

LA CONSTRUCTION DE L'ÉCHANTILLON DE LA FUTURE ENQUÊTE EMPLOI EN CONTINU A PARTIR DU RECENSEMENT DE 1999.

M. CHRISTINE^(*)

^(*)INSEE, Unité Méthodes Statistiques

AVERTISSEMENT.

L'ensemble des travaux de constitution de l'échantillon Emploi ont été conçus et pilotés par l'Unité Méthodes Statistiques, sous la direction de Marc CHRISTINE, en collaboration avec la division Emploi. Le test et la validation des spécifications de "découpage" et leur mise en oeuvre dans les Directions régionales de l'INSEE ont été accompagnés par un groupe de représentants des établissements régionaux¹.

Ont plus particulièrement collaboré à ce travail :

- Jonathan BOSREDON, pour l'assistance à la conception.
- Georges BOURDALLE, pour la mise en oeuvre des outils d'assistance cartographique aux directions régionales, le tirage des unités "de 1ère phase", le suivi et les conseils aux régionaux pour la réalisation du découpage et la validation des dossiers, la constitution du fichier final des aires.
- Martial MAILLARD, pour le suivi de l'avancement de la réalisation terrain et la validation des dossiers.
- Gérard MORVAN, pour l'écriture de l'algorithme de regroupement des petites communes.
- Philippe RAVALET, pour le calcul de l'allocation de l'échantillon par strate.
- Pascal ARDILLY, Bernard BIANCHERI, Christine CASTILLO, Maryse JEGOU, Pierre LEOSTIC, François MARRAS, Agnès PEREZ, pour les tests et la validation des procédures de "découpage" sur le terrain, auxquels se sont joints Louis CHAVAGNON et Georges GARNIER pour la formation des régionaux.

¹ Directions régionales de Bourgogne, Centre, Ile de France, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Pays de la Loire, Rhône-Alpes.

La nouvelle enquête Emploi en continu, dont un prototype est actuellement sur le terrain, sous le nom de “ dispositif léger ”, doit entrer en vigueur en vraie grandeur à compter du 3ème trimestre de 2001.

Cette nouvelle enquête présentera des changements très importants par rapport aux séries précédentes, qui étaient fondées sur une interrogation ponctuelle, au mois de Mars de chaque année². Néanmoins, comme par le passé, son échantillonnage continuera d’obéir à des règles spécifiques, distinctes de celles relatives aux autres enquêtes auprès des ménages réalisées par l’INSEE.

En particulier, le caractère **aréolaire** de l’échantillon a été conservé dans la nouvelle enquête : les logements interrogés ne sont pas tirés aléatoirement dans une zone géographique donnée, mais regroupés au sein de petites entités, les *aires*. Ce sont les aires qui sont sélectionnées aléatoirement, comme unités finales de tirage, et tous les logements qui les constituent font alors partie de l’échantillon³.

Ceci implique une méthodologie particulière de constitution de l’échantillon, distincte de la logique de l’échantillon-maître. Comme pour ce dernier, néanmoins, chaque nouveau recensement est l’occasion de redéfinir l’échantillon Emploi pour la période intercensitaire suivante⁴. Ainsi, le précédent échantillon avait été construit en 1991-1992, pour la première introduction lors du tiers entrant de l’enquête annuelle de 1992.

Mais, si la constitution du nouvel échantillon relève des mêmes principes généraux que celui de 1992, *elle en diffère dans la mise en oeuvre sur des points notables*. Elle a en particulier bénéficié de l’introduction de nouveaux concepts de géographie infra-communale issus du RP 1999 (notion d’IRIS⁵), la logique du découpage mise en oeuvre est systématiquement « descendante » et de nombreuses innovations en matière de cartographie ont été employées.

L’ensemble des travaux méthodologiques de constitution de ce nouvel échantillon ont été conduits entre Novembre 1999 et Juin 2000. La mise en oeuvre sur le terrain (opération dite de “ *découpage des aires* ”) se fera en deux temps : une moitié au cours du 2ème semestre 2000 et l’autre moitié au cours du 2ème semestre 2001.

² La dernière enquête de ce type doit avoir lieu en mars 2002, ce qui impliquera une période de recouvrement avec la nouvelle enquête. Néanmoins, les échantillons de ces deux types d’enquêtes sont totalement disjoints.

³ A la nuance près du statut des logements neufs. Cf. infra.

⁴ Cf. infra pour le calibrage temporel de cet échantillon.

⁵ Îlots regroupés pour l’information statistique.

1. Présentation générale

1.1 Une enquête en continu

L'enquête Emploi en continu se caractérise par le fait que la collecte est étalée tout au long de l'année. Le concept central de l'enquête est celui de *semaine de référence*. C'est en regard de l'activité de la personne interrogée vis-à-vis du marché du travail, au cours de cette semaine, qu'est définie sa classification selon les normes du Bureau International du Travail (chômeur, en emploi, inactif). Le règlement européen qui régit et harmonise les différentes enquêtes sur les Forces de Travail prévoit que les semaines de référence couvertes par l'enquête doivent s'étaler uniformément sur l'année.

Ceci implique une *collecte également étalée sur l'année*. Dans la pratique, pour la fraction d'échantillon interrogée relativement à une semaine de référence donnée, la collecte s'étale sur les deux semaines plus deux jours suivant celle-ci⁶.

L'un des objectifs de cette nouvelle enquête étant de fournir des résultats conjoncturels sur le marché du travail (au minimum trimestriels), l'accent a été mis sur la fourniture d'estimations de variations trimestrielles⁷. Afin de minimiser la variance de ces estimations, on a choisi de conserver une fraction importante de l'échantillon d'un trimestre sur l'autre.

Ainsi, l'échantillon retenu est-il un échantillon de *logements*⁸ dont le **taux de renouvellement est d'1/6 par trimestre**, chaque logement étant donc interrogé 6 trimestres consécutifs.

1.2 Calendrier de mise en œuvre et durée de vie

La mise en œuvre de la nouvelle enquête en vraie grandeur se fera à partir de Juillet 2001. A cette occasion, le nouvel échantillon entrera également en vigueur ; mais, compte tenu du caractère rotatif de l'échantillon, cet échantillon doit se substituer progressivement à l'échantillon actuel : seule une fraction d'1/6 sera donc introduite au 3ème trimestre 2001. Les 5/6 restants de l'échantillon seront des réinterrogations de l'échantillon utilisé dans le cadre du dispositif léger⁹. Au 4ème trimestre 2001, une deuxième fraction du nouvel échantillon sera introduite, et la

⁶ Une période de collecte un peu plus longue est prévue en période d'été.

⁷ On verra cependant au § 4.1 que la détermination du nombre d'unités à enquêter résulte de la prise en compte de deux types antinomiques de contraintes de précision imposées par EUROSTAT, l'une en termes de *variation trimestrielle nationale*, l'autre en termes de *niveau annuel moyen régional*.

⁸ Et non de ménages, ni d'individus : on ne suit pas les ménages ni les individus qui changent de logements.

⁹ Celui-ci est issu soit des enquêtes annuelles de 1997 et suivantes, soit des aires de réserve construites à partir du RP 1990.

fraction analogue (en taille, si ce n'est en localisation géographique) de l'ancien échantillon sera éliminée.

Ainsi, pendant une période transitoire de 5 trimestres, coexisteront des fractions du nouvel échantillon et de l'ancien ; ce n'est qu'à partir du 4^{ème} trimestre 2002 que l'enquête sera entièrement réalisée sur le nouvel échantillon.

Dans le passé, l'échantillon était calibré pour pouvoir être utilisé jusqu'au recensement suivant¹⁰. Pour le présent échantillon, on a considéré qu'il devait durer jusqu'à ce que le recensement rénové de la population soit en régime de croisière et puisse alimenter (d'une manière qui reste à définir, mais qui est en dehors du champ du présent travail¹¹) les nouveaux échantillons des enquêtes " Ménages ".

Néanmoins, par précaution, pour éviter un épuisement prématuré de l'échantillon tiré du recensement de 1999 (qui obligerait à en constituer un second si le nouveau recensement ne pouvait être mis en oeuvre comme on le prévoit actuellement¹²), le calibrage de cet échantillon s'appuie sur une date de fin de vie relativement lointaine : *mi-2010, soit 9 années (36 trimestres) à compter de la date de première introduction.*

1.3 Un échantillon aréolaire

Comme pour ses prédécesseurs, le nouvel échantillon Emploi est **aréolaire**.

Les avantages d'un échantillon aréolaire sont bien connus :

- on obtient un meilleur taux de réponse que dans les enquêtes construites à partir de l'échantillon-maître ; ceci est dû à l'effet d'entraînement, en général positif¹³, entre ménages d'une même aire ; ce taux, il est vrai, est aussi lié à la relative simplicité de l'enquête et à l'intérêt que portent les ménages pour le thème ;

- le ratissage exhaustif de l'aire avant l'enquête permet de n'oublier aucun logement et, en particulier, d'interroger les logements les plus précaires, dont les occupants peuvent avoir une relation particulière au marché du travail ;

¹⁰ Et même au-delà, grâce à la constitution d'une *réserve*. Ce choix a été prudemment et judicieusement fait en 1992, puisqu'à l'époque, le recensement suivant devait avoir lieu en 1997 et qu'en définitive, l'enquête Emploi annuelle sera alimentée par des échantillons issus du RP de 1990 jusqu'à son édition de 2002, qui devrait être la dernière.

¹¹ En particulier, doit-on attendre qu'un cycle complet de collecte du RRP soit achevé pour tirer des échantillons ou commencera-t-on dès qu'une première année de collecte sera validée ?

¹² 1^{ère} collecte prévue en 2004.

¹³ Certes, l'argument inverse existe aussi, mais il semble plus exceptionnel : cas de l'enquête réalisée dans une « barre » d'immeubles et sujette à un refus collectif...

- ce même ratissage permet de gérer les *logements neufs*, c'est-à-dire dont la construction a été achevée depuis le dernier RP : actuellement, les logements neufs apparaissant sur l'aire et issus de permis de construire de moins de 10 logements sont intégrés à l'échantillon RP et interrogés comme les autres logements de l'aire, tandis qu'un échantillon spécial ("grappes spéciales Emploi") est tiré du fichier des permis de construire pour couvrir les permis collectifs de 10 logements ou plus. Pour l'avenir, on étudie actuellement l'hypothèse de constituer une base de sondage à partir de tous les logements neufs ratissés et d'y puiser un échantillon pour l'enquête.

- enfin, la concentration des déplacements des enquêteurs permet d'obtenir un meilleur rapport rentabilité/coût et rend possible la réalisation d'une enquête de taille beaucoup plus importante que celle des autres enquêtes auprès des ménages (environ 100.000 fiches-adresses pour l'enquête Emploi annuelle).

En contrepartie, les inconvénients de l'aréolaire sont également connus : l'effet de grappe détériore la précision. Par corollaire, on a intérêt à avoir des aires de taille réduite et hétérogènes. On verra cependant que cette dernière contrainte est difficile à mettre en oeuvre en pratique.

La taille des aires a ainsi été fixée à **20 logements¹⁴ en moyenne par aire¹⁵**, quelle que soit la tranche de taille d'unité urbaine des communes auxquelles appartiennent les aires (donc, aussi bien en zone rurale qu'urbaine¹⁶).

1.4 Nécessité d'un échantillonnage en plusieurs phases

La difficulté fondamentale de la constitution de l'échantillon Emploi réside dans le fait que l'aire, en tant que regroupement d'une vingtaine de logements "voisins les uns des autres", n'est pas une unité géographique standard. Il n'existe aucun fichier contenant les aires dans lequel on pourrait effectuer un tirage aléatoire.

