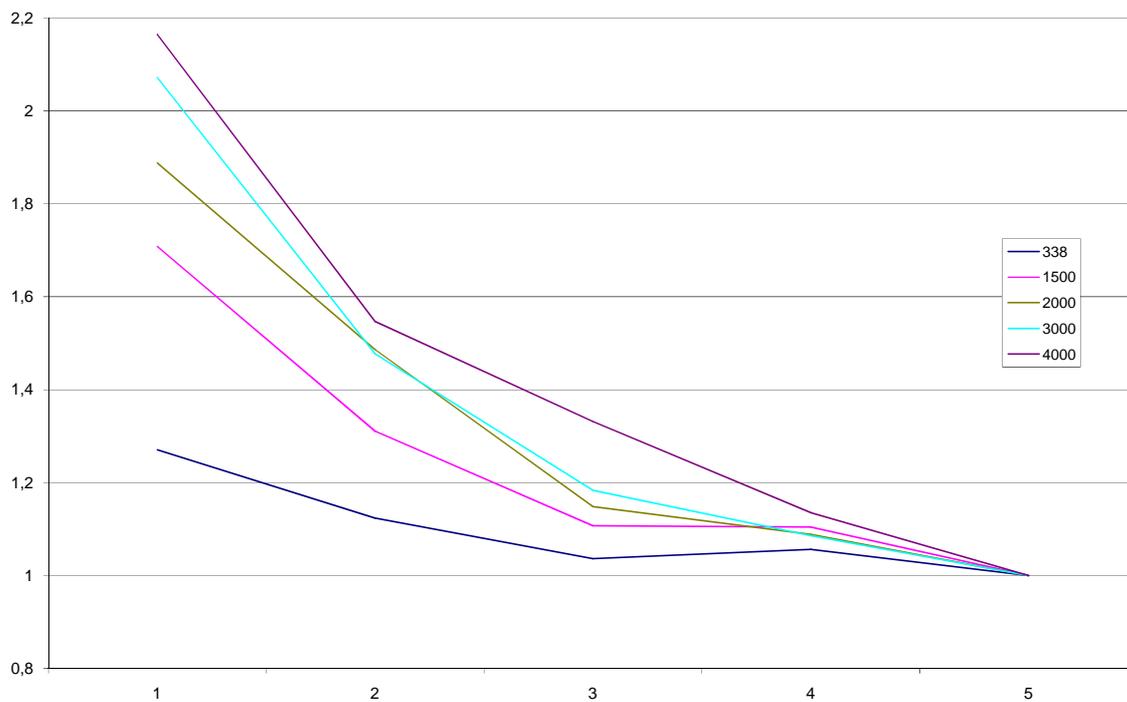
Rapports des écarts-types de l'estimation du nombre d'actifs occupés sur les grandes communes de Guadeloupe, en fonction du nombre d'EAR mobilisées



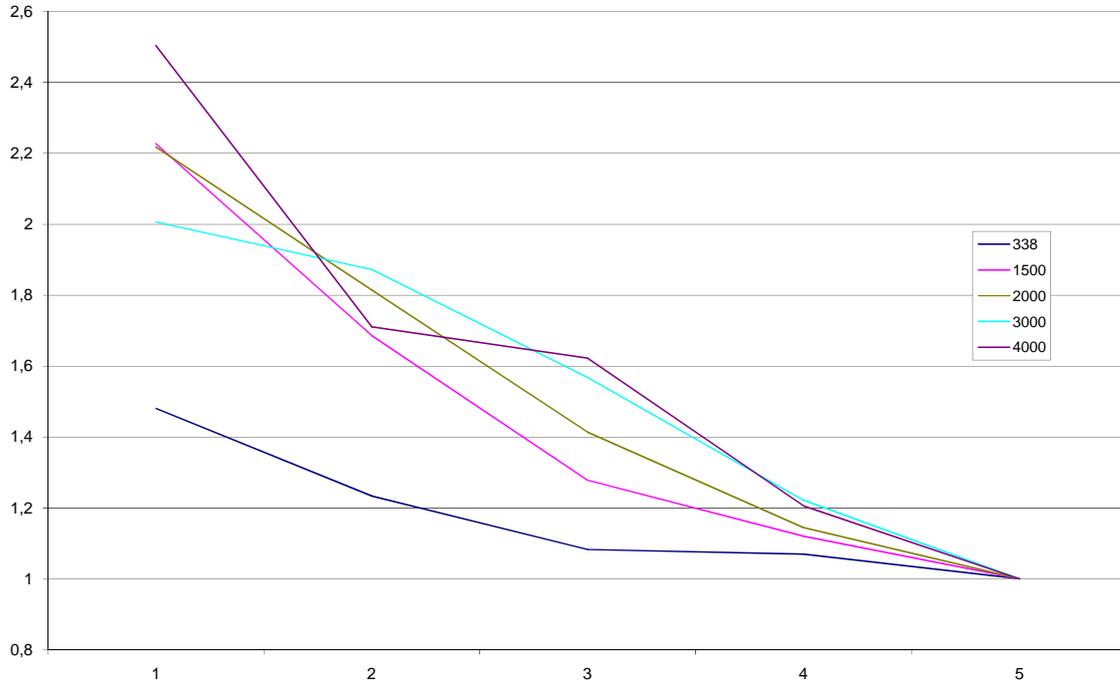
Rapports des écarts-types de l'estimation du nombre de chômeurs sur les grandes communes de Martinique, en fonction du nombre d'EAR mobilisées



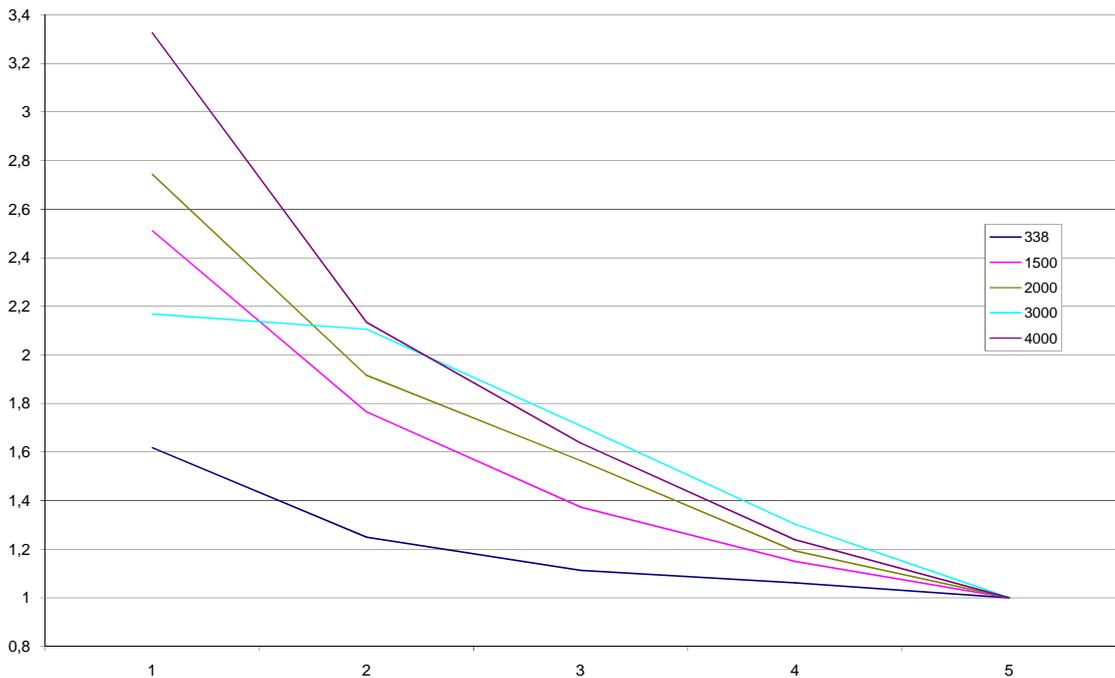
Rapports des écarts-types de l'estimation du nombre d'actifs occupés sur les grandes communes de Martinique, en fonction du nombre d'EAR mobilisées



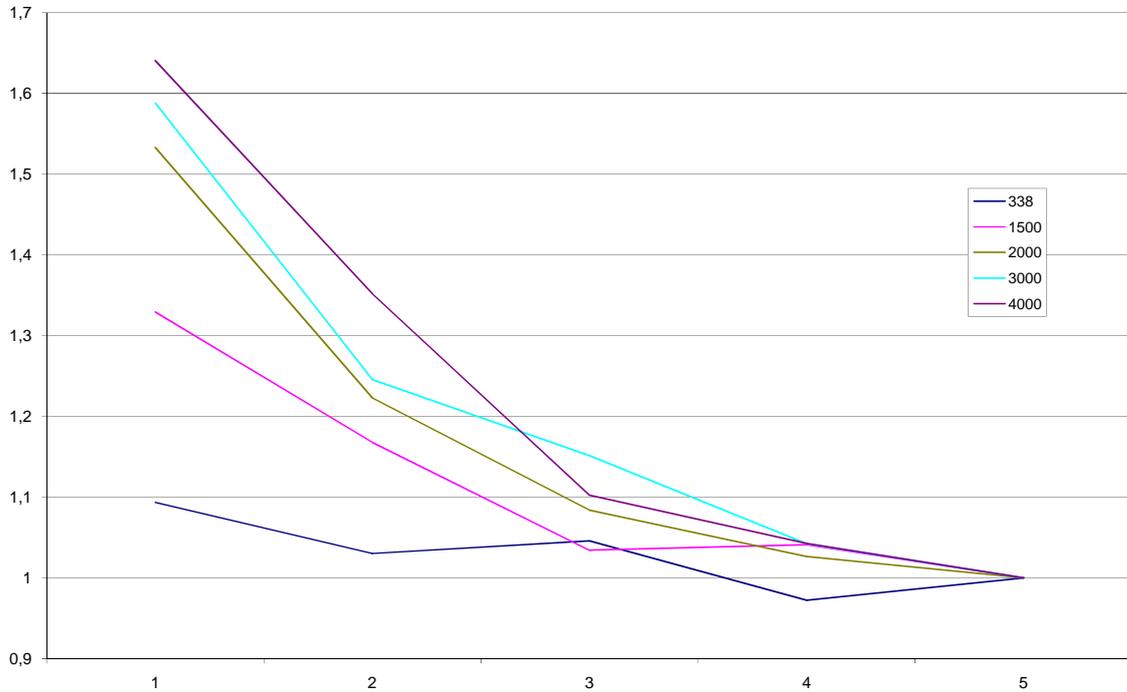
Rapports des écarts-types de l'estimation du nombre de chômeurs sur les grandes communes de Guyane, en fonction du nombre d'EAR mobilisées



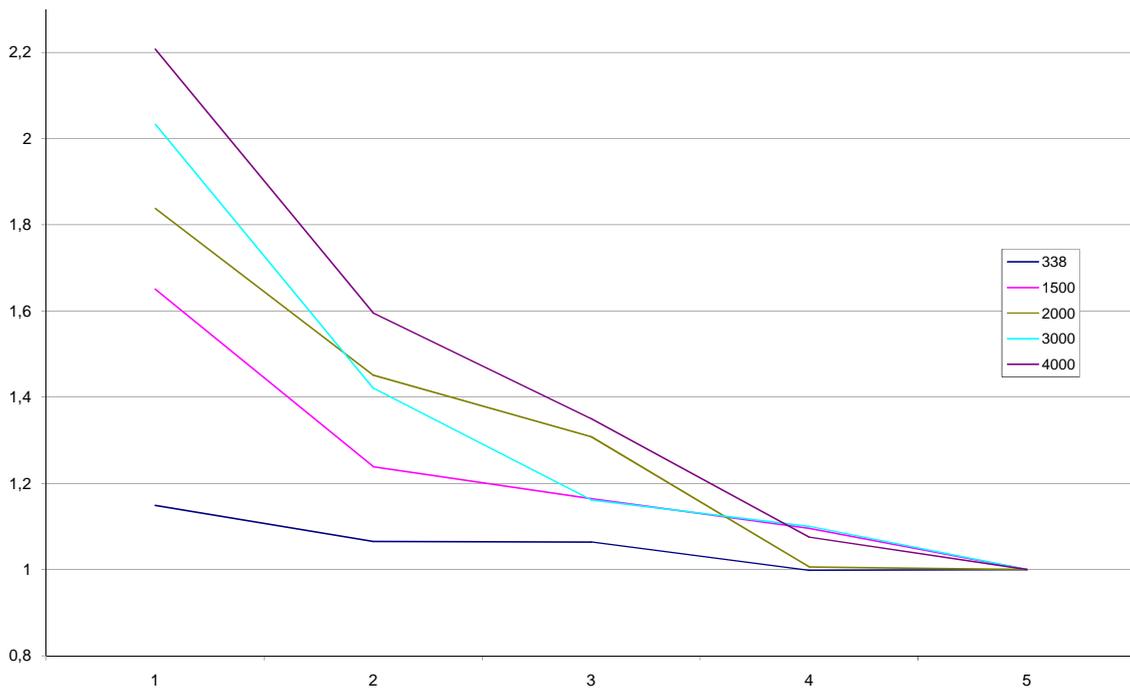
Rapports des écarts-types de l'estimation du nombre d'actifs occupés sur les grandes communes de Guyane, en fonction du nombre d'EAR mobilisées



Rapports des écarts-types de l'estimation du nombre de chômeurs sur les grandes communes de La Réunion, en fonction du nombre d'EAR mobilisées



Rapports des écarts-types de l'estimation du nombre d'actifs occupés sur les grandes communes de La Réunion, en fonction du nombre d'EAR mobilisées



Toutes les représentations graphiques ne représentent pas aussi « purement » les unes que les autres la situation théorique¹⁰, mais on a ainsi la confirmation des intuitions qui ont guidé ce travail de simulation :

- *comparativement à un sondage en une passe*, à nombre d'EAR utilisées fixé, on perd en précision avec une taille d'échantillon plus grande

- *comparativement à un sondage en une passe*, à taille d'échantillon fixée, on gagne en précision en utilisant plusieurs EAR.

L'amélioration de précision en passant de une à deux EAR est moins sensible sur de petits échantillons du type de l'EEC-DOM. Elle l'est par contre assez nettement pour des tailles d'échantillon plus importantes¹¹.

Bien que, pratiquement, l'utilisation de deux EAR conduise à utiliser un volume de FAs plus faible que la somme des volumes de deux EAR, il semble que l'utilisation de deux EAR soit un bon compromis pour les échantillons jusqu'à 1500 ou 2000 FAs. Le gain marginal de l'utilisation d'une troisième EAR est comparativement faible (sauf peut-être en Guyane) et ne justifie ainsi pas de prendre le risque de mobiliser des informations éventuellement « un peu trop » anciennes.

La conclusion est moins facile pour des échantillons volumineux, le gain marginal de l'utilisation d'une troisième EAR pour le tirage est nettement plus important, et pourrait justifier de passer outre les scrupules à utiliser des données anciennes.

A partir de ces enseignements, on peut préconiser l'utilisation de deux EAR pour le dispositif standard. Si c'est possible, la programmation du dispositif d'échantillonnage pourrait ménager la possibilité d'utilisation d'une troisième EAR pour des cas d'enquêtes particulièrement volumineuses.

3.4. Les déséquilibres des groupes de rotation du recensement en grande commune

La simulation proposée par Pascal Ardilly s'intéresse, pour ainsi dire, à l'effet « pur » de première phase. Il ne s'appuie donc pas sur les groupes de rotation existants, mais en produit à chaque étape, afin d'étudier la précision du « cadre » du dispositif.

Or, pour le recensement des quatre DOM « historiques », les groupes de rotation des grandes communes sont constitués d'îlots et sont donc « fixés » géographiquement ; ces groupes avaient été équilibrés à partir du recensement de 1999, mais l'équilibrage n'est plus respecté : les dynamiques de construction et de peuplement ont modifié les équilibres, à la fois du point de vue des volumes simples des groupes mais également du point de vue de leurs structures internes. La réallocation des logements neufs dans les groupes de rotation limite ces problèmes en métropole, mais les DOM n'ont pas de tel dispositif.

Une note du Criem, essentiellement descriptive, est consacrée à ce sujet¹² et présente de façon extrêmement détaillée les variations des volumes et des variables de structure département par département, et commune par commune pour toutes les EAR disponibles au moment de sa rédaction. Les quatre DOM « historiques » sont concernés par ce déséquilibre des groupes de rotation, qui accentue encore plus la non-comparabilité des résultats de deux recensements consécutifs.

Du point de vue des enquêtes auprès des ménages, la première phase d'échantillonnage (qui correspond à la réalisation du recensement pour un dispositif appuyé sur lui) n'est donc pas vraiment satisfaisante.

¹⁰ Les estimateurs de variance ont évidemment également une variance. En utilisant non plus une simulation de 1000 éléments, mais une simulation plus importante, on aurait réduit les phénomènes d'oscillation qui « pollue » un peu les résultats et éloigne les représentations graphiques obtenues de la situation de référence théorique donnée avant.

¹¹ Sur le principe cela pourrait signifier que l'EEC pourrait être tirée dans une seule EAR ; cependant, les considérations de déséquilibres développées par ailleurs n'incitent pas à développer cette option.

¹² note N° 373/DR974-SES/ d'avril 2012 « Structure des groupes de rotation du recensement des Dom dans la perspective de l'échantillonnage des enquêtes auprès des ménages »

3.4.1. Des déséquilibres en volumes

3.4.1.1. Récapitulatif

En Guadeloupe :

	RP - 2006					RP - 2007					RP - 2008				
	GR - 1	GR - 2	GR - 3	GR - 4	GR - 5	GR - 2	GR - 3	GR - 4	GR - 5	GR - 1	GR - 3	GR - 4	GR - 5	GR - 1	GR - 2
97101	18,9%	19,6%	21,1%	20,3%	20,0%	19,8%	21,3%	20,5%	20,2%	18,4%	21,0%	20,2%	19,9%	18,1%	20,9%
97103	16,2%	20,8%	20,3%	21,5%	21,2%	20,0%	19,5%	20,7%	20,4%	19,4%	18,5%	19,5%	19,3%	18,3%	24,5%
97105	19,8%	22,4%	19,3%	22,1%	16,4%	22,5%	19,3%	22,2%	16,5%	19,5%	19,7%	22,6%	16,8%	19,9%	21,1%
97107	18,7%	21,4%	20,0%	20,1%	19,9%	21,1%	19,7%	19,8%	19,7%	19,7%	19,5%	19,6%	19,5%	19,5%	21,9%
97113	18,1%	17,8%	18,9%	24,9%	20,4%	17,4%	18,5%	24,4%	20,0%	19,7%	18,3%	24,1%	19,8%	19,5%	18,3%
97115	15,7%	25,0%	20,7%	19,9%	18,7%	24,7%	20,4%	19,7%	18,5%	16,7%	20,4%	19,7%	18,5%	16,8%	24,6%
97116	17,0%	17,5%	20,8%	20,9%	23,8%	17,2%	20,4%	20,5%	23,4%	18,6%	19,9%	20,0%	22,8%	18,1%	19,2%
97117	18,8%	18,4%	19,5%	21,9%	21,5%	17,9%	18,9%	21,3%	20,9%	21,0%	18,6%	20,9%	20,6%	20,6%	19,3%
97118	19,5%	17,9%	21,0%	20,4%	21,2%	17,5%	20,6%	19,9%	20,7%	21,3%	19,3%	18,7%	19,4%	19,9%	22,6%
97120	17,4%	19,1%	19,4%	21,7%	22,4%	18,4%	18,6%	20,8%	21,5%	20,8%	18,5%	20,7%	21,4%	20,7%	18,7%
97124	18,8%	19,5%	17,8%	20,4%	23,6%	19,3%	17,6%	20,1%	23,3%	19,7%	17,5%	20,1%	23,3%	19,7%	19,3%
97125	15,3%	19,6%	19,6%	26,7%	18,7%	18,8%	18,8%	25,5%	17,8%	19,2%	18,2%	24,7%	17,3%	18,6%	21,3%
97128	18,1%	21,7%	20,1%	21,4%	18,7%	21,5%	19,9%	21,2%	18,5%	18,9%	19,3%	20,6%	17,9%	18,3%	23,8%
97129	15,7%	17,8%	23,3%	24,0%	19,1%	17,7%	23,2%	23,9%	19,0%	16,3%	23,1%	23,8%	18,9%	16,3%	18,0%

En Martinique :

	RP - 2006					RP - 2007					RP - 2008				
	GR - 1	GR - 2	GR - 3	GR - 4	GR - 5	GR - 2	GR - 3	GR - 4	GR - 5	GR - 1	GR - 3	GR - 4	GR - 5	GR - 1	GR - 2
97207	19,2%	19,5%	18,7%	20,3%	22,3%	19,2%	18,4%	20,0%	22,0%	20,5%	17,9%	19,4%	21,4%	19,9%	21,4%
97209	19,1%	20,1%	20,4%	21,7%	18,7%	19,9%	20,2%	21,5%	18,6%	19,8%	20,0%	21,3%	18,4%	19,6%	20,6%
97210	18,5%	21,3%	21,7%	17,8%	20,6%	20,7%	21,1%	17,3%	19,9%	21,0%	21,0%	17,2%	19,8%	20,8%	21,2%
97212	24,3%	22,6%	17,7%	18,5%	16,8%	22,5%	17,6%	18,4%	16,7%	24,9%	17,4%	18,2%	16,5%	24,6%	23,3%
97213	17,2%	20,0%	22,4%	19,9%	20,5%	19,9%	22,3%	19,8%	20,4%	17,6%	21,9%	19,5%	20,0%	17,3%	21,2%
97220	20,2%	21,2%	19,0%	20,3%	19,2%	21,0%	18,8%	20,1%	19,1%	20,9%	18,5%	19,8%	18,7%	20,5%	22,6%
97221	17,5%	19,4%	20,2%	21,3%	21,5%	19,2%	20,0%	21,0%	21,2%	18,5%	19,8%	20,9%	21,1%	18,4%	19,7%
97222	17,5%	17,3%	20,0%	27,8%	17,4%	16,8%	19,5%	27,1%	16,9%	19,7%	19,3%	26,8%	16,8%	19,5%	17,6%
97224	18,8%	18,8%	20,0%	21,8%	20,5%	18,5%	19,7%	21,5%	20,2%	20,1%	19,5%	21,2%	20,0%	19,9%	19,3%
97228	19,6%	20,0%	20,1%	19,9%	20,5%	19,7%	19,8%	19,6%	20,3%	20,5%	19,8%	19,5%	20,2%	20,4%	20,1%
97229	23,1%	17,9%	20,5%	18,9%	19,6%	17,8%	20,4%	18,8%	19,5%	23,6%	19,7%	18,2%	18,9%	22,8%	20,4%
97230	18,6%	17,7%	17,9%	24,9%	20,9%	17,6%	17,8%	24,8%	20,8%	19,0%	17,2%	23,9%	20,1%	18,4%	20,5%

En Guyane :

	RP - 2006					RP - 2007					RP - 2008				
	GR - 1	GR - 2	GR - 3	GR - 4	GR - 5	GR - 2	GR - 3	GR - 4	GR - 5	GR - 1	GR - 3	GR - 4	GR - 5	GR - 1	GR - 2
97302	20,9%	18,4%	19,3%	21,1%	20,3%	18,4%	19,3%	21,1%	20,4%	20,8%	19,4%	21,2%	20,4%	20,9%	18,1%
97304	18,2%	17,3%	20,9%	21,6%	21,9%	16,4%	19,9%	20,5%	20,8%	22,4%	19,6%	20,2%	20,5%	22,1%	17,6%
97307	20,4%	20,5%	19,1%	19,7%	20,3%	20,2%	18,7%	19,4%	19,9%	21,8%	18,0%	18,6%	19,1%	20,9%	23,4%
97309	21,0%	18,6%	21,8%	18,6%	20,1%	17,8%	20,9%	17,8%	19,3%	24,1%	20,2%	17,3%	18,7%	23,3%	20,5%
97311	24,9%	16,4%	18,1%	20,9%	19,6%	16,2%	17,8%	20,5%	19,3%	26,2%	16,9%	19,5%	18,4%	24,9%	20,3%

A La Réunion :

	RP - 2006					RP - 2007					RP - 2008				
	GR - 1	GR - 2	GR - 3	GR - 4	GR - 5	GR - 2	GR - 3	GR - 4	GR - 5	GR - 1	GR - 3	GR - 4	GR - 5	GR - 1	GR - 2
97404	22,9%	20,0%	18,8%	18,7%	19,6%	20,0%	18,8%	18,7%	19,6%	22,9%	18,2%	18,0%	18,9%	22,2%	22,7%
97405	19,2%	19,4%	18,9%	19,5%	23,0%	18,5%	18,0%	18,6%	21,9%	23,1%	17,5%	18,1%	21,3%	22,5%	20,5%
97407	19,9%	18,6%	19,5%	20,4%	21,6%	18,6%	19,4%	20,3%	21,5%	20,2%	18,9%	19,8%	21,0%	19,7%	20,6%
97408	21,4%	16,9%	21,9%	16,1%	23,8%	16,3%	21,1%	15,5%	22,9%	24,3%	19,5%	14,3%	21,1%	22,4%	22,7%
97409	19,2%	19,1%	18,2%	22,0%	21,5%	18,6%	17,8%	21,5%	21,0%	21,2%	17,3%	20,9%	20,4%	20,6%	20,8%
97410	21,1%	20,4%	19,7%	18,7%	20,1%	20,3%	19,5%	18,6%	19,9%	21,7%	19,0%	18,1%	19,4%	21,1%	22,4%
97411	19,0%	18,1%	20,5%	22,3%	20,0%	17,6%	19,9%	21,7%	19,4%	21,3%	19,3%	21,0%	18,8%	20,7%	20,2%
97412	19,6%	20,4%	20,0%	21,2%	18,9%	19,3%	19,0%	20,1%	17,9%	23,7%	18,5%	19,6%	17,5%	23,2%	21,1%
97413	18,8%	18,6%	20,6%	19,2%	22,8%	17,8%	19,6%	18,3%	21,7%	22,5%	18,9%	17,6%	20,9%	21,7%	20,9%
97414	19,8%	18,7%	20,8%	19,0%	21,6%	18,2%	20,3%	18,5%	21,1%	21,8%	19,7%	18,0%	20,5%	21,2%	20,6%
97415	18,8%	20,4%	20,4%	20,2%	20,2%	19,9%	19,9%	19,7%	19,7%	20,8%	19,3%	19,1%	19,1%	20,2%	22,4%
97416	19,5%	18,5%	20,9%	21,3%	19,7%	18,0%	20,3%	20,7%	19,1%	22,0%	19,5%	19,9%	18,3%	21,1%	21,1%
97418	20,4%	21,1%	20,4%	19,4%	18,6%	20,2%	19,5%	18,5%	17,8%	24,0%	19,1%	18,1%	17,4%	23,5%	21,8%
97420	16,0%	22,4%	18,2%	22,4%	21,0%	22,1%	17,9%	22,0%	20,7%	17,3%	17,4%	21,3%	20,0%	16,7%	24,5%
97422	18,0%	19,7%	20,5%	20,0%	21,7%	19,2%	20,0%	19,5%	21,1%	20,2%	19,0%	18,5%	20,1%	19,2%	23,2%

Lecture des tableaux :

Pour chacun des trois recensements, on a la proportion de logements effectivement enquêtés lors de chaque EAR réalisée sur un groupe de rotation ; à l'intérieur de chaque groupe de cinq colonnes correspondant à un chaque cycle de recensement, les groupes, i.e. les EAR, sont classés du plus ancien mobilisé ou plus récent.

