

Du nouveau recensement au futur système d'échantillonnage des enquêtes ménages

Nathalie CARON et Marc CHRISTINE

Insee, DSDS, Unité Méthodes Statistiques

I. Introduction

Les échantillons des enquêtes nationales auprès des ménages réalisées par l'Insee sont sélectionnés dans des listes de logements ad hoc constituées à partir de chaque recensement de la population. Ces listes sont en général complétées par des sources annexes permettant la couverture des logements construits postérieurement au dernier recensement disponible.

Pour assurer le tirage des échantillons des enquêtes, des systèmes standardisés ont été conçus et mis en œuvre au moins depuis la décennie 60. Ceux-ci se sont perfectionnés au fil du temps :

- d'abord parce que les besoins en matière d'enquêtes ont évolué, surtout au cours des 15 ou 20 dernières années,
- ensuite parce que les progrès de l'informatique permettent de manipuler des fichiers de plus en plus complexes et volumineux avec des gains de temps et de stockage considérables et de disposer d'outils nouveaux (cartographie, saisie optique ...),
- et enfin aussi parce que les méthodes statistiques ont elles-mêmes évolué et se sont améliorées.

La philosophie d'ensemble des différents systèmes d'échantillonnage qui se sont succédé après chaque recensement est néanmoins peu modifiée. Elle résulte de compromis entre d'une part, des objectifs statistiques de précision et d'optimisation des plans de sondage à coût donné et, d'autre part, les contraintes induites par le choix de réaliser la plupart de ces enquêtes en face à face, en recourant à un réseau d'enquêteurs localisé à proximité des logements enquêtés et relativement stable dans le temps.

Le nouveau recensement de la population mis en œuvre depuis janvier 2004 substitue au comptage traditionnel organisé exhaustivement tous les 7 à 9 ans une technique d'enquêtes annuelles de recensement, basée sur un échantillonnage de communes ou d'adresses qui est différent selon que les communes ont plus ou moins de 10 000 habitants. Révolution dans l'histoire des recensements, elle implique de ce fait **une réflexion approfondie sur le nouveau système d'échantillonnage de l'ensemble des enquêtes ménages à mettre en place. L'objectif est de mettre en place ce nouveau système d'échantillonnage à l'horizon de 2008.**

Après avoir rappelé les principales caractéristiques du système d'échantillonnage actuel ainsi que celles du nouveau recensement, nous présentons dans le cadre de cette communication la synthèse de nos réflexions actuelles, stratégiques et méthodologiques, dans l'élaboration de l'architecture générale du futur système d'échantillonnage ainsi que les orientations privilégiées. Le cas des DOM sera évoqué en dernière partie.

II. Le système d'échantillonnage actuel et ses points faibles

II.1. Le système d'échantillonnage actuel¹

Le système actuel d'échantillonnage des enquêtes ménages nationales de l'Insee est organisé autour d'un **échantillon-maître** (EM) qui constitue la base de sondage principale dans laquelle sont sélectionnés la plupart des échantillons des enquêtes ménages nationales² **à l'exception de l'enquête Emploi.**

Après le premier échantillon-maître élaboré par M. Chartier en 1962, cinq autres ont été mis au point successivement après chaque recensement. Tous les échantillons-maîtres possèdent de grandes similitudes entre eux en ce qui concerne leur plan de sondage : stratification selon le type de communes permettant une distinction selon le degré d'urbanisation et, notamment, la séparation entre communes rurales et communes urbaines, plan de sondage à différents degrés selon la strate, constitution d'unités primaires à partir de divisions administratives comme le canton en zone rurale, utilisation d'unités géographiques comme l'unité urbaine... Les unités finales sont des logements, concentrés dans des unités primaires afin de ne pas disperser les lieux d'enquêtes et limiter les frais de déplacement des enquêteurs (surtout dans la partie rurale).

L'enquête Emploi constitue une exception notable au système de l'échantillon-maître : un système d'échantillonnage spécifique, dit *aréolaire*, a été mis en place depuis 1968 ; les unités finales d'échantillonnage sont des entités géographiques composées d'environ 20 logements voisins, appelées « aires », ce qui permet d'obtenir un « rendement important » de la collecte dans un laps de temps donné.

Ces deux systèmes d'échantillonnage et les bases de sondage sur lesquelles ceux-ci sont articulés sont renouvelés après chaque recensement de la population et restent fixes pendant la période intercensitaire (de l'ordre de 7-9 ans).

Afin d'assurer la représentativité du parc des logements lors des tirages d'échantillons des enquêtes ménages, ces bases de sondage sont complétées par un système permettant de sélectionner des logements « neufs », c'est-à-dire construits après le dernier recensement. Ainsi, pour compléter l'EM, a été mise en place une **Base de Sondage des Logements Neufs (BSLN)**, constituée à partir des listes de permis de construire gérées par le Ministère de l'Équipement (fichiers SITADEL). En ce qui concerne l'échantillon Emploi actuel, le complément de logements neufs échantillonnables est obtenu directement sur le terrain grâce à une opération confiée aux enquêteurs, baptisée **ratissage des aires**, qui permet le repérage et l'identification de tous les logements de chaque aire, non seulement ceux présents au dernier recensement (y compris ceux oubliés) mais aussi ceux construits après ce dernier. Il y a alors interrogation de la totalité ou d'une partie de ces logements, en fonction du nombre total de logements neufs repérés dans l'aire.

Le système d'échantillons (maître et Emploi) utilisé actuellement a été mis en place après le recensement de 1999. Parmi différentes innovations, on retiendra plus particulièrement l'utilisation systématique de la saisie optique des feuilles de logement du recensement lors de l'impression des fiches-adresses destinées aux enquêteurs et la création d'une base complémentaire (appelée **Echantillon Maître pour les EXTensions régionales - EMEX**), construite pour le tirage des enquêtes bénéficiant d'extensions régionales. Quant à l'échantillon Emploi, sa constitution s'est adaptée à la modification importante entrée en vigueur en 2001, à savoir le passage à une enquête « en continu », et, en particulier, aux implications qui en résultent en matière de rotation des aires.

¹ Pour plus de détails, nous invitons le lecteur à se référer à la description du système actuel en annexe 1.

² Il existe quelques exceptions, comme l'enquête Handicap-Incapacités et Dépendances, l'enquête auprès des Sans-domicile ou encore l'enquête mensuelle CAMME de conjoncture auprès des ménages, mais celles-ci sont en faible nombre.

II.2. Les points faibles du système actuel

Le système actuel d'échantillonnage de type EM donne globalement satisfaction aux différents acteurs concernés : responsables d'enquêtes, DR et utilisateurs de la chaîne de tirage. Signalons cependant que plus la base de sondage vieillit (autrement dit, plus on s'éloigne du recensement), plus les « déchets » lors de la collecte sont importants, ce qui augmente par conséquent les coûts. De même le système retenu pour le tirage de l'échantillon Emploi est qualifié de satisfaisant par les responsables de l'enquête Emploi, bien que l'opération de constitution des aires (baptisée « découpage ») demeure particulièrement lourde et coûteuse en moyens humains.

Cependant, plusieurs points sont susceptibles d'être améliorés dont voici les principaux :

- **BSLN**

La mécanique générale de constitution de la BSLN est très complexe et très lourde. En particulier, la base de sondage BSLN est alimentée à partir d'une source administrative, SITADEL, gérée en dehors de l'Insee, qui est utilisée ici à des fins statistiques. Cette utilisation crée des difficultés : par exemple, certaines informations importantes pour la constitution de la base de sondage (comme le nombre de logements) ne sont pas toujours fiables. Les délais de remontée de l'information à partir des permis de construire peuvent être plus ou moins longs, il y a donc un décalage entre la situation réelle de la construction neuve et l'image qu'en donne la base à un instant donné.

De plus, la construction de la base BSLN proprement dite nécessite un suivi sur le terrain des logements sélectionnés dans les permis de construire au moyen du réseau d'enquêteurs de l'Insee. Les informations recueillies sont ensuite saisies dans un logiciel de gestion spécifique implanté dans chaque direction régionale (DR). Toutes ces opérations de suivi et de saisie gérées par les DR sont coûteuses. Le contrôle de ces opérations est quasiment impossible de façon centralisée.

Au final, l'Insee n'a pas une maîtrise entière des différents éléments de la qualité de la BSLN et celle-ci reste inégale même si elle s'améliore au fil du temps (la durée d'achèvement des logements semble correcte mais le volume de la construction neuve estimé à partir de la BSLN elle-même reste en dessous des estimations de la construction neuve réalisées par le Ministère de l'Équipement).

- **Des « dimensions » non (ou mal) prises en compte dans le plan de sondage de l'EM/EMEX**

1. Une demande croissante a émergé au cours des derniers mois pour prendre en compte les Zones Urbaines Sensibles (ZUS) dans les analyses issues des enquêtes ménages, notamment pour analyser les disparités de comportements en ZUS / hors ZUS et, au minimum, pour avoir une représentation contrôlée des logements en ZUS par rapport à l'ensemble.

En l'état, cette demande entraîne des complications pour le tirage des échantillons des enquêtes concernées, puisque, par définition, la constitution de l'échantillon-maître de 1999 avait totalement ignoré la présence ou non de ZUS. Pour pallier cette difficulté, une base de sondage de logements spécifique sur les ZUS a été créée à partir du RP 99 en juin 2004. Comme l'échantillon-maître, cette base ne concerne donc pas la construction achevée après mars 1999.

Cette base a été utilisée pour la première fois à l'occasion de l'enquête « Information et Vie Quotidienne » d'octobre 2004 : cette enquête a donné lieu à un échantillonnage particulier dans les ZUS suite aux demandes de la Délégation Interministérielle à la Ville, un des financeurs de l'enquête. Un ensemble de programmes informatiques non intégré dans la chaîne de tirage permet de pratiquer un échantillonnage adapté aux demandes. Un échantillon tiré dans la base ZUS complète en général un tirage dans l'échantillon-maître. Les pondérations finales de l'échantillon doivent alors être « retravaillées » afin de tenir compte des deux parties d'échantillons issues de bases différentes. Notons toutefois qu'il est pratiquement impossible de couvrir correctement les logements neufs construits dans les ZUS.

2. L'utilisation de la tranche d'unité urbaine (code TU99) pour stratifier les unités urbaines et construire une strate d'unités primaires rurales a été fortement critiquée en 1999. La critique la plus importante porte sur le fait que le code TU99 mélange, au sein d'une même strate rurale, des communes très hétérogènes dans leurs comportements : les vraies rurales isolées et les rurales sous influence urbaine dont les habitants (c'est notamment le cas en Ile-de-France) ont des modes de vie, des contraintes ou des aspirations très liés à la proximité d'une grande agglomération et très similaires à ceux des résidents de cette dernière.