De surcroît, quand bien même pourrait-on penser constituer dès le recensement ces regroupements de logements au moment du passage de l'agent recenseur¹⁷, cela ne résoudrait pas la difficulté entièrement. En effet, derrière le concept d'aire repose aussi l'idée de territoire géographique, en relation avec les opérations de "ratissage" préalables à l'enquête : on repère tous les logements présents sur cette portion de territoire, existant au recensement ou construits postérieurement à celui-ci. Se contenter de définir ex-ante une aire comme un groupe de 20 logements repérés au moment du recensement ne suffit donc pas.

¹⁴ Indiquons une fois pour toutes que les nombres de logements auxquels il sera toujours fait référence ici sont les **nombres totaux de logements issus du RP 1999**.

¹⁵ Cf. infra pour les règles de tolérance.

¹⁶ Dans les enquêtes Emploi annuelles fondées sur l'échantillon issu du Recensement de 1990, les aires étaient de 20 logements en moyenne en zone urbaine, mais de 40 en zone rurale. Afin de minimiser les effets de grappe, on a souhaité réduire la taille des aires à 20 logements dans tout type de zone.

¹⁷ Hypothèse d'école ; cela n'a pas été envisagé pour le RP 1999.

Il faut donc définir une procédure ad hoc pour construire les aires et réaliser un échantillon. Mais le découpage de l'ensemble du territoire en aires de 20 logements est évidemment totalement irréaliste. On va donc définir une procédure en plusieurs étapes en constituant au préalable un échantillon d'unités géographiques, dites "unités échantillonnées de 1ère phase", au sein desquelles seront découpées les aires.

Au total, l'échantillon d'aires sera **un échantillon en plusieurs phases** :

- **1ère phase : tirage d'unités géographiques ;**
- **2ème phase : découpage des aires à l'intérieur des unités de 1ère phase.**

Cette 2ème phase se subdivise elle-même en sous-phases. En effet, les unités de 1ère phase ont des tailles relativement variables, qui pourraient engendrer, selon les cas, des procédures de découpage longues et coûteuses. Aussi passe-t-on par des unités de niveau intermédiaire, les *secteurs*, dont la définition sera donnée ci-après.

1.5 Champ géographique

Les travaux de construction de l'échantillon de l'enquête Emploi s'appliquent exclusivement à la France métropolitaine ; **les DOM sont exclus**. L'enquête Emploi qui y est réalisée suit en effet une méthodologie différente, notamment en termes d'échantillonnage et de fréquence (l'enquête reste pour l'instant annuelle, avec une collecte étalée néanmoins sur un trimestre).

2. Les différentes phases de l'échantillonnage.

2.1 Echantillon de première phase : nature des unités

Les unités à échantillonner en première phase ont été choisies comme unités géographiques ou administratives standards afin de faciliter leur définition, leur délimitation et leur découpage.

Leur nature découle de deux principes fondamentaux :

a) il faut éviter de travailler sur des unités trop grandes au départ, de façon à alléger la charge de découpage.

b) on privilégie toujours *une logique descendante*. On préfère partir d'une unité trop grande qu'on découpe, au lieu de l'inverse : partir d'unités trop petites qu'on réagrège ensuite. *Il faut noter que ceci constitue un changement profond par rapport aux méthodes de constitution du précédent échantillon Emploi, en 1992.*

Les unités de 1ère phase sont, selon les cas précisés ci-dessous, des *communes, IRIS ou districts*, ou des *agréats formés à partir de ces unités de base*.

Cet échantillon d'unités de première phase sera stratifié selon les caractéristiques des communes auxquelles elles sont rattachées, **par région et tranche de taille d'Unité Urbaine (UU)¹⁸ et, indirectement, par zone d'emploi¹⁹** (cf. *infra*, § 3.1).

2.2 Organisation des aires à l'intérieur d'une unité d'échantillonnage

Une même unité d'échantillonnage va donner naissance à plusieurs aires. En effet, compte tenu du caractère rotatif de l'échantillon Emploi (cf. *supra*), chaque aire sera interrogée 6 trimestres consécutifs, à la suite de quoi elle doit être remplacée par une aire ayant des caractéristiques similaires. Ceci permettra d'éviter les fluctuations aléatoires dues à un changement d'échantillon et de minimiser les pertes de précision dans l'estimation des variations trimestrielles.

¹⁸ On s'est appuyé sur la nouvelle classification des communes selon la tranche d'unité urbaine à laquelle elles appartiennent, **issue du RP 1999**, qui a été disponible au printemps 2000.

¹⁹ Contrairement au cas de la variable TU, toutes les fois que seront évoquées les zones d'emploi, ce sera par référence à la **définition en vigueur actuellement, issue du RP 1990**. La redéfinition des zones d'emploi à l'issue du RP 1999 n'était pas disponible au moment de ce travail.

Dans la pratique, l'aire remplaçante sera une *aire voisine issue de la même unité*. Chaque unité de 1ère phase doit donc être découpée de manière à constituer, in fine, plusieurs aires de **20 logements** en moyenne, qui constitueront l'échantillon Emploi proprement dit (ou de deuxième phase), composé des logements devant être enquêtés.

Chaque trimestre, une et une seule aire dans chaque unité échantillonnée (soit une vingtaine de logements par unité) sera mobilisée. Ceci a pour objectif d'assurer la fourniture de résultats trimestriels représentatifs de la France entière.

Par corollaire, comme on calibre actuellement l'échantillon trimestriel final de l'enquête à environ 51.000 logements à enquêter (soit 204.000 par an, à comparer aux 100.000 fiches-adresses de l'enquête annuelle), on en déduit le nombre approximatif d'unités à tirer en première phase : **2.550 unités**²⁰.

Le nombre d'aires associées à une même unité d'échantillonnage est calculé en fonction de la durée de vie probable de l'échantillon Emploi tiré du RP 1999 : les aires issues de ce nouvel échantillon seront introduites progressivement à partir du 3ème trimestre 2001 et l'on fera vivre l'échantillon jusqu'à la fin du 2ème trimestre 2010 (date à laquelle l'ensemble des aires construites à partir des unités d'échantillonnage introduites initialement s'épuiseront)²¹, soit une durée de vie totale de 9 années (soit 36 trimestres).

Chaque aire interrogée étant conservée 6 trimestres consécutifs dans l'échantillon, **il faut donc, pour assurer cette durée de vie, construire 6 aires (36/6) associées à une même unité échantillonnée. Il en résulte que les unités d'échantillonnage devront comporter au moins 120 logements (6 * 20).** En revanche, les unités n'ont pas de limite supérieure de taille fixée a priori ; celle-ci pourra atteindre dans certains cas quelques milliers de logements.

²⁰ En réalité, notamment pour se prémunir contre des aléas budgétaires, on tirera un nombre d'unités excédentaires qui constitueront une *réserve*, destinée, par exemple, à alimenter des tests, ou à se prémunir contre des difficultés ultérieures de collecte (cas d'une baisse du taux de réponse, zones difficiles à enquêter du fait de refus massifs, remplacements liés à des causes diverses...).

²¹ Cf. encadré.

2.3 Constitution de secteurs à l'intérieur d'une unité d'échantillonnage

Pour la construction effective des aires, on passe par la notion de *secteur* : un secteur regroupera les 6 aires issues d'une même unité échantillonnée. Mais, du fait de la logique descendante énoncée ci-dessus, on ne découpera pas les aires pour les regrouper en secteurs mais, au contraire, on découpera l'unité échantillonnée en secteurs, puis ceux-ci (ou du moins l'un d'entre eux) en aires.

Plus précisément, chaque unité doit être découpée en secteurs de 120 à 240²² logements (si sa taille est comprise dans ces limites, elle constitue directement un secteur). Un de ces secteurs sera tiré au sort, avec une probabilité proportionnelle à sa taille (en nombre total de logements) et le secteur sélectionné sera lui-même découpé en aires de 20 logements en moyenne. 6 d'entre elles seront tirées au sort pour constituer l'échantillon final.

Cette notion de secteur est à entendre au sens de portion de territoire qui sera visitée au cours de la durée de vie de l'échantillon Emploi, les aires sélectionnées s'enchaînant les unes aux autres à l'intérieur du secteur de manière séquentielle. Sans être nécessairement contiguës, ces aires seront circonscrites dans une zone géographique d'amplitude limitée, correspondant à la zone d'action d'un enquêteur.

Ainsi, l'intérêt de passer par la notion de secteur est-il multiple :

- diminuer la charge du découpage, en ne découplant en aires que des unités de taille réduite (et relativement homogène) ;

- assurer une certaine stabilité géographique du réseau d'enquêteurs pendant la durée de vie de l'échantillon : on pourra utiliser les mêmes enquêteurs lors du renouvellement des fractions d'échantillon au sein d'un même secteur ;

- minimiser les fluctuations aléatoires lors du changement d'aires, puisqu'une aire sortante est remplacée par une aire voisine.

²² Cf. infra pour la justification de ce seuil maximal.

Résumé de la méthodologie

Pour schématiser, la constitution de l'échantillon Emploi nécessite trois étapes :

- construction des unités à échantillonner en 1ère phase : communes ou agrégations de communes, IRIS ou agrégations d'IRIS, districts²³. Ces unités (quel qu'en soit le type) doivent comporter au moins 120 logements (*nombre total de logements*).
- tirage d'un échantillon de ces unités, selon un plan de sondage stratifié selon les caractéristiques des communes auxquelles elles sont rattachées, **par région et tranche de taille d'UU** (regroupée en 5 modalités) **et, indirectement, par zone d'emploi**.
- découpage de chaque unité échantillonnée en un ou plusieurs secteurs, sélection aléatoire d'un seul de ces secteurs, découpage de celui-ci en au moins 6 aires de 20 logements et constitution des listes de logements de chaque aire.
- tirage final de 6 aires exactement qui seront retenues pour faire partie de l'échantillon final ; chaque trimestre, une seule aire de chaque secteur sera soumise à enquête.

²³ On peut dire, schématiquement, que, dans les communes découpées en IRIS, l'unité d'échantillonnage de 1^{ère} phase sera l'IRIS ou le district, et dans les autres communes, la commune elle-même ou un regroupement de communes.

La rotation des sous-échantillons. Règle " 6, 6, 13 ".

Une aire donnée est enquêtée 6 trimestres consécutifs puis remplacée par une aire échantillonnée du même secteur. Les aires d'un même secteur se différencient par leur *numéro de série*²⁴, allant de 1 à 6. Les aires portant le même numéro de série constitueront une *série d'aires*. Pour un secteur donné, l'appartenance d'une aire à l'une des six séries sera définie aléatoirement, après découpage de l'ensemble des aires.

L'ensemble des secteurs sera lui-même partitionné en 6 fractions, repérées par un *n° de sous-échantillon* allant également de 1 à 6²⁵. Ces sous-échantillons se distinguent notamment par leur date d'introduction dans l'échantillon Emploi. La décomposition en sous-échantillons permet d'assurer le renouvellement de l'échantillon par 1/6 chaque trimestre.

Ainsi, par exemple, au 3ème trimestre 2001, les aires issues des secteurs appartenant au sous-échantillon n°1 sont introduites. Ce sont d'abord les aires de la série 1, puis, une fois celles-ci épuisées (au bout de 6 trimestres), elles sont remplacées par des aires de la série 2 issues des mêmes secteurs, donc voisines de celles de la série 1.

Au 4ème trimestre 2001, ce sont les aires issues des secteurs appartenant au sous-échantillon n°2 qui sont introduites. On commence par les aires portant le n° de série 1, qui, une fois épuisées (au bout de 6 trimestres), seront remplacées par des aires de la série 2 issues des mêmes secteurs et voisines de celles de la série 1. Et ainsi de suite...