Exemple : dernière ligne du dernier tableau, commune 97422 : pour le RRP 2006, 18,0 % des observations utilisées sont issues du GR-1 (soit l'EAR 2004), le plus ancien mobilisé pour le RRP 2006, 19,7 % issues du GR-2 (soit l'EAR 2005), 21,7 % issues du GR-5 (soit l'EAR 2008), le plus récent mobilisé pour le RRP 2006.

Réserves sur les tableaux précédents :

L'utilisation des fichiers de diffusion du RP, à la différence de l'utilisation des Bases de Sondage d'Adresses (BSA), ne permet pas d'éliminer des effets éventuels du tirage des adresses à recenser. On ne dispose pas – pour l'heure – des BSA des années 2004 et 2005, éliminant *de facto* l'utilisation parallèle de cette source, meilleures pour les questions de volumes générales, mais qui n'identifie pas, par contre, les résidences principales.

Comme d'une part le tirage de ces adresses est constitué de manière à couvrir 40 % des logements et que d'autre part les taux de FLNE sont plus liés à la commune qu'à l'EAR, on considère que les résultats issus des fichiers de diffusion fourniront une indication fiable.

3.4.1.2. Mouvements erratiques des volumes entre les groupes de rotation

Dans une commune sans nouvelle construction sur l'ensemble du cycle, chaque EAR devrait représenter exactement 20 % des logements effectivement recensés de l'ensemble du cycle.

Dans une commune où la construction neuve serait également répartie dans les cinq groupes de rotation, on serait dans la situation suivante (exemple avec croissance régulière de 5 % sur 5 ans) :

année n	année n+1	année n+2	année n+3	année n+4
100	105	110,25	115,7625	121,550625
18,10%	19,00%	19,95%	20,95%	22,00%

Le dynamisme général de la construction Outre-mer devrait conduire à observer régulièrement une situation du type précédent, *si les groupes continuaient d'être comparables en*

nombre de logements : décroissance (régulière ou non) de la proportion de logements effectivement recensés dans le groupe en fonction de l'ancienneté de l'EAR (et donc de la mobilisation du groupe de rotation).

Or, dans les tableaux précédents, on ne peut repérer aucune évolution du type précédent. Les mouvements sont essentiellement erratiques. Les exemples les plus nets sont les communes de Sainte-Suzanne à La Réunion (97420) sur le recensement de 2007 (les contributions des groupes du plus ancien au plus récent (GR-2, GR-3, GR-4, GR-5, GR-1) sont ainsi 22,1 %, 17,9 %, 22,0 %, 20,7 %, 17,3 %), de Saint-Laurent-du-Maroni en Guyane (97311) en 2008, du Gros-Morne en Martinique (97212) en 2007, de Basse-Terre (97105) en Guadeloupe en 2006.

3.4.1.3. Des différences de volume majeures entre les EAR

On note également des différences majeures sur les contributions de chacune des EAR, i.e. de chacun des groupes, à chacun des trois recensements. Les cellules surlignées dans les tableaux précédents sont celles qui présentent une différence de deux points et plus à la contribution « régulière » de 20 %¹³.

Sur la commune du Robert en Martinique (97222) par exemple, le GR-4 contient de l'ordre de 27 % des logements recensés pour chacun des RRP, tandis que le GR-5 contient environ 17 % des logements. La commune du Lamentin en Guadeloupe (97115) présente une situation similaire avec ses GR-2 et GR-1 réciproquement, de même pour la commune de Saint-Laurent-du-Maroni en Guyane (97311) avec les GR-1 et GR-2 et Sainte-Marie à La Réunion (97418) avec les GR-1 et GR-5.

3.4.1.4. GR-1 et GR-2 : des différences d'évolution de volume

Au moment de l'analyse, en 2012, seuls deux groupes de rotation ont été mobilisés deux fois et sont disponibles, le GR-1 en 2004 et 2009 et le GR-2 en 2005 et 2010.

Les tableaux suivants présentent l'évolution sur cinq ans du nombre de logements effectivement recensés dans ces deux groupes.

	évolution sur 5 ans										
	GR - 1	GR - 2		GR - 1	GR - 2		GR - 1	GR - 2		GR - 1	GR - 2
97101	-3,3%	7,0%	97207	8,1%	14,4%	97302	-0,9%	-2,0%	97404	0,2%	17,5%
97103	24,4%	29,8%	97209	4,3%	4,6%	97304	30,2%	8,8%	97405	26,2%	14,1%
97105	-1,9%	-7,9%	97210	16,5%	2,7%	97307	8,8%	21,2%	97407	26,2%	14,1%
97107	6,6%	4,8%	97212	3,3%	4,8%	97309	19,4%	18,8%	97408	1,8%	13,7%
97113	11,0%	5,8%	97213	2,8%	8,2%	97311	7,1%	32,1%	97409	17,9%	50,9%
97115	7,9%	-0,7%	97220	4,4%	9,4%				97410	12,8%	15,1%
97116	11,4%	14,4%	97221	7,0%	3,2%				97411	3,6%	13,4%
97117	14,8%	9,4%	97222	15,6%	5,7%				97412	15,5%	18,1%
97118	11,6%	37,9%	97224	8,5%	5,7%				97413	27,6%	11,9%
97120	24,6%	2,3%	97228	6,0%	2,2%				97414	26,1%	22,4%
97124	6,3%	0,3%	97229	2,5%	18,3%				97415	12,8%	16,2%
97125	30,9%	17,0%	97230	2,9%	20,2%				97416	13,4%	15,9%
97128	5,5%	14,3%							97418	16,1%	22,1%
97129	4,8%	2,1%							97420	23,0%	10,5%
									97422	15,1%	27,1%

Peu de communes possèdent des dynamiques similaires sur ces deux groupes.

Sur les 46 grandes communes de l'Outre-mer, 16 ont un différentiel de plus de 10 points entre les deux groupes (dont 8 dans le seul département de la Réunion), et 13 un différentiel de 5 à 10 points.

Les plus importants différentiels concernent les communes de Petit-Bourg en Guadeloupe (97118) avec 26,4 points, de La Trinité en Martinique (97230) avec 17,4 points, de Saint-Laurent-du-Maroni en Guyane (97311) avec 25 % et de Saint-André à La Réunion (97409) avec 33 points.

¹³ Le choix de la borne de deux points est arbitraire. Néanmoins, c'est celle qui correspond à une croissance annuelle régulière de 5%, c'est-à-dire une augmentation conséquente du parc de logements.

3.4.1.5. Mouvement de l'ensemble des grandes communes par DOM

	RP - 2006					RP - 2007					RP - 2008				
	GR - 1	GR - 2	GR - 3	GR - 4	GR - 5	GR - 2	GR - 3	GR - 4	GR - 5	GR - 1	GR - 3	GR - 4	GR - 5	GR - 1	GR - 2
total GC 971	17,9%	19,7%	20,3%	21,7%	20,4%	19,4%	20,0%	21,4%	20,1%	19,2%	19,6%	21,0%	19,7%	18,8%	21,0%
total GC 972	19,2%	19,7%	20,3%	21,2%	19,6%	19,4%	20,1%	21,0%	19,4%	20,1%	19,8%	20,7%	19,1%	19,8%	20,6%
total GC 973	21,1%	18,3%	19,6%	20,6%	20,4%	18,0%	19,3%	20,3%	20,1%	22,5%	18,9%	19,9%	19,7%	22,0%	19,5%
total GC 974	19,3%	19,2%	20,2%	20,6%	20,6%	18,7%	19,7%	20,0%	20,1%	21,5%	19,0%	19,3%	19,4%	20,8%	21,5%

Sur l'ensemble des grandes communes, il apparaît nettement moins de différences dans les contributions des EAR ; même si les mouvements erratiques constatés au niveau communal ne sont pas absents au niveau départemental, les contributions des EAR sont plus concentrées. Cependant, en Guyane (département qui ne contient que cinq grandes communes), le GR-1 contribue de façon assez importante et le GR-2 de façon assez faible ; en Guadeloupe, le GR-1 est visiblement trop faible et en Martinique le GR-4 trop fort. A La Réunion, on constate le mouvement attendu global de contribution décroissante en fonction de l'ancienneté de l'EAR

3.4.2. Des déséquilibres en structures

Pour apprécier les écarts entre les proportions des modalités des variables, on rappelle que la variance d'une proportion p peut être exprimée par l'expression suivante pour un sondage aléatoire simple : $(1-f) \times \frac{p \times (1-p)}{n-1}$, f étant le taux de sondage et n la taille de l'échantillon. Le rayon de

l'intervalle de confiance est (à 95 %) environ le double de l'écart-type. Le terme $p \times (1-p)$ est maximal quand $p = 0,5$; $1-f$ peut être assimilé à 0,6 pour les EAR. On peut ainsi proposer un majorant grossier de l'estimation de l'écart-type de la proportion, en fonction de la taille de l'EAR.

Le tableau suivant propose dans sa deuxième colonne le rayon de l'intervalle de confiance maximal (proportion autour de 50 %), qui sera le majorant des rayons qu'on peut garder principalement en tête, et les rayons pour quelques autres proportions :

taille de l'échantillon	majorant (proportion à 50%) des rayons des intervalles de confiance à 95% :	rayons des intervalles de confiance à 95% pour des proportions de :			
		40 % / 60 %	30% / 70%	20% / 80%	10% / 90%
250	4,90%	4,80%	4,49%	3,92%	2,94%
500	3,46%	3,39%	3,17%	2,77%	2,08%
750	2,83%	2,77%	2,59%	2,26%	1,70%
1000	2,45%	2,40%	2,24%	1,96%	1,47%
1250	2,19%	2,15%	2,01%	1,75%	1,31%
1500	2,00%	1,96%	1,83%	1,60%	1,20%
1750	1,85%	1,81%	1,70%	1,48%	1,11%
2000	1,73%	1,70%	1,59%	1,39%	1,04%
2250	1,63%	1,60%	1,50%	1,31%	0,98%
2500	1,55%	1,52%	1,42%	1,24%	0,93%
2750	1,48%	1,45%	1,35%	1,18%	0,89%
3000	1,41%	1,39%	1,30%	1,13%	0,85%
3250	1,36%	1,33%	1,25%	1,09%	0,82%
3500	1,31%	1,28%	1,20%	1,05%	0,79%
3750	1,26%	1,24%	1,16%	1,01%	0,76%
4000	1,22%	1,20%	1,12%	0,98%	0,73%
4250	1,19%	1,16%	1,09%	0,95%	0,71%
4500	1,15%	1,13%	1,06%	0,92%	0,69%

Les proportions seront nécessairement différentes (à 95 %) entre deux EAR d'une même commune si leur différence est supérieure à la somme des rayons correspondants à chaque EAR. Par exemple, pour une même commune où deux EAR contiendraient 1500 logements, si une proportion diffère de plus de quatre points entre ces deux enquêtes, on pourra considérer que les proportions seront significativement différentes¹⁴.

A titre illustratif des situations de déséquilibres (en structure), les tableaux suivant répertorient les variations de l'aspect du bâti (variable importante pour l'échantillonnage et le calage), et le statut d'occupation du ménage (nécessaire pour l'enquête « Loyers et Charges » ou pour l'enquête « Logement ») pour quelques grandes communes de Guadeloupe.

Entre autres, on peut citer des écarts très importants sur les taux de propriétaires aux Abymes (10,4 points sur le premier cycle), à Capesterre-Belle-Eau (10,6 points sur le premier cycle), à Petit-Bourg (10,3/10,4 points sur les trois cycles), à Pointe-à-Pitre (10,7 points sur le premier cycle). Ils dépassent les 20 points au Gosier (21,6 points sur le premier cycle) et à Sainte-Rose (20,4 points sur le premier cycle). Des écarts importants existent également sur les types de logement (individuel versus collectif) : pour sept « grandes » grandes communes, ces écarts dépassent les 10 points (Les Abymes, Baie-Mahault, Capesterre-Belle-Eau, Le Gosier, Petit-Bourg, Pointe-à-Pitre, Sainte-Rose) et peuvent frôler les 20 points (Baie-Mahault).

Les variations sur l'aspect du bâti sont plus faibles. On constate ainsi une variabilité maximale pour la modalité « maisons ou immeubles en dur » de 4 à 10 points selon la commune ; cependant, plus de 80% des logements sont généralement de ce type. Les intervalles de confiance se recouvrent peu entre EAR d'une même commune. Le lecteur pourra utiliser le tableau précédent pour apprécier les écarts présentés ci-dessous.

Pour l'exemple, raisonnons rapidement sur la commune. Sur le premier cycle de recensement, la taille moyenne d'une EAR est de 1906 logements pour une proportion moyenne de 86% de « maisons ou immeubles en dur », ce qui nous conduit à un rayon approximatif de l'intervalle de confiance de l'ordre de 1,2 point, donc d'une différence significative entre deux proportions issues de deux EAR différentes de l'ordre de 2,4 points. Sur le premier cycle de recensement Aux Abymes, cinq des huit différences entre deux EAR sont supérieures à ce seuil.

Des variations de même type ont lieu pour de nombreuses autres variables. Le lecteur pourra se reporter à la note dédiée du Criem pour une analyse plus détaillée.

			Année de collecte						
			2004	2005	2006	2007	2008	2009	2010
97101 - Les Abymes	Aspect du bâti	Habitation de fortune	61	25	12	15	12	11	7
			3.39	1.34	0.60	0.77	0.63	0.63	0.35
		Case traditionnelle	169	116	167	135	146	85	75
			9.39	6.20	8.30	6.97	7.65	4.89	3.75
		Maison ou immeuble en bois	74	63	78	105	139	165	68
		4.11	3.37	3.87	5.42	7.28	9.48	3.40	
		Maison ou immeuble en dur	1496	1667	1756	1682	1612	1479	1852
		83.11	89.10	87.23	86.84	84.44	85.00	92.51	
		Total	1800	1871	2013	1937	1909	1740	2002
	Statut d'occupation du ménage	Propriétaire	956	861	928	881	815	890	825
		53.11	46.02	46.10	45.48	42.69	51.15	41.21	
Locataire d'un logement loué vide		746	924	1008	969	980	745	1106	
		41.44	49.39	50.07	50.03	51.34	42.82	55.24	
Locataire d'un logement loué meublé		28	25	30	27	33	30	18	
	1.56	1.34	1.49	1.39	1.73	1.72	0.90		
	Logé gratuitement	70	61	47	60	81	75	53	
	3.89	3.26	2.33	3.10	4.24	4.31	2.65		
	Total	1800	1871	2013	1937	1909	1740	2002	

¹⁴ En toute rigueur, le tirage en jeu ici est un tirage d'adresses qui représentent 40% des logements et non un tirage direct de logements avec un taux de sondage de 40%, pour lequel les expressions de calcul de précision utilisées ici s'appliqueraient directement. Cette première phase de tirage d'adresse augmente la variance des estimations ; ainsi, l'approximation qu'on propose ici conduit à sous-estimer un peu le rayon de l'intervalle de confiance.

		Année de collecte							
		2004	2005	2006	2007	2008	2009	2010	
97103 - Baie-Mahault	Aspect du bâti	Habitation de fortune	9	9	19	2	9	5	7
		Case traditionnelle	49	25	20	10	19	21	4
		Maison ou immeuble en bois	15	34	16	34	27	13	22
		Maison ou immeuble en dur	562	748	743	798	779	751	1026
		Total	635	816	798	844	834	790	1059
	Statut d'occupation du ménage	Propriétaire	360	399	402	417	474	431	497
		Locataire d'un logement loué vide	233	363	338	395	316	310	516
		Locataire d'un logement loué meublé	8	14	25	9	8	12	15
		Logé gratuitement	34	40	33	23	36	37	31
		Total	635	816	798	844	834	790	1059
97107 - Capesterre-Belle-Eau	Aspect du bâti	Habitation de fortune	9	9	19	2	9	5	7
		Case traditionnelle	49	25	20	10	19	21	4
		Maison ou immeuble en bois	15	34	16	34	27	13	22
		Maison ou immeuble en dur	562	748	743	798	779	751	1026
		Total	635	816	798	844	834	790	1059
	Statut d'occupation du ménage	Propriétaire	360	399	402	417	474	431	497
		Locataire d'un logement loué vide	233	363	338	395	316	310	516
		Locataire d'un logement loué meublé	8	14	25	9	8	12	15
		Logé gratuitement	34	40	33	23	36	37	31
		Total	635	816	798	844	834	790	1059
97113 - Le Gosier	Aspect du bâti	Habitation de fortune	4	1	0	6	9	1	4
		Case traditionnelle	24	28	22	19	22	10	18
		Maison ou immeuble en bois	21	33	41	57	34	41	31
		Maison ou immeuble en dur	760	734	781	1031	848	846	789
		Total	809	796	844	1113	913	898	842
	Statut d'occupation du ménage	Propriétaire	537	501	473	499	569	528	515
		Locataire d'un logement loué vide	197	194	296	535	290	298	252
		Locataire d'un logement loué meublé	40	70	48	39	18	31	43
		Logé gratuitement	35	31	27	40	36	41	32
		Total	809	796	844	1113	913	898	842

		Année de collecte							
		2004	2005	2006	2007	2008	2009	2010	
97118 - Petit-Bourg	Aspect du bâti	Habitation de fortune	9	4	2	5	3	10	11
		Case traditionnelle	16	23	24	22	14	20	11
		Maison ou immeuble en bois	43	37	26	37	35	25	51
		Maison ou immeuble en dur	538	492	602	570	606	621	694
		Total	606	556	654	634	658	676	767
	Statut d'occupation du ménage	Propriétaire	362	378	402	438	462	403	422
		Locataire d'un logement loué vide	197	152	219	161	143	235	301
		Locataire d'un logement loué meublé	6	1	5	6	17	11	8
		Logé gratuitement	41	25	28	29	36	27	36
		Total	606	556	654	634	658	676	767
97120 - Pointe-à-Pitre	Aspect du bâti	Habitation de fortune	13	2	7	3	6	0	1
		Case traditionnelle	6	21	24	8	19	7	9
		Maison ou immeuble en bois	27	37	32	49	36	24	34
		Maison ou immeuble en dur	516	557	561	638	662	669	587
		Total	562	617	624	698	723	700	631
	Statut d'occupation du ménage	Propriétaire	106	150	121	95	144	131	116
		Locataire d'un logement loué vide	423	452	476	572	539	535	493
		Locataire d'un logement loué meublé	18	12	14	19	20	19	12
		Logé gratuitement	15	3	13	12	20	15	10
		Total	562	617	624	698	723	700	631
97129 - Sainte-Rose	Aspect du bâti	Habitation de fortune	1	6	8	9	9	6	4
		Case traditionnelle	6	9	10	5	7	17	10
		Maison ou immeuble en bois	46	58	26	22	33	17	43
		Maison ou immeuble en dur	408	450	642	670	513	443	477
		Total	461	523	686	706	562	483	534
	Statut d'occupation du ménage	Propriétaire	377	433	428	464	415	360	421
		Locataire d'un logement loué vide	69	77	227	220	107	77	84
		Locataire d'un logement loué meublé	3	2	3	8	13	10	8
		Logé gratuitement	12	11	28	14	27	36	21
		Total	461	523	686	706	562	483	534

3.4.3. La piste de l'empilement pour lisser les déséquilibres

Les résultats « qualitatifs » précédents pointent un certain nombre de difficultés que soulève la méthode de recensement utilisé Outre-mer, dans le cadre d'une utilisation de ce dispositif pour asseoir un dispositif d'échantillonnage pour les enquêtes auprès des ménages, en particulier pour obtenir une stratification sur les grandes communes sur une EAR ; la difficulté n'est d'ailleurs pas moindre pour effectuer un tirage non stratifié en considérant l'ensemble des grandes communes - La Réunion posant certes de ce point de vue moins de problème, car il y a moins de variabilité quand on considère l'ensemble des grandes communes.

Deux impacts directs sur les plans de sondage sont à prendre en considération.

D'une part, cette variabilité dans le recensement (qui constitue la première phase du point de vue de l'échantillonnage des enquêtes auprès des ménages) induit une contribution certaine à la variance globale des estimations dans les enquêtes.

D'autre part, les structures des EAR étant, à cause de cette grande variabilité, éloignées de la structure générale qu'on va utiliser pour le calage de l'enquête, les poids de tirage seront mécaniquement éloignés des poids calés. Le calage pourrait être bien plus qu'un ajustement marginal. En filigrane se dessine, au-delà de la question de pouvoir effectuer un calage satisfaisant ou non, la question de simplement pouvoir caler l'enquête. La question n'est pas rhétorique : par le passé, certains calages de l'EEADOM en Guyane ont par exemple nécessité d'accepter des rapports de poids initiaux/finaux très importants au calage.

Une piste possible est la mobilisation de deux (ou plus) EAR pour la construction des bases (annuelles) de sondage.

La mobilisation de plusieurs EAR apporte nécessairement un gain du point de vue de la variance de première phase puisqu'on en élargit le socle. La question des structures des EAR n'est abordée, dans ce qui suit, que dans la perspective des difficultés de calage.