La notion de tranche d'unité urbaine avait cependant été conservée d'une part pour la simplicité de ce code (fondé exclusivement sur des critères de densité du bâti) et d'autre part parce qu'il n'était pas si simple de mettre en évidence une typologie alternative reconnue par le plus grand nombre et robuste. Signalons toutefois que la sélection des unités primaires a été réalisée par un tirage équilibré³ sur le revenu et l'âge des individus (en trois tranches) au niveau de regroupement de régions ; ceci a permis de prendre en compte et refléter une certaine hétérogénéité de la strate rurale.

- ***Des points faibles signalés par les enquêteurs***

Lors de la première réunion du groupe de réflexion sur les conditions de travail des enquêteurs qui s'est déroulée en novembre 2004, ceux-ci ont évoqué les points suivants directement liés au système d'échantillonnage :

- les logements détruits ou en cours de destruction (« barres » partiellement murées, par exemple) : peut-on incorporer cette information en amont pour éviter de tirer dans les échantillons d'enquête des logements de cette catégorie ?
- l'impact de zones à forte concentration de résidences secondaires (sur le littoral en particulier). Jusqu'à présent, on sélectionne dans les échantillons (en particulier ceux issus de l'EM) des résidences secondaires avec un taux de sous-représentation dans l'optique de « couvrir » les cas de changement de destination d'une résidence depuis le dernier recensement. Cette remarque renvoie aussi à celle de la concentration de l'échantillon. En effet, dans l'EM actuel, le tirage intermédiaire de districts à l'intérieur des unités urbaines de plus de 20 000 habitants conduit à accroître le risque de tomber, par hasard, sur une zone exclusivement constituée de résidences secondaires, alors que d'autres parties de l'unité urbaine peuvent être occupées par des logements principaux ; cette remarque vaut aussi pour l'échantillon Emploi, dans le cas où des aires entières tombent dans des zones purement touristiques.
- la sélection de sous-populations. De plus en plus d'enquêtes ont pour champ des populations ciblées : les « jeunes », les « vieux »... De ce fait, selon les unités primaires, les enquêteurs peuvent avoir très peu de logements dans le champ de l'enquête ; en effet, ce nombre n'est pas contrôlé a priori et par conséquent peut être très variable d'une unité primaire à une autre.

- ***Une chaîne de tirage manquant de souplesse***

De manière générale, la chaîne de tirage actuelle est rigide et s'adapte peu aux nouveaux besoins, que ce soit pour l'inclusion de nouvelles variables issues du RP99 dans la base de sondage, mais aussi pour la désignation et la gestion d'échantillons « de réserve » qui sont mobilisés en cours de collecte si le taux de non-réponse s'avère trop important. De même, l'application est inadaptée à la gestion des panels.

³ Pour plus de détails sur le tirage équilibré, voir par exemple la référence bibliographique [4].

- **La lourdeur des opérations de découpage des aires Emploi**

Les unités finales d'échantillonnage dans le système Emploi sont des aires de 20 logements. Comme aucune base de sondage d'aires n'existe, il est nécessaire de procéder au préalable à leur construction (voir annexe 1). Par conséquent, le plan de sondage de l'enquête Emploi nécessite des opérations de découpage des aires qui sont réalisées dans les directions régionales (DR). Ces opérations sont lourdes, complexes et nécessitent des moyens humains très importants en DR mais aussi au niveau central pour le suivi et le contrôle de la qualité des découpages.

III. Les transformations imposées par le nouveau recensement

Les bases de sondage actuelles EM/EMEX et l'échantillon Emploi constitués en 2001 (à partir des données du recensement de 1999) sont les derniers à être construits sur la base d'un recensement classique, avec une durée de vie comparable à celle de leurs prédécesseurs, soit environ 8 ans. La mise en place du nouveau recensement de la population entraîne en effet une modification radicale de la méthodologie employée en apportant de « l'information fraîche » chaque année mais non exhaustive sur l'ensemble du territoire.

Détaillons ce point en rappelant les principales spécifications du plan de sondage du nouveau recensement. Celui-ci distingue deux strates selon la taille des communes : la strate des petites communes (moins de 10 000 habitants) et la strate des grandes communes (10 000 habitants et plus).

- Les petites communes sont réparties en cinq groupes de rotation par un tirage aléatoire simple équilibré⁴. Chaque année, toutes les communes appartenant à un groupe de rotation donné sont enquêtées exhaustivement. En cinq ans, toutes les petites communes sont donc enquêtées.
- Pour les grandes communes, la constitution de la base de sondage se fait à partir du répertoire d'immeubles localisés (RIL). Dans chaque grande commune, trois strates sont construites : la strate des « grandes adresses », la strate des « adresses nouvelles » et celle des « autres adresses ».
 - Dans la strate « autres adresses » définie comme complémentaire des autres strates, cinq groupes de rotation sont constitués dans chaque commune (par tirage équilibré sur des variables démographiques). Dans le groupe de rotation retenu pour l'année, un échantillon d'adresses est sélectionné, représentant environ 8% des logements de la ville considérée. Au bout des cinq ans, 40% des logements des grandes communes auront été recensés.
 - Les « grandes adresses » qui correspondent aux adresses de plus de 60 logements sont introduites dans les cinq groupes de rotation (définis précédemment pour la commune considérée). Elles sont enquêtées exhaustivement au cours d'un cycle de cinq ans.
 - Pour les « nouvelles adresses »⁵ repérées comme telles dans le RIL lors de sa mise à jour, l'information disponible est limitée et pas toujours très fiable. Ces « adresses nouvelles » d'une année donnée sont réparties par tirage équilibré sur le nombre de logements (variable disponible dans le RIL) dans les cinq groupes de rotation et seront enquêtées exhaustivement au cours de chacune des années du cycle de cinq ans considéré à compter de la première année d'apparition. Dès qu'une « adresse

⁴ L'équilibrage est effectué au niveau régional en tenant compte du nombre de logements (individuels/collectifs), de la pyramide des âges en cinq classes et de la répartition de la population par sexe, données issues du recensement de 1999.

⁵ Dans les « nouvelles adresses », on retrouve en particulier les logements « neufs ». Plus précisément, une adresse neuve est soit une adresse qui n'est pas connue dans la base de sondage de l'année n-1 et qui est connue dans la base de sondage de l'année n soit une adresse dont le nombre de logements a suffisamment varié d'une année sur l'autre pour que des soupçons pèsent sur le bâti décrit en n-1 à cette adresse. On notera que la définition ne fait à aucun moment référence à la date d'achèvement du logement.

nouvelle » est recensée, elle rejoint la strate des « autres adresses » du groupe de rotation auquel elle appartient et participera au tirage de l'échantillon des autres adresses lors du prochain cycle du recensement.

En résumé, avec cette méthode de collecte, il y aura chaque année un recensement exhaustif dans environ 7 000 communes de moins de 10 000 habitants et une enquête de recensement par sondage dans toutes les communes de 10 000 habitants ou plus (soit environ 900 communes)⁶. Au bout de cinq ans, c'est-à-dire fin 2008, l'ensemble des petites communes auront été recensées ainsi que 40% des logements des grandes communes.

IV. Les orientations « stratégiques » du futur système d'échantillonnage : rémanence et évolutions

La nouvelle méthodologie du recensement rend nécessaire une refonte générale de celle de l'EM et de l'échantillon Emploi. Le lancement de cette refonte ne pouvait se faire sans une réflexion préalable d'ensemble conduisant à s'interroger sur la pérennité des principes sur lesquels étaient fondés les systèmes d'échantillonnage depuis 40 ans et sur l'émergence de besoins nouveaux à prendre en considération.

Cette phase indispensable, conduite entre l'automne 2003 et l'été 2004, s'est appuyée sur la consultation d'une trentaine de personnes (utilisateurs, producteurs, experts, régionaux...) et a permis de définir les grandes orientations pour le nouveau système en faisant émerger plusieurs idées fortes.

IV.1 Organisation de la collecte

Des idées très novatrices en matière d'organisation des enquêtes sont apparues (perspectives de collecte par Internet, utilisation de panels individuels de type « échantillon démographique permanent » comme base de sondage...).

Le mode de collecte en face à face constitue un des éléments qui garantissent la qualité des enquêtes et qui permettent la réalisation d'enquêtes complexes et longues. C'est la raison pour laquelle l'INSEE a développé et entretenu un réseau d'enquêteurs qui a acquis au fil du temps un véritable professionnalisme et une bonne connaissance de l'Insee. C'est aussi grâce à ce réseau très compétent que les enquêtes en face à face sont réalisées avec la qualité requise. A l'inverse, on voit mal comment la plupart des enquêtes ménages de l'INSEE, complexes et longues, pourraient être réalisées **autrement qu'en face à face**.

Certes, ce système comporte des inconvénients :

- l'interrogation en face à face génère des coûts très importants qui pourraient justifier la recherche de modes de collecte moins coûteux. D'ailleurs, il existe des enquêtes ménages peu complexes, comme par exemple les parties fixes des enquêtes Permanentes sur les conditions de vie des ménages, qui pourraient s'accommoder d'un mode de collecte différent et d'autres (comme l'enquête CAMME) qui utilisent déjà le mode de collecte par téléphone.
- le mode de collecte en face à face soulève dans certains cas des problèmes pratiques : l'accès au logement (sécurité accrue dans les programmes de logements neufs), l'interrogation dans des quartiers difficiles, des petits problèmes de déplacement (zones de montagne, îles ...).

⁶ Chaque année, on recense environ 3,1 millions de logements dans les petites communes et 1,2 million dans les grandes communes.

Il a donc été reconnu que serait toujours maintenu un noyau dur d'enquêtes réalisées en face à face et nécessitant un système d'échantillonnage du type échantillon-maître ou échantillon aréolaire (pour l'enquête Emploi), ces systèmes devant prendre en compte les contraintes liées au réseau d'enquêteurs : à la fois minimiser les coûts de déplacement et bénéficier de la faculté de disposer d'un réseau expérimenté.

C'est donc sur ce noyau dur que nos réflexions se sont concentrées.

IV.2 Utilisation de l'information la plus fraîche disponible

Pour autant, un certain nombre de besoins, fonctionnels ou en termes d'outils, doivent être pris en considération pour la construction d'un nouveau système.

Du point de vue des utilisateurs des enquêtes, les perspectives offertes par le nouveau recensement portent surtout sur la **fraîcheur de la base de sondage** obtenue en réduisant au maximum le décalage temporel qui existe entre la date de collecte de l'enquête et la date de recensement.

La qualité d'une base de sondage (et par conséquent celle des échantillons qui en découlent) dépend en effet directement de sa « fraîcheur ». Plus la base de sondage vieillit, moindre est sa qualité : déchets liés à la disparition ou la transformation de logements plus nombreux, changements des occupants du logement qui peuvent entraîner des difficultés de repérage et qui rendent moins pertinentes les caractéristiques connues au recensement sur le ménage pour cibler au mieux la population d'intérêt (cas des enquêtes en deux phases), augmentation corrélative de la proportion de logements neufs, tirés dans une base de moindre qualité que le recensement.