Au cours d'un même trimestre, des aires sont interrogées dans chacun des sous-échantillons. On voit qu'il y a toujours au moins un sous-échantillon pour lequel on change de série au trimestre suivant. Ceci illustre le renouvellement de l'échantillon par 1/6 chaque trimestre.

On notera enfin que chaque secteur est caractérisé par un *n° de semaine* allant de 1 à 13. Celui-ci fixe, pour toute la durée de vie de l'échantillon et pour toutes les aires du même secteur, le rang, au sein de chaque trimestre, de la semaine d'interrogation (ou plus précisément de la semaine de référence, au sens des concepts du Bureau International du Travail). Ainsi, au cours des interrogations successives d'un même logement, ses occupants sont-ils suivis au regard de leur position sur le marché du travail sur des intervalles d'exactement 13 semaines.

Les partitions de l'ensemble des unités échantillonnées en 6 sous-échantillons, d'une part, et en 13 semaines, d'autre part, seront constituées de manière aléatoire, indépendamment l'une de l'autre (cf. infra, §6.4).

²⁴ Variable dénommée SA dans l'échantillon Emploi.

²⁵ On notera que les deux facteurs 6 n'ont aucun rapport entre eux : l'un caractérise le secteur (*n° de sous-échantillon*) et différencie les secteurs entre eux selon le trimestre d'introduction dans le nouvel échantillon ; sa valeur est liée au taux de renouvellement de l'échantillon. L'autre caractérise l'aire à l'intérieur d'un même secteur (*n° de série*) et définit la séquence d'interrogation des aires au cours de la période ; sa valeur est liée à la durée de vie de l'échantillon et au nombre d'interrogations successives d'une même aire (égal à l'inverse du taux de renouvellement trimestriel).

	Sous- échantillon n°1	Sous- échantillon n°2	Sous- échantillon n°3	Sous- échantillon n°4	Sous- échantillon n°5	Sous- échantillon n°6
	N° de série	N° de série	N° de série	N° de série	N° de série	N° de série
3T-2001	1					
4T-2001	1	1				
1T-2002	1	1	1			
2T-2002	1	1	1	1		
3T-2002	1	1	1	1	1	
4T-2002	1	1	1	1	1	1
1T-2003	2	1	1	1	1	1
2T-2003	2	2	1	1	1	1
3T-2003	2	2	2	1	1	1
4T-2003	2	2	2	2	1	1
1T-2004	2	2	2	2	2	1
2T-2004	2	2	2	2	2	2
3T-2004	3	2	2	2	2	2
4T-2004	3	3	2	2	2	2
1T-2005	3	3	3	2	2	2
2T-2005	3	3	3	3	2	2
3T-2005	3	3	3	3	3	2
4T-2005	3	3	3	3	3	3
1T-2006	4	3	3	3	3	3
2T-2006	4	4	3	3	3	3
3T-2006	4	4	4	3	3	3
4T-2006	4	4	4	4	3	3
1T-2007	4	4	4	4	4	3
2T-2007	4	4	4	4	4	4
3T-2007	5	4	4	4	4	4
4T-2007	5	5	4	4	4	4
1T-2008	5	5	5	4	4	4
2T-2008	5	5	5	5	4	4
3T-2008	5	5	5	5	5	4
4T-2008	5	5	5	5	5	5
1T-2009	6	5	5	5	5	5
2T-2009	6	6	5	5	5	5
3T-2009	6	6	6	5	5	5
4T-2009	6	6	6	6	5	5
1T-2010	6	6	6	6	6	5
2T-2010	6	6	6	6	6	6
3T-2010		6	6	6	6	6
4T-2010			6	6	6	6
1T-2011				6	6	6
2T-2011					6	6
3T-2011						6

3. Les unités échantillonnées de 1^{ère} phase

3.1 Constitution des unités d'échantillonnage.

La constitution des unités d'échantillonnage a obéi à trois grands principes :

- d'une part, leur taille devait être la plus faible possible (pour éviter des opérations de découpage longues et fastidieuses sur des zones trop étendues), mais néanmoins supérieure au seuil requis (120 logements).
- d'autre part, elles devaient s'appuyer sur des découpages géographiques existants.
- enfin, on a privilégié la méthode consistant à partir d'une unité trop grande, quitte à la découper ensuite, plutôt que sélectionner des unités trop petites²⁶. Cette dernière solution obligerait en effet à effectuer, après tirage, une agrégation manuelle difficile à contrôler et à définir de manière systématique²⁷.

Ainsi, en règle générale, dans les communes de moins de 10000 habitants, les unités d'échantillonnage sont soit des *communes*, soit des *agrégations de communes*. Pour les plus grandes communes (plus de 10000 habitants), on passe à l'unité plus désagrégée naturellement disponible, l'*IRIS 2000*²⁸, voire le *district du recensement*. Dans certaines communes de taille comprise entre 5000 et 10000 logements, une partition a été opérée en IRIS à la suite du RP 1999. On s'est appuyé dessus toutes les fois que c'était possible.

a) communes non découpées en IRIS.

Ces communes ou des regroupements de celles-ci serviront d'unités d'échantillonnage.

b) communes découpées en IRIS.

Dans ce cas, les unités d'échantillonnage seront soit des districts, soit des IRIS ou des regroupements d'IRIS.

²⁶ A la nuance près du regroupement des petites unités, cf. infra.

²⁷ Il faudrait en effet définir une unité " adjacente " à laquelle agréger l'unité examinée. Mais de telles règles sont difficilement paramétrables, surtout quand il s'agit d'agréger des districts. La solution mise en oeuvre dans le précédent échantillon consistait à sélectionner une petite unité, à lui associer une unité plus grande l'englobant (section cadastrale) et à découper cette grande unité ; elle n'a pas été non plus retenue : on préfère partir d'emblée de la " grande " unité.

²⁸ Le fait de tirer des IRIS, plutôt que des communes, proportionnellement à leur taille, ne change évidemment pas la probabilité d'atteindre une commune donnée.

Pour la définition des unités d'échantillonnage, deux cas de figure sont à considérer :

- soit un IRIS donné ne contient que des districts de 120 logements ou plus. Ces districts pourront tous servir directement d'unités d'échantillonnage. *Ce cas n'est valable toutefois que dans les communes de plus de 10000 habitants, compte tenu de la validité du découpage en districts. Les autres communes relèvent d'office de la 2ème stratégie ci-dessous.*
- soit il contient au moins un " petit " district ou bien il est situé dans une commune de 5000 à 10000 habitants. Dans ce cas, le découpage en districts ne peut pas être utilisé pour la définition de l'unité d'échantillonnage, car cela nécessiterait d'agréger des districts trop petits. *C'est alors l'IRIS tout entier qui sert d'unité d'échantillonnage.*

Les IRIS ont donc été partitionnés automatiquement en ces deux catégories.

Au total, on dispose d'un ensemble d'unités statistiques mixtes :

- dans les communes de plus de 10000 habitants : des districts comprenant au moins 120 logements inclus dans des IRIS où il n'y a que des districts de taille supérieure ou égale à ce seuil ;
- dans les autres cas : des IRIS.

Encore une fois, cette manière de procéder est destinée à éviter de passer par une procédure de regroupement de districts lorsque ceux-ci ont une taille inférieure à la taille requise.

3.2 Agrégation des petites unités.

Pour disposer au départ d'unités ayant au moins 120 logements, la première étape a consisté à procéder au **regroupement préalable des " petites " communes et des " petits " IRIS**, c'est-à-dire de taille inférieure à ce seuil.

a) regroupement des petites communes.

Pour cette opération, on a mis au point un algorithme d'agrégation automatique des communes. Les résultats de cette procédure ont été envoyés ensuite aux DR, qui devaient valider les regroupements proposés, ou modifier leurs contours et en proposer de nouveaux, pour tenir compte, notamment, de contraintes d'accessibilité ou de non-contiguïté sur le terrain qui n'auraient pu être prises en compte par la procédure automatique²⁹.

Le choix d'une telle procédure a été dicté par le souci de concentrer les moyens humains en DR plutôt sur les opérations de découpage proprement dit que sur celles du regroupement, dans la mesure où ces dernières pouvaient être plus facilement automatisées.

Les principes de regroupement étaient les suivants :

- * les agrégats devaient être constitués *à l'intérieur de chaque strate définie par la tranche de taille d'unité urbaine* d'appartenance des communes considérées.
- * les agrégats devaient *respecter les limites des zones d'emploi* (donc des régions), mais on pouvait tolérer un chevauchement des limites départementales.
- * les agrégats devaient être *connexes* : on ne pouvait agréger une commune donnée qu'à une commune contiguë ou à un groupe de communes comportant une commune contiguë à celle en cours d'analyse.

La procédure itérative a visé principalement à agréger des petites communes entre elles, de façon à former des agrégats dont la taille était la plus proche du seuil retenu. Pour cela, les communes ont été examinées par ordre croissant de taille dans l'algorithme. Par exception, on a pu être amené, à la fin de l'opération, à agréger une petite commune à une grande, mais, en cas de candidats concurrents, la procédure cherchait toujours à minimiser la taille de l'agrégat obtenu.

²⁹ L'algorithme repose sur le *fichier de contiguïté des communes* qui s'appuie sur la contiguïté des limites de communes sur les plans. Cependant, dans certains cas, des regroupements proposés de manière automatique peuvent ne pas convenir du fait de la situation géographique réelle (communes séparées par un massif montagneux, une rivière ou un estuaire important, ou dont les chefs-lieux sont éloignés alors que leurs territoires sont adjacents, avec, par exemple, présence d'un massif forestier important...).

Le résultat de la procédure automatique de regroupement a été fourni aux DR sous forme d'une liste des agrégats formés et des communes qu'ils contenaient³⁰. On leur a fourni également un outil cartographique (dérivé de celui développé pour la constitution des Unités primaires de l'échantillon-maître), afin de les aider à visualiser ces regroupements (cf. contribution de G. BOURDALLE).

Les DR devaient valider les regroupements proposés. Elles devaient, notamment, vérifier que les agrégats délimités n'étaient pas coupés par des barrières géographiques importantes (massifs montagneux...), n'avaient pas de forme trop atypique (chapelet de petites communes s'étendant sur une grande distance...) et qu'il n'y avait pas, à leur sens, de regroupements plus pertinents que ceux proposés, notamment du point de vue des déplacements terrain³¹.

Le cas échéant, elles devaient alors proposer des modifications.

b) regroupement des petits IRIS.

Pour toutes les communes découpées en IRIS, il convenait de regrouper, s'il en existait, les IRIS de taille inférieure au seuil de 120 logements. Comme ils constituaient une exception³² et étaient en faible nombre, ils ont été **traités par procédure manuelle.**

On n'a pas édicté de règle générale. Les DR, sur la base de la cartographie existante, devaient proposer des agrégations de ces très petits IRIS, *en privilégiant la règle de la contiguïté* : un IRIS de ce type devait être agrégé à un IRIS voisin (*de la même commune obligatoirement*), de telle sorte que l'ensemble ait une taille suffisante (d'au moins 120 logements). Il était recommandé de ne pas créer d'agrégats trop gros afin de ne pas augmenter la charge de leur découpage ultérieur s'ils étaient tirés.

Pour aider les Directions régionales dans ce travail, la liste des communes concernées et des IRIS³³ qui en faisaient partie (ainsi que leurs tailles en nombres de logements) leur a été fournie. En retour, les DR devaient compléter cette liste par l'indication des **IRIS de rattachement**³⁴.