Les tableaux suivants donnent, pour toutes les grandes communes de Guadeloupe, les variations, pour chacun des trois cycles d'EAR, et pour chacun des deux « cycles » d'empilement d'EAR deux à deux les variations des proportions des deux modalités principales des variables de « type de logement » et de « statut d'occupation » du logement, qui sont mobilisées pour l'échantillonnage. Pour faciliter la lecture de ces tableaux, une illustration est disponible ci après.

	proportion de ménages propriétaires de leur logement					proportion de ménages locataires de logements loués vide				
	variation sur 5 EAR consécutives			variation sur 5 empilements consécutifs de deux EAR consécutives		variation sur 5 EAR consécutives			variation sur 5 empilements consécutifs de deux EAR consécutives	
	Cycle 1	cycle 2	cycle 3	cycle 1	cycle 2	Cycle 1	cycle 2	cycle 3	cycle 1	cycle 2
97101	10,42	8,46	9,94	5,40	2,63	9,90	8,52	12,42	5,18	3,40
97103	7,93	7,93	9,90	6,10	6,10	10,11	8,91	10,84	6,09	6,13
97105	14,58	14,58	14,58	10,54	10,54	11,93	11,93	11,93	9,45	9,45
97107	10,66	7,78	9,89	4,77	4,77	8,30	9,04	9,82	5,25	5,25
97113	21,55	18,11	17,49	15,01	10,91	23,72	23,70	18,14	18,10	12,58
97115	33,27	33,27	29,99	27,17	27,17	36,19	36,19	27,36	28,12	28,12
97116	22,17	18,96	14,49	14,18	8,00	21,25	20,37	18,32	13,87	8,42
97117	9,44	9,44	9,44	5,81	5,54	10,18	10,18	10,18	7,61	7,61
97118	10,47	10,59	15,19	5,98	12,49	11,76	13,03	17,51	7,13	13,62
97120	10,70	10,70	6,31	5,50	5,50	8,69	8,69	7,40	5,06	4,50
97124	12,00	15,31	15,31	9,98	12,51	20,22	20,22	15,72	15,76	14,62
97125	18,66	9,26	9,26	10,74	6,63	16,23	10,83	10,69	10,38	7,15
97128	3,91	3,91	12,27	2,88	7,26	7,58	7,58	15,33	5,30	8,42
97129	20,40	20,40	16,45	18,24	12,71	18,37	18,37	17,36	17,27	16,28

	proportion de ménages habitant une maison individuelle					proportion de ménages habitant un logement collectif				
	variation sur 5 EAR consécutives			variation sur 5 empilements consécutifs de deux EAR consécutives		variation sur 5 EAR consécutives			variation sur 5 empilements consécutifs de deux EAR consécutives	
	Cycle 1	cycle 2	cycle 3	cycle 1	cycle 2	Cycle 1	cycle 2	cycle 3	cycle 1	cycle 2
97101	6,35	7,38	10,24	3,25	3,25	9,20	7,99	10,97	4,95	3,07
97103	19,08	8,62	13,95	9,62	8,05	20,23	8,90	13,84	9,24	7,27
97105	20,87	19,94	20,42	10,55	9,48	20,23	20,23	20,24	9,32	9,37
97107	11,59	9,33	8,98	5,75	5,75	10,67	9,00	10,16	5,66	5,66
97113	17,01	16,37	15,40	11,83	9,78	20,24	16,77	15,13	13,30	9,26
97115	28,90	28,90	29,62	22,65	22,65	28,21	28,21	28,41	22,75	22,75
97116	23,10	23,10	20,90	13,11	11,46	23,15	23,15	22,20	13,82	12,28
97117	9,10	9,10	9,10	5,78	5,78	8,37	8,37	8,37	5,39	5,39
97118	12,85	12,85	17,68	6,70	11,61	12,86	12,86	17,63	6,90	11,54
97120	4,68	8,28	8,28	5,72	7,02	9,65	12,96	12,96	8,25	9,47
97124	22,84	22,84	21,97	16,28	15,65	23,52	23,52	26,29	15,39	16,56
97125	21,37	17,24	17,24	12,06	10,16	23,34	17,89	17,89	13,51	10,66
97128	4,68	5,14	9,00	3,27	4,17	5,05	5,05	9,84	3,24	4,87
97129	16,57	16,52	15,64	16,26	15,27	16,36	16,36	15,58	15,97	15,25

Le lissage des structures pour empilement est toujours réalisé, les variations sont toujours moindres sur un cycle de cinq « bases » quand on empile deux enquêtes.

Cependant, le gain de variabilité n'est pas toujours de la même ampleur selon les communes. Il dépend en effet de l'ampleur des distorsions des structures dans l'ordre des EAR.

Soit l'illustration suivante :

Deux communes A et B ont des groupes de rotation de même taille, 100 logements. Dans deux groupes, la proportion d'une modalité X d'une variable est de 80 % et dans les trois autres de 20 % (soit une proportion de 50 % sur l'ensemble de la commune).

Sur un cycle de recensement, EAR par EAR, la variation de la proportion de cette modalité est de 60 points, quel que soit l'ordre d'interrogation des EAR.

Il n'en va pas de même pour les variations de cinq empilements de deux EAR consécutives.

Si les EAR où la modalité X est surreprésentée sont les GR-1 et GR-2, on est dans la première des deux situations suivantes. Si par contre les EAR où la modalité X est surreprésentée sont les GR-1 et GR-3, on est dans la deuxième des situations suivantes.

	empilement de deux EAR consécutives											
	GR1	GR2	GR3	GR4	GR5	variation sur un cycle	GR1/2	GR2/3	GR3/4	GR4/5	GR5/1	variation sur un cycle
Commune A	80	80	20	20	20	60 points	160	100	40	40	100	60 points
	20	20	80	80	80		80%	50%	20%	20%	50%	
	100	100	100	100	100		40	100	160	160	100	
							20%	50%	80%	80%	50%	
Commune B	80	20	80	20	20	60 points	100	100	100	40	100	30 points
	20	80	20	80	80		50%	50%	50%	20%	50%	
	100	100	100	100	100		100	100	100	160	100	
							50%	50%	50%	80%	50%	
						200	200	200	200	200		

Cet exemple explicite l'influence de l'ordre des groupes de rotation sur le lissage des distorsions quand on empile deux EAR : au mieux, on améliore considérablement la structure de la base de tirage, au pire les distorsions restent ce qu'elles sont.

Avec un empilement de trois EAR pour constituer une base de sondage annuelle, les données disponibles des recensements (2004 à 2010) permettent seulement d'illustrer la variation des proportions de modalités de variables avec une seule valeur pour un cycle de cinq bases. On peut cependant reprendre le même exemple numérique que précédemment pour un empilement de trois enquêtes.

		GR1	GR2	GR3	GR4	GR5	variation sur un cycle	empilement de trois EAR consécutives					
		GR1/2/3	GR2/3/4	GR3/4/5	GR4/5/1	GR5/1/2	variation sur un cycle						
Commune A		80	80	20	20	20	60 points	180	120	60	120	180	40 points
		20	20	80	80	80		60%	40%	20%	40%	60%	
		100	100	100	100	100		120	180	240	180	120	
								40%	60%	80%	60%	40%	
		300	300	300	300	300		300	300	300	300	300	
Commune B		80	20	80	20	20	60 points	180	120	120	120	120	20 points
		20	80	20	80	80		60%	40%	40%	40%	40%	
		100	100	100	100	100		120	180	180	180	180	
								40%	60%	60%	60%	60%	
		300	300	300	300	300		300	300	300	300	300	

Avec l'empilement de trois EAR, même dans la situation la plus défavorable pour l'empilement de deux EAR, on limite l'expression des distorsions de structures entre les groupes de rotation. En effet, dans ce cas, un empilement de trois sur le cycle de cinq mobilise nécessairement les deux enquêtes qui réalisent les valeurs minimale et maximale de la variable.

La faiblesse de l'utilisation d'un empilement d'EAR est le problème de la fraîcheur de la base.

La dernière EAR permet d'utiliser des données vieilles d'un an, empiler deux EAR (respectivement trois EAR) oblige à utiliser des données vieilles de deux ans (respectivement trois ans).

4. Le nouveau dispositif en septembre 2013 pour les DOM « historiques »

4.1. Présentation générale

La base de sondage complète contient les cinq dernières EAR. Le nombre d'EAR à choisir pour le tirage dans les grandes communes est ouvert ; dans les tirages classiques, il est finalement arrêté à deux, mais la structure de la base de sondage permet de recourir aux EAR plus anciennes. Dans les petites communes, l'ensemble du cycle est conservé.

Nous savons maintenant dans quelle base de sondage tirer nos échantillons ultra-marins. Reste maintenant à étudier *comment* les tirer. Le nouveau dispositif d'échantillonnage doit pouvoir être utilisé pour l'essentiel des enquêtes auprès des ménages réalisées par l'Insee dans les quatre DOM « historiques », mais au sein du dispositif statistique, l'enquête sur l'emploi (EECDOM dans sa forme moderne, appliquée à l'Outre-mer) tient une place tellement centrale que les spécificités de collecte de cette enquête ont été intégrées très en amont de la réflexion, profitant du passage des DOM « historiques » dans le dispositif métropolitain « en continu ». La mise en place du nouveau cadre d'emploi des enquêteurs (NCEE) a également contraint la réflexion.

4.2. Le paradoxe de l'EECDOM : disperser le tirage, regrouper la collecte

Le rapport de l'Inspection Générale sur la stratégie d'enquête auprès des ménages dans les Dom diffusé en 2010 proposait plusieurs scénarii de l'enquête Emploi dans les 4 Dom. Celui qui a été retenu s'appuie sur l'interrogation chaque année d'environ *8 000 logements par Dom*, sur le même schéma que l'enquête Emploi en continu métropolitaine (panel rotatif, six vagues d'interrogation trimestrielle) ; ces 8000 logements devant être tirés *de manière dispersée sur le territoire*. Ce scénario permet d'assurer une représentativité trimestrielle pour l'ensemble des Dom respectant la contrainte européenne, ainsi qu'une représentativité annuelle pour chacun des Dom raisonnablement diminuée par rapport à la représentativité issue des enquêtes Emploi annuelles qui ont eu cours de 2001 à 2012.

Le tirage de l'EECDOM devait ainsi répondre à deux logiques opposées¹⁵ :

- D'une part, neutraliser l'effet de grappe au niveau du logement, c'est à dire interroger des logements dispersés sur le territoire. En métropole, des grappes de logements contigus sont constituées pour la réalisation de l'EEC à partir des sources fiscales, mais la dispersion des logements devait être conservée pour l'enquête ultra-marine pour ne pas trop dégrader la précision.
- D'autre part, permettre une collecte dans des délais très contraints. Rappelons que la semaine de référence est la période sur laquelle porte l'essentiel des questions de l'enquête emploi. La collecte doit avoir lieu pendant les 2 semaines et 2 jours qui suivent cette semaine de référence. Il s'agissait ici d'optimiser la répartition des logements tirés sur l'ensemble du territoire, répartition au sein des semaines de référence et répartition au sein du réseau d'enquêteurs.

8000 logements interrogés par an = 2000 logements interrogés par trimestre = 333 logements entrants chaque trimestre. Chaque trimestre étant découpé en 13 semaines, nous obtenons environ 26 logements entrants (333 / 13) à enquêter chaque semaine de référence.

1 vague = 338 logements = 13 semaines de référence * 26 logements.

A défaut de pouvoir regrouper ces 26 logements dans l'espace en constituant des grappes, l'idée est de les regrouper dans le temps. Les deux possibilités les plus immédiates sont

- a) soit 13 charges de 26 logements à réaliser chaque semaine de référence,
- b) soit 26 charges de 13 logements.

Une solution mixte pour obtenir un nombre de charges intermédiaire entre 13 et 26 consisterait à créer des secteurs de collecte à charge « double » (26 logements) et des secteurs à charge « simple » (13 logements), en fonction de l'accessibilité des logements sur le territoire.

En 2012, la faisabilité des charges de 13 et de 26 logements entrants par enquêteur a été testée sur le terrain. Il s'est avéré que cela était trop compliqué voir impossible pour un seul enquêteur d'enquêter 26 logements entrants en deux semaines et demi de collecte (solution a). La décision a donc été prise d'opter pour des charges de 13 logements (solution b). La taille moyenne des charges est donc de 13 logements et il y a deux nouvelles charges (dites « entrantes ») par semaine.

Le tirage a donc été pensé de manière à ce que l'échantillon entrant de 338 logements soit *par construction réparti en 26 secteurs réputés équivalents et réalisant une partition du territoire*, chacun de ces secteurs étant associé à une semaine de référence.

1 vague = 338 logements = 26 secteurs * 13 logements « regroupés au mieux ».

On aboutit ainsi à une charge « enquêteur » associée à une semaine de référence de 13 logements « regroupés au mieux ». Cette charge reste nettement inférieure à celle de l'EEC métropole (environ 24 logements « concentrés ») et à celle de l'ancienne EEADOM (comprise entre 15 et 25 logements « regroupés » ; quelques libertés étaient prises pour réaliser ces regroupements avec les semaines de référence de la collecte, ce qui était interdit dans le processus EEC) pour tenir compte des distances beaucoup plus importantes entre les logements à enquêter malgré les efforts de regroupement.

4.3. En réponse au basculement vers l'EECDOM et au NCEE : un double découpage en Secteurs d'Activité des Enquêteurs (SAE) et en strates de tirage

Chaque département d'outre-mer a donc été découpé en 26 SAE (**Secteurs d'Activité des Enquêteurs**). Ce zonage a tout d'abord été mis en place pour organiser la collecte de l'enquête emploi en continu dans les DOM. Les SAE s'inspirent à la fois des secteurs définis dans l'EEC métropolitaine et des Zones d'Activité des Enquêteurs (ZAE) métropolitaines. Un SAE est ainsi construit pour que la collecte EEC y soit dédiée à un enquêteur, à l'exception peut-être de certains secteurs dont l'étendue pourrait nécessiter le recours à plusieurs enquêteurs. En tout état de cause, cela signifie *qu'un minimum de 26 enquêteurs doivent contribuer à l'EEC pour que le système soit optimisé*. On verra plus loin que ce n'est actuellement pas le cas dans chaque DOM (§5.1).

Le découpage en SAE peut aussi servir de cadre de collecte aux autres enquêtes. En effet, les SAE ont aussi été créés pour répondre à une nouvelle contrainte de fonctionnement, la mise en place du NCEE en 2013. Cet outil peut ainsi être utilisé pour les prévisions de charge des enquêteurs et l'organisation de la collecte de toute enquête.

La partition du territoire était compliquée par le fait que plusieurs logiques étaient à l'œuvre :

- une logique théorique de construction de strates de tirage : les strates doivent abriter des populations aussi homogènes que possible ;
- une logique de construction de secteurs de collecte : les secteurs, appelés SAE, doivent être aussi homogènes que possible en charge et en accessibilité.

La solution mise en œuvre tente de concilier les deux logiques : des strates de tirage et des SAE coexistent, mais ont été pensés pour fonctionner aussi bien que possible ensemble, comme un maillage. Les SAE découpent le territoire en zones homogénéisées (du point de vue de la collecte) et contenant environ le même nombre de résidences principales (ainsi, *en moyenne*, les SAE offrent des

charges équivalentes). Les strates de tirage découpent le territoire en zones homogènes du point de vue du tirage, et à un niveau géographique suffisamment fin pour garantir un nombre de fiches-adresse proche dans chaque SAE.

La logique mise en l'œuvre en 2012 pour construire les SAE et les strates a été la suivante :

1 - 26 secteurs (fixes) ont été définis en respectant une contrainte d'équilibrage : au sein de chaque microrégion, chaque secteur contient un nombre proche de résidences principales d'après le recensement de 2008. Cette contrainte d'équilibrage au sein de chaque microrégion a été fixée car :

- Construire des secteurs de même taille permet de limiter les fluctuations qu'engendrera le tirage sur la charge par SAE ;
- On considère le découpage en microrégions comme explicatif des comportements. Les taux de chômage par zone d'emploi sont ainsi contrastés. La microrégion apparaît donc comme une variable de stratification améliorant l'échantillonnage.

Le nombre de SAE par micro-région et la taille des SAE suivant la micro-région ont été fixés de la manière suivante :

micro-région	nombre de résidences principales (RP08) (1)	nombre de logements attendus dans une vague EEC	nombre de SAE = nombre de logements / 13 (2)	taille moyenne du SAE (1) / (2)
Pointe à Pitre (11a)	54 497	117	9	6 055
Nord et est Grande terre (11b)	36 170	78	6	6 028
Nord et Est basse_Terre (11c)	36 708	79	6	6 118
Basse-Terre (12)	24 649	53	4	6 162
Marie-Galante (13)	4 954	11	1	4 954
Guadeloupe	156 978	338	26	6 038
Centre-Agglomération (21)	69 029	148	11	6 275
Sud-Caraïbes (22)	25 180	54	4	6 295
Sud (23)	21 200	46	4	5 300
Centre-Atlantique (24)	25 472	55	4	6 368
Nord-Caraïbes (25)	9 216	20	2	4 608
Nord-Atlantique (26)	7 233	16	1	7 233
Martinique	157 330	338	26	6 051
Cayenne (31)	37 993	222	17	2 235
Kourou (32)	8 572	50	4	2 143
St Laurent (33)	11 362	66	5	2 272
Guyane	57 928	338	26	2 228
Saint Pierre (41)	100 553	122	9	11 173
Saint Benoit (42)	37 424	46	4	9 356
Saint Denis (43)	71 890	88	7	10 270
Le Port (44)	67 785	83	6	11 298
Réunion	277 652	338	26	10 679

Pour l'enquête emploi en continu (EEC), il a été décidé de relâcher la contrainte d'équilibrage sur le nombre de résidences principales pour 3 SAE guyanais où l'habitat est isolé, les distances y étant trop importantes. Ces 3 SAE comportent donc moins de résidences principales que les autres et sont donc surreprésentées pour l'EEC. Ce choix a l'avantage aussi de fidéliser des enquêteurs sur des zones où le recrutement n'est pas aisé à cause des distances importantes à parcourir. Cette surreprésentation a été reproduite pour l'enquête Logement de 2013-2014 mais pas pour Cadre de Vie et Sécurité de 2015.

2 - Il est cependant apparu que la contrainte d'équilibrage des SAE par microrégion était insuffisante pour limiter les fluctuations engendrées par un tirage stratifié par microrégion : ainsi, le tirage d'un échantillon trimestriel EEC (338 logements) pouvait aboutir à des tailles de charge allant de 5 à 20 logements, ce qui apparaissait trop aléatoire (la charge recherchée est de 13 logements en moyenne).

Une stratification supplémentaire a donc été réalisée. Les contraintes, liées à un mode de tirage dans le RP différencié selon la taille des communes étaient cette fois les suivantes :

- 1 grande commune = 1 strate a minima,
- l'ensemble des petites communes d'une microrégion = 1 strate a minima.

Suivant leur taille, ces strates « minimales » ont été affinées. Par exemple, sur les 338 logements d'une vague, on attend 79 logements dans la commune de Fort-de-France en Martinique. Cette grande commune a été découpée en plusieurs strates au final, en respectant les contraintes suivantes :

1. chaque strate devait être suffisamment grosse pour qu'on n'ait jamais 0 logement répondant à l'intérieur de la strate : on évite ainsi de descendre en dessous de 5 logements à enquêter ;
2. au sein d'une strate, il ne devait pas y avoir des populations très (trop) différentes ;
3. plus on a de strates, moins on a de fluctuation sur la charge d'enquête de chacune des 26 SAE.

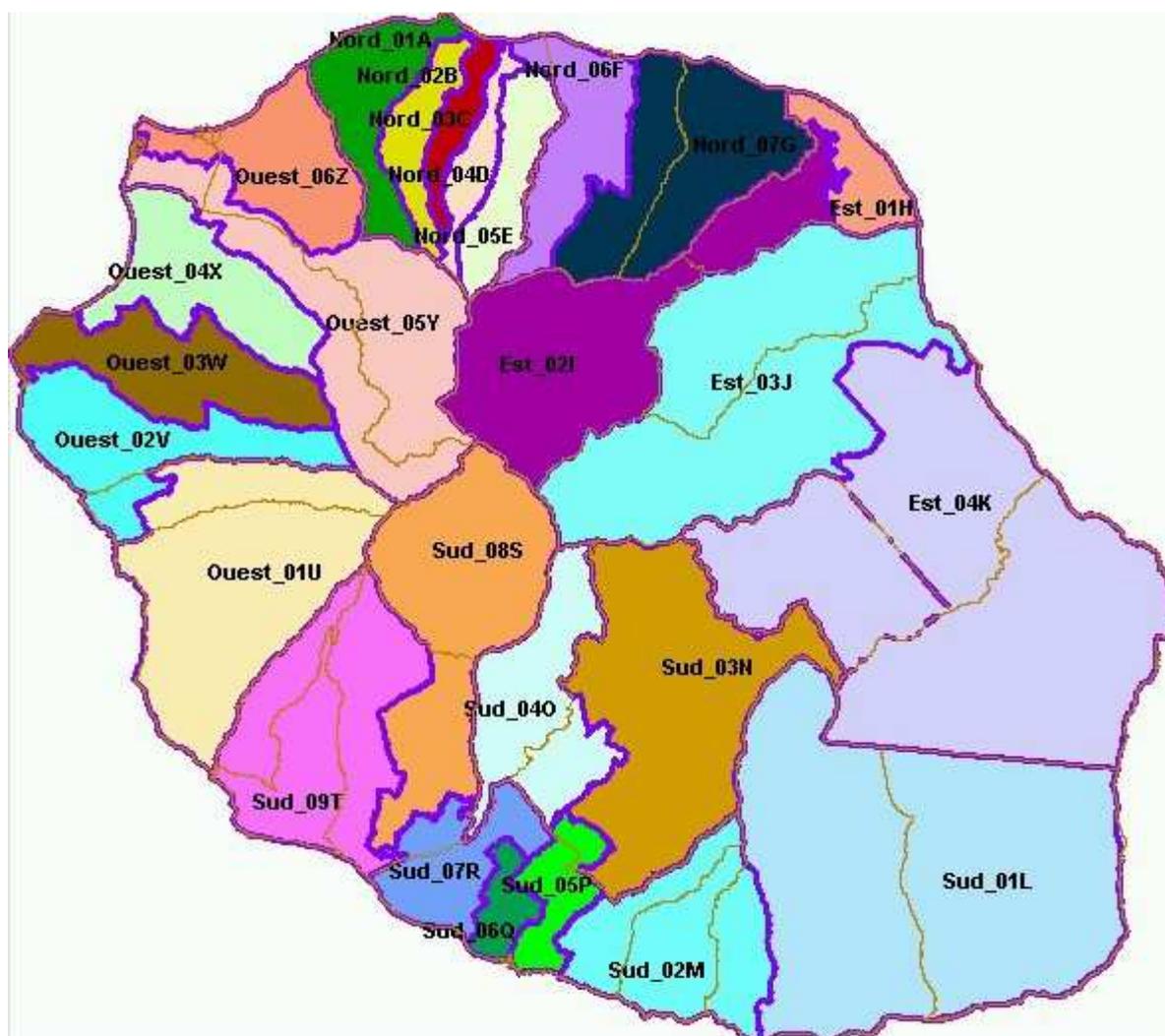
Les tirages stratifiés réalisés à l'aide de ces découpages ont montré leur efficacité à limiter les fluctuations autour de 13 logements. Dans les échantillons envoyés aux UEM, ces tailles ne varient « plus que » entre 9 et 17 logements.

4.4. Illustration à La Réunion : « du battant des lames au sommet des montagnes »¹⁶

A La Réunion, le découpage en SAE a été fait selon une logique « Hauts » et « Bas », autrement dit en partant depuis le littoral (mer) vers l'intérieur de l'île (remparts). Dans la terminologie locale, les Bas correspondent aux zones les plus urbanisées des plaines côtières et des basses pentes. Les Hauts par opposition aux Bas se caractérisent par une empreinte humaine plus faible, un habitat moins dense et plus dispersé et un caractère rural plus ou moins prononcé selon les zones.