L'hypothèse privilégiée consiste donc à utiliser comme base de sondage d'une année donnée les logements du dernier groupe de rotation enquêté au titre du recensement.

Un des corollaires importants de ce choix est de **pouvoir s'affranchir d'un système complémentaire de type BSLN pour l'échantillonnage des logements « neufs »**, qui constitue l'un des principaux points faibles du système actuel mentionnés dans la partie II.2.

Cette hypothèse doit être compatible avec les autres aspects de la méthodologie et, notamment, les contraintes induites par la gestion du réseau d'enquêteurs en relation avec le renouvellement des unités (unités primaires, aires Emploi...) (voir partie V.2.).

IV.3 Prise en compte des besoins locaux

Le nouveau système devra permettre une meilleure prise en compte des besoins régionaux ou localisés.

Jusqu'à présent, selon la nature de la demande régionale, le service d'échantillonnage proposé était différent. En effet, l'EMEX créé à l'occasion du dernier renouvellement de l'échantillon maître (basé sur le recensement de 1999) ne répond par définition qu'aux demandes d'extensions régionales. L'avantage d'un tel système est d'homogénéiser la conception, le tirage et le traitement des extensions régionales (donc de favoriser les comparaisons inter-régionales et avec le niveau national), alors qu'auparavant chaque DR gérait son extension à peu près comme elle l'entendait et l'intégration des données de l'extension dans les résultats nationaux n'était pas assurée.

En revanche, si la demande n'est pas celle d'une extension régionale d'une enquête nationale mais d'une enquête purement locale, le tirage est en principe actuellement réalisé par la direction régionale concernée et il n'existe aucun outil (programmes informatiques par exemple) facilitant le tirage des échantillons.

La prise en compte des besoins locaux conduira notamment à intégrer au nouveau système un ensemble de fonctionnalités permettant :

- le tirage d'extensions régionales d'enquêtes nationales (« EMEX ») comme c'est déjà le cas avec le système actuel, mais avec la différence fondamentale que, cet EMEX étant conçu de manière concomitante à l'EM, un certain nombre des difficultés théoriques et pratiques rencontrées en 2001-2002 n'auront plus lieu d'être (voir [2])⁷,
- la fourniture d'outils aux DR pour le tirage d'enquêtes locales,
- et la prise en compte des Zones Urbaines Sensibles dans la construction des bases de sondage (voir partie II.2.).

IV.4 Amélioration des critères de stratification

Les critères de stratification géographique seront également revus pour mieux décrire la diversité du territoire, tant au niveau supra communal qu'en infra-communal.

L'objectif d'une stratification est essentiellement d'améliorer la précision des estimateurs ; il faut pour cela trouver des critères de stratification ayant un pouvoir explicatif important vis-à-vis des principaux comportements analysés dans les enquêtes auprès des ménages et permettant la constitution de segments de la population relativement homogènes.

D'une manière générale, toute une gamme de critères de stratification s'appuient sur les aspects géographiques ; ceci conduit naturellement à la notion de *zonage*. Dans le cadre de cette réflexion, il ne s'agit pas de faire une typologie de tous les zonages possibles : il faut distinguer les zonages concernant la problématique de l'échantillonnage à travers la *stratification des zonages d'étude*.

Les caractères attendus d'un zonage sont sa *robustesse* vis-à-vis de paramétrages ou de choix techniques, sa *pertinence* (c'est-à-dire un pouvoir explicatif suffisant) vis-à-vis des phénomènes étudiés dans les enquêtes et son *actualisation régulière* (avec une certaine stabilité : un zonage ne doit pas être redéfini chaque année et un cycle de 5 ans, voire 10 ans, pourrait être une bonne périodicité).

- **Les zonages supra-communaux**

La segmentation régionale sera vraisemblablement maintenue, d'une part parce que la région constitue « l'unité de gestion » du réseau d'enquêteurs et d'autre part parce qu'elle est aussi une *unité d'analyse*. En effet, la région constituant une unité de décision et de (très) relative homogénéité socio-économique justifie l'obtention de certains résultats régionaux pour certaines enquêtes, a fortiori dans les cas de demandes d'extension régionale auxquelles l'EMEX est destiné à répondre (voir partie IV.3).

L'alternative à l'utilisation de la tranche d'unité urbaine (voir partie II.2) pourrait être le **zonage en aires urbaines et aires d'emploi dans l'espace rural (ZAUER)** qui est fondé à la fois sur des données de démographie des logements mais aussi sur les comparaisons des effectifs des actifs occupés au lieu de résidence et au lieu de travail. Ce zonage qui conduit à une typologie de l'espace, distinguant les espaces à dominante urbaine (aires urbaines et communes multipolarisées) et les espaces à dominante rurale (rural sous faible influence urbaine, pôles ruraux et périphérie de ces pôles, rural isolé) présente l'avantage, notamment, de **bien distinguer le rural « vrai » du rural sous influence urbaine, ce que ne permet pas le code TU99** (voir partie II.2.).

⁷ Ces difficultés ont principalement comme origine le fait que la constitution de l'EMEX s'est déroulée postérieurement à celle de l'EM99.

Des critères complémentaires pourraient être utilisés comme ceux fondés sur les accès aux services et l'organisation territoriale des équipements. Ainsi, la notion de « bassin de vie »⁸ se définit comme le plus petit territoire sur lequel les habitants ont un accès aux principaux services et à l'emploi.

Une autre solution pourrait être de construire des typologies communales ad-hoc à partir d'autres critères, par exemple : taux de personnes âgées, importance du caractère touristique (à définir et mesurer, par exemple par le poids de la capacité hôtelière par rapport aux logements principaux d'une zone ou d'une commune, ou à partir d'une nomenclature de « lieux » touristiques, ceci permettant de mieux prendre en compte les remarques des enquêteurs sur la concentration de certains points d'enquête dans des zones touristiques, voir supra, partie II.2), orientation socio-économique dominante (typologie des postes de travail au lieu de travail, source possible DADS...).

A l'inverse, d'autres zonages supra-communaux du territoire, soit administratifs comme les entités de coopération intercommunale (EPCI, communautés de communes), soit d'étude ou ne présentant pas de garantie concernant leur maintenance, comme les zones d'emploi⁹, ne sont pas retenus par la suite.

• **Les zonages infra-communaux**

La question des zonages infra-communaux se pose principalement pour les villes ou unités urbaines de grande taille, a fortiori si elles sont retenues d'emblée dans l'échantillon.

La stratification infra-communale doit permettre de définir un découpage à l'intérieur des communes (ou de tout autre agrégat) en zones au sein desquelles seront sélectionnés des logements, de manière à avoir la meilleure représentation possible de l'hétérogénéité urbaine.

Pour définir ces zones en infra-communal, **la « maille de base » la plus naturelle est l'IRIS**, l'îlot constituant une maille beaucoup trop fine)

On peut envisager plusieurs critères infra-communaux ad-hoc corrélés à la richesse et basés par exemple sur les revenus déclarés (total ou moyenne) ou le prix moyen au m² du logement dans la zone considérée. D'autres critères existent, comme le **code TABARD** : il s'agit d'une variable composite, abstraite, résultant d'une AFC, mais il n'y a pas de perspective claire sur sa maintenance ou son actualisation.

En ce qui concerne les ZUS, on a évoqué ci-dessus l'émergence des besoins en la matière. L'importance de la demande dans le contexte de l'analyse de la précarité ou de la pauvreté conduit à réfléchir à l'introduction de cette typologie dans la construction des futurs échantillons. Une manière de les prendre en compte serait par exemple :

- au niveau supra, de distinguer les communes (ou tout autre agrégat) selon la présence ou non de ZUS en leur sein.
- au niveau infra, de partitionner les IRIS selon qu'ils sont ou non inclus dans une ZUS ; évidemment, la notion de ZUS est orthogonale à la partition en IRIS, puisqu'une ZUS peut chevaucher plusieurs IRIS et même plusieurs îlots. Des conventions seraient nécessaires pour décider d'affecter ou non tout un IRIS à une ZUS qui le couvrirait partiellement. Ensuite, il faudrait vraisemblablement compléter la typologie en définissant une stratification des IRIS hors ZUS, qui restent majoritaires¹⁰.

⁸ Le découpage du territoire en bassins de vie est conçu comme outil d'analyse de l'espace rural. Les 1715 bassins de vie identifiés regroupent 36% de la population française.

⁹ Ce zonage a été créé dans les années 80 par le Ministère du travail et revu en 1994 par la DARES. A notre connaissance, aucune maintenance n'est actuellement prévue.

¹⁰ Le nombre total de logements situés dans des ZUS et celui des logements situés dans les communes impactées par les ZUS sont respectivement actuellement de 1,8 millions et de 9 millions.

Notons que la liste des ZUS est fixée par décret¹¹ et est par conséquent susceptible d'évoluer (de nouvelles ZUS peuvent apparaître, le périmètre d'une ZUS peut être modifiée). Cependant, leur intérêt est suffisant pour que l'on les prenne en considération.

IV. 5 Automatisation du processus

Enfin, certaines phases du processus seront **automatisées** : découpage des aires Emploi, constitution des unités primaires...

Pour alléger la charge matérielle du découpage des aires Emploi, une piste pourrait être d'automatiser leur constitution, au moins pour les communes de 10 000 habitants ou plus, grâce au RIL. Il s'agirait alors de constituer des groupes de 20 logements - taille standard d'une aire -, en s'appuyant sur la notion d'adresse et sur la répartition des logements au sein des adresses. Des regroupements automatiques d'adresses « contiguës » (concept à définir) devraient être réalisés dans certains cas.

La constitution des unités primaires devra également être automatisée au maximum avec l'utilisation de la cartographie mais aussi avec des outils d'aide au regroupement tout en vérifiant la dispersion géographique de l'échantillon obtenu et sa « compatibilité » avec le réseau d'enquêteurs.

V. Les travaux méthodologiques

Les travaux méthodologiques concernant les plans de sondage des échantillon-maître et Emploi sont actuellement en cours et vont se poursuivre durant l'année 2005.

V.1 L'échantillon-maître

Les réflexions ont débuté plus particulièrement sur les critères de stratification possibles ainsi que sur la façon de construire les unités primaires dans le rural.

En zone urbaine, quelques décisions sont d'ores et déjà prises : en particulier, il est envisagé *d'abandonner l'étape du tirage de districts*, ce qui devrait permettre de limiter un des problèmes soulevés par les enquêteurs concernant la concentration des résidences secondaires dans certaines zones d'enquête (voir partie II.2).