³⁰ Identifiés par un identifiant de la forme : n° de région, n° de la zone d'emploi, code TU99, lettre A, n° séquentiel dans la région à 4 positions.

³¹ Sur la base de la question suivante : si un agrégat donné est tiré, sera-t-il facile d'organiser les déplacements pour définir et valider le découpage en aires de l'ensemble de son territoire ?

³² Il y a 1721 communes découpées en IRIS, soit 15024 IRIS. Parmi ceux-ci, 728 ont une taille inférieure à 120 logements (données issues des fichiers du RP 1999 à la date de début Mars 2000). Il s'agit d'IRIS dont le territoire est composé principalement de zones d'activité ou de zones non bâties.

³³ Identifiés par les codes département, commune et un code alphanumérique à 4 positions.

³⁴ Par convention, le plus gros IRIS de l'agrégat obtenu.

4. Constitution de l'échantillon des Unités de 1^{ère} phase

4.1. Détermination de la taille et de la répartition de l'échantillon des Unités de 1^{ère} phase.

Les unités échantillonnées en première phase sont stratifiées selon deux critères explicites : *région et catégorie d'agglomération (code TU99)* (certaines modalités étant regroupées³⁵).

Le nombre d'unités à tirer dans chaque strate a été déterminé par un *calcul d'allocation optimale* de l'échantillon, visant à minimiser la variance de l'estimation nationale trimestrielle du nombre de chômeurs sous les contraintes suivantes :

- contraintes budgétaires, fixant le nombre d'unités échantillonnées de 1^{ère} phase (donc le nombre d'aires mobilisables au cours d'un trimestre donné) : **soit environ 2550 unités.**

- contraintes de précision européennes, imposées par le règlement européen régissant les enquêtes « Forces de travail ».

La rédaction du règlement européen.

"1. Pour un groupe de chômeurs représentant 5% de la population d'âge actif, l'erreur-type relative pour l'estimation des moyennes annuelles (...) ne doit pas excéder 8% au niveau de la NUTS II³⁶ de la sous-population en cause.

2. Dans le cas d'une enquête continue, pour des sous-populations représentant 5% de la population d'âge actif, l'erreur-type relative pour l'estimation des variations entre deux trimestres successifs, au niveau national, ne doit pas excéder 2% de la population en cause"

³⁵ Les regroupements utilisés, à partir du code TU99, sont les suivants : 0, 1-2, 3-4, 5-6, 7-8.

³⁶ Terminologie européenne pour désigner le niveau « région ». La Corse, en tant que région de taille inférieure à 300.000 habitants, n'est pas assujettie à ces contraintes de précision régionale.

Remarques sur l'interprétation du règlement européen.

La première contrainte porte explicitement sur une population de *chômeurs*.

Il est clair que, si la lettre du règlement n'est pas ambiguë, sa mise en oeuvre au sens strict pose problème : d'une part, comment interpréter le critère dans les cas où la part du chômage dans la population en âge de travailler est plus forte que le seuil de 5 % ? En France, ainsi, d'après l'enquête Emploi de Janvier 1999, les chômeurs représentaient 8.0 % de la population des 15-64 ans. Et comment l'interpréter dans un contexte où elle décroîtrait tendanciellement ?

D'autre part, la référence au chômage disparaît complètement dans le point 2 qui parle de n'importe quelle sous-population représentant 5% d'une population de référence. Ce point pose problème puisque les calculs de précision sur une variation trimestrielle font intervenir le taux de renouvellement de l'échantillon (fixé une fois pour toutes), mais aussi la *corrélation inter-trimestrielle de la variable d'intérêt*, qui est spécifique au choix de celle-ci (pour le chômage BIT : de l'ordre de 0.7). De ce point de vue, il ne semble pas possible de satisfaire une règle valable pour toute sous-population d'un poids donné.

On retrouve aussi une illustration de l'antagonisme entre les deux systèmes de contraintes : la précision annuelle régionale sera d'autant meilleure que la corrélation entre deux trimestres est plus faible et c'est l'inverse pour la précision de la variation trimestrielle nationale.

Les contraintes européennes, particulièrement les contraintes régionales, étant très coercitives et semblant difficiles à satisfaire dans leur intégralité compte-tenu des contraintes budgétaires, plusieurs scénarii ont dû être élaborés, pour parvenir à la solution la plus satisfaisante, sous le double point de vue national et européen.

Ces scénarii reposent sur un modèle d'allocation de Neyman, dans lequel on a fait l'hypothèse simplificatrice d'un tirage aléatoire simple sans remise des aires au sein de chaque strate. Le nombre d'aires déterminé par le modèle fournit alors le nombre d'unités échantillonnées de 1ère phase³⁷.

On a cherché à *optimiser la précision d'une estimation nationale trimestrielle du nombre de chômeurs*. Les paramètres intervenant dans ce calcul (essentiellement, la variance au sein de chaque strate des totaux des nombres de chômeurs par aire) ont été estimés à partir des enquêtes annuelles les plus récentes, après avoir artificiellement coupé les aires de 40 logements en deux pour se rapprocher des conditions du futur échantillonnage. On a fait naturellement comme si l'estimation issue d'une enquête annuelle ponctuelle pouvait se transposer à une

³⁷ Il est clair que ces calculs, réalisés avant toute donnée d'enquête réelle, demeurent approximatifs. Ils devront être refaits lorsque l'on disposera de données réelles et en tenant compte de l'ensemble des spécifications du plan de sondage mis en oeuvre.

enquête trimestrielle continue. Notamment, on n'a pas tenu compte de possibles différences des valeurs de ces paramètres entre les différents trimestres (l'enquête annuelle concernant seulement le mois de Mars).

Ce critère à optimiser est lié à la contrainte européenne en termes de variation trimestrielle nationale. En effet, le passage du coefficient de variation pour l'estimation du niveau trimestriel national au « pseudo-CV » utilisé dans la contrainte EUROSTAT pour les variations trimestrielles nationales peut se faire approximativement en multipliant par $\sqrt{2[1 - (1 - \alpha)\rho]}$, où ρ est le coefficient de corrélation entre les valeurs individuelles de la variable d'intérêt relatives à la même unité échantillonnée entre deux trimestres consécutifs (supposé constant sur toutes les strates) et α le taux de renouvellement de l'échantillon (1/6).

Les premières données du dispositif léger donnent une valeur de ρ égale à 0.62, tandis que la norme prise par EUROSTAT est de 0.8.

Le coefficient évoqué ci-dessus vaut 1 si ρ vaut 0.6 et 0.82 si ρ vaut 0.8. Il pourrait être amélioré si, au lieu de faire la différence simple de deux niveaux trimestriels pour estimer une variation, on pondérerait de manière adéquate la fraction d'échantillon constante et la partie renouvelée. On obtiendrait alors respectivement les valeurs de 0.94 et de 0.68 pour le coefficient précédent.

Pour la simulation d'une estimation nationale régionale, intervenant dans la 1ère contrainte européenne, on a dû convertir fictivement les estimations ponctuelles régionales issues des enquêtes annuelles en une estimation annuelle moyenne mobilisant plusieurs trimestres.

En prenant l'option de ne retenir, parmi les sous-échantillons trimestriels interrogés au cours d'une année, que des sous-échantillons *indépendants* entre eux (soit 9), on passe de la taille trimestrielle à son équivalent annuel en multipliant par un facteur égal à 1.5.

La mobilisation de tous les sous-échantillons trimestriels d'une année pour évaluer le niveau annuel moyen, comme moyenne simple des estimations trimestrielles (à défaut de rechercher une pondération optimale de ces différents sous-échantillons en fonction de leur nombre de répétitions), conduit à un coefficient de passage de la taille trimestrielle à la taille annuelle équivalente égal à :

$$\frac{48}{12 + 15\rho_1 + 8\rho_2 + 3\rho_3}$$
, où ρ_i désigne le coefficient de corrélation entre les valeurs individuelles de la variable d'intérêt relatives à la même unité échantillonnée entre les trimestres t et $t + i$. Ces paramètres sont estimés à partir du dispositif léger.

Pour $\rho_1 = 0.6$, $\rho_2 = 0.5$, $\rho_3 = 0.4$, ce coefficient vaut 1.83.

Pour $\rho_1 = 0.8$, $\rho_2 = 0.7$, $\rho_3 = 0.6$, il vaut 1.53.

Il est intéressant de donner quelques éléments sur les simulations effectuées pour illustrer les différents scénarii et l'arbitrage délicat entre les différentes contraintes.

	Allocation optimale sans contrainte régionale	Seuil uniforme à 8.4 %	Seuils différentiels (8.4 et 10 % pour les 3 plus petites régions)	Seuil à 8.4 %, pas de contrainte pour les 3 petites régions	Pour mémoire : estimation des CV régionaux de l'enquête annuelle
ILE DE FRANCE	508 (3.7)	409 (4.1)	438 (4.0)	466 (3.9)	3.9
CHAMPAGNE-ARDENNE	64 (10.3)	95 (8.4)	93 (8.4)	95 (8.4)	8.3
PICARDIE	74 (7.8)	64 (8.4)	64 (8.4)	68 (8.2)	6.7
HAUTE-NORMANDIE	89 (7.6)	71 (8.4)	77 (8.1)	81 (7.9)	7.0
CENTRE	77 (7.3)	62 (8.2)	66 (7.9)	70 (7.6)	6.0
BASSE-NORMANDIE	50 (12.1)	102 (8.4)	72 (10.0)	45 (12.6)	8.2
BOURGOGNE	65 (10.4)	98 (8.4)	96 (8.4)	98 (8.4)	7.8
NORD-PAS-DE-CALAIS	220 (4.7)	177 (5.2)	190 (5.1)	202 (4.9)	5.5
LORRAINE	86 (8.0)	78 (8.4)	76 (8.4)	79 (8.4)	7.7
ALSACE	62 (12.0)	126 (8.4)	123 (8.4)	126 (8.4)	9.5
FRANCHE-COMTE	43 (13.3)	108 (8.4)	76 (10.0)	40 (13.9)	8.5
PAYS DE LA LOIRE	140 (6.4)	113 (7.1)	121 (6.9)	128 (6.7)	6.4
BRETAGNE	92 (7.6)	74 (8.4)	79 (8.2)	84 (7.9)	6.0
POITOU-CHARENTES	65 (9.9)	90 (8.4)	88 (8.4)	90 (8.4)	7.6
AQUITAINE	118 (6.5)	95 (7.3)	102 (7.0)	109 (6.8)	6.4
MIDI-PYRENEES	99 (7.8)	85 (8.4)	86 (8.4)	91 (8.1)	7.4
LIMOUSIN	26 (18.1)	119 (8.4)	84 (10.0)	23 (18.9)	8.7
RHONE-ALPES	221 (5.7)	178 (6.3)	191 (6.1)	203 (5.9)	5.4
AUVERGNE	43 (11.5)	79 (8.4)	78 (8.4)	79 (8.4)	7.0
LANGUEDOC-ROUSSILLON	123 (7.1)	99 (7.9)	106 (7.6)	112 (7.4)	6.8
PACA-CORSE ³⁸	284 (4.7)	228 (5.2)	245 (5.1)	260 (4.9)	5.2
TOTAL ³⁹	2550	2550	2550	2550	40
CV NATIONAL (%)	2.01	2.13	2.08	2.05	1.50

³⁸ La Corse, non assujettie aux contraintes de précision régionale, a été, par commodité de gestion, regroupée avec PACA.

³⁹ Il peut y avoir de légères erreurs d'arrondi.