¹⁶ Extrait du rapport final Maiol UEM DOM, note d'octobre 2012 N° 835/DR974-SES/

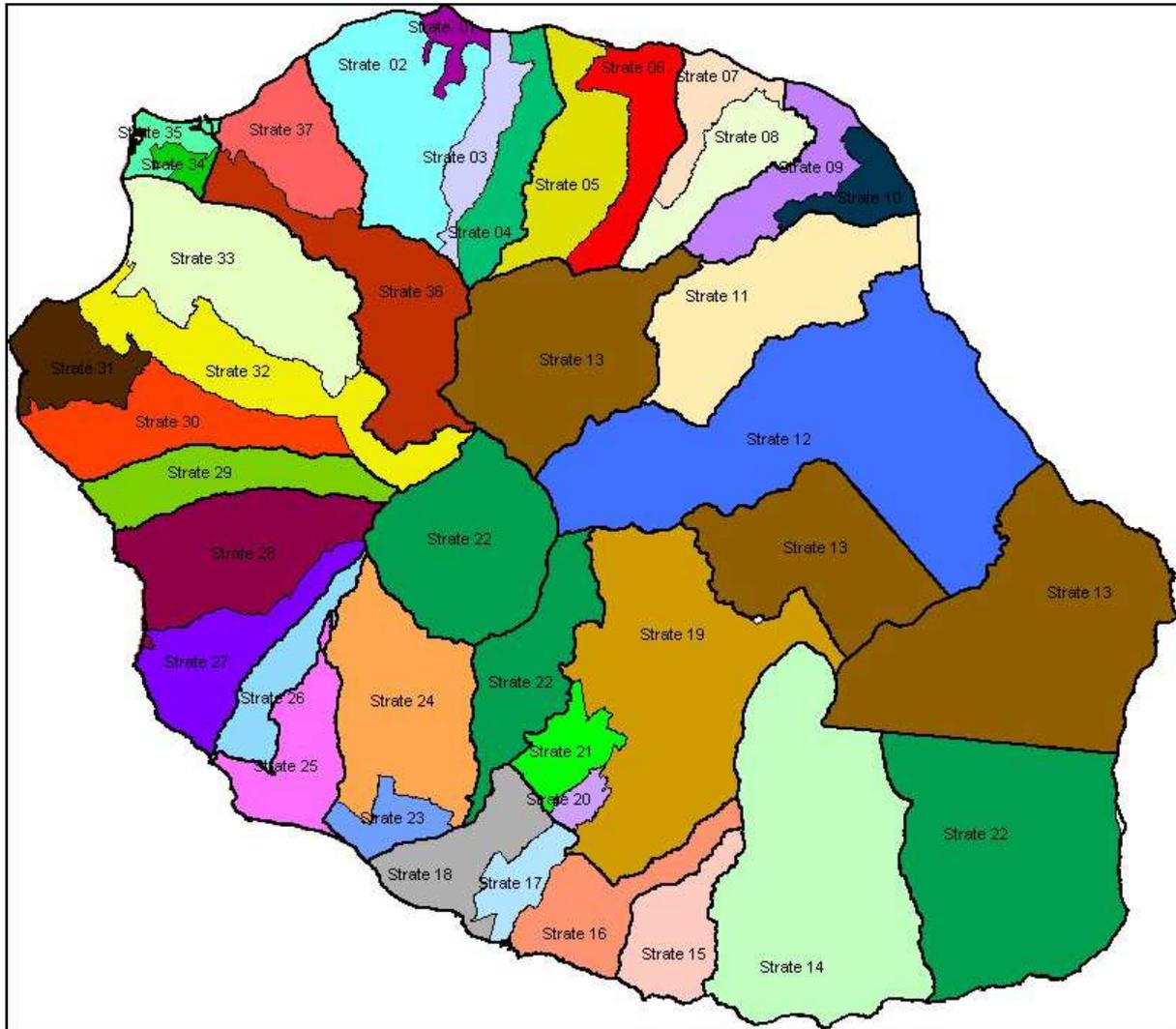
La cartographie des 26 SAE de La Réunion



L'ensemble du territoire a été pris en compte dans le découpage, y compris les zones des hauts les plus reculées ou difficiles d'accès comme par exemple le cirque de Mafate. La surface de certains SAE peuvent paraître importante en raison justement de l'importance des zones naturelles (forêts, volcan, etc.), qui ne sont pas habitées. Ainsi la superficie du secteur **L** (Saint-Philippe à Saint-Joseph) fait près de 31 hectares dont les 3/4 sont des espaces vides, correspondant en partie à la zone du volcan de la Fournaise.

Le découpage en strate reste relativement proche du découpage en SAE, comme on le voit en comparant les deux cartes réunionnaises. On visualise cependant que le cirque de Cilaos forme une strate à part entière, la strate 22, alors qu'il est associé à la ville de Saint Louis dans le SAE **S**. Les populations vivant à Cilaos, cirque accessible par une seule et longue route, et à Saint Louis, ville des Bas, sont en effet des populations très différentes, qu'il ne convenait pas de laisser ensemble au sein d'une strate de tirage.

La cartographie des 37 strates de tirage à La Réunion



4.5. Associer SAE, vague et semaine de référence

Ce travail a été réalisé en concertation avec les UEM, en particulier pour anticiper les questions liées aux absences et remplacements dans le Nouveau Cadre d'Emploi des Enquêteurs.

Une répartition concentrée dans le temps a été privilégiée : chaque enquêteur travaille 6 semaines de référence d'affilée, 2 semaines étant associée à des charges « visite » (1^{ère} et dernière interrogation), les 4 autres à des charges « téléphone ». De plus, afin d'éviter que deux secteurs voisins ne soient enquêtés simultanément, la collecte y est espacée de 6 semaines. Sur l'exemple ci dessous, l'enquêteur du secteur B pourra ainsi plus aisément venir en renfort sur le secteur A si besoin.

Secteur A

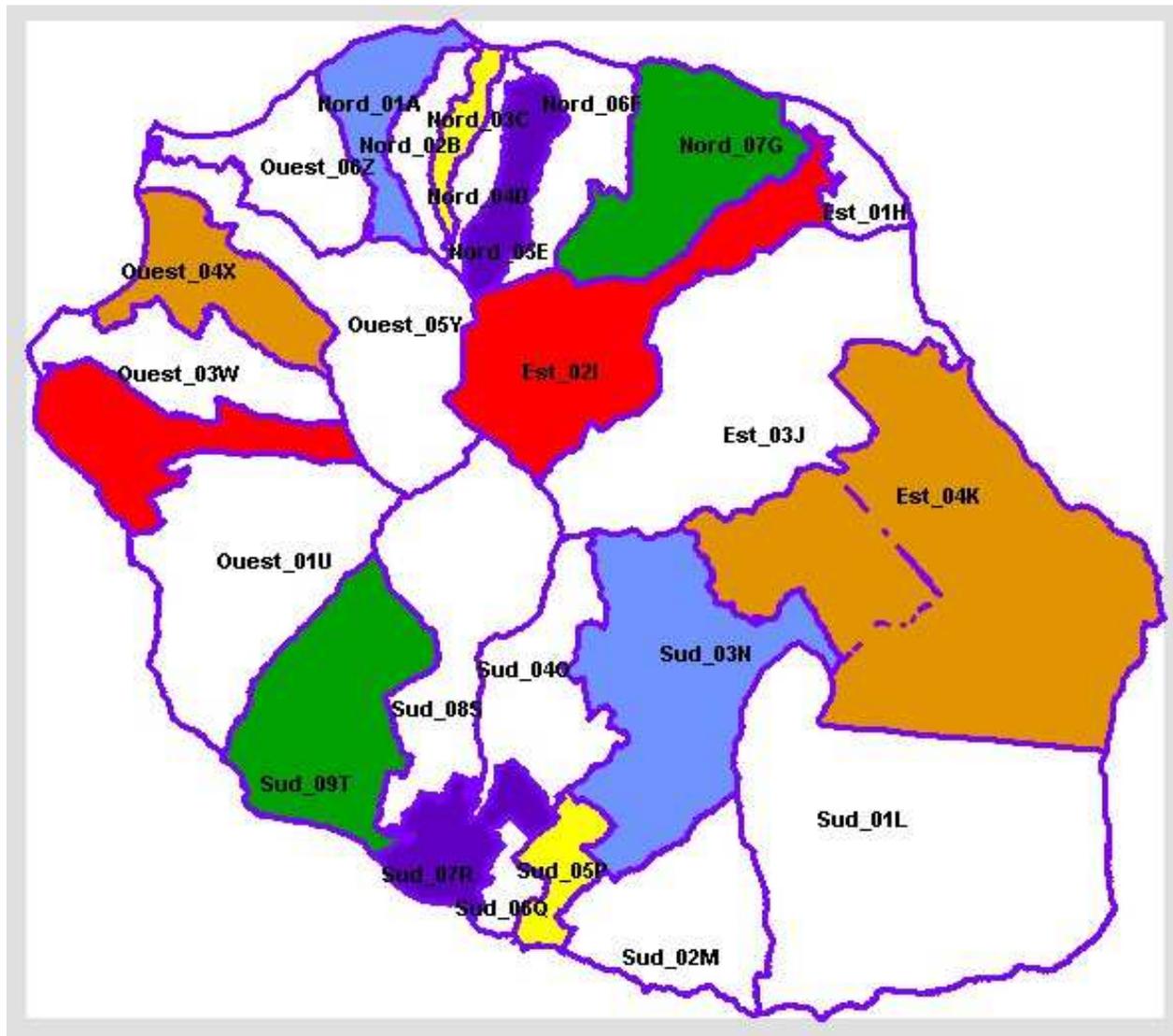
		Semaine												
		1	2	3	4	5	6	7	8	9	10	11	12	13
V a g u e	1				■	■	■							
	2		■											
	3					■								
	4			■										
	5						■							
	6		■											

Secteur B

		Semaine												
		1	2	3	4	5	6	7	8	9	10	11	12	13
V a g u e	1										■	■	■	
	2									■				
	3											■		
	4										■			
	5												■	
	6													■

Enfin la répartition est étudiée pour éviter qu'un enquêteur ait deux semaines de référence consécutives correspondant à une charge « visite ». La collecte est espacée de deux ou trois semaines entre deux charges « visite » (sur l'exemple ci dessus, les charges « visite » sont de couleur verte pour la 1ère interrogation, rose pour la dernière)¹⁷.

Au final, lors d'une semaine de référence donnée, la collecte a lieu sur 12 SAE, comme l'illustre la carte ci dessous¹⁸. La collecte reste « dispersée dans l'espace » mais est « regroupée au mieux dans le temps ».



¹⁷ Le rapport final Maiol UEM DOM, note d'octobre 2012 N° 835/DR974-SES, examine cette question en détail

¹⁸ Illustration sur la semaine 1 du T2 2014

4.6. Présentation de la base de sondage

Comme nous l'avons dit plus haut, la base de sondage constituée chaque année au Criem contient les cinq dernières EAR auxquelles on adjoint un certain nombre d'informations utiles pour le tirage, notamment le découpage en SAE et en strates décrit ci-dessus. Cette base de sondage est dite « complète ». Depuis la mise en place du nouveau dispositif d'échantillonnage, les tirages ont été faits dans l'ensemble du cycle d'EAR dans les petites communes et dans les 2 dernières EAR pour les grandes communes. Cette base de sondage est dite « réduite ».

Chaque EAR est mise à disposition en décembre de l'année où elle a été collectée. Au 1^{er} trimestre, le Criem peut ainsi constituer la base de sondage « complète ». Début 2015, la base de tirage dite "RP 2012" a donc été constituée, empilant les EAR 2010 à 2014¹⁹. Cette base est utilisée pour le tirage des enquêtes ménages sur le terrain entre avril 2015 et mars 2016.

La base de tirage RP 2012, est constituée :

- des 4/5^{ème} de la base de tirage de l'année précédente, alias base de tirage RP 2011, ou plus exactement des logements recensés entre 2010 et 2013 présents dans la base RP 2011,
- des logements recensés en 2014 (EAR 2014).

Les étapes de construction de la base sont ici résumées.

a) Sélection de la nouvelle EAR, ici l'EAR 2014.

On ne prend, dans les fichiers de logements EAR, que ceux qui correspondent à des résidences principales de logements ordinaires.

Dans les DOM, certaines communes sont hors champ des enquêtes ménages, car difficiles d'accès et peu peuplées. On exclut les îles guadeloupéennes des Saintes et de la Désirade, ainsi que sept petites communes dans l'intérieur guyanais : Ouanary (97314), Saul (97352), Maripasoula (97353), Camopi (97356), Grand-Santi (97357), Saint-Élie (97358) et Papaïchton (97362). D'après le RP de 2009, ces restrictions correspondent à 1.1% des résidences principales en Guadeloupe et à 7.2% en Guyane.

Dans les étapes b) à e), la nouvelle EAR est mise au format requis par les besoins de l'échantillonnage.

b) Ajout des informations utiles issues de sources externes, infra communales.

On marque les îlots trop éloignés et peu peuplés que l'on ne souhaite pas "visiter", à nouveau parce qu'ils sont peu peuplés et difficiles d'accès. C'est le cas de certaines zones de forêt en Guyane ou de montagne à La Réunion. On ajoute l'information sur les ZUS (zones urbaines sensibles) pour le cas d'enquête à sur échantillonner selon ce critère ainsi que pour réviser à la hausse les temps de collecte des FA localisées en ZUS. On ajoute le code postal pour les envois postaux, notamment de l'EEC²⁰.

c) Ajout d'information d'adressage, issues de la base cartographique

L'objectif de cette étape est d'ajouter des informations d'adressage tirées des enquêtes cartographiques menées avant chaque collecte du recensement. La proportion d'adresses non

¹⁹ La base de tirage RP 2012 n'est pas strictement identique à la base du RP 2012 diffusée. En effet, avant de diffuser les données du RP, non seulement les données de 5 EAR sont empilées mais d'autres traitements sont effectués, notamment sur les pondérations.

²⁰ Pour la Guyane et les Antilles, on considère qu'il y a un seul code postal par commune, même si ce n'est pas toujours le cas, comme par exemple pour les Abymes (97401). A La Réunion, certaines grandes communes sont découpées en plusieurs codes postaux, dont les périmètres peuvent être reconstitués par regroupement d'Iris.

normalisées restant importante dans les DOM (exception de La Réunion), il a été jugé indispensable au moment du basculement à des tirages dans le recensement, de conserver la possibilité d'une identification cartographique du logement à enquêter. Grâce à un travail coordonné avec les équipes SIG des DOM, les enquêteurs continuent donc de disposer de cartes (plans papier pour les DFA, carto embarquée pour La Réunion) pour faciliter – voire permettre – le repérage.

Ces bases cartographiques, une par petite commune recensée l'année en question et une base globale pour l'ensemble des grandes communes, sont disponibles en novembre de l'année précédent la collecte du recensement, soit en novembre 2013 pour l'EAR qui débute en janvier 2014. Les communes sont traitées différemment selon leur taille.

Pour les grandes communes, le rang d'adresse issu de la carto a été conservé dans les fichiers des EAR, la fusion avec le recensement peut donc se faire sans difficulté avec l'identifiant cartographique (commune - ilot - rang d'adresse dans l'ilot).

Pour les petites communes, les adresses sont malheureusement renumérotées au moment du recensement. Il est donc nécessaire d'avoir une « table de passage » entre les anciens identifiants (issus de la cartographie) et les nouveaux (affectés par le recensement). Ces tables de passage sont construites par le Criem à partir des fichiers Suivi de Collecte RP (SCRIP). La pratique a montré qu'ils sont de bonne qualité.

d) Ajout des variables de stratification et de SAE

e) Disjonction

L'Insee, notamment sous l'impulsion d'Eurostat et de son Code de bonnes pratiques, a pour habitude d'éviter un re-tirage rapproché de logements, un intervalle de cinq ans étant considéré comme acceptable. Il s'agit donc ici de marquer les logements recensés en 2014 qui ont été enquêtés entre 2010 et 2014.

Les logements les plus anciens de l'ancienne base de sondage (ici ceux recensés en 2009), vont être remplacés dans la nouvelle base de sondage par les logements nouvellement recensés (ici en 2014). Ces logements appartiennent au même groupe de rotation du recensement. On peut donc reporter les marquages de l'EAR 2009 sur l'EAR 2014. Pour les adresses collectives, la disjonction est sûrement peu fiable au niveau du logement car les logements ont pu y être numérotés de manière différente à cinq années d'intervalle.

Les variables de marquage utilisées sont les suivantes :

- UTILISE prend 4 valeurs
 - 0 si le logement n'a pas été tiré
 - 1 s'il a été tiré dans le RP
 - 2 s'il a été tiré dans la carto
 - 5 s'il a été tiré mais non enquêté : cas des enquêtes sur champ ciblé ou des enquêtes avec surreprésentation. Pour ne pas détériorer la base de tirage, on tire un échantillon à probabilités égales et on « jette » la partie de l'échantillon qui ne correspond pas au champ de l'enquête. Exemple Conditions de Travail 2012 : des logements n'abritant que des inactifs de plus de 60 ans ont été tirés mais pas enquêtés.
- ECHENQ : les 2 premières positions correspondent à l'année de l'enquête et les suivantes au code de l'enquête. Seuls les marquages datant de moins de 5 ans sont conservés.

f) Empilement. On empile l'EAR 2014 avec les EAR 2010 à 2013 présentes dans la base RP2011.

g) Une dernière étape programme est nécessaire pour la Guadeloupe pour le moment, seul DOM concerné par l'enquête Loyers et Charges.

La méthode d'échantillonnage utilisé pour Loyers et Charges est décrite de façon détaillée au paragraphe 2.2.3. Chaque trimestre, onze UP entrantes sont tirées à probabilités égales par micro-région, et équilibrées sur le nombre de logements du champ. Pour ce faire, on crée dans un premier temps une base d'unités primaires qui servira au tirage du premier degré. Le tirage du premier degré (unités primaires entrantes dans chacun des quatre sous-échantillons de l'année) est réalisé dès la constitution de la base de tirage générale de la Guadeloupe.

4.7. Présentation de la procédure d'échantillonnage

Le Criem dispose d'un processus et d'un modèle de programmation à utiliser pour chaque tirage dans le recensement. Le tirage de l'enquête emploi en continu (EECDOM) 2013 a fourni un premier squelette, qui a été retravaillé lors des tirages de Conditions de Travail 2012, de l'EEC 2014 et 2015, de Logement 2013-2014, d'IVQ Martinique 2014 et enfin de CVS Antilles et Guyane 2015.

Comme pour la constitution de la base de sondage, les étapes du tirage sont ici résumées.

a) Préparation de la base de tirage.

On choisit tout d'abord le nombre d'EAR à mobiliser en grande commune, les 2 dernières dans la grande majorité des cas jusqu'à présent.

On élimine de la base de tirage les logements difficiles d'accès. S'ils ne représentent qu'une part négligeable pour La Réunion, en Guyane en revanche, cette exclusion concerne 9% des logements de la base RP usuelle.

Enfin, on élimine de la base de tirage les logements enquêtés au cours des 5 dernières années. Ces suppressions résultent de tirages à probabilités égales (pour une enquête donnée, les logements tirés ont des probabilités d'inclusion dans l'échantillon proches), ce qui ne déséquilibre pas la base de tirage "en moyenne". Cependant, l'enquête Loyers et Charges (L&C) en Guadeloupe et certains tests ne sont pas issus de tirages à probabilités égales :

- pour L&C, seuls des logements a priori occupés par des locataires sont tirés ;
- pour certains tests, on tire sur des catégories de population ciblées.

Les volumes de ces échantillons sont faibles, ils représentent autour de 2% de la base usuelle pour L&C en Guadeloupe par exemple. Plutôt que de prendre le risque de retirer des logements déjà enquêtés - et donc de s'opposer à un refus légitime de la part de l'enquêté -, ils sont actuellement enlevés de la base de tirage et des contrôles sont faits sur la proportion de locataires dans les échantillons tirés en Guadeloupe. D'autre part, pour limiter le déséquilibre de la base de tirage, il a été décidé de tirer dès la constitution de la base de tirage le premier sous-échantillon qui doit être fourni rapidement, mais le tirage des 3 autres sous-échantillons est décalé, pour être réalisé au maximum après le tirage des autres enquêtes de l'année.

Cette suppression des logements déjà enquêtés est loin d'être anodine sur ces petits territoires où les taux de sondage sont très élevés : en Guyane, plus de 20% des logements de la base de tirage usuelle sont supprimés pour cette raison, et autour de 10% dans les trois autres DOM.

b) Un tirage stratifié

Pour toutes les enquêtes (hors tests), le tirage est un tirage stratifié. On tire proportionnellement à la taille de la strate dans la micro-région.

Les allocations par strate sont calculées de la manière suivante :

- Calcul du nombre de logements que l'on souhaite obtenir dans chaque SAE (n_{sae}) = taille de l'échantillon / 26, que l'on arrondit à l'unité la plus proche ;
- Les allocations sont d'abord calculées par microrégion, proportionnellement à la taille de la microrégion au recensement de 2008. Puis elles sont arrondies au multiple de n_{sae} , nombre de logements que l'on souhaite obtenir dans chaque SAE.
- Ces allocations par microrégion sont ensuite distribuées par strate en fonction de leur taille dans la microrégion.

Les arrondis effectués pour calculer ces allocations ne sont pas toujours anodins. Prenons l'exemple de l'enquête Cadre de Vie et Sécurité de 2015 en Martinique.

- Il était demandé un total de 3 000 logements. Dans la pratique, cette taille est ramenée à 2 990 pour obtenir un multiple de 26. On souhaite obtenir 115 logements dans chaque SAE.

- Les allocations sont d'abord calculées par microrégion. La Martinique est découpée en 6 micro-régions, de tailles très inégales. Celle du Centre-Agglomération (zone autour de Fort-de-France) est la plus grande, elle représente 44% des résidences principales martiniquaises d'après le recensement de 2008. Les 2 micro-régions du Nord de la Martinique ne contiennent elles que 6% des résidences principales pour la zone Nord-Caraïbes et 5% pour la zone Nord-Atlantique. Si l'on tirait strictement proportionnellement à la taille de la micro-région, on devrait tirer 1316 logements dans la zone de Fort-de-France (44% de 3000 logements). Comme on l'a vu plus haut, ces chiffres sont en réalité arrondis au multiple de 115, et on y a donc tiré 1265 logements. L'effet de ces arrondis est encore plus impressionnant sur les 2 plus petites micro-régions : on a tiré 230 logements dans la zone Nord-Caraïbes et 115 dans la zone Nord-Atlantique, ce qui laisserait penser que la première est deux fois plus peuplée que la deuxième !

L'effort visant à obtenir une répartition égale par SAE aboutit ici à une sur-représentation de la zone Nord-Atlantique au détriment de la zone Nord-Caraïbes. Le jeu n'en vaut peut être pas la chandelle, nous y reviendrons plus tard (§5.2).

Le poids final de chaque logement tiré est :

$$POIDS_FINAL = \frac{Nbre_résidences_princ_ds_strate_au_RP2008}{Nbre_Logts_tirés_ds_strate}$$

Ce sont actuellement les chiffres du recensement de 2008 qui sont utilisées pour calculer allocations et poids du logement : la taille des strates de tirage est totalement indépendante de la base de tirage, et fortement liée à l'image des territoires en 2008. Nous y reviendrons plus tard (§5.2).

c) Dans chaque strate, un tirage systématique

Le tirage dans chacune des strates est un tirage systématique à probabilités égales. Tout tri sur une variable du recensement liée aux variables d'intérêt de l'enquête considérée est possible pour améliorer la précision de l'échantillon. Pour l'EEC par exemple, chaque strate de la base est triée par nombre de chômeurs du ménage puis par nombre d'actifs occupés du ménage. Pour l'enquête Logement 2013-2014, par type de construction (individuel / collectif), aspect du bâti (matériau en dur / cases traditionnelles) et surface. Pour l'enquête IVQ Martinique 2014, par diplôme puis âge de la personne de référence. Pour l'enquête CVS Antilles-Guyane 2015, par type de logement (individuel / collectif) et diplôme de la personne de référence.

Le pas de tirage correspond au poids du logement dans la base de tirage :

$$Pi_tirage = \frac{Nbre_Logts_tirés_ds_strate}{Nb_Logts_strate_ds_base_de_tirage_}$$

d) Marquage des adresses de l'échantillon dans la base générale de l'année

4.8. Les logements neufs

Le nouveau dispositif d'échantillonnage, déjà assez abouti, présente des limites. La principale, c'est qu'il ne prend pas en compte les logements neufs ; l'utilisation des deux EAR les plus fraîches limite le manque de logements neufs pour les grandes communes, mais dans les petites, le déficit est plus difficilement négligeable quand on utilise des données non actualisées depuis cinq ans.