Pour la partie rurale de l'échantillon-maître, l'état actuel de nos réflexions est synthétisé dans ce qui suit.

Afin de concentrer les enquêtes et par conséquent limiter les frais de déplacements des enquêteurs, le plan de sondage du nouvel échantillon-maître sera, tout comme les précédents, un plan de sondage stratifié à plusieurs degrés. Celui-ci nécessite, d'une part, de définir les critères de stratification adéquats (qui permettent de distinguer en particulier les communes à caractère « rural » des autres, ces dernières étant classées selon un certain degré d'urbanisation) et, d'autre part, de constituer des unités primaires, en particulier dans le rural, car c'est là que se pose principalement le problème de la concentration géographique en liaison avec le plan de rotation du recensement.

La piste privilégiée concernant les critères de stratification consiste, comme nous l'avons signalé ci-dessus, à *partir du zonage en aires urbaines* qui permet en particulier de séparer l'espace à dominante rurale de l'espace à dominante urbaine.

¹¹ La première liste des ZUS a été fixée en 1996. Elle a été modifiée une première fois par décret en 2000 (ajout d'une ZUS) et une seconde fois en 2001 (modification du périmètre d'une ancienne ZUS).

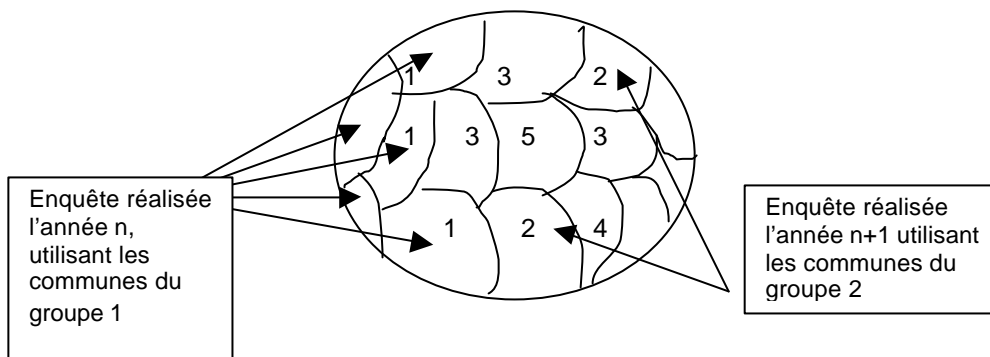
Pour la constitution des unités primaires, l'hypothèse privilégiée consiste à *sélectionner les échantillons des enquêtes d'une année donnée dans la partie recensée la plus « récente » et, si possible, celle de l'année précédente*. Alors, la répartition spatiale des communes impactées chaque année est imposée et résulte de la constitution aléatoire des groupes de rotation. Comme chaque unité primaire doit avoir a priori des contours fixes, la difficulté suivante apparaît : en définissant une unité primaire comme une zone fixe choisie ex-ante indépendamment de la composition en communes des différents groupes de rotation, celle-ci comporterait un nombre aléatoire de communes de chaque groupe de rotation (de tailles également aléatoires).

Ceci est problématique pour les deux raisons suivantes :

- les unités primaires ainsi constituées peuvent ne pas comporter des communes de tous les groupes de rotation
- même si c'est le cas, la réserve dans chaque groupe peut ne pas être suffisante en cas de non ré-interrogation des logements (quand la fraction fraîchement recensée d'une unité primaire va s'épuiser, on aura des problèmes pour assurer le tirage de l'allocation demandée).

D'autres problèmes statistiques (en particulier pour assurer l'équipondération finale des logements tirés) sont également présents.

Pour pallier cette difficulté, l'idée serait de pouvoir construire des unités primaires (dans la partie rurale du territoire) par agrégation de communes issues des 5 groupes de rotation, en assurant si possible que chaque groupe de rotation possède à peu près le même nombre de logements dans l'unité primaire, tout en respectant le principe de contiguïté des communes. Ces unités primaires seraient donc du type représenté dans le schéma ci-dessous, où les communes composant l'UP sont indiquées et où le chiffre correspond au groupe de rotation de la commune.



Une année donnée n , les enquêtes sont réalisées dans la partie la plus récemment recensée, disons pour simplifier dans les communes du groupe de rotation 1 ; l'année suivante, les enquêtes sont réalisées dans les communes du groupe de rotation 2 et ainsi de suite.

Cette optique permettrait donc de pouvoir assurer une certaine concentration de la collecte sur des zones géographiques fixes, définies avec une certaine durée de vie, tout en s'assurant de sélectionner les échantillons dans la partie la plus « fraîche » de ces zones ; une fraction de ces zones serait ainsi mobilisée chaque année en termes de base de sondage et chaque fraction aurait approximativement le même poids en nombre de logements au sein de la zone.

Il est à noter que les zones ainsi constituées devraient probablement être plus « étendues » que les unités primaires actuelles qui correspondent en général à des cantons, des regroupements ou parfois des fractions de cantons.

Ces zones qui joueront le rôle d'unités primaires dans le nouveau système sont appelées **Zones d'Action des Enquêteurs (ZAE)**. En effet, vis-à-vis du réseau d'enquêteurs, ces zones vont jouer un rôle similaire à celui des unités primaires actuelles. Mais elles devraient permettre aussi d'introduire une plus grande souplesse dans la gestion spatiale du réseau - à taille constante - tout en poursuivant l'objectif de limiter les frais de déplacement. En effet, une solution pourrait être qu'un enquêteur donné (en zone rurale) ne soit plus affecté à une unité primaire fixée mais plutôt considérer qu'un pool d'enquêteurs soit attaché à plusieurs ZAE, l'affectation finale des enquêteurs dépendant des communes effectivement mobilisées pour les enquêtes d'une année donnée. Un autre paramètre à prendre en considération est la taille du réseau d'enquêteurs. Celle-ci est liée à la charge moyenne d'enquêtes par enquêteur et à la longueur moyenne de leurs déplacements. Ces paramètres n'ont a priori pas d'impact sur le mode de constitution des ZAE, si ce n'est sur leur « étendue », mais ces questions devront être débattues dans le cadre du dialogue social habituel avec les enquêteurs.

Un des premiers travaux à réaliser va consister à regarder si la constitution des ZAE telles qu'elles ont été décrites dans cette partie est techniquement possible dans la partie rurale, c'est-à-dire si l'on peut simultanément satisfaire les contraintes de taille, de composition en communes des 5 groupes de rotation et de connexité des ZAE. Si c'est le cas, il faudra ensuite vérifier si les ZAE ainsi construites ne risquent pas de générer des déplacements en moyenne trop longs pour les enquêteurs. Dans ce but, il sera nécessaire d'évaluer, à partir de simulations de tirages d'enquêtes, la distance moyenne parcourue par un enquêteur donné (en partant par exemple de la configuration actuelle du réseau d'enquêteurs). En effet, le choix du plan de sondage est très dépendant **du rayon d'action envisageable pour les enquêteurs**.

Si la constitution des ZAE telle que décrite ci-dessus n'est pas possible, il sera nécessaire de privilégier une autre solution. Celle-ci pourrait être d'utiliser comme base de sondage l'empilement des bases annuelles relatives aux 5 derniers recensements et donc d'abandonner le principe d'un tirage dans la partie la plus « fraîche » (ce qui posera la question des logements neufs).

Les travaux de simulation effectués devraient permettre de répondre à cette question au cours de l'année 2005.

Une fois que les unités primaires sont définies, il est alors nécessaire de s'interroger sur la *périodicité de leur renouvellement*. Pour conserver une certaine stabilité temporelle au réseau des enquêteurs, il semble irréaliste de changer d'unités primaires trop souvent et, par suite, de sélectionner un nouvel échantillon-maître chaque année. Les unités primaires (ou ZAE) doivent donc être fixées pour une certaine période qui pourrait par exemple être 5 ans, voire même 10 ans¹² (ces chiffres correspondant à des multiples de cycle du recensement).

Outre la constitution des unités primaires dans la partie rurale, il existe de nombreux points méthodologiques sur la partie EM qui restent à résoudre. En particulier, il est nécessaire de :

- poursuivre la réflexion sur le plan de sondage pour les communes ne figurant pas dans la catégorie « rurale ». Dans la strate « petit urbain », il faudra en particulier tenir compte, lors de la constitution du plan de sondage, de la distinction obligatoire induite par le recensement entre petites communes et grandes communes. En effet, dans le plan de sondage du nouveau recensement, la notion d'unité urbaine présente dans la constitution des échantillons-maîtres depuis 1962 n'apparaît pas directement puisque le recensement distingue deux types de communes (les 10 000 habitants et plus et les autres) et s'appuie sur un plan de sondage rotatif sur cinq ans différencié selon ces types. De plus, les grandes communes ne sont interrogées que partiellement. Or les unités urbaines au sens précédent comportent à la fois des petites et des grandes communes.
- choisir la stratification infra-communale pour les communes urbaines.

¹² Actuellement, l'échantillon-maître étant fixé une fois pour toutes entre deux recensements, les unités primaires sont renouvelées tous les 10 ans.

- en matière de tirage des unités de 1^{er} ou de 2nd degré, la technique de l'*équilibrage* (utilisant la macro CUBE), déjà utilisée pour le tirage des unités primaires de l'échantillon-maître 1999, devrait pouvoir l'être à nouveau pour le tirage des ZAE. En ce qui concerne le tirage des logements, c'est aussi une perspective envisageable, à condition de définir les variables d'équilibrage adéquates. Cela a d'ailleurs été mis en œuvre pour le tirage des adresses dans les nouveaux échantillons des DOM (voir partie VI).

V.2 L'échantillon Emploi

Comme on l'a déjà signalé ci-dessus, l'échantillon Emploi a la particularité d'être *aréolaire* (voir annexe 1). Cette caractéristique essentielle ne semble pas être remise en cause sur le plan stratégique (en particulier, le maintien de l'aréolaire est demandé par le principal « client », la division Emploi) ; néanmoins elle doit être analysée de près pour vérifier sa compatibilité avec les autres principes énoncés à la partie IV.

Les deux difficultés majeures de la méthodologie relative à cet échantillon sont donc d'une part le caractère *aréolaire* de l'enquête et d'autre part son caractère *rotatif* (avec un renouvellement actuel d'1/6 par trimestre). Ces difficultés sont accrues par le caractère lui-même rotatif du recensement.

Bien que les travaux méthodologiques soient loin d'être finalisés, voici quelques lignes directrices.