⁴⁰ Le nombre d'aires total de l'enquête annuelle issues du RP 1990 est de 3270 mais les aires sont de tailles différentes (20 logements en urbain, 40 en rural). Néanmoins, les CV ont été ici recalculés après partage artificiel des aires de 40 en 2 aires de 20. Les allocations régionales correspondantes ne sont donc pas strictement comparables à celles des autres scénarii.

Lecture du tableau :

Les différents scénarii présentés correspondent à des allocations régionales évaluées en fonction des contraintes (relâchées ou pas) de seuil de coefficient de variation (CV) pour chaque région. Le seuil EUROSTAT de 8 %, quand il est explicitement utilisé, est élargi à 8.4 %, assimilable à un 8 % par arrondi.

Dans chaque case, figurent :

- le nombre d'aires à interroger trimestriellement dans chaque région (sur la base d'un **total national de 2550 aires** correspondant à la traduction de la contrainte budgétaire, soit un taux de sondage national trimestriel moyen de 1 /563)

- le **CV correspondant (en %) de l'estimation du niveau régional annuel moyen du nombre de chômeurs**⁴¹.

La dernière ligne indique, pour chaque scénario, le **CV de l'estimation du niveau national trimestriel du nombre de chômeurs**. On rappelle que l'exercice d'allocation régionale de l'échantillon vise à minimiser la variance de cette dernière estimation⁴².

En multipliant par un coefficient évalué en 1ère approximation à 0.91⁴³, ce CV donne la valeur à comparer au seuil de précision imposé par EUROSTAT pour l'estimation de la **variation nationale trimestrielle du nombre de chômeurs (seuil de 2%)**.

Cependant, ces calculs demeurent approximatifs dans la mesure où il ne s'agit que de simulations et non d'un calcul exact : celui-ci devrait tenir compte, d'une part, du vrai plan de sondage des unités et, d'autre part, de toutes les procédures de correction et de redressement qui seront mises en oeuvre dans l'enquête Emploi en continu, en particulier les possibilités de calage pour les estimations régionales. Il faudra étudier notamment si des pyramides des âges régionales pourront être obtenues, afin d'examiner, à terme, l'apport de précision sur les niveaux régionaux que peut constituer un calage sur une telle pyramide (cf. infra, § 6.5). Enfin, les calculs, menés actuellement par extrapolation de l'enquête annuelle, devront être refaits en utilisant les données réelles de l'enquête en continu.

⁴¹ La conversion de la taille de l'échantillon trimestriel en un équivalent annuel qui tient compte de la répétition des échantillons et d'estimations de la corrélation inter-trimestrielle pour la variable étudiée (chômage BIT) se fait en utilisant le coefficient multiplicatif 1.7.

⁴² Le résultat d'une allocation sans contrainte de seuil régional conduit, grosso modo, à un nombre d'unités à interroger par strate proportionnel à la taille de la strate et à l'écart-type de la variable d'intérêt dans la strate. Si celui-ci est le même partout, le nombre d'unités à tirer sera proportionnel à la taille de la strate. Un écart-type plus faible dans une strate conduira, par exemple, à diminuer corrélativement ce nombre d'unités et inversement.

⁴³ Sur la base d'un coefficient de corrélation sur la variable « chômage BIT » de 0.7 entre 2 trimestres consécutifs.

L'influence de la taille des régions.

Les différences entre les résultats de l'allocation sans contrainte et de celles où une contrainte est explicitement introduite peuvent s'interpréter par l'importance des écarts de taille entre les régions extrêmes. En effet, en France par exemple, l'Ile de France a un poids très important. Avoir une bonne précision sur un niveau trimestriel national entraîne d'avoir une bonne précision sur cette grande région. En revanche, on peut « sacrifier » sur la précision des plus petites régions (Limousin, par exemple).

Si l'on impose des seuils de précision homogènes ou même différenciés, on doit réallouer une partie des aires des plus grosses régions au profit des plus petites, et ce au détriment de la précision des grosses régions et de la précision nationale.

Le scénario retenu in fine.

On a retenu finalement un scénario proche de la 3^{ème} colonne du tableau précédent : CV régional maximal de 10 %. Il apparaît alors que 8 régions sont à ce seuil et 10 en dessous de 8%. Le CV régional minimal est de 3,9 pour l'Ile de France et le CV global est de 2,04% pour le trimestriel national, soit quasiment la cible.

On a fixé un *nombre d'unités à découper supérieur au nombre d'aires devant être introduites au démarrage dans l'échantillon (soit 3002)*. Ceci constitue un parti de prudence, pour se prémunir contre les risques que la précision réelle soit inférieure à la précision souhaitée ou que, par exemple, la non-réponse soit plus importante que prévu, et pour s'autoriser, dans la mesure où le budget le permettrait, une augmentation ultérieure de la taille de l'échantillon. Les unités excédentaires constitueront une *réserve*⁴⁴.

L'échantillon utile a été finalement calibré à 2554 unités⁴⁵. On a d'abord constitué l'échantillon complet (les 3002) (chaque unité étant tirée avec une probabilité proportionnelle à sa taille), puis on a sélectionné l'échantillon utile par un tirage aléatoire simple dans chaque strate au sein de l'échantillon complet.

⁴⁴ Dans la pratique, cette réserve a servi à remplacer des unités échantillonnées pour diverses causes : zone déjà interrogée récemment dans le cadre du « dispositif léger », modification complète de la désignation de la voirie rendant impossible tout repérage, zone sinistrée à Toulouse suite à l'explosion de l'usine AZF en septembre 2001, commune des Vosges s'étant signalée de manière particulière lors du RP 99...

⁴⁵ Ce chiffre est à comparer aux 3270 de l'enquête annuelle ; mais la comparaison est biaisée, car les aires de l'enquête annuelle sont soit de 40, soit de 20 logements. On peut dire néanmoins que l'échantillon de l'enquête Emploi en continu présente une moindre dispersion spatiale que celui de l'enquête annuelle, compensée par l'étalement temporel.

Chaque unité échantillonnée s'est vu attribuer un *n° de dossier* (n° séquentiel national) ; on lui a également affecté de manière aléatoire un *rang de traitement* (de 1 à 7), la réserve étant versée dans le rang 7. Ces rangs indiquent l'ordre de traitement par les Directions régionales et constituent un n° de sous-échantillon provisoire.

Les tableaux suivants donnent la répartition finalement adoptée des unités échantillonnées de 1ère phase⁴⁶, strate par strate.

Un examen géographique fin montre que les 167 unités urbaines de plus de 32000 habitants font toutes partie de cet échantillon.

⁴⁶ Ces tableaux sont classés par région de rattachement des unités. Pour des raisons d'organisation, certaines DR sont cependant susceptibles de superviser en partie les travaux de terrain relevant du territoire d'une autre DR.

**TABLEAU 1 : NOMBRE d'UNITÉS ECHANTILLONNÉES de 1ère PHASE
dans chaque STRATE
(REGION * TRANCHE de TAILLE d'UNITÉ URBAINE AGRÉGÉE).
RÉSERVE INCLUSE.**

	Communes rurales (TU = 0)	UU de moins de 10000 habitants (TU = 1,2)	UU de 10000 à moins de 50000 hab. (TU = 3,4)	UU de 50000 à moins de 200 000 hab. (TU = 5,6)	UU de 200000 hab. ou plus (TU = 7,8)	TOTAL
ILE DE FRANCE	16	8	19	5	429	477
CHAMPAGNE- ARDENNE	29	11	32	30	17	119
PICARDIE	31	19	20	34	-	104
HAUTE-NORMANDIE	29	7	18	8	47	109
CENTRE	30	18	15	24	29	116
BASSE-NORMANDIE	36	13	15	27	-	91
BOURGOGNE	43	18	28	8	21	118
NORD-PAS-DE-CALAIS	18	13	16	53	104	204
LORRAINE	21	16	16	32	16	101
ALSACE	13	12	11	3	41	80
FRANCHE-COMTE	26	8	13	31	-	78
PAYS DE LA LOIRE	32	14	18	40	31	135
BRETAGNE	40	18	21	14	20	113
POITOU-CHARENTES	47	12	18	32	-	109
AQUITAINE	33	12	17	21	39	122
MIDI-PYRENEES	35	14	18	8	35	110
LIMOUSIN	40	9	6	34	-	89
RHONE-ALPES	49	21	30	42	94	236
AUVERGNE	34	11	11	11	18	85
LANGUEDOC- ROUSSILLON	33	31	20	37	20	141
PACA	25	17	33	25	146	246
CORSE	8	2	1	8	-	19
TOTAL	668	304	396	527	1107	3002

**TABLEAU 2 : NOMBRE d'UNITÉS ÉCHANTILLONNÉES de 1ère PHASE
dans chaque STRATE
(RÉGION * TRANCHE de TAILLE d'UNITÉ URBAINE AGRÉGÉE).
RÉSERVE EXCLUE.**

	Communes rurales (TU = 0)	UU de moins de 10000 habitants (TU = 1,2)	UU de 10000 à moins de 50000 hab. (TU = 3,4)	UU de 50000 à moins de 200000 hab. (TU = 5,6)	UU de 200000 hab. ou plus (TU = 7,8)	TOTAL
ILE DE FRANCE	15	8	17	5	387	432
CHAMPAGNE- ARDENNE	22	9	24	23	13	91
PICARDIE	24	15	16	26	-	81
HAUTE-NORMANDIE	22	5	14	6	36	83
CENTRE	24	14	12	19	23	92
BASSE-NORMANDIE	29	11	12	22	-	74
BOURGOGNE	33	14	22	6	16	91
NORD-PAS-DE-CALAIS	16	12	15	48	94	185
LORRAINE	17	13	13	26	13	82
ALSACE	11	10	9	2	33	65
FRANCHE-COMTE	21	7	10	25	-	63
PAYS DE LA LOIRE	29	13	16	36	28	122
BRETAGNE	33	14	17	11	16	91
POITOU-CHARENTES	36	9	14	25	-	84
AQUITAINE	29	10	15	19	35	108
MIDI-PYRENEES	30	11	15	7	30	93
LIMOUSIN	32	7	5	27	-	71
RHONE-ALPES	44	19	27	38	85	213
AUVERGNE	27	9	9	9	15	69
LANGUEDOC- ROUSSILLON	29	27	17	32	18	123
PACA	23	17	29	23	132	224
CORSE	7	2	1	7	-	17
TOTAL	553	256	329	442	974	2554

4.2 Tirage de l'échantillon des Unités de 1ère phase.

Une fois constituées les unités de 1ère phase et définies la taille et l'allocation de l'échantillon, l'UMS a procédé au tirage de l'échantillon.

Les différentes unités (sans distinction selon leur nature) ont été triées par strate, puis, à l'intérieur de chaque strate, par zone d'emploi, commune et IRIS (le cas échéant). Le tri par zone d'emploi a pour effet d'assurer une stratification indirecte selon ce critère (avec allocation proportionnelle). La zone d'emploi est en effet une délimitation pertinente en regard de l'objet de l'enquête et la procédure permet une dispersion maximale des unités tirées entre les zones d'emploi d'une même strate, assurant ainsi la meilleure représentation de celles-ci.

Les unités ont été alors tirées par un tirage systématique, strate par strate, sur ce fichier, de façon à leur assurer une probabilité de tirage proportionnelle à leur taille, en nombre total de logements.