Cette question a deux facettes : d'une part il faut *trouver* des logements neufs échantillonnables ; d'autre part, il faut savoir *quel poids de sondage leur affecter*, et c'est là qu'est en fait l'enjeu du questionnement.

La question a été étudiée en 2012/2013 au sujet de l'enquête « Logement », un protocole spécifique étant en place en métropole pour prendre en compte les logements « très neufs » alors même que le dispositif tout juste mis en place pour l'Outre-mer était déficient de ce point de vue.

4.8.1. La question des logements neufs : le tirage

Cette question des logements neufs n'est pas anodine dans le contexte de l'Outre-mer. En effet, la dynamique de construction est beaucoup plus élevée qu'en métropole. La part annuelle des logements neufs est ainsi de l'ordre de 2% en Guadeloupe et en Martinique (soit deux fois plus qu'en métropole) et de 3% en Guyane et à La Réunion (soit trois fois plus qu'en métropole).

En Guadeloupe, en Martinique et à La Réunion, des tirages de logements neufs sont possibles, en utilisant les fichiers issus du rapprochement de Sitadel et des fichiers mensuels de la Direction Générale des Impôts (DGI)²¹. Ces fichiers sont de qualité satisfaisante dans les DOM et constituent une source de tirage de logements neufs utilisable.

A noter que cette source n'est disponible que depuis janvier 2010 en Martinique et à la Réunion et janvier 2011 en Guadeloupe. Pour l'ENL, tirée dans la base de tirage RP 2010, il a fallu trouver une autre source de tirage pour les logements neufs des petites communes recensées entre 2008 et 2010 (2011) : ce sont les fichiers de la taxe d'habitation qui ont dû être mobilisés.

Les sources fiscales (Sitadel-DGI ou TH) ne sont actuellement pas mobilisables en Guyane. Pour l'heure, la seule façon de repérer des logements nouvellement construits est d'opérer par différentiel de cartographie. Un premier travail de ce type avait été prévu pour l'enquête « Logement » : un échantillon d'une vingtaine de logements neufs devait être ajouté en cours de collecte. Cet ajout a été abandonné, non pas parce qu'il était difficile de créer cet échantillon mais parce que l'enjeu a été jugé mineur face aux grosses difficultés de collecte rencontrées.

4.8.2. La question des logements neufs : le calage

La taxe d'habitation, utilisée pour le calage des données métropolitaines, n'est pas encore suffisamment fiable pour constituer la source de calage sur les logements neufs en Guadeloupe, mais aussi en Martinique et à La Réunion, car elle s'avère fragile sur les logements achevés au cours des trois dernières années. Pour le calage de l'ENL, plusieurs sources ont été comparées pour constituer la marge sur la date de construction des logements :

- Les fichiers de la Taxe d'habitation ;
- Les données sur les Permis de construire ;
- L'évolution du parc de logements selon le recensement.

Les marges retenues suite à cette comparaison sont les suivantes :

	Guadeloupe		Martinique		Guyane côtière		La Réunion	
TOTAL	175 103		171 334		65 544		312 767	
dt const. Avant 2009	158 468	90,50%	155 914	91,00%	57 679	88,00%	275 235	88,00%
dt const. de 2009 à 2013	16 635	9,50%	15 420	9,00%	7 865	12,00%	37 532	12,00%
Constr annuelle récente	3 327		3 084	%	1 573		7 506	

D'autres pistes pourraient être explorées, en particulier l'exploitation de fichiers de distributeurs en réseau (eau, électricité). Un nombre de nouveaux raccordements aux réseaux dans l'année doit fournir une estimation relativement fine du nombre de logements neufs « habitables ».

²¹ depuis la mise en place de la RGPP, le SOES reçoit mensuellement de la DGI et de façon centralisée la liste des locaux achevés dans le mois précédent (achevés donc pas forcément occupés). Les impôts récupèrent en contre partie toutes les autorisations de construire mensuelles pour envoyer leurs agents vérifier sur le terrain (en vue de la taxation future). Plus exactement, la DGI renvoie au SOES la liste des identifiants de permis qu'ils constatent achevés mois après mois permettant de faire le lien avec Sitadel2 pour récupérer les adresses et autres informations.

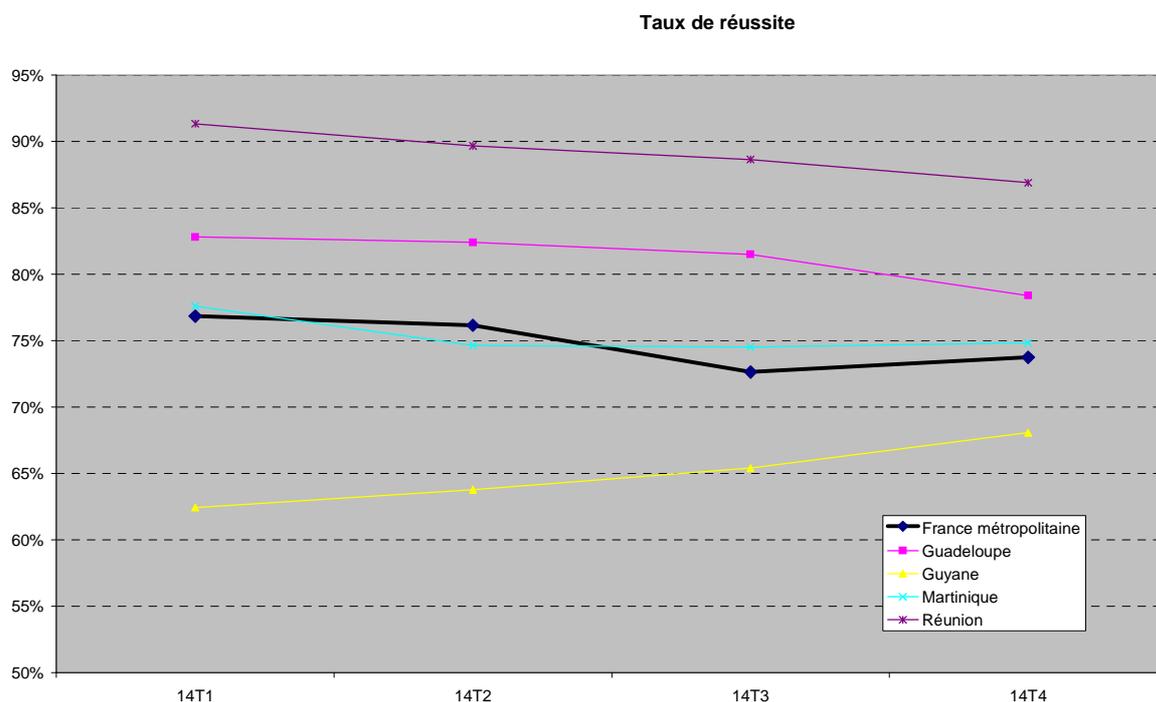
5. Les premières années du nouveau dispositif et pistes d'amélioration

5.1. Deux ans plus tard, quel bilan des SAE et du tirage dans le recensement ?

Début 2015, soit 2 ans après le lancement de l'EEC et la mise en place du NCEE, quel est le bilan du nouveau système d'échantillonnage ?

5.1.1. Bilan sur l'EEC

Après une phase de montée en charge progressive en 2013, l'EECDOM a atteint depuis le 2^{ème} trimestre 2014 son fonctionnement courant. Comparativement à la métropole, le taux de réussite²² en 2014 s'avère meilleur pour la Guadeloupe et La Réunion, La Réunion étant même depuis son introduction en tête du palmarès régional. Le taux de réussite martiniquais équivaut celui de la métropole. En revanche, les résultats guyanais sont en deçà. Le T4 2013 a été marqué par une surcharge importante et plusieurs grappes ont été abandonnées. Ce trimestre catastrophique a eu des répercussions sur le début d'année 2014, et la situation s'est nettement améliorée tout au long de l'année 2014.

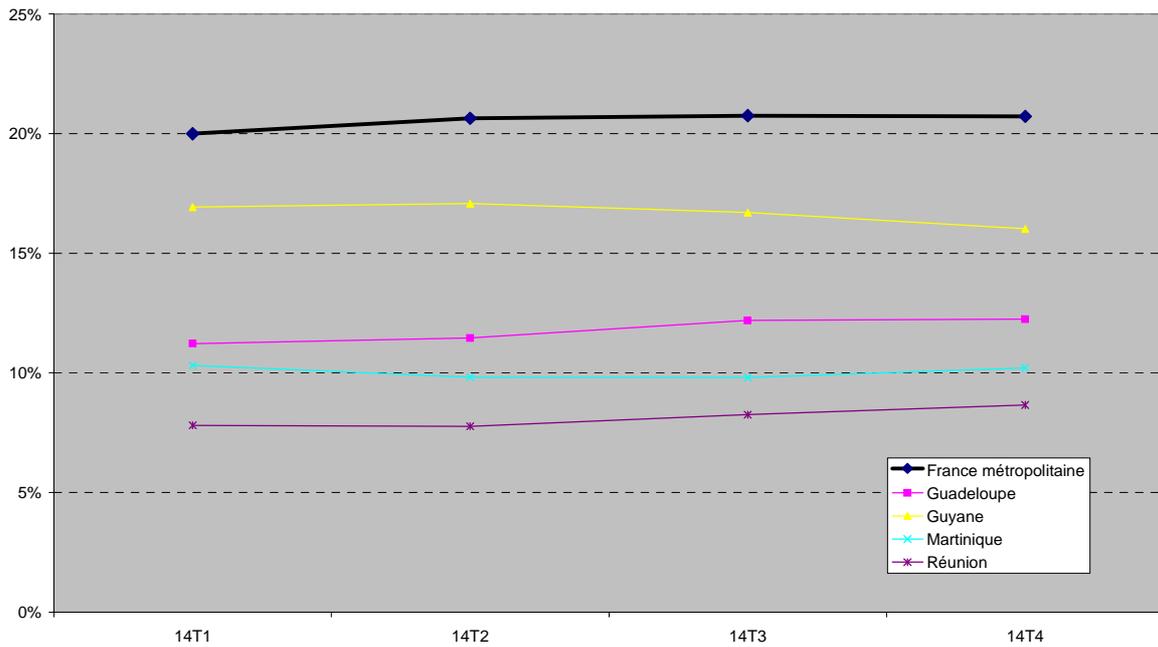


Le taux de hors champ²³ est sans surprise bien inférieur dans les DOM. En effet, seuls des logements recensés en tant que résidences principales sont sélectionnés dans les échantillons ultramarins, alors que le tirage métropolitain dans la taxe d'habitation sélectionne sur un champ plus large.

²² taux de réussite = nombre de FA réalisées / nombre de FA dans le champ (résidences principales)

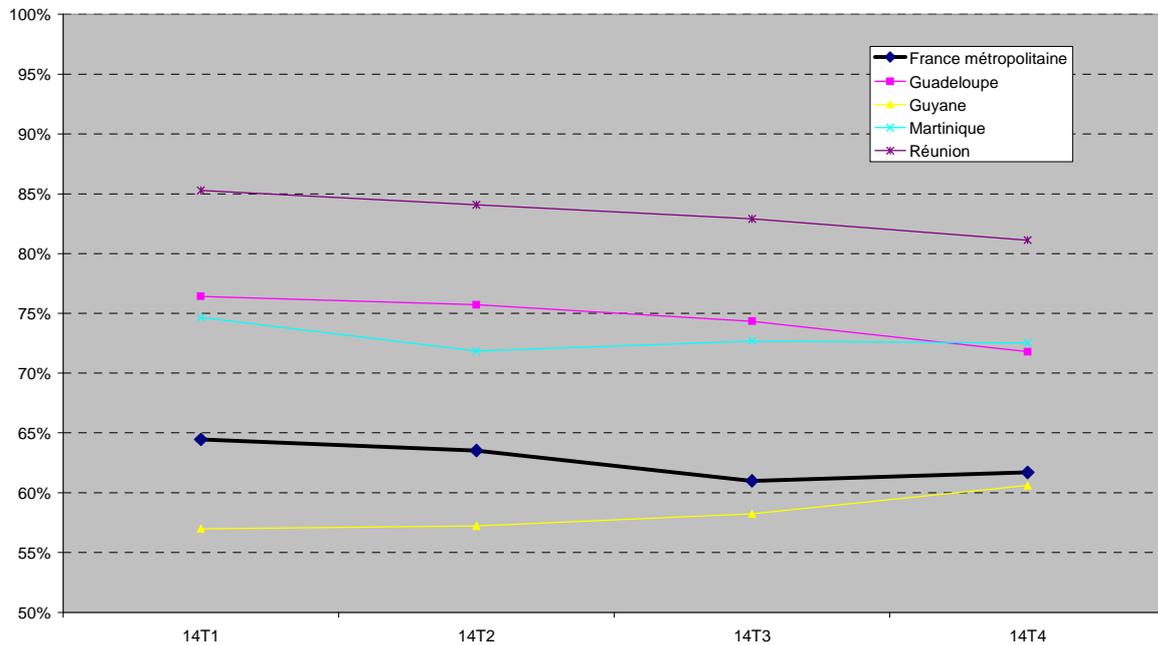
²³ taux de hors champ = nombre de FA hors du champ (logements vacants, résidences secondaires...) / nombre de FA total collectées

Taux de hors champ



C'est ce qui explique que le taux de collecte²⁴ guyanais réussit quasiment à atteindre le niveau métropolitain.

Taux de collecte



Du point de vue de gestion de la collecte, à La Réunion et en Guadeloupe, le bilan est largement positif. Les enquêteurs sont maintenant familiarisés avec leur secteur de collecte, ce qui leur facilite le repérage. La concentration sur 8 semaines consécutives de la charge EEC (les 6 semaines de référence + les 2.5 semaines de collecte) permet aux gestionnaires de fournir aux enquêteurs un planning annuel standard et cette visibilité facilite la planification des congés, par les gestionnaires comme par les enquêteurs. Un petit bémol toutefois : les enquêteurs travaillant sur les

²⁴ taux de collecte : nombre de FA réalisées / nombre de FA total collectées

SAE actifs en fin de trimestre ne peuvent aisément poser de congés pour les Fêtes de Noël. D'autre part, ce « trou » de 5 semaines chaque trimestre permet aussi aux enquêteurs de limiter la lassitude d'une enquête répétitive. Concernant la multi-enquête, l'effet de la concentration de la charge EECDOM est mitigé : positif pour la gestion des enquêtes sur longue-durée (l'enquêteur collecte essentiellement de l'EECDOM pendant 8 semaines, puis de la 2^{ème} enquête les 5 semaines suivantes), négatif pour les enquêtes de courte durée (l'enquêteur ne peut attendre pour débiter la collecte de la 2^{ème} enquête et doit donc ajouter à sa charge EECDOM déjà importante une charge supplémentaire).

Enfin, les remplacements, même s'ils restent des opérations chronophages, sont effectivement facilités par l'alternance des périodes d'activité sur l'EEC entre 2 SAE voisins.

En Martinique, où le nombre d'enquêteurs affectés à l'EEC est de 23 enquêteurs, soit bien moins que le nombre de 26 nécessaire à l'optimisation du système (voir §3.3), le bilan est beaucoup moins positif. Chaque trimestre, le BEM doit résoudre un casse-tête pour affecter les FA des 3 SAE dits orphelins. A La Réunion, seul 1 SAE est orphelin. Une gestion routinière s'est mise en place : les 6 enquêteurs les plus proches du SAE orphelin récupèrent chacun la charge correspondant à l'une de ses 6 semaines de référence. Pour autant, la situation est déjà jugée inconfortable.

Pourquoi donc seule la Guadeloupe a effectivement un enquêteur sur chacun des SAE pour l'EEC ? Quelques années avant la mise en place du NCEE, les réseaux enquêteurs des DOM étaient bien supérieurs à 26. 2010 et 2011 notamment ont été des années chargées, avec la grosse enquête « Migrations – Famille – Vieillesse », 4500 FA / DOM, qui avait donné lieu à des recrutements importants. C'est certainement pourquoi à la mise en place du NCEE, ce seuil minimal de 26 enquêteurs pour l'EECDOM n'a pas été identifié comme un risque à éviter, alors même que le mot d'ordre a été d'augmenter les petites quotités (<50 %) afin d'éviter de laisser un enquêteur sans protection sociale.

En Guyane par exemple, 28 enquêteurs ménages ont été recrutés, notamment pour être sûr de pouvoir affecter chaque enquêteur à un seul SAE. Mais les aléas du réseau, des départs et la volonté d'avoir des quotités entre 50 et 70% ont réduit la taille du réseau. Cette situation devrait rapidement s'améliorer, des recrutements sont en cours pour aboutir à 25 enquêteurs ménages et un seul SAE orphelin, à Cayenne, qu'il est assez facile de redistribuer. A La Réunion, un recrutement est aussi envisagé rapidement pour retrouver 26 enquêteurs affectés à l'EECDOM.

5.1.2. Quelle utilisation des SAE pour la répartition – enquêteurs des autres enquêtes ?

Le premier échantillon à avoir été transmis aux UEM avec le découpage par SAE a été Conditions de Travail (CdT) en 2012. Il a été suivi par l'enquête nationale sur le Logement (ENL) en 2013, Information et Vie Quotidienne (IVQ) Martinique en 2014 et tout récemment par Cadre de Vie et Sécurité (CVS) dans les DFA en 2015.

L'utilisation des SAE pour la répartition par enquêteur varie selon les DOM.

A part en Guadeloupe, une première répartition est réalisée avec le SAE. Des ajustements sont ensuite réalisés avec les SAE voisins pour que la quotité affectée à chaque enquêteur soit respectée. Contrairement aux ZAE métropolitaines, aucun SAE n'est isolé. Les frontières des secteurs de collecte peuvent ainsi être ajustées à chaque enquête tout en conservant des déplacements enquêteurs raisonnables. Pour l'enquête Patrimoine de 2015, le tirage a eu lieu comme en métropole dans la source fiscale, rendant impossible la livraison d'un échantillon réparti par SAE. La DEM de La Réunion s'est rapprochée de l'équipe SIG locale pour enrichir son échantillon du découpage en SAE, afin de faciliter son travail de répartition enquêteur. Cet exemple montre à quel point le découpage par SAE est maintenant intégré comme secteur de collecte standard sur ce territoire.

En Guadeloupe, la répartition des enquêtes hors EECDOM et Loyers et Charges (enquête dont l'organisation est très spécifique, voir §2.2.3) ne s'appuie pas sur le SAE, mais directement sur la commune, puis sur l'îlot pour les communes sur lesquels plusieurs enquêteurs vont travailler. Le BEM

a construit un tableau d'assemblage commune-ilot et enquêteurs, lui permettant d'affecter à chaque enquêteur sa zone de collecte habituelle. Par exemple, l'échantillon de CVS va être réparti sur 6 ou 7 enquêteurs sur la commune Des Abymes, alors que cette commune n'est découpée qu'en 4 SAE.

5.1.3. Quelle utilisation des SAE pour la planification des enquêtes ?

Cette utilisation des SAE pour la répartition enquêteur est la même lorsque l'UEM travaille sur les échantillons livrés dans l'outil de gestion de la charge des enquêteurs. Rappelons que depuis la mise en place du NCEE en 2013, les UEM utilisent un outil (« Chester » puis « Opale ») pour anticiper les charges d'une année donnée ou assurer la gestion courante des charges. Cet outil est livré en septembre N-1, et contient les échantillons des enquêtes qui seront sur le terrain au cours de l'année N. Ces échantillons sont définitifs si leur collecte démarre lors du 1^{er} semestre de l'année N ; et sont prévisionnels sinon.

Les échantillons prévisionnels n'ont pas encore été tirés car les bases de sondage correspondantes ne sont pas encore constituées, dans l'attente de l'EAR la plus récente. Dans ce cas, les tirages sont simulés en se basant sur une répartition parfaitement équilibrée entre les SAE. Quand le SAE est une commune ou une fraction de commune, l'estimation de charge sur ce SAE est immédiate : elle est le rapport du nombre fiches-adresses de l'enquête par le nombre de SAE (26). Quand un SAE regroupe plusieurs communes (cas des petites communes), le recensement de 2009 a été exploité pour donner la part de chaque commune (en terme de résidences principales) à l'intérieur du SAE : l'estimation de la charge sur le croisement SAE x commune s'en déduit alors immédiatement (au pro rata)²⁵.

Cette estimation est particulièrement grossière, surtout pour les grandes communes.

Pour les petites communes, l'introduction d'une nouvelle EAR peut modifier la part d'une commune à l'intérieur du SAE qui la contient. L'estimation pourrait donc être fragilisée par une évolution importante d'une petite commune, mais un impact fort est assez improbable.

Pour les grandes communes par contre, le déséquilibre des groupes de rotation et leurs traces inégales sur les SAE peut générer un impact fort lors de l'introduction d'une nouvelle EAR dans la base de sondage, si bien que l'estimation de charge fournie (vraie « en moyenne » sur cinq ans) peut être assez éloignée de la « vraie » charge annuelle.

Il est inévitable que lors de la livraison de l'échantillon définitif, un SAE contienne plus, ou moins, de FA qu'annoncé. Même si le mode de tirage est conçu pour aboutir à des charges équilibrées par SAE, on a vu plus haut que pour les charges EECDOM de 13 logements par SAE en théorie, on varie en pratique entre 9 et 17. Prenons un deuxième exemple, celui d'IVQ en Martinique : l'échantillon Chester livré en septembre 2013 était prévisionnel, chaque SAE contenait 96 ou 97 (2500 / 26) logements. Au final, le nombre de logements par SAE variait en réalité entre 77 et 118.

En métropole, des variations de cet ordre entre l'échantillon prévisionnel et définitif n'ont pas été jugées acceptables, à cause de la présence de ZAE isolées, et de l'impossibilité de transférer une partie des FA d'un enquêteur à un enquêteur « voisin » sans augmenter considérablement les temps de déplacement. Il a donc été décidé d'introduire en allocations de tirage le nombre de FA par ZAE présent dans l'échantillon prévisionnel, même si ces allocations ne sont plus complètement optimales.

Dans les DOM, cette dégradation de l'échantillon d'origine a pour le moment été évitée, car les ajustements peuvent être faits au moment de la réception de l'échantillon définitif pour que la zone de collecte (élargie ou rétrécie par rapport aux délimitations du SAE) contienne un nombre de logements compatible avec les quotités des enquêteurs. Ces ajustements sont gérés par les UEM comme ceux liés à un congé maladie par exemple.

Ce réajustement des frontières des SAE suivant les enquêtes et les plannings des enquêteurs a pour le moment aussi permis d'éviter l'utilisation de la procédure de sous-allocation en cas d'absence

²⁵ Concrètement, seules les variables SAE et commune sont mises à disposition dans les échantillons prévisionnels, pas la variable IRIS.

d'un enquêteur en début de collecte, procédure mise en place par la Division Sondages. Ici aussi, si un SAE se retrouve « orphelin » en début de collecte, l'UEM a la possibilité de revoir à la hausse la répartition des enquêteurs voisins, sans abandonner de FA, et donc sans dégrader la précision de l'échantillon. Cette opération reste toutefois chronophage pour les gestionnaires, et certainement inconfortable pour les enquêteurs s'ils sont avertis tardivement d'un alourdissement de leur charge.

Un travail pourrait être mené afin d'anticiper ces changements d'équilibre si cela s'avérait trop contraignant en terme de gestion.

- En petites communes, où la plupart des tirages se font dans 5 EAR, il s'agirait « d'actualiser » les volumes des croisements SAE x commune de la plus ancienne des EAR (qui sera remplacée l'année suivante par la nouvelle EAR qui viendra compléter la base de sondage) en utilisant les différentiels entre les cartographies n-5 et n ; ces différentiels fournissant ainsi une approximation de l'évolution du parc de logements du croisement SAE x commune.
- En grandes communes, où la plupart des tirages se font dans les 2 dernières EAR, une méthode analogue à celle utilisée en métropole pourrait être utilisée, en estimant cette fois les volumes des SAE infra-communales de l'EAR à intégrer grâce aux volumes de l'EAR plus ancienne correspondant au même groupe de rotation.