- **Articulation avec le recensement le plus récent.**

Le point méthodologique le plus délicat consiste à articuler la rotation de l'échantillon (et par conséquent des aires) avec celle des groupes de rotation du recensement. En maintenant comme principe de base la sélection de l'échantillon Emploi dans la partie recensée la plus récente, les problèmes suivants apparaissent :

- en zone rurale pour les communes de moins de 10 000 habitants, les aires seront découpées à l'intérieur d'une commune donnée, qui appartient à un groupe de rotation du RP donné ; si l'on veut conserver une certaine concentration des points d'impact de l'échantillon, de façon que, d'une année sur l'autre il n'y ait pas de variation trop brutale de la localisation géographique (aussi bien en termes de gestion du réseau d'enquêteurs que pour minimiser la variance due à une variation trop forte de l'échantillon), les communes des 5 groupes de rotation devront être regroupées pour former une entité impactée au cours d'un cycle de 5 ans commune par commune. Ces entités, bien que plus petites, seraient donc construites sur des principes analogues à ceux des ZAE définies ci-dessus (voir partie V.1).
- dans les grandes communes (au sens du RP), l'éclatement des adresses en 5 groupes de rotation entraîne que, soit il est possible de constituer une même aire au sein d'une même adresse, soit il faut relâcher la contrainte de connexité des aires en fabriquant une aire à cheval sur deux ou plusieurs adresses, proches mais non nécessairement contiguës : dans un « grain connexe » de 20 logements, on risque en effet de trouver des logements de différents groupes de rotation.
- dans les communes de moins de 10.000 habitants (donc recensées exhaustivement tous les 5 ans) qui ne se situent pas en zone rurale, le problème est plus complexe, car l'association de communes contiguës appartenant aux différents groupes de rotation, comme il est proposé en rural, ne sera plus possible.

• **Contraintes dues à la rotation de l'échantillon**

Actuellement, l'échantillon Emploi est renouvelé par 1/6 tous les trimestres. Ce renouvellement est assuré de la manière suivante : chaque trimestre, une aire interrogée sort de l'échantillon et est remplacée par une aire voisine, concrètement appartenant à un même secteur au sein d'une même unité « de 1^{ère} phase »¹³. Le système des unités de 1^{ère} phase est fixe, des groupes de 6 aires sont constitués en leur sein et une seule aire de chaque unité est interrogée chaque trimestre. Les aires interrogées en 1^{ère} interrogation le sont par visite.

Comment concilier cette règle avec la rotation propre du RP, dont le rythme naturel est quinquennal ?
Regardons par exemple un mécanisme possible en zone rurale.

		GR1				GR2				GR3				GR4				GR5			
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
ANN E 1	T1	X																			
	T2	X	X																		
	T3	X	X	X																	
	T4	X	X	X	X																
ANN E 2	T1	X	X	X	X	X															
	T2	X	X	X	X	X	X														
	T3		X	X	X	X	X	X													
	T4			X	X	X	X	X	X												
ANN E 3	T1				X	X	X	X	X												
	T2					X	X	X	X	X											
	T3						X	X	X	X	X										
	T4							X	X	X	X	X									
ANN E 4	T1								X	X	X	X	X								
	T2									X	X	X	X	X							
	T3										X	X	X	X	X						
	T4											X	X	X	X	X					
ANN E 5	T1												X	X	X	X	X				
	T2													X	X	X	X	X			
	T3														X	X	X	X	X		
	T4															X	X	X	X	X	
ANN E 6	T1	X														X	X	X	X	X	
	T2	X	X														X	X	X	X	
	T3	X	X	X														X	X	X	
	T4	X	X	X	X														X	X	
ANN E 7	T1	X	X	X	X	X														X	
	T2	X	X	X	X	X	X														
	T3		X	X	X	X	X	X													
	T4			X	X	X	X	X													
ANN E 8	T1				X	X	X	X													
	T2					X	X	X													
	T3						X	X													
	T4							X													

¹³ On reprend ici la terminologie employée lors de la constitution de l'échantillon Emploi 2000-2001. Cette terminologie peut être contestée, le terme d'Unité primaire semblant plus approprié, mais elle est conservée :

- par continuité avec le précédent échantillon Emploi
- pour les différencier des UP de l'EM (qui seront elles-mêmes rebaptisées ZAE)
- parce que ces unités, une fois tirées, sont ensuite assujetties à une opération constituant une 2^{ème} phase : acquisition d'information, notamment cartographique, en vue de son découpage géographique, pour constituer des secteurs, puis tirage de secteurs à nouveau redécoupés en aires ; l'étape finale de tirage des aires correspond à un véritable dernier degré.

GR correspond au groupe de rotation du RP, S au numéro de sous-échantillon et T au numéro de trimestre.

Chaque croix représente une interrogation d'une aire, pendant 6 trimestres consécutifs. Les sous-échantillons sont introduits de manière séquentielle. Quand une aire du sous-échantillon 1 du GR1 a terminé son cycle, elle est remplacée par une aire du sous-échantillon 3 du GR2 (cases surlignées). L'unité contenant cette nouvelle aire doit être géographiquement proche de celle ayant alimenté la première, d'où l'intérêt de constituer des groupes de communes connexes appartenant aux différents groupes de rotation.

Ainsi, chaque trimestre, 6 fractions d'échantillon sont interrogées. Les aires entrantes sont introduites chaque trimestre, de façon à toujours être sélectionnées dans le groupe de rotation qui aura été recensé l'année précédente. Cela conduit à constituer 4 sous-échantillons d'aires au sein de chaque commune sélectionnée comme unité de 1^{ère} phase (S1 à S4). Dans ce schéma, ce ne sont pas nécessairement les mêmes communes qui alimentent les sous-échantillons introduits respectivement aux 4 premiers trimestres de l'année 1, mais ce sont toutes des communes du GR1.

Les communes seront donc regroupées par 5 (une par GR) et ces groupes de communes seront répartis en 4 paquets.

Au bout de 5 ans (cycle du RP), on est amené à revenir sur les communes ayant alimenté le 1^{er} trimestre de la 1^{ère} année. Soit on conserve la même commune et on interroge une nouvelle aire qui va constituer une 2^{ème} série, soit on renouvelle les communes.

Ce schéma n'a qu'une valeur indicative ; il conviendrait de le généraliser pour voir comment il s'adapte à un rythme de réinterrogation quelconque, étant entendu que toutes les solutions ne sont pas possibles. En particulier, l'équilibre général impose des contraintes au rythme de réinterrogation, faute de quoi un «hiatus » au moment du passage à la 6^{ème} année risque de se produire, le paquet de communes remobilisé n'étant pas nécessairement le même que celui ayant alimenté le trimestre 4 de l'année 5.

- **Constitution et tirage des « unités de 1^{ère} phase »**

Comme on l'a dit ci-dessus, il serait judicieux, au moins en zone rurale, de regrouper les communes par groupe de 5, une de chaque GR. La question devra être approfondie pour les communes non rurales ; bien entendu, il faudra au préalable définir une *stratification* ad hoc de l'espace, à l'instar vraisemblablement de ce qui sera fait pour l'échantillon-maître. Ensuite, il faut *définir les unités statistiques* qui vont servir d'unités de première phase et les sélectionner.

Le tirage pourrait être réalisé en imposant des conditions d'équilibrage, ce qui n'avait pas été fait en 1999-2000. En particulier, des gains en précision devraient pouvoir être obtenus (ou, de manière alternative, des économies de moyens rendues possibles par une réduction de l'échantillon à précision constante) grâce à un équilibrage adaptant l'échantillon aux principales variables mesurées dans l'enquête Emploi : chômage, activité. *Aussi cherchera-t-on à utiliser soit les données issues du recensement le plus récent, soit les estimations détaillées en cumul, en matière de marché du travail, pour les introduire comme variable d'équilibrage.*

- **Calibrage des aires**

Il est aussi nécessaire de définir si le calibrage des aires doit se faire sur le seul nombre de logements principaux (et non totaux). Si oui, cela risquerait de compliquer les opérations de découpage, les contours géométriques d'une aire intégrant nécessairement des résidences secondaires dont il faudrait néanmoins tenir compte dans la description, voire dans l'étendue géographique de l'aire.

- ***Tirage des aires***

Il faudra définir le mode de tirage le plus adéquat. L'idée d'un équilibrage pour ce tirage n'est pas nécessairement opérationnelle, dans la mesure où il faudrait pouvoir disposer de données exogènes (principalement du RP) à un niveau de détail extrêmement fin.

De plus, en fonction des décisions prises en matière de renouvellement (voir infra), il faudrait voir quelle référence temporelle on utilise pour le recensement. Ainsi, si le découpage des aires n'est pas annuel, cela veut dire clairement que leur sélection s'appuiera sur des données disponibles au moment du tirage et que ce sont ces données qui gouverneront la constitution de l'échantillon d'aires pendant toute leur durée de vie. Par ailleurs, si l'un des objectifs essentiels de l'enquête Emploi est la mesure des variations infra-annuelles de l'activité et du chômage, on aura intérêt à stabiliser relativement l'échantillon dans le temps afin de réduire la variance des estimations de variations. Ceci plaide plutôt pour un tirage d'aires non renouvelé chaque année et s'appuyant en conséquence sur des données cumulées.

- ***Ratissage des aires***

Une autre question importante est relative à l'opportunité du maintien et aux modalités de réalisation de la phase de *ratissage préalable des aires*. Cette opération est liée traditionnellement à l'idée d'un contrôle de l'exhaustivité de la base de sondage. On vérifie en particulier que les logements les plus précaires n'ont pas été oubliés. Mais elle permet également l'actualisation de la base, de façon à couvrir la construction « neuve ».

Si la base de sondage dans laquelle seront puisés les logements enquêtés est celle du recensement de l'année précédente, cette phase ne se justifie sans doute plus. Si on le maintient tout de même, comment intégrer ses résultats du ratissage dans la construction de la base de sondage ou dans le processus de pondération ?

- ***Rythme de renouvellement des aires et des procédures de découpage.***

Dans le cadre du système aréolaire se pose inévitablement la *question de la périodicité des opérations de découpage des aires* qui sont réalisées dans les directions régionales (la constitution des aires résultant de travaux dont une bonne partie restera manuelle). Ce point n'est pas tranché à ce stade.

Il faudra donc préciser s'il est préférable d'envisager des redécoupages complets des aires à périodicité régulière (par exemple tous les 5 ans) ou bien un étalement chaque année portant sur une fraction annuelle à renouveler. On peut dire d'ores et déjà qu'un découpage plus fréquent qu'antérieurement permettrait d'assurer une meilleure continuité entre les méthodes et une meilleure mémoire des tâches à accomplir, chez les personnes en charge du travail, sans parler d'une meilleure conservation des consignes.

Comme indiqué ci-dessus (partie IV.5), il sera en tout cas souhaitable de ***chercher à réduire la charge matérielle du découpage.***

Nombre des questions évoquées devront bien entendues être expertisées. Il faudra également rester tout à fait attentif aux évolutions de court terme de l'enquête Emploi qui pourraient avoir une incidence sur l'échantillonnage et, notamment, celles relatives au plan de rotation. Et sans doute sera-t-il nécessaire de faire différentes études méthodologiques pour arbitrer entre plusieurs solutions dans certains cas.