Il n'a pas été jugé utile, au moment de ce tirage, d'introduire des conditions d'équilibrage : assurer par exemple que l'échantillon des unités de 1ère phase conduise à une estimation exacte de totaux connus sur l'ensemble de la population. Ce choix se justifie par plusieurs raisons :

a) dans la mesure où l'unité de 1ère phase n'est pas l'unité finale mais que des tirages et des découpages sont ensuite réalisés en son sein, ces conditions d'équilibrage perdent de leur intérêt (surtout si les unités comportent une certaine hétérogénéité : rien n'indique que la partie tirée et découpée sera un estimateur exact des mêmes totaux exogènes au sein de l'unité de 1ère phase)

b) les données exogènes disponibles étaient de toute façon assez pauvres, on ne disposait au moment du tirage que des données de la population légale au RP 1999.

c) ces données n'étaient disponibles qu'au niveau communal, alors que l'on devait tirer dans certains cas des unités infra-communales.

L'annexe 1 montre un extrait du fichier de tirage des unités de 1ère phase, tel qu'il a été remis aux Directions régionales. Chaque dossier est relatif à une unité particulière, dont apparaissent le rang, le type, la région, la zone d'emploi et la commune de rattachement, ainsi que la tranche de taille d'unité urbaine à laquelle appartient celle-ci.

Par commodité, le fichier se présente sous forme hiérarchisée, détaillant la liste des districts du recensement composant l'unité ou les différentes communes d'un regroupement de communes. Ceci permet de faciliter le travail de découpage en grandes zones ou en secteurs et le recoupement avec les bordereaux du RP et la cartographie cadastrale, sans néanmoins se substituer aux opérations nécessaires au découpage (incluant la visualisation des différentes portions de territoire sur les plans et le calcul de leur taille). Les unités sont séparées par un *délimiteur* et la totalisation du nombre de logements au niveau de chaque unité est effectuée, ainsi que leurs nombres de logements. Enfin, un *nombre aléatoire* est indiqué, qui sert au tirage des grandes zones ou des secteurs (cf. infra, § 5.2.d).

**ANNEXE 1 : EXEMPLE DE PRESENTATION
DU FICHIER DES UNITES ECHANTILLONNEES DE 1ERE PHASE.**

----- RANG=1 DOSSIER=2326 -----										
Code Commune	Code Région	Code Commune	TYPE	TU99	Zone				Logts total	Nombre aléatoire
					Emploi	Iris	DISTRICT	LOGT		
2B096	94	CORTE	Iris	2	9414	0101	AI02	26	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE06	46	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE07	62	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AI01	73	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE02	121	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE03	153	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AB01	171	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE04	178	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AB02	207	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AB03	357	1808	688
2B096	94	CORTE	Iris	2	9414	0101	AE01	414	1808	688
----- RANG=1 DOSSIER=2328 -----										
Code Commune	Code Région	Code Commune	TYPE	TU99	Zone				Logts total	Nombre aléatoire
					Emploi	Iris	DISTRICT	LOGT		
2A004	94	AJACCIO	Iris	5	9411	0201	BX08	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX09	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX16	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX24	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX26	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX27	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX29	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX45	0	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX25	1	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX42	1	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX18	2	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX49	4	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX47	8	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX34	9	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX28	12	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX30	13	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX04	16	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX23	20	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX15	43	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX48	43	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX17	56	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX14	65	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX50	65	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX44	74	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX13	88	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX22	92	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX46	111	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX02	150	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX32	201	1284	209
2A004	94	AJACCIO	Iris	5	9411	0201	BX33	210	1284	209

5. Le découpage des aires.

Cette opération constitue la plus grande originalité du processus de constitution de l'échantillon Emploi. Elle nécessite un travail "manuel" important en Direction régionale, complété, si nécessaire, par des vérifications ou des compléments d'information recueillis sur le terrain.

L'élaboration et la validation des instructions de découpage se sont appuyées sur les travaux d'un groupe composé de représentants des Directions régionales de l'INSEE. Des tests ont été réalisés courant Avril 2000, sur un petit échantillon d'unités de différents types. Une formation nationale des responsables régionaux du découpage a eu lieu en Juin 2000. L'opération sur le terrain a débuté début Juillet 2000.

5.1. Principes généraux du découpage.

Les opérations de découpage consistent en plusieurs étapes :

- chaque unité échantillonnée est découpée en *secteurs* de 120 à 240 logements, ou reste à l'identique si elle a une taille comprise entre 120 et 240.
- un de ces secteurs est tiré proportionnellement à sa taille (en nombre total de logements).
- pour les très grosses unités, découpage préalable en *grandes zones*, puis tirage de l'une d'entre elles qui sera découpée en secteurs et poursuite de l'opération comme ci-dessus.
- le secteur sélectionné est découpé en au moins 6 aires de 20 logements.
- on établit un descriptif complet des aires retenues dans l'échantillon, à la fois en termes de délimitation géographique et en termes de contenu, sous forme de liste des identifiants des immeubles et des logements (au recensement de 1999) les constituant. La synthèse de ce travail est décrite dans le *dossier de découpage*.

5.2 Découpage des communes ou agrégats de communes.

Les unités échantillonnées qui seront des communes ou des agrégats de communes seront dénommées plus simplement par la suite *communes-échantillons*.

L'objectif à atteindre est de découper ces unités en aires comportant 20 logements en moyenne, ayant des limites facilement repérables sur le terrain par l'enquêteur qui aura à réaliser ultérieurement les enquêtes Emploi.

a) *cas des communes ou agrégats de communes ayant de 120 à 240⁴⁷ logements.*

Ces unités sont directement découpées en aires.

b) *cas des communes ou agrégats de communes ayant plus de 240 logements.*

Comme un tel découpage serait très lourd pour ces communes, on passe par l'intermédiaire d'un pré-découpage en *secteurs*. Il faut découper chaque commune-échantillon en un certain nombre de secteurs, limités soit par des rues, des routes, éventuellement des voies ferrées ou des cours d'eau, soit par les limites de la commune. **Le nombre de logements de chaque secteur doit être compris entre 120 et 240 logements.** Une commune ou un agrégat de communes de 120 à 240 logements peuvent donc être directement assimilés à un secteur (cas a)).

Il était cependant fortement conseillé de se rapprocher des limites (120), afin de simplifier le travail ultérieur sur le terrain. On a ainsi intérêt à avoir le plus grand nombre de secteurs, les plus petits et les plus compacts possibles (de surface minimale).

Parmi ces secteurs, on en tirera ensuite un avec une probabilité proportionnelle au nombre de logements recensés, qui sera découpé en aires de 20 logements en moyenne.

Le tirage au sort d'un secteur a pour but d'éviter tout risque de biais pouvant résulter d'un choix raisonné. Ainsi, la constitution des secteurs n'a pas à obéir à des choix a priori, "favorisant" telle ou telle partie de la commune-échantillon.

c) *cas particulier des communes comportant plus d'un millier de logements.*

Afin d'éviter un découpage fastidieux de l'ensemble du territoire de ces communes en secteurs de 120 à 240 logements, on procède au découpage préalable de la commune en *grandes zones de taille comprise approximativement entre 500 et 1000 logements*, ce qui revient plus ou moins à constituer des « pseudo-IRIS ».

La constitution de ces grandes zones devait idéalement utiliser un ou plusieurs districts de la zone agglomérée et un ou plusieurs districts de la zone éparsée.

Cependant, cette règle pouvait être difficile à mettre en pratique et la connexité des zones devait être recherchée avant tout.

⁴⁷ La limite à 240 peut se justifier comme suit : si l'on abaissait ce seuil, on risquerait d'avoir de petits secteurs qu'il serait difficile de découper en aires sans transgresser les règles de tolérance en matière de taille des aires. Exemple, si la limite était à 200 : une commune-échantillon de 210 logements serait découpée en 2 secteurs de 105 logements en moyenne ; mais si l'un d'entre eux est de 95 et l'autre de 115, on ne pourra plus découper le petit secteur en 6 aires sans passer en dessous d'une taille de 16 logements pour au moins une aire.

L'une des zones est ensuite tirée au sort avec une probabilité proportionnelle au nombre de logements recensés puis découpée en secteurs de 120 à 240 logements. On est ramené ensuite au cas précédent.

5.2.1 Rassemblement de la documentation pour le découpage en secteurs.

Le premier travail a consisté à rassembler les plans et cartes qui couvrent les communes-échantillons. Les cartes IGN au 1/25 000 permettent d'avoir une vue d'ensemble du territoire de la commune, incluant la localisation des limites communales et des terrains non bâtis (mais susceptibles d'être un jour construits et qui doivent, à ce titre, être affectés aux différents secteurs ou aires constitués). Ces cartes permettent aussi de bien localiser les lieux-dits inclus dans la commune.

Les autres plans utiles sont les plans du cadastre, plans établis par la commune ou plans du syndicat d'initiative, s'il y en a un. Les DR devaient voir quels documents cadastraux elles possédaient déjà et quelle était leur date de mise à jour. Elles devaient prendre contact avec la mairie pour voir si celle-ci elle disposait de plans numérisés (SIG), voire de photographies aériennes (utiles dans certains cas, mais cependant difficilement accessibles et maniables..).

Les *tableaux d'assemblage* des planches cadastrales sont très utiles pour visualiser la partie bâtie de la commune et identifier les planches particulières nécessaires⁴⁸. L'inconvénient de tous ces plans est leur hétérogénéité : on utilise des plans de toute sorte et de différentes échelles.

Le but était d'obtenir des plans sur lesquels on puisse repérer les voies principales qui traversent la commune et à partir desquels un enquêteur réalisant ultérieurement l'enquête Emploi pourrait se repérer facilement dans la commune. A ce stade, on ne se préoccupait pas des chemins, sentiers etc ... ; on s'intéressait seulement aux rues, routes et autres voies principales (cours d'eau, voies ferrées) et aux points de repère fondamentaux (bâtiments publics...).

Il était souhaitable, évidemment, que ces plans comportent les noms des voies, que ce soit des noms de rues ou des numérotations de routes, qu'il distingue les voies selon leur importance, qu'ils soient le plus récents possibles⁴⁹ et à une échelle suffisamment grande.

Il était recommandé de griser au crayon les voies repérées sur le plan, de *vérifier l'orientation* des plans (nord, sud...) et de la compléter éventuellement à la main, même approximativement, sur ceux où elle manquerait.

⁴⁸ Toutes les planches d'une commune ne sont pas utiles, certaines peuvent couvrir des portions de territoire totalement non bâties.

⁴⁹ L'un des problèmes pratiques de l'opération réside en effet dans le fait que les plans cadastraux ne sont pas toujours bien mis à jour. En particulier, mais ce n'était pas le plus grave, certains numéros de routes nationales modifiés dans les années 1970 (départementalisation et renumérotation) n'étaient pas corrigés sur les plans cadastraux, même en cas de mise à jour postérieure de ceux-ci.

5.2.2 Liste des immeubles du RP99.

L'autre matériau de base fondamental pour le découpage est la liste des immeubles au RP99 des communes concernées (*bordereaux de district M6*). Ces documents ont servi à établir une liste de tous les logements recensés dans le secteur sélectionné et à affecter les immeubles aux aires découpées. On pouvait aussi avoir recours aux fichiers de saisie des M6, comportant les pages intérieures du bordereau.

5.2.3 Découpage en secteurs.

Le découpage de la commune-échantillon ou de la grande zone en secteurs se fait principalement en Direction Régionale à l'aide des plans et cartes et des bordereaux M6 du RP. Le déplacement sur le terrain n'est pas nécessaire à ce stade sauf si les documents existants se révèlent tout à fait insuffisants.