5.1.4. Un repérage facilité

A l'unanimité, gestionnaires comme enquêteurs font le constat d'un repérage facilité avec le passage au tirage dans le recensement. A part en Guyane, l'utilisation des cartes (plans papier pour les DFA, carto embarquée pour La Réunion) est de moins en moins systématique, même si les gestionnaires continuent à les mettre à disposition des enquêteurs. L'envoi de lettres-avis par courrier est dorénavant possible.

Il a cependant fallu un temps d'adaptation pour intégrer les nouvelles consignes de repérage. Les premiers tirages dans le recensement ont eu lieu en Guadeloupe. Pour IVQ en 2008, Patrimoine en 2009 et Loyers et Charges depuis 2010. Depuis 2012, ils sont devenus la norme et ont pris le pas sur les tirages dans la base cartographique. Les consignes de repérage ont évolué avec le changement de base de sondage. Enquêteurs comme gestionnaires avaient pris l'habitude de centrer le repérage sur la localisation d'un bâti sur une carte, grâce au code îlot puis aux rangs d'adresse et de logement. Avec l'arrivée des tirages dans le recensement, tous se sont demandé s'ils devaient dorénavant axer le repérage plutôt :

- a) sur le repérage « physique » c'est-à-dire îlot + rang d'adresse physique dans l'îlot (issu du Système d'Information Géographique) + rang du logement issu de la règle de numérotation censée prévaloir auprès des agents recenseurs et des enquêteurs ménages ?

INFORMATIONS DANS LES BASES DE SONDEGE										
Département et commune								Année de collecte		
BRAS-PANON (LA REUNION)								2011		
RG	DOMAINE	ZAE	DEP	COM	CIL	IRIS ou îlot	FIL	Rg Adresse	Rg Logement	Zus
04	T	Z00032	974	402	—	A118	—	042	009	Non

- b) sur les informations du logement recensé, issues du scan de la feuille logement du recensement ?

Nom de l'occupant principal..... **BOURGIN Anne-Aimée**

Localisation du logement

Numéro **4** Bis, Ter... Rue, Av, Bd... Nom de la voie (ou lieu-dit) **impasse des améthystes**

Résidence, hameau, lotissement... **Résidence Vanillier** Bâtiment **Bât D**

Escalier **0 2 appartement 35** N° de porte ou position par rapport à l'escalier (exemple : couloir de droite, 1^{ère} porte à gauche)

Code postal **97412** Bras-Panon Commune (et arrondissement pour Paris, Lyon et Marseille)

Continuez !

Si en métropole, le repérage est centré sur les informations du logement recensé (b), des discussions ont été nécessaires pour se convaincre de la pertinence de faire le même choix pour les DOM. Les arguments favorables à l'uniformisation des pratiques de repérage sur celles de métropole ont été les suivants :

- Le tirage est réalisé d'après les caractéristiques du logement recensé, ce qui a encore plus d'importance pour les enquêtes ciblées sur une population particulière. Il s'agit de retrouver sur le terrain le logement tiré pour conserver les qualités statistiques de l'échantillon tiré.
- Les informations apparaissant sur le scan de la FA papier sont souvent suffisantes pour retrouver le logement tiré (un peu moins vrai dans les DFA qu'à La Réunion).
- Enfin, les conséquences du choix b) sur la disjonction sont limitées. Dans les grandes communes, il peut y avoir risque de retraitage si une adresse a été recensée dans le mauvais groupe de rotation. Ce risque paraît faible dans la mesure où cela implique que l'agent recenseur a fait un « franchissement d'îlot » (il a été recensé dans un îlot qui n'avait pas à être recensé cette année-là)²⁶.

Toutefois, dans les cas où les informations du scan seraient insuffisantes pour identifier le logement à enquêter, il est demandé aux enquêteurs ménages de se baser sur le repérage « physique ».

Le repérage « physique » conserve évidemment sa pertinence, notamment lors de la première phase de repérage, en permettant aux enquêteurs d'identifier rapidement le « pâté de maisons » dans lequel figure le logement à repérer.

Cette décision d'axer le repérage des logements tirés dans le recensement sur les informations issues du scan signifie que l'enquêteur doit aller chercher le logement recensé, même si l'agent recenseur a fait une erreur :

- soit qu'il ait recensé un logement qui n'avait pas été tiré au sondage EAR dans les grandes communes (le rang d'adresse carto ne correspond pas au libellé d'adresse recensé, l'agent recenseur a recensé le logement voisin du logement tiré par exemple)
- soit qu'il se soit trompé en écrivant l'adresse (le nom de l'occupant est juste mais il a écrit sur la feuille logement RP n°12 au lieu de n°14).

Les enquêteurs des enquêtes ménages DOM sont aussi souvent des enquêteurs carto, et il a été difficile au début de les convaincre de ne pas corriger l'erreur faite par l'agent recenseur le cas échéant. 2 ans plus tard, ces consignes sont totalement intégrées en Guadeloupe et à La Réunion, un peu moins en Martinique et en Guyane.

²⁶ Toujours dans les GC, la disjonction est assurée au sein des logements d'un immeuble (collectif) car l'ensemble des logements sont recensés la même année par le même agent recenseur : on ne devrait pas retrouver 2 feuilles logement (FL) ayant les mêmes informations de localisation. Dans les PC, un risque de retraitage existe quel que soit le choix effectué puisque le mode de tirage dans les DOM s'appuie sur l'utilisation de 5 EAR cumulées pour les PC. Ainsi, les logements d'une PC donnée vont être tirés dans l'EAR n-5 pour une enquête sur le terrain pendant l'année n et seront tirés dans l'EAR n pour une enquête sur le terrain pendant l'année n+1. Assurer la disjonction entre les EAR n-5 et n ne peut être garanti, ne serait-ce qu'au moment du tirage.

5.1.5. Tirer dans 3 EAR plutôt que 2 : si c'était à refaire sur Cadre de Vie et Sécurité 2015 en Guyane

La partie 3 a conclu à l'utilisation de 2 EAR pour constituer la base de sondage utilisée dans le système standard. Depuis la mise en place du nouveau dispositif d'échantillonnage, c'est ce que le Criem a retenu, dans chacun des 4 DOM. Une difficulté a cependant émergé suite au tirage de l'enquête Cadre de Vie et Sécurité 2015 (CVS) en Guyane. Le volume global de la base de tirage, nettement suffisant pour permettre le tirage de 3000 logements, cachait des disparités importantes suivant les strates de tirage. Le nombre de logements s'est avéré insuffisant dans une strate de grande commune, à Kourou. 130 logements devaient y être tirés, seuls 94 logements étaient à disposition dans la base de tirage dite « réduite »²⁷. Une autre strate de grande commune, à Cayenne cette fois, se retrouve quasiment vidée suite au tirage de CVS : sur les 170 logements présents dans la base de tirage, 157 sont enquêtés sur CVS.

A posteriori, il aurait été plus judicieux de tirer cette enquête dans 3 EAR pour la Guyane, à la fois pour des questions de volume (d'autant plus que ce problème nous suivra l'an prochain lors de la constitution de la nouvelle base de sondage), mais aussi de précision. La synthèse du paragraphe §3.3 allait d'ailleurs en ce sens : « La conclusion est moins facile pour des échantillons volumineux (au-delà de 2000 logements), le gain marginal de l'utilisation d'une troisième EAR pour le tirage est nettement plus important, et pourrait justifier de passer outre les scrupules à utiliser des données anciennes. »

5.2. Assouplir les contraintes sur les SAE ?

Comme on l'a vu plus haut (§4.7), les allocations de tirage sont actuellement calculées sur le nombre de résidences principales au recensement de 2008. Entre le recensement de 2008 et celui de 2011, le nombre de résidences principales a augmenté d'environ 4% aux Antilles et de 7% en Guyane et à La Réunion. Cette croissance est loin d'être homogène sur le territoire. En Guyane en particulier, Cayenne perd des habitants alors que deux de ses communes frontalières, Matoury et Macouria, croissent de plus de 20%. Ne pas prendre en compte ces différences d'évolution dégrade la précision de l'échantillon tiré.

On a vu aussi que les allocations de tirage sont pensées pour s'approcher au maximum d'un nombre identique de logements à enquêter dans chaque SAE. Or, dès leur construction basée sur les chiffres du recensement de 2008, les SAE n'étaient pas tous de la même taille : les SAE de Baie Mahault et de l'île de Marie-Galante en Guadeloupe contenaient par exemple moins de 5000 résidences principales en 2008, alors que les plus grands SAE en contenaient près de 7000. Ces écarts s'expliquent notamment par le fait que le découpage en SAE s'est fait à l'intérieur du découpage en micro-régions, ce qui a mené à construire des SAE de tailles différentes selon la micro-région. Si l'on reprend l'exemple des 2 plus petites micro-régions martiniquaises vu au §4.7 : alors qu'elles contiennent un nombre de résidences proches, l'une a été scindée en 2 SAE, l'autre, un peu moins peuplée, constitue 1 SAE à elle seule. La taille de l'échantillon va mécaniquement être 2 fois plus élevée dans la première micro-région, ce qui n'est pas représentatif du volume de résidences principales. Cet objectif d'allocation égale par SAE dégrade lui aussi la précision de l'échantillon tiré.

Une simulation a été réalisée sur les échantillons de Cadre de Vie et Sécurité 2015 pour les DFA et de l'enquête logement 2013-2014 pour La Réunion pour comparer les échantillons réellement tirés avec ce qu'ils auraient été si l'on avait utilisé le recensement de 2011 comme référence et si l'on avait réalisé un tirage stratifié à allocation proportionnelle, sans contrainte d'objectif d'allocation égale par SAE :

- Aux Antilles, la contrainte d'objectif d'allocation égale par SAE crée de fortes sur ou sous représentation des petites micro-régions. Par exemple, 21 logements sur les 115 réellement tirés sur l'île de Marie-Galante n'auraient pas été tirés.

²⁷ Dans la pratique, les 94 logements de la strate ont été intégrés à l'échantillon. Le poids des FA de la strate a été revu en conséquence. Les allocations des autres strates n'ont pas été modifiées.

- En Guyane, les évolutions contrastées entre 2008 et 2011 provoquent une surreprésentation de Cayenne et de Kourou, au détriment de Macouria, Matoury et Saint-Laurent. Par exemple, il aurait été tiré 17 logements de plus à Macouria par rapport aux 129 réellement tirés.
- A La Réunion, les deux effets cumulés aboutissent à une surreprésentation de trois strates, le centre-ville de Saint Denis, la strate de Saint-Benoît et une des deux strates de Saint-André. Par exemple, 16 logements sur les 164 tirés dans le centre-ville de Saint-Denis n'auraient pas été tirés.

Dans la mesure où les gestionnaires des Unités Enquêtes Ménages (UEM) réajustent systématiquement les frontières des SAE pour constituer les zones de collecte des enquêtes autres que l'EECDOM (§4.1), dégrader la précision de l'échantillon pour s'approcher au maximum d'un nombre identique de logements à enquêter dans chaque SAE ne paraît pas opportun. Cette contrainte pourra être levée lors des prochains tirages autres que l'EECDOM.

Concernant la mise à jour des chiffres du recensement pour le calcul des allocations de tirage, il semble raisonnable de la programmer tous les 5 ans, une fois le cycle complet du recensement mis à jour. Cela signifie qu'en 2016, lorsque le recensement de 2013 sera disponible, le Criem pourra proposer aux UEM de réviser les frontières des SAE pour retrouver des SAE plus proches en terme de nombre de résidences principales. D'autre part, le Criem calculera les allocations de tirage par strate à partir des chiffres de 2011.

5.3. Les tirages dans les sources fiscales : disjonction de bases

Les bases fiscales ne sont pas seulement un outil à perfectionner pour la problématique des logements neufs dans les enquêtes auprès des ménages. Elles sont également des bases de sondage à part entière, de plus en plus utilisées, dans les Dom comme en métropole. Se posent ainsi naturellement les questions d'articulation entre les deux bases de sondage, en particulier la disjonction, et ce de manière d'autant plus complexe que le recensement et les fichiers fiscaux ne reposent pas sur la même logique, ni sur les mêmes briques géographiques. Le recensement Outre-Mer s'appuie sur les îlots, sur lesquels sont positionnées des adresses (rappelons qu'une adresse est ici un bâti, un immeuble, pouvant contenir un seul logement – une maison – ou plusieurs – immeuble collectif). Les fichiers fiscaux s'appuient sur les sections cadastrales, elles mêmes découpées en parcelles.

5.3.1. Quid d'une enquête entièrement échantillonnée à partir des fichiers fiscaux ?

Jusqu'à récemment, seules l'EEC métropolitaine et l'enquête « Patrimoine » étaient tirées dans la source fiscale. Mais ces sources sont de mieux en mieux connues par les méthodologues de l'Insee et séduisent de plus en plus les maîtrises d'ouvrage pour trois raisons majeures :

- elles contiennent des informations très précises sur les revenus des foyers, d'où son utilisation pour Patrimoine ;
- elles contiennent des adresses au format informatique (et non uniquement des scans d'image comme pour les tirages dans le recensement), ce qui permet des envois postaux de masse. C'est pour cette raison par exemple que les sources fiscales sont utilisées comme base de sondage des enquêtes internet ou de l'enquête postale « Vie Quotidienne et Santé » (VQS) ;
- elles permettent de constituer une base de sondage d'individus (majeurs) car on y dispose d'informations nominatives, d'où son utilisation pour l'enquête « Formation et Qualification professionnelle » (FQP), enquêtes d'individus permettant un appariement avec d'autres sources administratives.

L'année 2014 a ainsi été une année riche en tirages dans la source fiscale, avec les trois enquêtes VQS, FQP et « Patrimoine ». Si La Réunion avait déjà expérimenté le tirage dans la source fiscale avec Patrimoine en 2009 et avec l'enquête internet « Qualité de vie au travail », couplée à l'enquête « Condition de travail » en 2012, c'était une première pour les DFA. Il est actuellement trop tôt pour faire le bilan de ces tirages, nous allons ici nous concentrer sur la question de la (non) disjonction.

En métropole²⁸, des travaux sont actuellement en cours afin d'apparier les identifiants des logements des deux bases de sondage et donc de pouvoir « marquer » dans les deux bases de sondage les logements tirés. Si ceci est déjà aujourd'hui possible au niveau des adresses, grâce à un identifiant créé à partir du numéro + type + libellé de voie²⁹, c'est plus complexe au niveau des logements situés à une même adresse : il n'est pas possible aujourd'hui d'apparier de façon sûre un logement à l'intérieur d'un immeuble dans la base fiscale et dans le recensement. Des travaux sont en cours pour améliorer l'appariement des logements collectifs. Ces travaux se basent sur les caractéristiques des individus qui composent le logement (sexe, date de naissance, commune de résidence). Les derniers travaux menés dans le cadre du Répertoire Statistique des Logements (RSL) montrent qu'il est possible d'apparier de façon quasi sûre 90% des logements individuels. Ces travaux sont en cours de validation à partir de l'échantillon de l'enquête Nationale sur les Revenus des Jeunes pour laquelle on les adresses du recensement ont été saisies. Si ce taux de recouvrement est confirmé, il deviendra possible d'améliorer les méthodes de disjonction qui sont à l'heure actuelle assez frustes.

Afin d'éviter au maximum le recouvrement d'échantillons entre les deux bases de sondage, les méthodes de disjonction utilisées actuellement en métropole sont les suivantes :

- dans la mesure du possible les enquêtes qui ne sont pas en face à face sont tirées hors échantillon maître, c'est-à-dire hors du périmètre des enquêtes tirées dans le recensement ;
- si le tirage doit se faire dans l'échantillon maître, les adresses entières sont marquées dans la base fiscale et dans le recensement. Ceci génère un biais dans la mesure où ceci revient à surreprésenter les logements individuels. En effet, la probabilité de supprimer une adresse de la base de sondage est d'autant plus forte qu'elle compte un nombre de logements important.

Dans les Dom, il n'y a pas d'échantillon maître, les SAE forment une partition du territoire, et non seulement un échantillon. La principale mesure mise en place en métropole pour éviter le retraitage ne peut donc être répliquée.

Quant au marquage des adresses grâce à un identifiant commun au recensement et à la source fiscale, il n'est pas encore en place. L'utilisation d'un identifiant numéro + type + libellé de voie comme en métropole s'avère peu adapté au vu de la proportion encore élevée d'adresses non normalisées dans les DOM. A La Réunion, l'équipe SIG locale a réalisé fin 2014 un premier test d'appariement par un autre moyen : 83% des 195 000 parcelles réunionnaises fournies par le cadastre de 2013 répertoriées dans la taxe d'habitation obtiennent une correspondance cartographique (îlot + rang d'adresse) unique. Les 17% restants correspondent à des parcelles sur lesquels cohabitent plusieurs adresses. Concrètement cela signifie que l'on pourrait marquer dans notre base recensement 83% des 2250 parcelles cadastrales tirées pour l'enquête « Patrimoine » à La Réunion, avec le biais sur les logements individuels évoqué ci dessus. Si l'on marque aussi les 17% restants, on marque des adresses qui n'ont pas été tirées.

Dans les Dom, les enquêtes récemment tirées dans la source fiscale, « Patrimoine » et « FQP », n'ont ainsi fait l'objet que de disjonction entre elles (VQS étant une enquête postale courte, elle ne fait pas l'objet de disjonction, y compris en métropole). Rien n'empêche donc qu'un logement tiré pour l'enquête « FQP » ait déjà été enquêté sur l'enquête Emploi par exemple. D'autant que les masses des bases de sondage ne sont évidemment pas du même ordre de grandeur qu'en métropole et que le risque de retraitage y est bien plus important. En Guadeloupe, environ 1 logement sur 55 a été tiré pour l'enquête Cadre de Vie et Sécurité en 2015, et 1 sur 66 pour l'enquête FQP en 2014. Le risque d'être tiré pour ces deux enquêtes est donc de l'ordre d'une chance sur 3600, ce qui est somme toute suffisamment faible pour s'en contenter actuellement. D'ailleurs, les UEM ne se sont pas faits écho de problème de retirages constatés sur le terrain.

Plus simple qu'un appariement « parcelle cadastrale de la source fiscale » et « adresse du recensement », il est possible de lier section cadastrale de la source fiscale et îlot du recensement, une section cadastrale étant schématiquement découpé en îlots. Ce travail a été fait pour les 4 DOM par les équipes SIG. On pourrait ainsi imaginer de réserver une partie de la base de sondage

²⁸ Voir note 2014 d'Olivier Haag n°5/DG75-L110 « Les avantages et les inconvénients des bases de sondage RP et fiscales »

²⁹ Identifiant dit CICN2 utilisable en grandes communes

recensement pour les tirages dans la source fiscale. Pour les petites communes, on peut tirer, lors de l'empilement de la nouvelle EAR, une sélection de sections cadastrales qui seront réservées au tirage dans les fichiers fiscaux dans l'année à suivre, et qui ne sera plus utilisée pour aucun tirage lors des cinq années suivantes. Chaque année, à l'empilement de la nouvelle EAR, un autre tirage de sections cadastrales serait effectué pour ne pas figer des zones géographiques pour telle ou telle base de sondage. L'ensemble des petites communes disposerait donc, chaque année, d'un ensemble de sections cadastrales (regroupement d'îlots) où on pourrait échantillonner. Un tel système serait néanmoins plus compliqué à mettre en œuvre dans les grandes communes, car une même section cadastrale peut contenir des îlots appartenant à différents groupes de rotation du recensement.

5.3.2. Point de vigilance sur les logements neufs

Les logements dits « neufs » sont les logements construits après le recensement sur les îlots du groupe de rotation concerné par le recensement d'une année donnée. La disjonction entre les logements « neufs » tirés dans la sources fiscale et « non-neufs » tirés dans le recensement serait donc *de facto* assurée si le tirage dans le recensement ne concernait qu'une seule EAR - comme en métropole dans le dispositif standard - et si on tirait les logements neufs *dans le même groupe de rotation que les logements non neufs*.

Or, dans le dispositif Dom, on tire les logements « non-neufs » dans deux EAR dans les grandes communes et dans cinq pour les petites communes. Ainsi les logements « neufs » tirés pour l'enquête « Logement » 2013-2014 ont des dates d'achèvement allant de 2008 à 2013. Ces logements ont de bonnes chances d'être recensés dans les années suivant l'enquête « Logement », et donc d'apparaître dans les bases de sondage recensement utilisées pour le tirage des enquêtes de 2014 et suivantes.

Une solution économique pour éviter ce risque de retraitage peut être de s'assurer de tirer les logements neufs à l'intérieur du groupe de rotation (des groupes de rotation) correspondant à la dernière EAR (aux deux dernières EAR ?) mobilisée(s) pour la construction de la base de sondage, c'est-à-dire de ne retenir comme susceptibles d'être tirés que l'ensemble des logements neufs construits très récemment. C'est naturel pour les grandes communes (le « retard » de la base de sondage RP pour les grandes communes n'est pas très important), mais l'est moins pour les petites communes. Dans les petites communes, il paraît plus satisfaisant d'assumer un risque de retraitage, tant qu'une solution pérenne de disjonction ne peut être mise en œuvre en toute généralité (qui permettrait des tirages articulés dans les deux sources).

5.4. Quid des « petites » enquêtes ?

Le cas des petites enquêtes, disons moins de 500 fiches adresses (FA), est une question non résolue à bien des niveaux dans les Dom. La grande majorité des enquêtes menées Outre-Mer sont des enquêtes dites « à extension », soit qu'elles répondent à une obligation européenne (c'est le cas de l'enquête Emploi), soit qu'elles soient permises par un financement de partenaires. Ces enquêtes à extension visent un minimum de 1500 FA, 1000 questionnaires recueillis étant le seuil minimum classiquement retenu pour une exploitation régionale.

Évidemment, toutes les enquêtes menées sur le territoire national ne peuvent faire l'objet d'une extension dans chaque Dom. Pour autant, l'Insee a le souci d'afficher des résultats « France entière », incluant donc des données domiennes. Ponctuellement donc, certaines enquêtes sont programmées dans les Dom avec des échantillons de petites tailles. Citons l'exemple de l'enquête Patrimoine.

Lors de sa première édition Outre-Mer, en 2009, elle a eu lieu dans un format d'enquête à extension à la Réunion et en Guadeloupe, y permettant des exploitations régionales. Des résultats France entière étant souhaités, une petite centaine de FA ont aussi été réalisées en Martinique, ce qui correspond à la contribution de ce territoire à l'échelle nationale. Les données conjuguées de la

Guadeloupe et de la Martinique ont ensuite été extrapolées lors du calage pour représenter les 3 DFA (ce qui pose d'autres questions sur la pertinence de cette extrapolation mais là n'est pas le sujet). Cette petite centaine de logements ont été tirés à l'époque sur l'ensemble de la Martinique, générant des temps et frais de déplacement jugés démesurément coûteux.

Lors de sa deuxième et récente édition, en 2014-2015, seule La Réunion réalise l'enquête au format extension. 300 FA sont par ailleurs réalisées en Guadeloupe, ces 300 FA ayant vocation à représenter les 3 DFA dans les résultats France entière. Cette fois, l'Unité Enquête Ménages de Guadeloupe a pu demander à restreindre le périmètre de tirage de ces 300 FA. 15 communes sur les 32 ont été conservées, représentant environ trois quarts de la population. Le périmètre des strates de tirage a été respecté, afin de ne pas détériorer les propriétés de notre système d'échantillonnage. Le revenu médian de cette sous-population de 15 communes s'avère néanmoins assez éloigné de celui de la population générale.

Avec cette pratique de sélection d'un périmètre de tirage pour les petites enquêtes, la logique de collecte pourrait l'emporter sur la logique statistique. Pour schématiser, on court le risque d'enquêter là où on a des enquêteurs disponibles. Pour remédier à ce risque, il serait judicieux d'aménager un protocole de tirage spécifique dans le cas d'enquête de petite taille. Le plus simple serait de réaliser un premier degré de tirage au niveau des strates, afin de restreindre le nombre de strates concernées par l'enquête. Cet aménagement pourra être étudié lors de la prochaine commande d'échantillon de petite taille dans les DOM.

5.5. Lorsque la stratification géographique ne peut être utilisée

L'échantillonnage de l'enquête sanitaire sur la pollution à la chlordédone en Guadeloupe et en Martinique en partenariat avec l'Agence Régionale de Santé Guadeloupe-Martinique, l'Institut National de Veille Sanitaire (INVS) et l'Agence Nationale de Sécurité Sanitaire (ANSES) a permis de mettre en évidence une autre limite du système en l'état. Pour cette enquête, trois strates de tirage étaient définies par des zones géographiques précises assises sur le découpage des territoires en îlots, îlots dit « contaminés » ou littoraux notamment. Ce découpage était orthogonal à la partition du territoire en SAE et strates standards de tirage, qui ont dû être abandonnés pour ces tirages d'échantillons.