VI. Des précurseurs du « nouveau système » : les DOM

Jusqu'à présent, il n'a existé aucun système général d'échantillon-maître pour les DOM. Les échantillons des enquêtes nationales ayant une déclinaison DOM, relativement rares dans le passé¹⁴, étaient tirés au cas par cas (exemple : l'enquête Budget de Famille de 2000) selon des méthodologies spécifiques. Autrement dit, les systèmes actuels présentés en annexe 1 ne s'appliquent qu'à la France métropolitaine.

Lors de la refonte du système d'échantillonnage de l'enquête Emploi DOM en 2001, le nouveau système qui a été mis en place pour cette enquête avait aussi pour objectif de servir de « réserve » pour d'autres enquêtes « sociales ». Cet échantillon, construit différemment de celui réalisé en métropole, allie une méthode de type échantillon-maître, avec constitution et tirage d'unités primaires fixes¹⁵, avec la mise en œuvre d'un « ratissage » ayant pour but de localiser les logements et de repérer la construction neuve. Cette dernière opération, fondée sur une enquête cartographique, est indispensable dans un contexte où le bâti est très dispersé, le système de repérage de la voirie fréquemment absent et la construction neuve hors permis importante.

La mise en œuvre du nouveau recensement en 2004, avec ses propres contraintes en termes de pré-enquête cartographique, a conduit à modifier dès que possible ce système pour éviter la juxtaposition de deux opérations cartographiques lourdes, portant sur des entités distinctes géographiquement, que les moyens humains disponibles ne permettaient plus d'assurer simultanément.

Dès janvier 2005, un système d'échantillonnage a donc été mis en place dans les DOM, selon une méthodologie unique alimentant à la fois l'enquête Emploi DOM et d'autres enquêtes sociales (la première étant l'enquête Budget de Famille).

Ce système place les DOM en « pôle position » dans l'utilisation des bases issues du nouveau recensement, bien avant que ce ne soit le cas en métropole et pourrait aussi être considéré comme un « laboratoire d'essai ». Cependant, le système retenu est différent de celui vers lequel on s'oriente pour la métropole.

Le principe retenu est de considérer comme base de sondage pour les enquêtes d'une année n, en régime de croisière¹⁶ :

1. pour les grandes communes (GC) : la base-adresses du groupe de rotation d'îlots devant être recensés l'année n (restreinte à la fraction de 60 % des adresses non sélectionnées pour le recensement de l'année considérée).
2. pour les petites communes (PC) : une base constituée par les adresses ratissées l'année n-1 sur 1/5^{ème} des îlots de chaque petite commune. Comme, chaque année, un groupe de communes représentant approximativement 1/5 de l'ensemble des petites communes est déjà concerné par des opérations cartographiques pour les besoins du recensement, ceci conduira à organiser annuellement des opérations cartographiques supplémentaires sur 1/5^{ème} de chacune des petites communes n'appartenant pas au groupe de rotation de l'année (soit, sur environ 4/25^{ème} du territoire couvert par l'ensemble des petites communes, à l'intérieur de chaque DOM).

Il en résulte que les PC appartenant au groupe de rotation devant être recensé l'année n sont aussi susceptibles d'être couvertes par une enquête la même année, tandis que les fractions des 4 autres groupes de rotation, susceptibles d'être enquêtées l'année n, auront été recensées une année antérieure.

¹⁴ La seule régulière depuis 1993 étant l'enquête Emploi DOM.

¹⁵ Les unités primaires constituées contiennent environ 200 logements en moyenne et sont donc beaucoup plus petites que celles constituées dans l'EM en métropole (qui ont entre 1800 et 3600 logements en strate 0).

¹⁶ En 2005, année de démarrage du nouvel échantillonnage, une solution transitoire consistera à n'utiliser, pour les petites communes, que la base des adresses relatives aux deux premiers groupes de rotation recensés en 2004 et 2005, les seuls disponibles.

Le choix des bases-adresses au lieu des bases de logements recensés s'explique principalement par des raisons de calendrier. En effet, l'objectif était de tirer les premiers échantillons issus du nouveau système pour les enquêtes démarrant en mars 2005, mais, au moment de la conception de ce système, les bases issues de la 1^{ère} campagne de recensement de 2004 n'étaient pas encore disponibles.

Ce choix conditionne une partie de la période future puisqu'il a été décidé de maintenir cette logique au moins pendant le 1^{er} cycle du nouveau recensement, de façon à ne pas perturber le déroulement des enquêtes par des modifications de la méthodologie en cours de route, surtout lorsqu'il s'agit d'enquêtes rotatives comme l'enquête Emploi.

On a donc la correspondance suivante entre l'année d'enquête et la date de constitution de la base de sondage « adresses » utilisée :

ANNEE D'ENQUETE	Date de constitution de la base-adresses (enquête cartographique)	Disjonction avec le recensement de l'année	
		PC	GC
2005	BA 2004	Complète sur 4/5 ^{ème} des communes concernées	Complète (base de sondage = 60 % des adresses non recensées)
2006	BA 2005		
2007	BA 2006		
2008	BA 2007		
2009	BA 2008		

Une modification très importante de la logique des échantillons dans les DOM par rapport aux systèmes existants a été l'**abandon du tirage d'un 1^{er} degré d'unités primaires**. En effet, outre l'effet négatif sur la variance, l'argument le plus en opposition avec ce système était que, dans un contexte d'évolution et de transformation rapide du bâti dans ces territoires, des unités primaires fixes, ayant une certaine durée de vie, ne seraient plus suffisamment représentatives du territoire au bout d'un certain temps.

En contrepartie, avec des tirages d'adresses sur l'ensemble du territoire des DOM, on risque d'avoir une plus grande dispersion de l'échantillon, génératrice de coûts de collecte a priori plus élevés. Ces inconvénients ont néanmoins été jugés acceptables, d'une part parce que les territoires concernés ne sont pas très étendus et, d'autre part, en raison d'une montée en puissance du réseau d'enquêteurs, permettant une meilleure couverture de l'étendue de chacune de ces îles.

Néanmoins, l'objectif est bien de *tirer un échantillon final de logements*. Cet objectif est facilité par le choix d'utiliser une base de sondage d'adresses du fait que, dans les DOM, une très grosse majorité des adresses (de l'ordre de 90 %) ne comportent qu'un seul logement. Pour les autres adresses toutefois, un échantillonnage à deux degrés a dû être réalisé en distinguant :

- une strate de « petites » adresses où l'ensemble des logements de l'adresse sont interrogés à partir du moment où l'adresse est tirée
- et une strate de « grandes adresses » dans laquelle un échantillon de logements est sélectionné (la terminologie est analogue à celle employée par le nouveau recensement mais la définition est ici différente et est précisée ci dessous).

Au final, le plan de sondage est stratifié selon, d'une part, la distinction petite / grande commune et, d'autre part, la taille de l'adresse ; la sélection des adresses dans chaque strate, ainsi que celle des logements à l'intérieur des adresses, est précisée dans le tableau ci-dessous.

Taille de l'adresse →	Adresses monologements	Petites adresses	Grandes adresses
Taille de la commune ↓			
Petite commune	Tirage d'adresses à probabilités égales	Tirage d'adresses à probabilités égales. Interrogation de tous les logements de l'adresse	
Grande commune	Tirage d'adresses à probabilités égales	Tirage d'adresses à probabilités égales. Interrogation de tous les logements de l'adresse	Tirage d'adresses à probabilités proportionnelles à leur taille, puis tirage aléatoire simple de 21 logements.

Le seuil en nombre de logements pour distinguer les petites adresses des grandes a été fixé empiriquement à 21. L'objectif était à la fois de limiter les effets de grappe en interrogeant trop de logements dans une même adresse tout en limitant les risques d'erreurs de repérage par l'enquêteur lors de l'interrogation de quelques logements au sein des adresses. Par souci de simplicité, le seuil a été choisi en fonction de la répartition globale des adresses par taille de logements et aussi afin d'éviter de devoir tenir compte des grandes adresses dans les petites communes ; les rares cas résiduels de telles adresses ont été artificiellement coupés en deux (de façon à créer deux petites adresses au lieu d'une grande).

Le calcul de l'allocation par strate du nombre de logements à tirer s'est fait en imposant une « équipondération » de chaque logement tiré, sur la base d'une taille totale d'échantillon fixée, et en tenant compte des probabilités d'inclusion des groupes de rotation retenus du recensement, qui constituent une 1^{ère} phase de sondage.

Dans le cadre de la stratification décrite ci-dessus, le tirage d'adresses¹⁷ est réalisé en imposant des conditions d'équilibrage sur les variables suivantes :

- l'aspect du bâti en 3 modalités,
- la typologie de quartier en 7 classes (aux Antilles, on a utilisé des classes d'évolution démographique sur la période 1990-1999),
- l'appartenance à une grande zone géographique en 3 modalités,
- et des conditions de dispersion géographique en équilibrant sur la position du centre de gravité des mairies et la dispersion autour de ce centre de gravité (voir référence bibliographique [3]).

Pour l'enquête Budget de Famille 2005 qui est la première enquête à « utiliser » cette nouvelle méthodologie, un échantillon de réserve, destiné à être mobilisé en cours de collecte si les taux de réponse s'avéraient insuffisants, a aussi été sélectionné ; de plus, l'échantillon final de logements a été scindé en 6 sous-échantillons (appelés vagues) interrogés à des périodes différentes, en faisant en sorte, pour des raisons pratiques de déplacements d'enquêteurs, qu'une commune figure dans un nombre réduit de vagues, à l'exception des plus grandes, réparties sur l'ensemble des vagues¹⁸.

A l'issue du 1^{er} cycle du recensement, il conviendra de décider si ce système autonome mis en place pour les DOM est destiné à perdurer ou si, en fonction notamment des évolutions propres du RP (et notamment la mise en place d'un RIL dans les DOM), il devra être inclus à terme dans le système national général.

¹⁷ Réalisé avec la macro CUBE

¹⁸ On trouvera en annexe 2 une fiche récapitulative résumant les spécifications et les résultats de ce tirage, ainsi que quelques résultats numériques sur les effectifs des bases.

Bibliographie

- [1] Bourdallé G., Christine M., Wilms L., 2002, «Echantillons maître et Emploi », actes des journées de méthodologie statistique de décembre 2000, Insee méthodes n°100.
- [2] Christine M. et Wilms L., 2002, « Problèmes théoriques et pratiques de la construction de l'EMEX », Journées de méthodologie statistique de décembre 2002.
- [3] Christine M, 2005, «Construction d'un échantillon dispersé géographiquement », Journées de méthodologie statistique de mars 2005.
- [4] Deville J-C et Tillé Y., 2001, « échantillonnage par la méthode du Cube, variance et estimation de variance » in *Enquêtes, modèles et applications*, 344-363, Dunod.