En effet, ce peut être le cas, par exemple, si les immeubles ne sont pas repérés sur les M6 par des noms de rue ni, a fortiori, par des numéros (qui peuvent fort bien ne pas exister sur le terrain) mais seulement par des indications vagues (à droite de tel bâtiment, 1ère à gauche dans la rue...). Dans ce cas, le recours aux feuilles de logements (FL) et le déplacement sur le terrain sont nécessaires pour repérer les noms des occupants et permettre un positionnement des immeubles sur les plans. Mais il n'est pas question, à ce stade, de faire un travail de cartographie très fouillé sur l'ensemble de la commune ; il suffit de disposer des éléments permettant le partage de la commune en secteurs. La cartographie doit être affinée plus tard et uniquement sur le secteur sélectionné aléatoirement.

A partir des voies principales qui traversent la commune, on répartit ainsi l'ensemble du territoire en plusieurs secteurs en respectant les principes suivants :

1. **Il faut partitionner l'intégralité du territoire de la commune.** Tout point de la commune doit évidemment appartenir à un secteur et un seul. En particulier, les secteurs doivent englober l'ensemble des terrains non bâtis de la commune, afin de permettre la prise en compte de la construction neuve. Selon le même principe, les terrains non bâtis doivent être répartis entre le plus grand nombre possible de secteurs (et, de préférence, la totalité), tout en respectant le principe de connexité.
2. Le découpage de la commune doit être extrêmement clair ; en aucun cas on n'utilisera de voie très secondaire. Il faut en effet pouvoir ensuite affecter sans ambiguïté les immeubles présents au RP aux différents secteurs, de façon à calculer la taille totale de chaque secteur ainsi délimité.
3. **Chaque secteur doit avoir un nombre de logements compris entre 120 et 240 logements.** Les cas d'exception sont cependant tolérés, à condition d'être signalés à la Direction Générale.

Exemple :

Une commune-échantillon de moins de 240 logements constitue d'emblée un secteur.

Une commune de 490 logements sera divisée de préférence en 3 secteurs, mais si les voies principales qui traversent la commune ne le permettent pas, en 2 seulement, dont la taille moyenne sera donc légèrement supérieure au seuil maximal recommandé.

Remarques.

1. Idéalement, les secteurs devraient se ressembler le plus possible du point de vue de l'habitat. Pour ce faire, il faudrait privilégier les découpages en étoile de telle sorte que chaque secteur comporte une partie du bourg ou du centre-ville et une partie des zones résidentielles et des habitations éparses.
2. Cependant, il n'est pas toujours possible de mettre cette règle en pratique et il peut être plus facile de s'appuyer sur le découpage en districts au moment du recensement même si celui-ci n'a pas été fait en étoile⁵⁰. De surcroît, cela facilite les dénombrements de logements. Ainsi, les secteurs peuvent être directement des districts, s'ils ont la taille requise, ou bien des regroupements de districts. Dans certains cas, néanmoins, il faut " amputer " une partie d'un district pour l'affecter à un autre, de façon à obtenir un secteur de taille requise.
3. La règle du découpage en étoile doit donc être considérée comme facultative et appliquée du mieux possible mais elle ne doit pas entraîner de complications inutiles.

5.2.4 Sélection d'un secteur.

A la suite du découpage de la commune (ou de la grande zone) en secteurs, on évalue le nombre total de logements au RP99 de chaque secteur.

Les secteurs sont numérotés dans un ordre prédéterminé (dans le sens des aiguilles d'une montre en commençant par le plus au Nord).

Le responsable local de l'opération à la Direction régionale procède au tirage d'un de ces secteurs avec une probabilité proportionnelle au nombre de logements qu'il contient. Le mode opératoire est décrit en annexe 2.

⁵⁰ Il arrive fréquemment qu'un district rassemble les logements agglomérés au chef-lieu et qu'un autre rassemble les habitations éparses.

5.2.5 Etablissement d'un plan détaillé du secteur sélectionné.

Pour chaque secteur sélectionné (un par commune-échantillon), on constitue un plan détaillé qui permettra le découpage en aires. Ce plan doit être maniable mais suffisamment grand pour que le gestionnaire puisse y faire figurer les constructions, routes, rues, chemins ...

Le travail du gestionnaire en DR consiste à repérer dans le secteur sélectionné et à reporter sur le plan tout ce qui peut constituer des limites d'aire.

Le déplacement sur le terrain est la plupart du temps souhaitable (dans la limite du budget alloué à l'opération) pour vérifier et compléter le plan.

* *Quadrillage du secteur sélectionné.*

Doivent figurer sur le plan les voies suivantes qui traversent la zone :

- rues, routes
- cours d'eau, voies ferrées
- allées, sentiers, chemins facilement repérables sur le terrain, permanents et d'utilisation courante.
- limites communales

Dans les zones non habitées à grande superficie, il suffit d'indiquer les voies desservant les habitations.

Les voies doivent être distinguées selon leur importance ; leurs noms, s'ils existent, doivent figurer sur le plan. Doivent aussi figurer les principaux points de repère (bâtiments publics, etc. ...) ainsi que tous ceux nécessaires pour repérer des portions de territoire dans le secteur.

* Dans chaque portion de territoire ainsi délimitée par ce quadrillage, il faut ensuite pouvoir déterminer les logements qui en font partie à partir de la liste figurant sur les bordereaux M6 et en s'aidant du plan du secteur sélectionné. ***En cas de doute sur la qualité des bordereaux M6, il peut être nécessaire de recourir aux feuilles de logement du RP pour aider à cette détermination.***

Tous les immeubles du RP et ***seulement ceux-ci***, y compris les logements ordinaires inclus dans des communautés, qui se trouvent à la dernière page du M6, sont à prendre en considération, ainsi que leur nombre de logements.

En cas d'ambiguïté sur l'appartenance des immeubles à chacune des portions de la commune délimitées par le quadrillage, un déplacement sur le terrain est nécessaire.

Néanmoins, les écarts entre le terrain et les bordereaux du RP ne doivent pas donner lieu à recherche approfondie : il peut s'agir de logements oubliés au RP, détruits ou construits depuis cette date. **Ce qui compte, c'est la référence au RP** (c'est à partir du nombre de logements au RP que sont calculées les tailles des secteurs et des aires) et la possibilité de pouvoir ultérieurement (i.e., au moment des enquêtes sur le terrain) affecter sans ambiguïté un immeuble nouveau à l'une des aires découpées ; en particulier, le changement de destination des immeubles par rapport à la situation du RP n'a pas d'importance à cette phase. De même, si le déplacement terrain a fait apparaître un logement oublié au RP, celui-ci ne doit pas être pris en compte dans les dénombrements. Toutefois, s'il pose des problèmes particuliers de repérage, il peut être signalé, ce qui constituera un renseignement utile lors de la réalisation de l'enquête.

Il n'est pas impératif de repérer finement la localisation de tous les immeubles : en effet, ce travail sera fait de toute façon avant le démarrage de l'enquête, lors de la phase de ratissage de l'aire ; de surcroît, dans certains cas, il n'y a pas d'ambiguïté sur l'appartenance d'un immeuble à une aire ou à une autre. C'est le cas en particulier des *lieux-dits* : si l'on sait que tel lieu-dit comporte *n* logements au RP, peu importe leur localisation exacte : il suffit de savoir à quelle aire appartient le lieu-dit.

5.2.6 Découpage du secteur sélectionné en aires.

Les secteurs ont un nombre de logements variable (de 120 à 240), générant de 6 à 12 aires de 20 logements. Il faut essayer de se rapprocher au maximum de cette limite de 20 logements par aire, mais pour des raisons de terrain, on peut être amené à définir des aires légèrement plus petites ou plus grosses : **la fourchette autorisée va de 17 à 23 logements. Ce chiffre est à entendre au sens de nombre total de logements au RP 1999.**

*Nombre indicatif d'aires à construire en fonction de la taille du secteur*⁵¹.

• de 120 à 130 FL	: 6 aires
• entre 131 et 150 FL	: 7 aires
• entre 151 et 170 FL	: 8 aires
• entre 171 et 190 FL	: 9 aires
• entre 191 et 210 FL	: 10 aires
• entre 211 et 230 FL	: 11 aires
• entre 231 et 240 FL	: 12 aires

⁵¹ Ces chiffres sont indicatifs ; l'essentiel est le respect des règles de tolérance pour la taille de chaque aire énoncées ci-dessus. On cherchera cependant à éviter de constituer des aires dont la taille soit systématiquement en-dessous ou en-dessus de la moyenne.

Ce découpage se fait à partir du plan du secteur sélectionné et du positionnement des immeubles tels qu'ils ont été réalisés en Direction Régionale et confirmés le cas échéant par le déplacement terrain. Il faut obtenir des portions de territoires (aires) ayant, si possible, des limites naturelles et satisfaisant la taille requise. Une aire est définie avant tout par son contour, par des caractéristiques géographiques et des limites de bâti ; ce n'est qu'ensuite qu'on la décrira par les logements qu'elle contient. On pourra avoir intérêt à commencer le découpage à partir du centre de la commune. Un hameau peut aussi constituer une aire s'il a la taille suffisante.

Les règles suivantes doivent être adoptées :

1. **Exhaustivité** : chaque secteur doit être intégralement découpé et tout point du secteur sélectionné doit appartenir à une aire et une seule.
2. **Nombre** : impérativement, au moins 6 aires par secteur.
3. **Taille** : cf. ci-dessus.
4. Les aires doivent être d'un seul morceau (**connexes**) et si possible **contiguës**.
5. Selon des critères purement statistiques, les aires devraient se ressembler le plus possible au point de vue de l'habitat et avoir une *hétérogénéité intra*, c'est-à-dire comporter chacune, dans la mesure du possible, des proportions équivalentes d'habitat concentré, clairsemé, d'habitations éparses, d'habitat collectif ou individuel, de résidences principales ou secondaires, ou de terrains non bâtis. La première condition assure, dans le cas idéal où toutes les aires sont parfaitement identiques, qu'on n'encourra aucune perturbation lors du renouvellement de l'échantillon (c'est-à-dire quand une aire, après avoir été interrogée 6 trimestres, est remplacée par une aire voisine à l'intérieur du même secteur). La seconde a pour but de diminuer les effets de grappe. *Cependant, ces principes étant difficiles à mettre en oeuvre en pratique, il faut les considérer comme facultatifs.*

5.2.7 Désignation littérale des aires en Direction Régionale.

Le responsable local de l'opération en Direction Régionale identifie de manière littérale (A, B,...) les aires constituées (*selon un principe systématique : dans le sens des aiguilles d'une montre en commençant par la plus au Nord*).

5.3 Découpage des IRIS et des districts.

Rappel : dans les communes découpées en IRIS (plus de 10000 habitants, parfois plus de 5000), l'unité échantillonnée est l'*IRIS 2000* (les très petits IRIS ayant été au préalable regroupés ou agrégés à un IRIS de plus grande taille). Dans le cas où un IRIS ne comporte que des districts de 120 logements ou plus, l'unité échantillonnée est alors directement le district.

Dans ces communes, le découpage de l'IRIS peut s'appuyer sur le concept de *district du recensement*, les districts étant de tailles variables.

La Direction régionale doit réaliser un plan représentant l'unité échantillonnée puis le découper en secteurs de 120 à 240 logements. Les IRIS étant des zones comportant de l'ordre de 800 logements en moyenne⁵², un IRIS pourra être découpé en 4 à 7 secteurs. Les plus gros IRIS (plus de 1000 logements) pourront, comme cela été dit à propos des grandes communes (cf. supra, § 5.2, point c)), être découpés au préalable en grandes zones de 500 à 1000 logements.

5.3.1 Rassemblement de la documentation.

Les plans associés aux bordereaux de districts et les planches cadastrales restent un outil nécessaire, comme pour le découpage des communes.