La collecte de cette enquête n'étant pas réalisée par l'Insee, la structure géographique de l'échantillon n'a pas eu d'influence « pratique » sur ce cas d'espèce puisque le zonage en SAE est d'utilisation interne à l'Insee. Mais il apparaît que le système ne serait pas directement opérationnel si une enquête voulant s'appuyer sur une stratification géographique orthogonale à celle utilisée dans le dispositif standard devait être mise en œuvre par l'Insee.

Une étude prospective sur cette question serait appréciable mais actuellement d'autres pistes d'investigation méthodologique sont prioritaires.

5.6. Possibilité d'une repondération *a priori* de la base de sondage

Dans les grandes communes, seules deux EAR sont mobilisées pour le tirage (dont la plus ancienne a déjà été ponctionnée l'année précédente). Or, on l'a vu, les groupes de rotation des grandes communes sont particulièrement déséquilibrés (cf. §3.4). Les échantillons tirés vont être affectés par ces déséquilibres, et ce d'autant plus s'ils sont volumineux (cf. §3.3). On ne peut pas exclure des calages acrobatiques. Une solution, pour pallier les déséquilibres en structure des EAR, pourrait être de repondérer *a priori* (selon des modalités à définir) la base de sondage annuelle pour qu'elle « colle » mieux à ce qu'on connaît par ailleurs; par exemple avec le recensement complet de la situation des départements, des communes, des SAE.

L'exploitation de cette piste pourrait également avoir des applications complémentaires heureuses. Quand des enquêtes ciblent des sous-populations spécifiques et qu'on souhaite conserver la base équilibrée, la méthode habituelle consiste à tirer un échantillon en population générale contenant suffisamment de ménages de la population cible puis de « jeter » parmi cet échantillon général les ménages qui ne font pas partie de la population cible. Le principe est aisé à appliquer

quand les bases de sondage sont volumineuses ou la population cible pas trop petite, ce qui n'est pas le cas dans les DOM. Les enquêtes ciblées ou à surreprésentation sont contraintes. Le compromis assumé pour l'enquête « Loyers et Charges » en Guadeloupe (léger déséquilibre de la base par la ponction sèche de ménages locataires d'un logement loué vide) s'inscrit dans cette thématique : la méthode classique n'était pas applicable, et l'option de créer une strate de locataires dans le dispositif général d'échantillonnage conditionnait celui-ci à une enquête périphérique spécifique. Là encore, une repondération *a priori* de la base est une piste à explorer.

La suite de cette section présente quelques éléments de réflexion sur ce thème.

Le problème à résoudre est le suivant.

On pourrait tirer dans une EAR (ou plusieurs ; on simplifiera le problème ici en n'en considérant qu'une seule) si, sous contrainte des poids, celle-ci reflétait la « vraie » structure de la commune, au moins sur des variables structurantes importantes. Or l'EAR ne donne pas cette « vraie » structure. Y a-t-il moyen de jouer sur les poids pour aboutir à la situation souhaitée ?

5.6.1. Distorsion sur un facteur

Soit pour illustration la situation suivante.

Dans une commune, la proportion « vraie » d'une modalité X d'une variable (proportion « vraie » donnée par un cycle de recensement) est a ; le volume de logements de la commune est Y. Sur une EAR de volume z, la proportion de cette modalité X est b. On doit tirer sur cette commune un échantillon de volume v selon un sondage aléatoire simple.

On considérera dans la suite l'absence de non-réponse totale.

5.6.1.1. Distorsion des poids au calage

La probabilité de tirage en première phase est z/Y , et celle du tirage en deuxième phase v/z . La probabilité de tirage sur l'ensemble du processus est v/Y .

Si $a = b$ (auquel cas l'EAR représente fidèlement la situation d'ensemble selon ce critère), l'échantillon va contenir (en moyenne) $v.a = v.b$ logements de la modalité X.

Si $a \neq b$, l'échantillon va contenir $v.b$ logements de la modalité X *alors que, si l'EAR n'était pas déformée, il devrait contenir $v.a$ logements de la modalité X.*

Après calage, les $v.b$ logements qui possèdent la caractéristique X devront en représenter $v.a$, ce qui revient à dire que le rapport entre les poids initiaux et les poids calés sera de b/a pour les ménages possédant la caractéristique X et de a/b pour les autres ; on réalise une post-stratification.

Les poids initiaux sont en effet Y/v et les poids calés pour les ménages possédant la caractéristique X $\left(\frac{b}{a}\right) \times \left(\frac{Y}{v}\right)$ et les poids calés des ménages ne la possédant pas $\left(\frac{a}{b}\right) \times \left(\frac{Y}{v}\right)$.

Plus la distorsion sera importante entre la base de tirage et l'empilement de cinq EAR qui correspond à un cycle, plus seront mécaniquement imposés des rapports de poids importants et plus augmenteront les risques de calage impossibles quand se multiplieront les contraintes de calage.

5.6.1.2. Utilisation d'une stratification et jeux sur les probabilités de tirage au deuxième degré

Une solution simple pour pallier à la situation est de stratifier le tirage selon la modalité X, et de tirer dans la strate qui satisfait la modalité X, non plus avec un taux de sondage de v/z , mais avec un taux de sondage corrigé à $\left(\frac{a}{b}\right) \times \left(\frac{v}{z}\right)$, et inversement dans la strate qui ne satisfait pas la modalité X.

On peut considérer ce tirage stratifié comme, plus généralement, un tirage à probabilités inégales. Partant d'un tirage équipondéré ($\pi_k = (v/z)$), le palliatif pour obtenir la proportion « vraie » de la modalité X dans l'échantillon est un sondage avec $\pi_k^* = (a/b) \times (v/z)$ pour les éléments remplissant la modalité X et $\pi_k^* = (b/a) \times (v/z)$ pour les autres.

Les poids initiaux seront donc pour les ménages possédant la caractéristique X : $(Y/z) \times (1/\pi_k^*) = (Y/z) \times [(b/a) \times (z/v)] = (b/a) \times (Y/v)$ soit le poids calé (et réciproquement pour les ménages ne possédant pas la caractéristique X).

Notons que, si l'échantillon obtenu à la première phase, est réputé avoir été tiré à probabilités égales, un plan à probabilités inégales ($(b/a) \times (Y/z)$ pour les ménages possédant la caractéristique X, et inversement pour les autres) aurait plus vraisemblablement fourni l'EAR telle qu'elle apparaît. Si c'était effectivement un tel plan à probabilités inégales qui avait été appliqué pour la première phase et si on appliquait la probabilité de tirage π_k^* lors de la deuxième phase, le poids de tirage résultant aurait été alors, pour les ménages possédant la caractéristique X, $(b/a) \times (Y/z) \times (1/\pi_k^*) = (Y/v)$, soit le poids d'un SAS en une seule passe. On « corrige » ainsi « l'erreur » d'échantillonnage de la première phase.

De fait, l'échantillon « en une passe » contient en moyenne a.v ménages avec la caractéristique X, et l'échantillon tiré à deux phases avec les π_k^* à la seconde phase contient également en moyenne a.v ménages possédant la caractéristique X.

5.6.1.3. Jeu sur les poids du « premier » degré de tirage

Dans les deux formulations précédentes, les poids du premier degré de tirage sont toujours les mêmes, *a priori* égaux à Y/z , et les poids finaux n'induisent la correcte - ou non - représentation de la modalité X dans l'échantillon que par le jeu du poids de tirage au deuxième degré.

On peut également prendre un autre point de vue, celui de reporter la déformation sur les poids de première phase (*a priori* tous égaux à Y/z).

Pour les éléments qui satisfont la modalité X :

$$\begin{aligned} \text{Poids_final} &= \text{poids_premier_degré_SAS} * \text{poids_deuxième_degré_modifié} \\ &= (Y/z) \times (1/\pi_k^*) \\ &= (Y/z) \times [(b/a) \times (z/v)] \\ &= [(b/a) \times (Y/z)] \times (z/v) \\ &= \text{poids_premier_degré_modifié} * \text{poids_deuxième_degré_SAS} \end{aligned}$$

Autrement dit, les poids finaux ne sont pas modifiés si le facteur correctif des poids de deuxième phase « passe sur » les poids de première phase. Or, faire passer ce facteur correctif sur les poids du premier degré de tirage revient très exactement à *caler* les poids du premier degré de tirage sur la structure de la commune sur la variable mal représentée.

Comme précédemment, les poids initiaux correspondent aux poids calés. Par contre, comme on effectue un SAS lors de la deuxième phase, l'échantillon va contenir en moyenne b.v ménages possédant la caractéristique X.

5.6.1.4. Conclusion

Dans les deux cas décrits, les poids de tirage finaux sont calés sur une valeur connue dans la population, mais non respectée par l'échantillon obtenu lors de la première phase du tirage.

Quand on joue sur les poids à la deuxième phase, on impose à l'échantillon une « bonne » composition du point de vue de la modalité X (intéressant si X est sous-représenté en première phase, avec une forte variabilité), mais l'introduction de probabilités de tirage inégales complique le dispositif, en particulier si plusieurs facteurs sont mal représentés à l'issue de la première phase.

Le jeu sur les poids à la première phase (simple calage des poids issus de la première phase) ne permet pas de rétablir des proportions brutes de modalités dans l'échantillon collecté, mais il

permet de ne mettre en œuvre, lors de la deuxième phase, qu'un basique sondage aléatoire simple, qu'on peut raffiner par un équilibrage, un sondage systématique.

Rapporté au problème qu'on a posé précédemment, une EAR, même déformée, peut être suffisante pour tirer correctement un échantillon à l'échelle communale si on peut effectuer un calage des poids issus du recensement sur la structure « vraie » de la commune. Le problème est ainsi déplacé à la connaissance de cette « structure « vraie » » de la commune.

Un inconvénient immédiat de cette procédure, c'est une complication dans les calculs de poids et de précision.

5.6.2. Distorsion sur plusieurs facteurs

Si la structure de l'échantillon de première phase est déformée selon plusieurs facteurs, on ne peut utiliser directement le parallèle entre stratification et calage de la partie précédente que si les totaux des différents croisements entre facteurs sont connaissables.

Soit donc le cas général suivant :

Soit une population A, dont on a accès à une sous-population P seulement. P munie des poids \tilde{n} représente A, mais la représentation est calibrée sur une famille de critères X avec les poids calés n.

Ainsi : $\sum_P n_i \times X_i^j = T_{X^j}$ On note: $r_i = \frac{n_i}{\tilde{n}_i}$

1 - A quelles conditions sur les poids issus de la première phase un SAS lors de la deuxième phase reflète la situation connue de la population générale (sur les critères X) ?

2 - Quel plan de sondage lors de la seconde phase permet, sans modification des poids issus de la première phase, permet à l'échantillon de refléter la situation connue de la population générale (sur les critères X) ?

Le poids global non calé, si on effectue un SAS et qu'on utilise les poids non calés de la première phase, est : $\frac{1}{\pi_i} \times \tilde{n}_i$. Les π_i étant égaux, on ne retrouve pas les contraintes sur les facteurs

X : il faut caler.

Les poids $r_i \times \frac{1}{\pi_i} \times \tilde{n}_i$, soit les poids globaux non calés multipliés par le rapport entre les poids initiaux et calés de la première phase, sont par contre des poids calés sur les critères X.

En effet, pour tout j, on a :

$$E\left(\sum_s r_i \times \frac{1}{\pi_i} \times \tilde{n}_i \times X_i^j\right) = \sum_P r_i \times \tilde{n}_i \times X_i^j = \sum_P \frac{n_i}{\tilde{n}_i} \times \tilde{n}_i \times X_i^j = T_{X^j} \quad (1)$$

En admettant ici que le poids calé exhibé ici est bien le poids calé optimum (au sens des fonctions de distance utilisées dans le calage), *si celui-ci existe*, il s'ensuit que le rapport des poids globaux initiaux et calés suit la même distribution que celui des poids initiaux et calés de la première phase.

Il suit directement de (1) que, si on considère un SAS à la deuxième phase parmi un échantillon de première phase dont les poids sont calibrés (par calage) sur un certain nombre de facteurs connus de la population général, les poids globaux initiaux sont alors *déjà calibrés* sur la population générale.

En utilisant l'expression (1), on voit également qu'un tirage à probabilités d'inclusion inégales définies par π_i / r_i en deuxième phase avec les poids initiaux de la première phase conduit de même à un échantillon calé sur les facteurs X. Si le rapport de poids r d'un élément de P est inférieur à 1 - ce

qui signifie qu'il est sous-représenté dans l'échantillon issu de la première phase -, alors il aura une probabilité supérieure à celle d'un SAS d'être tiré lors de la deuxième phase (et inversement).

Les conclusions sont du même type que dans le cas d'un seul facteur déformé. En introduisant le rapport des poids initiaux et calés de la première phase, on obtient un échantillon dont la pondération est calée *a priori* sur des totaux connus dans la population entière. Soit on fait alors un sondage à probabilités d'inclusion inégales lors de la deuxième phase, et on obtient un échantillon dont, par ailleurs, la composition brute est réalignée en moyenne sur celle de la population générale (mais les calculs de précision sont compliqués par ces probabilités inégales), soit on met en œuvre un plan de seconde phase tout simplement appuyé sur un SAS (SAS pur, tirage équilibré, tirage systématique....) pour lequel les calculs de précision sont simples - mais la structure brute de l'échantillon ne sera pas réalignée.

Le problème se ramène donc toujours à celui de la connaissance d'une « structure « vraie » » de la commune.

6. Les enquêtes auprès des ménages à Mayotte.

6.1. Introduction

La réalisation d'enquêtes auprès des ménages par l'Insee à Mayotte est récente. L'Institut est en effet implanté dans le territoire depuis 1996 seulement, sous la forme d'une Antenne puis d'un Service Régional sous l'autorité de la Direction Régionale de La Réunion – Mayotte implantée à Saint-Denis-de-La-Réunion.

Une enquête « budget des familles » est réalisée en 2005, puis une Enquête sur l'Emploi ponctuelle en 2009 et une nouvelle enquête « BdF » en 2010. Les deux premières enquêtes ont mobilisé le recensement de population et la troisième un dispositif mixant recensement de population, cartographie et recensement agricole³⁰, mais aucun dispositif pérenne n'était en place.

Le changement de statut de Mayotte, devenu *département* en 2011, et les travaux à prévoir – à la fois du point de vue de la mise aux standards du territoire par rapport aux autres régions française et du point de vue de sa reconnaissance souhaitée en tant que Région Ultra-Périphérique (RUP) de l'Union Européenne – rendaient inéluctable la mise en place d'un dispositif systématique d'enquêtes, avec au moins une enquête sur l'emploi à un rythme régulier à préciser.

Deux sources étaient *a priori* envisageables pour la mise en place d'un dispositif pérenne d'échantillonnage des enquêtes auprès des ménages à Mayotte : le recensement (exhaustif, le dernier ayant été réalisé en 2012 et dont une édition est prévue tous les cinq ans) et une cartographie des bâtis.

Dans le contexte mahorais, il est difficile, pour ne pas dire impossible, d'exhiber une comparaison de ces deux sources selon des standards confortables et objectifs – comparaisons de variances, de design effects... Dans la suite, on donnera des éléments quantitatifs autant que possible, mais la mise en balance des éléments qualitatifs de connaissance du terrain est inéluctable.

Cette partie va mettre en balance dans un premier temps ces deux sources possibles pour l'échantillonnage. Dans un second temps, le dispositif en place depuis 2014 sera explicité.

6.2. Quelle source privilégier ? Recensement versus cartographie

En toute généralité, au delà des éléments quantitatifs de comparaison délicats à produire (variance...), plusieurs éléments sur les deux sources possibles sont à prendre en considération dans l'arbitrage : la puissance de la source (ciblage de population, déchets...), la qualité de repérage offerte, le coût, l'exhaustivité, le volume et enfin les capacités de gestion informatique et méthodologique des sources (par exemple les possibilités de disjonction). La question du repérage, particulièrement centrale, va impacter ces différentes thématiques.

Il faut enfin ne pas négliger les contraintes induites par le dynamisme de la construction à Mayotte. Des éléments mis à jour pour l'instruction de la gestion des logements neufs pour l'enquête « Logement » à Mayotte en 2013 laissent en effet imaginer une croissance du parc de l'ordre d'environ 3% par an. Aucun élément clair sur la construction neuve en tant que telle (ou sur la destruction) n'est disponible pour le territoire mahorais dans la mesure où les sources annexes habituelles (permis de construire, TH etc.) sont trop partielles ou inexistantes (pas de TH) à l'instruction en 2013 ; « à dire d'expert » et compte-tenu de l'estimation de la croissance du parc, on peut imaginer assez légitimement un taux de construction à 4% voire à 5%.

³⁰ BdF976 en 2005 a été réalisée par l'Antenne de Mayotte. Toutes les métadonnées n'ont pas été retrouvées. Pour l'EE976 de 2009 : tirage de 75 districts agrégés puis tirage de 2000 logements dans les 75 districts agrégés. Pour BdF976 de 2010 : tirage de 60 districts agrégés proportionnellement au nombre de logements au RP 2007 puis tirage de 25 logements / district agrégé au sein d'une base cartographique constituée par la mise à jour de la cartographie utilisée pour le recensement agricole

6.2.1. Quelle utilisation possible d'un recensement à Mayotte ?

A cause des contraintes inévitables induites par la gestion des logements neufs, il est clair qu'on ne peut pas utiliser comme tel le recensement de 2012 jusqu'au suivant en 2017 pour échantillonner les enquêtes auprès des ménages – celui-ci sera très rapidement obsolète – sans coupler son utilisation avec un dispositif identifiant des logements neufs et quantifiant le nombre global de ceux-ci.

Par ailleurs, le bâti mahorais, dense, en évolution rapide même en zone urbaine et assez anarchique³¹ (certains quartiers n'ont pas de rues au sens propre, il faut parfois passer par le terrain du voisin pour atteindre son logement), ne permet pas aux enquêteurs de se passer de plans précis, même avec une enquête appuyée sur le recensement : les adresses déclarées et les noms des habitants (récupérés dans les scans du RP) ne seront très certainement pas suffisants pour identifier de façon certaine le logement à enquêter : les enquêteurs *doivent* être munis de plans (à jour) du district où se trouve ce logement.

Mayotte ne dispose pas des sources administratives habituelles pour mettre en place une BSLN « enquêtes-ménage » comparable à celle qui existait en métropole jusqu'à la mise en place du dispositif OCTOPUSSE. La seule option raisonnable est la mise en place d'une enquête cartographique complémentaire (aux modalités à définir) qui permettrait l'identification par différentiel des logements neufs du département. Construire une BSLN par cartographie présente l'avantage « collatéral » de mettre à jour la cartographie utilisée pour le recensement afin que les enquêteurs se repèrent mieux sur le terrain et d'identifier précisément les logements (neufs) à enquêter, mais également pour les « autres » logements dans les zones où la cartographie est mise à jour.

Quelles sont dès lors les options pour définir une cartographie complémentaire permettant l'identification et la quantification globale des logements neufs ainsi que le repérage précis sur le terrain ? Celle-ci doit être annuelle (pour la fraîcheur des données) et porter sur une partie conséquente du territoire (pour fournir une information robuste sur l'évolution du bâti, en qualité et en quantité). Sous les hypothèses d'un programme d'enquête à 4 500³² FAs annuelles et de la recherche d'équipondération des logements (50 000 logements principaux environ au RP 2012), la cartographie à réaliser annuellement devrait être au moins de l'ordre d'un huitième du territoire. Enfin, il faudrait qu'elle soit très précisément articulée avec le recensement (pour les identifiants des logements) pour que le repérage soit efficace.

Le couplage du recensement avec une cartographie limite l'utilisation du recensement : celui sera(it) exploitable (du point de vue de l'identification correcte des logements à enquêter) là seulement où la cartographie annuelle sera(it) réalisée : la cartographie devrait alors concerner un cinquième du territoire *a minima* (cf. paragraphe suivant).

On va discuter dans la suite l'option « recensement », avec l'hypothèse d'une cartographie tournante complémentaire d'un cinquième, au coût *in fine* quasi comparable à celui d'une cartographie un peu plus réduite (coûts avec rendement d'échelle décroissants), qui alignerait la cartographie du SR sur les procédures habituelles par cinquième induites par le recensement tournant en vigueur dans les autres départements d'Outre-mer et qui, enfin, permettrait une identification satisfaisante des logements.

6.2.2. Quelle utilisation possible d'une cartographie à Mayotte ?

C'était cette même problématique de l'obsolescence rapide du recensement (et des problèmes induits par la nécessité d'identifier et de quantifier les logements neufs) qui avait poussé à

³¹ Cette construction anarchique se développe certes en général par « poches », mais un peu partout sur le territoire, jusque dans les centres villes.

³² L'hypothèse de 4 500 FAs correspond à 4 500 FAs en face à face (l'enquête-emploi semi-panéalisée type EEADOM contient environ deux tiers de réinterrogés par téléphone).

avoir dans les autres DOM à partir de 2004 un dispositif cartographique pour les enquêtes auprès des ménages : très rapidement le RP99 s'était révélé déficient pour réaliser les enquêtes (et en particulier l'EEADOM dont l'échantillonnage ne fonctionnait plus alors qu'elle était à peine arrivée en régime de croisière avec trois sous-échantillons). A la différence des DOM « historiques » en 2004, le recensement mahorais reste sur un modèle exhaustif quinquennal ; le recensement mahorais « n'offre » pas une logistique cartographique dont les enquêtes auprès des ménages peuvent se saisir.

Un dispositif appuyé sur la cartographie doit cependant avoir des vertus comparables à celui qui était en vigueur dans les DOM « historiques » sur la période 2004-2013 : en plus de fournir une base de sondage (de première phase pour les enquêtes auprès des ménages), il doit fournir annuellement une vision exhaustive d'une partie suffisamment importante du territoire pour qu'elle permette des extrapolations à l'ensemble pour le calage des enquêtes de l'année.

Une cartographie tournante des bâtis (ou points-RIL : un bâti peut contenir plusieurs logements, et l'enquête cartographique doit indiquer, entre autres caractéristiques, le nombre supposé de logements du bâti) sur des cinquièmes équilibrés de territoire (dispositif DOM 2004-2013³³) semble une solution assez robuste et qui permet en outre une disjonction des enquêtes entre années. C'est l'option de base qu'on va discuter ici ; on peut concevoir des variantes (enquêter plus du cinquième, par exemple), mais il n'est en tout état de cause pas possible de couvrir moins de territoire annuellement : 60 000 PIM environ ont été identifiés lors de la cartographie préalable au RP de 2012 pour environ 50 000 résidences principales, ce qui signifie en moyenne environ 10 000 résidences principales mobilisables annuellement pour, parallèlement, un volume probable d'environ 4 500 FAS pour deux enquêtes (EEA + autre enquête sociale).

6.2.3. Inconvénients et avantages de ces options

Les questions de coût sont de facto éliminées de la réflexion : dans les deux cas, une enquête cartographique comparable doit être réalisée. Une cartographie d'un cinquième du territoire (10 000 points-RIL) est estimée à un ETP d'enquêteur (plusieurs enquêteurs travailleraient un ou deux mois en tant qu'enquêteurs-carto) et serait intégrée à l'annualisation et la globalisation du travail d'enquête induites par le NCEE.

Les questions liées à l'exhaustivité sont également traitées : il n'y a pas d'avantages d'un dispositif sur l'autre de ce point de vue. Dans les deux cas, la source utilisée pourra être réputée représenter convenablement l'ensemble du territoire : la cartographie complémentaire « complète » le recensement des logements neufs et la base cartographique contient de fait tous les logements (dont les neufs) du territoire cartographié.

Les questions de volume semblent donner un certain avantage au recensement, mais celui-ci n'est que de façade. Le recensement contient l'ensemble des bâtis des deux îles (50 000 résidences principales environ ainsi que les logements neufs ajoutés annuellement) alors que la cartographie annuelle à mettre en place n'en contiendrait qu'une certaine fraction, environ 10 000 résidences principales chaque année.