ANNEXE 1 : Le système d'échantillonnage actuel.

Le système actuel d'échantillonnage des enquêtes nationales métropolitaines auprès des ménages réalisées par l'Insee distingue l'enquête Emploi des autres enquêtes ménages nationales, l'enquête Emploi reposant sur un système aréolaire¹⁹, alors que les autres enquêtes reposent sur l'échantillon-maître.

1. L'échantillon-maître et la base de sondage des logements neufs

La grande majorité des échantillons des enquêtes ménages nationales métropolitaines²⁰ sont sélectionnés dans la même base de sondage qui est renouvelée après chaque recensement de la population depuis 1962. Celle-ci est constituée de deux parties : l'Echantillon-Maître (EM) qui se définit comme une « réserve » localisée de logements construite à partir du dernier recensement et la base de sondage des logements neufs qui complète l'EM pour couvrir les logements achevés après la date du dernier recensement. Par construction, l'EM est une base de sondage constituée une fois pour toutes après le recensement ; il a donc eu, jusqu'à présent, une durée de vie correspondant à la durée de la période inter-censitaire (7/9 ans).

La base actuelle (mise en service pour la première fois en 2001) est composée d'environ 2 millions de logements provenant du recensement de la population de 1999, soit environ 7% des logements recensés. En revanche, la Base de Sondage de Logements Neufs (BSLN) a une taille qui augmente chaque année en proportion du volume de construction neuve. En septembre 2004, elle contenait environ 68 000 logements. La constitution de ces deux bases est détaillée ci-dessous.

1.1. L'échantillon-maître

Pourquoi un échantillon-maître?

Le mode de collecte des données issues des enquêtes ménages est en général²¹ le face à face. **Par conséquent, l'existence même et le mode de constitution de l'EM reposent en grande partie sur la volonté de ne pas disperser les lieux d'enquêtes afin de limiter les frais de déplacement (surtout dans la partie rurale) en recourant à un réseau d'enquêteurs stable travaillant à proximité de leur domicile.**

A l'origine, il existait une autre raison justifiant la mise en place d'un EM. Il permettait de limiter la taille de la base de sondage à manipuler, et par conséquent la taille des fichiers informatiques ainsi que le nombre de feuilles de logements du recensement à manipuler pour compléter la partie adresse des fiches-adresses (document de 4 pages remis aux enquêteurs qui sert principalement à identifier le logement enquêté sur le terrain). Ces contraintes ne s'imposent plus aujourd'hui grâce aux progrès de l'informatique et au fait que l'adresse n'est plus, depuis 1994, recopiée à la main à partir des bordereaux du recensement (on bénéficie de la saisie optique du recensement et on « récupère » directement les images « scannées » des adresses de l'ensemble des logements recensés).

Description du plan de sondage

Le plan de sondage de l'EM est déterminé afin de concentrer les enquêtes sur le territoire. L'EM résulte donc d'un plan de sondage à plusieurs degrés stratifié par catégorie de communes, le nombre de degrés étant variable selon la strate. Les communes rurales sont distinguées des communes urbaines, celles-ci étant classées selon la taille de l'unité urbaine à laquelle elles appartiennent.

¹⁹ Pour plus de détails techniques sur la constitution du système actuel, le lecteur peut se référer aux références bibliographiques [1] et [2].

²⁰ Le cas des enquêtes réalisées dans les DOM fait l'objet d'une partie spécifique.

²¹ Il existe quelques enquêtes Insee réalisées par téléphone mais celles-ci sont peu nombreuses (enquête CAMME et ré-interrogations dans le cadre de l'enquête Emploi en continu et de l'enquête Loyers et charges).

Plus précisément, cinq strates (appelées communément strates de gestion) ont ainsi été constituées :

- strate 0 : les communes rurales,
- strate 1 : les communes appartenant à une unité urbaine de moins de 20 000 habitants,
- strate 2 : les communes appartenant à une unité urbaine ayant entre 20 000 et 100 000 habitants,
- strate 3 : les communes appartenant à une unité urbaine de plus de 100 000 habitants,
- strate 4 : les communes de l'unité urbaine de Paris.

Les unités primaires sont des cantons, des regroupements ou, parfois, des fractions de cantons dans les zones rurales et des unités urbaines ou des regroupements d'unités urbaines ailleurs. Dans le rural, le petit urbain et le moyen urbain (strates 0, 1 et 2), les unités primaires retenues pour l'EM ont été sélectionnées dans chaque région par un tirage à probabilités inégales proportionnellement au nombre de résidences principales²². Toutes les « grandes » unités urbaines (strates 3 et 4) sont retenues au niveau de ce premier degré tandis qu'un échantillon d'unités urbaines ou rurales est sélectionné dans les autres strates. Cependant, un tirage supplémentaire de districts avec un taux de sondage 1/20 est réalisé à l'intérieur des grandes unités urbaines afin de concentrer encore plus les enquêtes. Il en est de même pour les unités urbaines retenues dans la strate 2.

L'EM, constitué de l'ensemble des logements appartenant aux unités ainsi obtenues, sert alors de « réserve » de logements pour tous les tirages des échantillons des enquêtes ménages entre deux recensements.

1.2. La base de sondage des logements neufs (BSLN)

Pour tenir compte des logements construits après le recensement (à défaut, la « représentativité » de l'échantillon-maître issu du RP diminuerait avec le temps), l'échantillon-maître est complété par la base de sondage des logements neufs (BSLN). Cette base est constituée à partir des fichiers SITADEL gérés par le ministère de l'Équipement qui contiennent en particulier l'ensemble des permis de construire autorisés.

On extrait de ce fichier un échantillon de logements²³ dans les permis de construire relatifs aux unités primaires de l'échantillon-maître (soit environ 5% des logements). Ces logements sélectionnés font l'objet d'un suivi par les directions régionales (DR) jusqu'à leur achèvement. Pour chacun d'entre eux, le travail demandé aux DR se décompose en deux temps :

- dans un premier temps, les DR sont amenées à faire sur le terrain une enquête d'achèvement concernant les permis de construire qui leur ont été confiés, en recourant au réseau des enquêteurs²⁴. Un enquêteur se déplace à l'adresse de chaque construction sélectionnée. Sur le terrain, il est chargé de compléter l'adresse, de repérer précisément le logement sélectionné dans le permis et de noter l'état d'avancement du logement. Si celui-ci est achevé, il sera introduit dans la BSLN, sinon il devra être à nouveau suivi par les enquêteurs lors de la campagne suivante.
- dans un second temps, les gestionnaires BSLN entrent les informations recueillies par les enquêteurs dans un logiciel de gestion (appelé logiciel BSLN).

Cette opération, appelée « campagne BSLN », est réalisée chaque année en DR ; elle concerne non seulement les nouveaux logements de l'année en question qui viennent d'être sélectionnés mais aussi ceux sélectionnés antérieurement mais qui ne sont pas encore terminés. Après chaque campagne, les logements qui sont déclarés achevés sont versés dans la BSLN nationale et sont donc susceptibles d'être sélectionnés lors des tirages d'enquêtes ultérieurs.

²² Et qui est équilibré selon le revenu et l'âge des individus en trois tranches d'âge au niveau de regroupement de régions.

²³ Repérés par un numéro séquentiel au sein des permis sélectionnés.

²⁴ Et les mairies et parfois les agents de la DR

1.3. L'Echantillon-Maître pour les EXTensions régionales (EMEX)

L'échantillon-maître n'a pas été construit dans l'optique de pouvoir obtenir des résultats régionaux à partir des parties régionales des échantillons nationaux, souvent de taille insuffisante²⁵. Pour établir des résultats régionaux avec une précision acceptable, il est par conséquent nécessaire de réaliser des extensions régionales d'enquêtes.

Jusqu'à une période récente, celles-ci étaient en général gérées directement par les directions régionales de l'Insee, de manière déconnectée de l'enquête nationale : échantillon spécifique tiré par la DR, mode de traitement (correction de la non-réponse, redressement, pondération...) autonome, exploitation distincte (avec des pondérations spécifiques et des risques d'incohérences avec les parties régionales de l'enquête nationale).

Dans le cadre de la dernière rénovation du système de tirage (mis en place après le recensement de 1999), l'Unité Méthodes Statistiques a construit un outil standard permettant une offre cohérente pour les régions, un traitement homogène des demandes d'extension régionale, l'intégration dans les travaux nationaux et une prise en charge globale des questions méthodologiques. L'étape amont de cette offre a été la constitution d'une base de sondage complémentaire appelée « Echantillon Maître pour les EXTensions régionales » (EMEX). Comme son nom l'indique, ce supplément de réserve de logements est fondé sur les mêmes principes d'ensemble que ceux de l'échantillon-maître (EM) : plan de sondage stratifié sur les mêmes critères que l'EM, à plusieurs degrés, concentration de la localisation des échantillons pour limiter les coûts de déplacement, souci d'une bonne « représentativité » régionale de l'ensemble EM + EMEX afin d'améliorer la qualité des résultats...

La base EMEX (et sa partie logements neufs) ne peut être mobilisée que pour améliorer la précision des résultats régionaux dans le cadre de l'extension d'une enquête nationale. En particulier, elle ne répond pas aux besoins d'échantillonnage pour les enquêtes purement locales.

Cette base a été utilisée pour la première fois en septembre 2002 pour le tirage des extensions régionales de l'enquête Santé²⁶.

2. L'échantillon Emploi

L'enquête Emploi, dont la première édition remonte à 1950, fonctionne sur un système d'échantillonnage aréolaire depuis 1968 qui est renouvelé comme l'EM après chaque recensement de la population. L'échantillon Emploi est défini une fois pour toutes lors du renouvellement du recensement et sa dernière édition, mise au point après le RP 1999, tient compte du passage de l'enquête Emploi annuelle à la nouvelle enquête Emploi en continu. Il a été mis en œuvre à partir du 3ème trimestre 2001.

Le choix d'un système aréolaire s'explique principalement par les raisons suivantes :

- un « rendement important » de la collecte grâce à la concentration des entretiens dans l'espace ; cette concentration permet la réduction des coûts, favorisée de surcroît par la relative simplicité de l'enquête, qui permet d'enchaîner plusieurs interviews au sein d'une même journée et qui est compatible avec une période de collecte courte,
- un effet de voisinage positif sur le taux de réponse à cette enquête (augmenté par une présence de l'enquêteur sur zone d'enquête pendant un laps de temps suffisant pour avoir l'occasion de rencontrer les enquêtés ou de prendre rendez-vous avec eux),

²⁵ L'EM n'a pas, a priori, une représentativité régionale suffisante, notamment parce que le nombre d'unités primaires tirées par région est trop faible. De plus, il n'y a pas de condition explicite de couverture du territoire de la région par l'échantillon-maître : ainsi, celui-ci peut impacter la région de manière plus ou moins bien répartie et il peut arriver que certains départements ne contiennent pas d'unités primaires de strate 0 ou de strate 1.