Dans la mesure où il était envisagé dans les années postérieures de réaliser deux enquêtes par an à Mayotte (EEADOM + une autre enquête) pour un volume total de 4 500 fiches-adresse (hypothèse maximale), une cartographie de 10 000 résidences principales paraît calibrée. Seule la perspective de la réalisation d'une seconde enquête particulièrement volumineuse dans l'année risquerait de mettre à mal la capacité en volume d'une base cartographique. La situation est ainsi plus favorable que celle de la Guyane dans les années 2004 – 2013 du point de vue des considérations de volume, et aucune solution *ad hoc* pour réagir à une limitation en volume n'avait dû être mise en place. En passant, la situation sur également plus favorable du point de vue de l'équilibre des groupes de rotation : ceux-ci ne sont pas « imposés » par le recensement ; ils peuvent être mis à jour à chaque recensement exhaustif tous les cinq ans, et les déséquilibres éventuels entre groupes peuvent être remis à plat régulièrement.

³³ Pour rappel, dans le dispositif qui était en vigueur dans les DOM « historiques » jusqu'en 2013, on utilisait la partie non recensée du GR des grandes communes et un cinquième de chacune des petites communes, ce cinquième des PC ayant été défini de manière analogue aux GR des GC

Par ailleurs, pour que l'identification des logements soit satisfaisante avec le recensement, il faut que les logements soient échantillonnés là où la cartographie a été mise à jour par l'enquête cartographique complémentaire nécessaire pour l'identification de logements neufs, à savoir dans la cartographie de l'année ou éventuellement dans celle de l'année précédente : les volumes mobilisables sont ainsi en fait du même ordre entre le recensement (+BSLN par cartographie) restreint aux zones où la cartographie est actualisée (qui sont les seules où une correcte identification serait possible (abstraction faite du problème technique de la déconnexion des identifiants entre cartographie et recensement, cf. infra) et la cartographie pure. De plus, il serait vain d'ajouter à la base uniquement les logements neufs, sans la corriger des destructions ou modifications (opérations de RHI nombreuses dans les années à venir). Aussi, le calage final portera dans un cas comme dans l'autre sur un nombre de logements estimé à partir de l'enquête cartographique partielle, et non pas sur le total surestimé « RP + logements neufs ».

La question du repérage donne un avantage net à la cartographie par rapport au recensement. Les éléments classiques du repérage (adresse « postale » appuyée sur une rivilisation, noms des habitants) dans le recensement doivent impérativement être associés à un plan localisant le logement à enquêter. Il est d'ailleurs à noter que dans les autres départements d'Outre-mer (où le recensement est maintenant la base du système d'échantillonnage), on considère que les éléments RP (scan inclus dans la FA) *précisent* le repérage cartographique joint au matériel des enquêteurs. Or l'articulation des identifiants cartographiques dans la cartographie et dans le recensement paraît faible : les expertises sur le RP exhaustif pour les travaux de redressement de celui-ci montrent que les appariements RP-carto ne sont pas encore satisfaisants³⁴ (parmi les 91% d'appariements réussis, seuls 82% assurent une cohérence sur l'aspect du bâti entre les sources cartographique et RP, i.e. environ **74% des résidences principales du RP ont à la fois une correspondance cartographique et une cohérence sur l'aspect du bâti**). Ces éléments ne présagent d'une qualité médiocre ni pour la cartographie ni pour le recensement : certes les travaux cartographiques n'ont pas encore atteint un régime de croisière à Mayotte (première « vraie » cartographie, mise en place progressive d'un réseau d'enquêteurs), mais on peut plus facilement douter de l'attention des agents recenseurs aux problèmes cartographiques (qui ne sont pas les leurs dans la perspective du recensement) : il est très probable que l'identifiant cartographique le plus sûr soit celui issu de la cartographie.

Par ailleurs, ces écarts entre la cartographie préalable au recensement et le recensement lui-même du point de vue des identifiants fins limite de fait la portée d'une enquête cartographique en vue d'une BSLN : les logements identifiés comme neufs dans la cartographie complémentaire pourraient déjà être répertoriés dans le recensement.

Une cartographie mise à jour tous les ans (et qui gagne en qualité d'année en année) sur laquelle les mêmes enquêteurs réaliseront des enquêtes auprès des ménages permet en revanche d'envisager une bien meilleure identification des logements.

Du point de vue des possibilités offertes par la source, l'avantage au recensement semble être évident ; en effet, le recensement permet de cibler des logements et non des adresses et de plus celui-ci offre une richesse d'information bien supérieure : non seulement on sait cibler l'enquête si nécessaire (profil d'individus ou de logements), mais de plus les déchets (logements non principaux (1 252 logements identifiés comme tels après les premiers redressements du RP) ou vacants (7 222 logements, soit un peu plus de 12%)...) sont de fait limités puisque le recensement permet de n'envoyer des enquêteurs que dans les résidences principales au moment du RP. Le ciblage des logements (contre les adresses) n'est pas déterminant pour le dispositif dans la mesure où les maisons individuelles (mono-logements) sont largement majoritaires à Mayotte (cf. infra). Quant à la richesse du recensement, celui-ci ne peut cependant pas s'appréhender indépendamment de la qualité du repérage sur le terrain. La déconnexion partielle (du point de vue des identifiants géographiques fins) de la cartographie actuelle réalisée pour le recensement et du recensement lui-même limite donc de facto la puissance du recensement comme source de sondage.³⁵

³⁴ Ces éléments sont plus détaillés dans le paragraphe « récapitulatif des problèmes d'appariement RP-cartographie »

³⁵ Un dispositif assis sur la cartographie plutôt que sur le recensement pose évidemment question pour les enquêtes devant cibler a priori des sous-populations. La mise en place progressive d'un dispositif statistique d'ensemble à Mayotte incite à penser que, dans les prochaines années, ce seront plutôt des enquêtes « structurantes » en population générale qui seront à mener en priorité. Par ailleurs, et c'est souvent le cas également dans les autres DOM, de nombreux type de sous-population régulièrement « à surreprésenter » dans les enquêtes, sont *de facto* surreprésentés par rapport à la métropole : jeunes, travailleurs non ou peu qualifiés... A titre d'exemple, l'échantillonnage de l'enquête « Information et Vie quotidienne » (IVQ) réalisée en 2012 en

Un léger avantage comparatif apparaît ainsi pour la cartographie du point de vue de la gestion des bases et de la méthode. De nombreux programmes (dont ceux utilisés pour l'EEADOM sur la période 2004-2013) seraient mobilisables au prix de minimes modifications. L'utilisation du recensement conjointement à une cartographie créant une BSLN nécessiterait à l'inverse l'écriture et les tests d'une chaîne entière de traitements.

Sur la simple question de la disjonction des enquêtes, un autre avantage comparatif apparaît également pour le dispositif par cartographie quand on aborde cette question de façon « pratique » (et non seulement du point de vue de la gestion informatique des fichiers) car il est lié aux possibilités de repérage ainsi qu'à la fiabilité de l'identifiant cartographique et à son articulation avec le RP. En effet, la disjonction est assurée de fait dans une cartographie tournante, le temps que les opérations cartographiques soient gérées avec un rythme de croisière au Service Régionale (SR) de Mayotte et que ces considérations soient devenues naturelles dans le réseau d'enquêteurs qui se met en place.

Enfin, la question se pose de la simple possibilité d'une véritable pérennité d'une utilisation du recensement comme source principale : si on tire dans le recensement avec un complément pour les logements neufs, après deux ans d'utilisation, la source principale risque fortement d'être devenue obsolète, et il pourrait être nécessaire de passer à une utilisation de la source cartographique. Mais on n'aurait pas plus la correspondance RIL des identifiants (rangs) RP divergents, et il serait quasi impossible d'opérer une disjonction entre les deux années de tirage.

6.2.4. Zoom sur les problèmes de repérage

Les agents recenseurs, munis de plans issus de la cartographie, doivent reporter dans les FL les identifiants cartographiques des bâtis et logements (de la même manière que les agents recenseurs dans les grandes communes de métropole ; dans les petites communes métropolitaines où l'opération de recensement est exhaustive comme à Mayotte, l'agent recenseur ratisse lui-même le terrain qui n'est pas RILisé). Un pont « RP-cartographie » (qui permettrait ainsi l'utilisation conjointe du RP et de la cartographie préalable à celui-ci) est ainsi en théorie possible, mais les premières exploitations montrent que ce pont est largement défaillant³⁶.

En unité de compte « logements » (hors communautés), on a les résultats suivants à partir des fichiers « Jouve » bruts du recensement (hors pondération mettant en concordance les données issues de SCRP et les fichiers « Jouve ») :

- nombre de logements recensés : 60 013
- nombre de résidences principales recensées : 50 995
- nombre de résidences principales recensées avec appariement carto : 46 479 (soit 91,1% des résidences principales)
- cohérence de l'aspect du bâti pour les résidences principales recensées appariées : 37 944 (81,6% des résidences principales appariées). *Près de 10% des logements ont été identifiés comme « habitat de fortune » dans une des sources et comme « logement en dur » dans l'autre.*

Guyane et à La Réunion, n'a pas retenu la surreprésentation de ménages avec des individus peu ou pas diplômés.

Cependant, des solutions *ad hoc* peuvent être mises en place s'il était nécessaire de mener, à partir d'une cartographie, des enquêtes ciblées (estimation de la part des ménages cibles dans l'ensemble des ménages, et estimation en conséquence du nombre de fiches-adresse à échantillonner par exemple). Ces solutions ne sont évidemment pas entièrement satisfaisantes car elles augmentent les déchets, mais c'est un dispositif robuste qui a fait ses preuves dans les DOM « historiques » quand le dispositif de sondage y était assis sur une cartographie ; et le cas mahorais nécessite un dispositif avant tout *robuste*.

³⁶ On peut citer comme éléments d'explication, en plus du repérage complexe à Mayotte déjà évoqué plusieurs fois, le fait que l'opération cartographique était une première au Service Régionale, et que, pour l'agent recenseur, la qualité « aval » de la cartographie n'est pas une priorité et il a souvent tendance à négliger de reporter et d'informer des créations et suppressions de rang (gestion des fiches navette). Enfin, il y a tout lieu de penser que la structure du recensement mahorais (cartographie articulée à l'exhaustivité) fasse simplement apparaître plus nettement des problèmes existant dans les recensements tournants mais qui n'ont pas de conséquences sur la conduite des enquêtes auprès des ménages : en métropole, la simple utilisation du scan RP avec les noms et adresse « postale » déclarée est suffisante pour les enquêteurs.

aspect du bâti issu de la cartographie	aspect du bâti issu du recensement					Total
		1	2	3	4	
1	204	10912	858	151	1806	13727
	.	23,87	1,88	0,33	3,95	30,02
2	8	189	85	12	207	493
	.	0,41	0,19	0,03	0,45	1,08
3	16	209	94	149	290	742
	.	0,46	0,21	0,33	0,63	1,62
4	528	2733	742	488	26798	30761
	.	5,98	1,62	1,07	58,61	67,28
Total	.	14043	1779	800	29101	45723
	.	30,71	3,89	1,75	63,65	100
Valeur(s) manquante(s) = 756						

74,4% des résidences principales sont appariées à la cartographie et présentent un aspect de bâti cohérent entre les deux sources.

En unité de compte « adresses », on a les résultats suivants :

- nombre d'adresses recensées (communautés incluses hors HMSA) : 50 890
- nombre d'adresses recensées (hors communautés) : 50 875
- nombre d'adresses cartographiées (communautés nécessairement incluses) : 54 328
- nombre d'adresses recensées (hors communautés) et appariées à la carto : 46 339
- nombre d'adresses cartographiées (mais non appariées au RP hors communautés) : 7 989

6.2.5. Bilan, première proposition de dispositif et versions alternatives éventuelles

Il semble que, si le recensement est un bon réflexe en général, la situation particulière de Mayotte invite à relativiser son intérêt par rapport à l'utilisation d'une base cartographique à l'heure de l'instruction. Quel que soit le système choisi, une opération cartographique est incontournable (indépendamment même des logements neufs puisque les « infos RP » ne suffisent pas à l'identification du logement). Les avantages *a priori* du recensement semblent régulièrement à relativiser dans le contexte mahorais ; l'utilisation d'une cartographie comme base de sondage, malgré ses faiblesses intrinsèques (sur la richesse de l'information notamment), paraît incontournable sur la période 2013 – 2017 au moins.

Le dispositif de sondage pourrait s'appuyer sur une cartographie par cinquième, un peu à l'image du précédent dispositif dans les DOM « historiques », mais sans articulation avec un recensement tournant dont la mise en place à Mayotte n'est pas à l'ordre du jour. Le tirage se ferait dans le cinquième venant d'être cartographié. Il conviendrait d'apporter une attention particulière à la stabilité de l'identifiant cartographique des bâtis.

Variante 1 : si l'identifiant cartographique est suffisamment robuste, on pourrait tirer dans le cinquième de l'année et le résidu du cinquième de l'année précédente (on ignore alors la construction neuve dans ce résidu). On aurait alors des bases de sondage un peu plus volumineuses.³⁷

Proposition optionnelle d'amélioration de qualité : la cartographie permet à la fois d'établir une base de sondage, mais également d'obtenir des informations qui serviront au calage (volume de logements, caractéristiques et répartition du bâti). Autrement dit, la cartographie réalise deux enquêtes en une : la première phase de l'échantillonnage des enquêtes auprès des ménages et une enquête

³⁷ Dispositif proche de celui justifiant un an sans enquête extension dans les PC après le recensement dans la constitution des bases de sondage cartographiques dans les DOM « historiques ».

ayant pour fin de fournir des informations diverses permettant d'améliorer la qualité de la première enquête d'intérêt (l'enquête-ménage) au moment du calage. Même si le tirage s'effectue sur un cinquième du territoire, une cartographie plus étendue et habilement pensée (sur un quart ou un tiers du territoire comprenant le cinquième où on effectuera le tirage) permettrait ainsi d'estimer mieux les données qui seront utiles au moment du calage, en augmentant la taille de l'enquête cartographique entendue comme « enquête de calage ». Par ailleurs, ce « surplus » cartographique permettrait également d'avoir des informations plus fines sur la construction et la destruction annuelle, et préparerait la mise à jour cartographique des districts quand ceux-ci appartiendront au cinquième à utiliser pour le tirage des enquêtes auprès des ménages. Le coût supplémentaire de ce surplus d'enquête « de calage » pourrait être en grande partie absorbée par les modifications dans l'employabilité des enquêteurs induites par le NCEE.

Ce dispositif (et ses variantes) permettrait de standardiser a priori les procédures en n'ayant pas à définir de BSLN puisque les logements neufs seront de facto intégrés à la base de sondage (cela simplifie *ipso facto* les questions de pondérations de logements issus d'une BSLN, et donc le calage des enquêtes, dans le contexte de Mayotte où les sources externes d'information sont peu nombreuses). De plus, on pourra réutiliser, à quelques retouches minimales près, les programmes déjà mis en place pour l'EEADOM (entre autres) sur la période 2004-2013.

On ne rencontrera pas avec ce dispositif les risques de déséquilibre à terme des GR-RP géographiques qu'on rencontre dans les autres DOM : si le RP de 2017 est encore exhaustif (hypothèse de loin la plus probable), les cinquièmes définis pour la période 2014 – 2017 pourront être revus à cette occasion pour retrouver un équilibre éventuellement perdu dans la période intercensitaire (il sera nécessaire que l'identifiant cartographique soit bien stabilisé à ce moment-là pour permettre la disjonction).

Ce dispositif et ses variantes présentent une externalité positive intéressante : d'un point de vue purement fonctionnel, ils permettent à la fois de lier plus étroitement les enquêteurs à leur zone (les mêmes enquêteurs font la cartographie et les enquêtes) et de mettre à jour continuellement le « RIL » mahorais.

6.2.6. Le découpage du territoire mahorais en cinq groupes de rotation équilibrés.

Quelle que soit l'option choisie *in fine*, il était impératif de découper le territoire mahorais en cinq groupes de rotation. La répartition des îlots du territoire entre les cinq groupes a été réalisée à partir des données du RP 2012. Pour aboutir à ces 5 groupes de rotation, 4 tirages d'îlots sont réalisés de manière successive et sans remise, selon un plan de sondage stratifié et équilibré³⁸.

Les 12 îlots contenant des grandes adresses (10 logements ou plus) ainsi que les 128 îlots dépeuplés ont été traités séparément.

6.2.6.1. Le cas standard : les îlots non-dépeuplés et sans « grandes » adresses

La procédure générale de tirage des îlots a été stratifiée géographiquement pour correctement couvrir les différentes micro-régions du territoire. Des variables d'équilibrage ont été mobilisées pour que, sur ces critères, les cinq groupes puissent être réputés équivalents.

Le tirage d'îlots est stratifié selon le groupement géographique suivant :

- 1 = communes 601, 602 et 612 (Acoua, Bandraboua, Mtsamboro - Nord Ouest);
- 2 = communes 613 et 617 (Mtsangamouji, Tsingoni - Ouest);
- 3 = communes 604, 606 et 609 (Boueni, Chirongui, Kani-Keli - Sud);
- 4 = communes 605, 614 et 616 (Chiconi, Ouangani, Sada - Centre);

³⁸ Propriété de l'équilibrage : après retrait d'un échantillon équilibré, la base de tirage conserve le même équilibre.

- 5 = communes 608 et 615 (Dzaoudzi, Pamandzi - Petite-Terre");
- 6 = communes 603 et 607 (Bandrelé, Dembeni - Sud est);
- 7 = commune 610 (Koungou);
- 8 = commune 611 (Mamoudzou)

Les variables d'équilibrage utilisées sont :

- la variable « un » contenant la valeur 1 pour que chaque groupe de rotation estime exactement le nombre total de logements dans la base de tirage.
- le nombre de logements individuels par îlot
- le nombre de logements collectifs par îlot
- le nombre d'individus de moins de 10 ans³⁹ par îlot
- le nombre d'individus de 10 à 19 ans par îlot
- le nombre d'individus de 20 à 39 ans par îlot
- le nombre d'individus de 40 à 59 ans par îlot
- le nombre d'individus de 60 ans et plus par îlot
- le nombre d'hommes par îlot
- le nombre de femmes par îlot
- la probabilité initiale de tirer le logement par groupement géographique pour avoir une taille fixe d'échantillon dans chacun des 8 groupements

La macro Cube a été utilisée, avec l'option d'atterrissage par relâchement successif des contraintes (dans l'ordre inverse de la présentation précédente). 1000 découpages en 5 groupes de rotation ont été effectués.

Pour choisir le « meilleur » découpage, on a d'abord sélectionné ceux pour lesquels un indicateur agrégé de qualité était le plus faible. Cet indicateur simple sommait les valeurs absolues des écarts (en pourcentage) entre les valeurs « vraies » (issues du recensement) et les valeurs estimées sur les cinq groupes de toutes les variables d'équilibrage. Pour ainsi dire, on a sélectionné à cette étape les découpages « les mieux équilibrés ». Vingt-deux découpages ont été alors retenus.

Le choix final parmi les vingt-deux découpages proposés a été pris au SR de Mayotte, à partir de considérations pratiques et de terrain.

6.2.6.2. Les cas spécifiques : les îlots dépeuplés et les îlots avec « grandes » adresses

Les 128 îlots dépeuplés ont été répartis afin d'anticiper un peuplement futur. Un protocole de tirage simplifié a été choisi, les variables d'équilibrage sont restreintes à :

- la variable « un » contenant la valeur 1 pour que chaque groupe de rotation estime exactement le nombre total de logements dans la base de tirage
- la probabilité initiale de tirer le logement par groupement géographique pour avoir une taille fixe d'échantillon dans chacun des 8 groupements.

Les 12 îlots contenant des « grandes adresses » (de 10 logements ou plus) ont été répartis manuellement de manière à ce qu'un nombre proche de logements en grande adresse soit assuré pour chaque groupe de rotation.

6.3. Le dispositif d'enquêtes à Mayotte depuis 2014

Le bâti mahorais est composé d'une très écrasante majorité de maisons individuelles (95% des adresses et 87% des logements, voir tableau suivant) et les quelques adresses collectives sont le plus souvent réduites à deux ou trois logements (les adresses de deux ou trois logements

³⁹ Âge révolu (AGER_R)

représentent 5% des adresses et 10% des logements ; les 283 adresses contenant plus de trois logements représentent 0,5% de l'ensemble des adresses et 3,5% des logements).

Il n'y a pas à redouter un effet de grappe particulièrement marqué si on échantillonne des adresses plutôt que des logements.

La base de sondage d'adresses est constituée par une cartographie d'un des groupes de rotation, réalisée l'année précédente.

On définit trois strates d'adresses pour l'échantillonnage :

- les mono-logements
- les petites adresses (2 à 9 logements) où l'enquêteur interroge l'ensemble des logements de l'adresse
- les grandes adresses (10 logements et plus) où l'enquêteur interroge 10 logements (tirés aléatoirement parmi l'ensemble des logements de l'adresse).

Les probabilités de tirage des adresses par strate sont ajustées pour approcher en moyenne l'équipondération des logements *in fine*.

Par strate, on peut alors réaliser un tirage équilibré ou un tirage systématique selon les critères suivants :

- micro-régions (petite-terre, nord, sud, métropole Mamoudzou-Koungou)
- type de bâti (bâti en dur vs. maisons en tôle/cases traditionnelles/bangas/...)

On aurait également pu faire un tirage systématique par taille des petites adresses dans la strate correspondante.

Pour « copier » le dispositif cartographique DOM 2004 – 2013, il resterait alors à construire une variable socio-démographique à l'IRIS, dans un esprit « Tabard » pour introduire une contrainte d'équilibrage supplémentaire.

nombre de logements de l'adresse	Fréquence	Pourcentage
1	51368	94.55
2	2310	4.25
3	363	0.67
4	127	0.23
5	26	0.05
6	41	0.08
7	15	0.03
8	27	0.05
9	3	0.01
10	10	0.02
11	6	0.01
12	11	0.02
14	3	0.01
15	1	0.00
16	2	0.00
18	1	0.00
19	1	0.00
20	2	0.00
21	1	0.00
22	1	0.00
24	1	0.00
26	1	0.00
27	1	0.00
28	1	0.00
29	1	0.00
30	2	0.00
48	1	0.00
81	1	0.00

Source : enquête cartographique préalable au RP

Bibliographie

Notes réalisées par le Criem :

« Quelques éléments de volumétrie pour la constitution de bases de sondage issues du recensement réalisé Outre-mer », N° 245/DR974-SES/ – 15 mars 2012

« Structure des groupes de rotation du recensement des Dom dans la perspective de l'échantillonnage des enquêtes auprès des ménages », N° 373/DR974-SES – 24 avril 2012

« Mise en place d'un dispositif pérenne de sondage pour les enquêtes auprès des ménages à Mayotte », N° 797/DR974-DR974-SES/ – 26 septembre 2013

« Méthode de constitution des 5 groupes de rotation mahorais », N° 39/DR974-SES/ – 23 janvier 2014

Documentation complémentaire :

« Les tirages des échantillons des enquêtes ménages dans les DOM à partir des bases du nouveau recensement de la population », Delphine Artaud, mémoire pour la *formation continue diplômante des attachés* - février 2006

« Rapport sur la stratégie d'enquête auprès des ménages dans les DOM », Pascal Ardilly et Philippe Domergue, Inspection générale, N° 84/DG75-B010/ML - 4 mai 2010

« Démarche Maiol – Unités des Enquêtes Ménages dans les DOM – Rapport sur l'analyse de l'existant », DR Insee La Réunion-Mayotte, n°609/DR974-SES/ - 12 juillet 2011

« Opération d'estimation de l'impact du tirage des ilots EAR sur l'EECDOM », Pascal Ardilly, DG Insee/DSDS/UMS, N° 2579/DG75-F401/ – 20 décembre 2011

« Démarche Maiol – Unités des Enquêtes Ménages dans les DOM – Rapport sur les organisations cibles », DR Insee La Réunion-Mayotte, n° 835/DR974-SES - octobre 2012

« Évaluation de la stratégie de l'Insee dans les DOM », Philippe Domergue et Jean-Pierre Le Gléau, Inspection générale, N° 41/DG75-B010/RP - 21 mai 2013

NB : les notes rédigés par les auteurs dans le cadre de leurs activités au Criem ont largement été reprises (et mises à jour) pour la rédaction de cet article.