²⁶ Elle a aussi été utilisée pour le tirage de l'enquête « *information et vie quotidienne* » (octobre 2004).

- et une réduction des imperfections de la base de sondage grâce au ratissage exhaustif de l'aire.

Par construction, les unités finales d'échantillonnage sont des groupes de logements voisins baptisés « aires ». Ces aires ont été calibrées en moyenne à 20 logements. Comme aucune base de sondage d'aires n'existe, il est nécessaire de procéder au préalable à leur construction ; mais comme le découpage de l'ensemble du territoire en aires est impossible, l'échantillon d'aires est obtenu en plusieurs étapes.

Dans un premier temps, un échantillon d'unités géographiques au sein desquelles seront découpées les aires est constitué. Les unités géographiques (communes, IRIS²⁷ ou districts) sont sélectionnées par un plan de sondage stratifié par région et tranche d'unités urbaines et, implicitement, par zone d'emploi, avec des probabilités proportionnelles au nombre total de logements. Ces unités primaires retenues sont ensuite découpées en secteurs de 120 à 240 logements ; on sélectionne un secteur par tirage aléatoire proportionnel à la taille ; celui-ci est lui-même découpé en aires de 20 logements dont on retient au final 6 par tirage aléatoire simple pour constituer l'échantillon.

Les découpages des aires sont réalisés en DR. Cette opération, menée traditionnellement après chaque recensement, est lourde, complexe et mobilise des moyens humains importants en DR. Elle nécessite le goût pour la cartographie, une grande minutie et une grande rigueur. Elle s'étale sur une période plus ou moins longue, dépendant de la planification des charges en DR²⁸.

L'échantillon de l'enquête Emploi est renouvelé par 1/6 chaque trimestre ; une aire sortante (après 6 trimestres d'interrogation) est ainsi remplacée par une aire entrante issue du même secteur. Comme 6 aires sont retenues pour chaque secteur, la durée de vie totale de l'échantillon est de 6*6 soit 36 trimestres. Depuis la mise en place de l'enquête Emploi en continu en 2001, les logements neufs construits depuis le dernier recensement sont identifiés directement dans les aires enquêtées une année donnée par les enquêteurs, lors de l'opération préalable de repérage des logements baptisée « ratissage »²⁹. En fonction du nombre de logements neufs repérés, la totalité ou seulement un échantillon d'entre eux est interrogé.

Il y a enfin disjonction de l'échantillon de l'enquête Emploi et de celui de l'EM (et EMEX) par élimination dans la base EM (et EMEX) des logements qui appartiennent aux aires Emploi.

3. Tirages d'échantillons.

Le tirage proprement dit des échantillons (pour la partie EM, adjointe éventuellement de l'EMEX) se fait au moyen d'une chaîne informatique produisant différents fichiers en sortie et permettant l'édition des fiches-adresses (voir supra) ; quant à l'échantillon Emploi, il est en fait constitué une fois pour toutes à partir du moment où l'on a construit - au prix d'opérations manuelles lourdes réalisées en DR - et sélectionné les aires-échantillons.

²⁷ Regroupement d'îlots selon des critères socio-économiques.

²⁸ Rappelons que, pour le dernier échantillon Emploi, le découpage des aires avait une durée prévue d'une année. Pour des raisons d'ordonnement des tâches en DR, il a été réalisé de façon discontinue : il a commencé en juillet 2000 pour s'achever en mai 2002, mais avec une interruption plus ou moins complète au 1^{er} semestre 2001.

²⁹ Antérieurement, les logements neufs issus de permis de construire de moins de 10 logements étaient repérés sur les aires enquêtées à l'occasion du ratissage et interrogés. Les autres logements neufs, issus de « gros » permis, n'étaient pas enquêtés avec l'ensemble des logements des aires auxquelles ils appartenaient mais remplacés statistiquement par un échantillon de logements tirés dans les permis de plus de 10 logements de la base SITADEL. Ceci avait pour but d'éviter des gonflements trop importants et non maîtrisés de la taille de certaines aires.

ANNEXE 2 : Exemple de fiche récapitulative du tirage dans les DOM³⁰.

DOM : REUNION

Enquête : BDF

Taille de l'échantillon souhaitée : 1 500 logements

Base adresses

La **base adresses** correspond à l'ensemble des adresses des 19 communes recensées en 2004 et 2005.

GC : RP05

PC : RP04 : Entre-Deux (403), Salazie (421),

RP05 : Saint-Philippe (417), Trois Bassins (423)

Champ statistique

Aucune exclusion

Base utile

La **base utile** contient l'ensemble des adresses avec logements « enquêtés ». Ont été exclues les adresses hors champ statistique.

Exclusion pour commodités d'enquête

Exclusion d'îlots d'accès difficile.

Strates : répartition des adresses dans la base utile

Effectif des strates de la base utile

	Mono logements	Petites adresses	Grandes adresses	Total
GC	22 981	1 122	76	24 179
PC	8 716	88	0	8 804
Total	31 697	1210	76	32 983

Dont adresses sélectionnées pour l'enquête emploi DOM 2004 ou exclues pour raison de coût de déplacement

	Mono logements	Petites adresses	Grandes adresses	Total
GC	472	7	0	479
PC	488	0	0	488
Total	960	7	0	967

Les adresses tirées pour une enquête antérieurement sont exclues du tirage.

Par la suite, la partie de la base utile, d'où les adresses enquêtées auparavant et les adresses d'accès difficile ont été exclues, s'appellera **la base de tirage**.

³⁰ Fiche élaborée par Delphine ARTAUD, Antenne statistique de Mayotte.

Variables d'équilibrage pour le tirage des adresses dans la base de tirage

- la variable un contenant la valeur 1 ;
- la probabilité initiale de tirer l'adresse par strate pour avoir une taille fixe d'échantillon dans chaque strate (en terme de nombre d'adresses) : 5 strates;
- l'aspect du bâti en 3 modalités (les modalités 1 « Habitation de fortune » et 2 « Case traditionnelle » ont été regroupées) ;
- la typologie de quartier en 7 classes ;
- l'appartenance à une grande zone géographique en 3 modalités ;
- le centre de gravité des mairies et la dispersion autour de ce centre de gravité.

Pour chaque modalité des variables catégorielles, les totaux s'ajustent sur le vrai nombre d'adresses dans la base de tirage. Une modalité est exclue à chaque fois pour éviter les colinéarités.

Echantillon de réserve

Besoin d'une réserve pour les vagues 3, 4, 5, 6 représentant 15 % de l'échantillon utile.

L'échantillon utile est de 5 000 Fiches Adresses (FA) pour les 4 DOM, réparti sur 6 vagues soit 833 FA par vague.

La réserve compte 750 FA pour les 4 DOM, répartis sur 4 vagues soit 188 FA par vague.

188 représente 22,5% de 833 (en nombre de logements).

L'échantillon a été gonflé de 22,5%, réparti en 6 vagues. Pour les deux premières vagues, 81,6 % (1/1.225) des adresses ont été tirées pour construire l'échantillon utile et le reste des adresses est remis dans la base de sondage.

Pour les vagues 3, 4, 5, 6, 81,6 % des adresses ont été tirées pour l'échantillon utile. Les adresses restantes composent la réserve.

Ainsi la réserve représente 15 % de l'échantillon utile.

La probabilité de tirage d'une adresse

Hypothèse de départ : tout logement tiré, quelle que soit la strate, a la même probabilité finale : PI.

PI = 0.0064078453

Cette probabilité est recalculée ensuite après avoir arrondi les allocations des strates à l'entier le plus proche.

Proba	Mono logements	Petites adresses	Grandes adresses
GC	0.0064070319	0.0064171123	0.0063015754
PC	0.0063927824	0.0138651687*	0

* Le calcul des allocations donne une seule adresse pour la strate PC-PA. Elle a été forcée à 3.

Allocations en terme d'adresses de l'échantillon « gonflé »

Nb d'adresses	Mono logements	Petites adresses	Grandes adresses	Total
GC	1 227	60	4	1 291
PC	137	3*	0	140
Total	1 364	63	4	1 431

* Le calcul des allocations donne une seule adresse pour la strate PC-PA. Elle a été forcée à 3.

Allocations en terme d'adresses de l'échantillon utile

Nb d'adresses	Mono logements	Petites adresses	Grandes adresses	Total
GC	1 000	50	3	1 053
PC	113	2	0	115
Total	1 113	52	3	1 168

Echantillon final utile en nombre de logements par vague et taille de la commune

Vagues	1	2	3	4	5	6	Total
GC	252	218	232	237	244	217	1 400
PC	0	31	26	19	12	31	119
Total	252	249	258	256	256	248	1 519

La vague 1 comprend 252 logements dans des grandes communes et aucun logement dans des petites communes. Une contrainte de la répartition par vagues est qu'un Iris est dans une vague et une seule.

Réserve en nombre de logements par vague et taille de la commune

Vagues	1	2	3	4	5	6	Total
GC	//	//	40	48	49	42	179
PC	//	//	5	4	2	12	23
Total	//	//	45	52	51	54	202

Répartition des logements de l'échantillon final utile par commune et par vague

Commune	Code commune	Zone d'enquête	Vague						Total
			1	2	3	4	5	6	
ENTRE DEUX	403	SP	0	0	0	0	0	31	31
ETANG SALE	404	SP	0	43	0	0	0	0	43
PETITE ILE	405	SP	0	0	12	0	0	0	12
LE PORT	407	LP	0	26	47	13	0	0	86
LA POSSESSION	408	LP	30	0	0	0	0	0	30
SAINT ANDRE	409	SD	0	18	32	53	0	0	103
SAINT BENOIT	410	SD	47	0	0	0	0	23	70
SAINT DENIS	411	SD	44	36	65	50	48	46	289
SAINT JOSEPH	412	SP	46	8	0	0	0	13	67
SAINT LEU	413	LP	0	0	0	0	27	27	54
SAINT LOUIS	414	SP	0	0	15	37	34	0	86
SAINT PAUL	415	LP	30	35	27	29	26	32	179
SAINT PIERRE	416	SP	22	27	19	24	28	18	138
SAINT PHILIPPE	417	SP	0	0	26	0	0	0	26
SAINTE MARIE	418	SD	0	0	0	0	30	43	73
SAINTE SUZANNE	420	SD	0	0	0	9	24	0	33
SALAZIE	421	SD	0	31	0	0	0	0	31
LE TAMPON	422	SP	33	25	15	22	27	15	137
TROIS BASSINS	423	LP	0	0	0	19	12	0	31
Total			252	249	258	256	256	248	1 519