Néanmoins, pour le découpage des IRIS ou districts, qui se font dans des zones d'urbanisation plus ou moins denses, on dispose de différents outils de cartographie numérisée. On peut, notamment, visualiser l'IRIS et, à partir de SEDDL, sortir le nombre de logements par IRIS et par îlot. Grâce à CICN, les fonds " Baseilots " peuvent être utilisés pour visualiser les îlots et leurs numéros, mais les numéros de voie ne figurent pas toujours et certaines voies internes aux îlots ne sont pas représentées. Avec REPLIC, on peut préciser l'appartenance d'une adresse à un îlot quand elle ne figure pas dans CICN.

Ces plans doivent comporter les voies principales qui traversent et entourent les districts représentés (identifiés par leur numéro RP), ainsi que les principaux points de repère.

Les plans doivent être accompagnés du nombre de logements par district (que l'on peut avoir directement à partir de la cartographie numérisée).

Les secteurs sont alors délimités facilement par regroupement d'îlots et leur taille est calculée à l'aide des bordereaux M6.

Pour les îlots ayant plus de 240 logements, on doit avoir recours aux M6 pour faire un découpage au niveau des immeubles en utilisant la numérotation des voies pour pouvoir les repérer facilement.

⁵² Soit 2000 habitants.

Dans le cas où les limites des districts sont des limites artificielles, il faut indiquer si possible les limites visibles les plus voisines de ces limites artificielles ainsi que le nombre de logements compris entre les limites artificielles et les limites visibles (il sera en général nécessaire d'aller sur le terrain).

5.3.2 Liste des immeubles au RP99.

Comme au § 5.2.b : il faut rassembler la liste de tous les immeubles recensés au RP99, avec leur nombre de logements. Dans les zones fortement urbaines, les bordereaux d'immeubles collectifs (DIC) sont utiles.

5.3.3 Découpage en secteurs.

Le découpage de l'unité se fait en Direction Régionale à l'aide du plan défini ci-dessus et en respectant les principes du paragraphe 5.2.c (supra).

5.3.4 Sélection d'un secteur.

À la suite du découpage en secteurs, on connaît le nombre de logements (au RP) correspondant à chaque secteur.

Le responsable local de l'opération à la Direction Régionale assure le tirage d'un secteur particulier, avec probabilité proportionnelle au nombre de logements qu'il contient (cf. supra).

5.3.5 Etablissement d'un plan détaillé du secteur sélectionné.

On tire un plan du secteur en format A3 avec Mapinfo.

Dans les secteurs constitués principalement d'un habitat clairsemé (pavillons...), on procède de la même manière que pour le découpage des communes (cf. supra). La même remarque s'applique aux secteurs où il n'y a pas de numérotation des immeubles dans les voies. En particulier, on cherchera à repérer ultérieurement la construction neuve sur les aires constituées ; on doit donc s'intéresser aux terrains non bâtis lors du quadrillage du secteur.

Dans les secteurs constitués d'un habitat dense, les pâtés de maisons ou pâtés d'immeubles contiennent souvent un nombre de logements supérieur au nombre de logements d'une aire ; par ailleurs, il est très difficile, pour les immeubles enchevêtrés, de repérer précisément la situation d'un logement sur un plan.

Vouloir construire un quadrillage en zone urbaine dense introduirait donc des difficultés importantes. Le découpage dans les secteurs à grande densité d'habitat consiste donc non pas tant à quadriller qu'à constituer des *grappes d'immeubles* définies par la localisation de leurs accès sur rue.

Sur le plan doivent figurer les voies qui traversent la zone (rues, routes, cours d'eau, voies ferrées, allées ...). Il est utile de savoir dans quel sens sont numérotées les voies, pour ne pas créer par erreur des portions d'aires définies par des tronçons de numéros, qui ne seraient pas contiguës.

Une fois ce repérage effectué, le gestionnaire doit être en mesure de déterminer, dans chaque portion de territoire délimitée, les immeubles RP désignés par leur adresse (au RP) et leur numéro d'identification.

Le déplacement terrain est nécessaire dès lors qu'il n'y a pas de repérage suffisant sur les M6, que la numérotation de la voirie est insuffisante ou ne peut se lire sur les M6, qu'on ne sait pas le sens de numérotation des voies (on peut cependant essayer de l'obtenir par la mairie)...

5.3.6 Découpage du secteur sélectionné.

Ce découpage se fait à partir du plan et de l'état descriptif du secteur sélectionné. En zone urbaine dense, les dossiers collectifs d'immeubles (DIC) sont un atout précieux. Faire attention qu'à une même adresse peuvent correspondre plusieurs DIC. En revanche, dans ces zones denses, le déplacement terrain n'apporte quasiment aucune information.

On s'efforce d'abord de limiter des aires par des voies.

Quand cela n'est pas possible, et qu'une limite d'aire doit couper un pâté d'immeubles, il faut faire en sorte que tous les immeubles ayant même accès sur rue appartiennent à la même aire.

Il faut faire attention au problème des *coins de rue* : si une aire comporte des immeubles ouvrant sur deux rues sécantes, il faut vérifier que l'aire est bien d'un seul morceau, compte tenu du sens de numérotation des rues.

Enfin dans les parties d'habitat très denses, il arrive qu'un immeuble doive appartenir à deux aires différentes : dans ce cas, on doit faire en sorte qu'une même cage d'escalier appartienne à une aire, puis, quand cela devient impossible (immeuble de grande hauteur), qu'un même étage appartienne à une seule aire.

Les règles suivantes doivent être adoptées :

- il faut au moins 6 aires par secteur.
- tout point du secteur sélectionné doit appartenir à une aire et une seule.
- enfin, on s'efforcera que chaque aire comprenne entre 17 et 23 logements (au sens du RP99) (cf. supra, pour le nombre indicatif d'aires à construire). Néanmoins, en zone urbaine très dense, on peut relâcher les seuils de tolérance *maximaux pour éviter d'avoir à couper un étage.*

5.3.7 Numérotation des aires.

Les aires sont numérotées de manière littérale (cf. supra).

5.4 Quelques remarques sur certaines difficultés du découpage.

5.4.1 Description des limites des secteurs ou des aires.

Il faut rappeler que la description des secteurs et des aires est avant tout *géographique* : elle doit permettre aux futurs enquêteurs de se repérer facilement sur le terrain pour identifier les logements à enquêter.

Il faut donc éviter des descriptions de secteurs ou d'aires faisant référence à des n°s de districts RP ou à des n° d'immeubles RP.

Par ailleurs, la configuration terrain peut évoluer au cours du temps à partir de la situation au RP. Il importe donc de faire en sorte que l'enquêteur, lorsqu'il fera le ratissage, soit en mesure d'affecter sans ambiguïté un nouveau logement à une aire donnée. Ainsi, par exemple, dire qu'une aire est formée des immeubles occupant les n°s 4, 6, et 10 de la rue X n'est pas correct, même si le n°2 est un établissement d'activité et l'emplacement du n°8 un terrain vague au moment du RP. En effet, plusieurs années après, l'établissement peut avoir été transformé en logements et, de toute façon, l'enquêteur devra vérifier qu'il n'abrite pas un logement ordinaire (gardien) et le terrain vague peut devenir bâti : l'enquêteur ne saura alors pas si ces logements font ou non partie de l'aire.

Ceci conduit aux cas de figure suivants, présentés par ordre de préférence.

a) délimitation à partir d'une ligne fermée : l'aire ou le secteur est alors l'ensemble du territoire « enfermé à l'intérieur » de cette ligne (cette notion topologique se comprend intuitivement). Tout terrain non bâti, mais susceptible un jour de l'être, appartient alors clairement au secteur ou à l'aire. On peut employer le vocabulaire “ *aire ou secteur délimités par...* ”.

La ligne fermée est décrite par l'énumération des voies ou autres types de limites (cours d'eau, limites communales...) la composant. L'indication de la numérotation des voies n'est pas du tout souhaitable.

Il faut veiller dans ce cas à désigner les voies ou les limites *dans l'ordre où elles se présentent*, de façon qu'on puisse les suivre sur le plan.

b) délimitation à partir d'un tronçon de voirie (surtout en zone urbaine) : le secteur ou l'aire sont alors composés des immeubles ouvrant sur un tronçon de voirie défini par son appellation, le côté de la voie qui est concerné (numéros pairs ou impairs), les extrémités du tronçon (angle d'une autre voie, carrefour...) et/ou ses numéros extrêmes. Utiliser alors le vocabulaire “ *aire ou secteur composés des immeubles ouvrant sur ...* ”. S'il y a plusieurs tronçons, chacun d'eux est supposé s'étendre du précédent au suivant. On dira donc, par exemple : “ *aire composée des*

immeubles de la rue Pierre Premier (côté pair) à partir de l'angle avec la rue Jean Gougeon, et de la rue St Eloi (côté impair), jusqu'au numéro 23 ". Il s'agit ici d'une aire en forme de " L " ; le premier tronçon (rue Pierre Premier, côté pair) va de l'angle de la rue Jean Gougeon à celui de la rue St Eloi, le deuxième (rue St Eloi, côté impair) va de l'angle de la rue Pierre Premier jusqu'au numéro 23.

Il faut faire attention dans ce cas au traitement de l' " arrière " des bâtiments : s'il y a des terrains vagues, il faut les inclure dans une aire et l'aire doit être alors délimitée par un contour fermé. Sinon, l'aire se limite aux immeubles ouvrant sur le tronçon considéré et les parties arrières, mêmes non bâties, sont supposées ouvrir sur un autre tronçon, donc faire partie d'une autre aire.

Ce n'est qu'après avoir décrit l'aire de façon géographique qu'il est demandé de lister les immeubles, *tels qu'existants au RP*, qui en font partie. Les immeubles repérés sur le terrain mais non recensés, soit qu'ils aient été construits depuis lors, soit qu'ils aient été oubliés au RP, n'ont pas à être listés ni pris en considération dans le calcul des tailles. Tout au plus, un oubli manifeste au RP pourra-t-il être signalé, comme élément additionnel destiné à faciliter le travail ultérieur de l'enquêteur.

5.4.2 Traitement des cas d'incohérences entre différents bordereaux RP.

Il peut se produire que les DIC ou FL conduisent à des nombres de logements différents de ceux figurant sur les M6.

Lors du découpage en zones ou en secteurs et du tirage de l'un d'entre eux, ce sont les chiffres du M6 qui sont utilisés. C'est d'ailleurs aussi le cas pour le tirage de l'unité échantillonnée elle-même, dont la probabilité d'inclusion est calculée à partir de la population légale, élaborée à partir des M6.

Ces chiffres ne doivent en aucun cas être modifiés car ils conditionnent le calcul des poids de sondage de chaque aire, par multiplication des pondérations affectées aux différentes phases du tirage.

Néanmoins, si la constitution des aires nécessite le recours aux DIC ou aux FL, on peut être amené à lister un nombre de logements différent de celui figurant dans le M6. Dans ce cas, on ne cherchera pas à corriger l'incohérence, ni à omettre dans la liste un logement de façon à assurer une cohérence artificielle, ni à refaire le découpage en amont en modifiant les chiffres du M6 : on prendra simplement la liste des logements RP telle que résultant des DIC ou FL pour décrire le contenu de l'aire. On ne tiendra pas compte d'éventuels franchissements de seuil, de toute façon minimes, auxquels ces divergences pourraient conduire (exemple : un secteur réputé de 230 logements au vu du M6 apparaît en fait comme un secteur de 250 au vu des FL : on laisse néanmoins le secteur en l'état, sans modifier ni son contenu ni sa probabilité de sélection qui a été calculée à partir du chiffre de 230).

ENCADRE

QUELQUES CONSIGNES PRATIQUES DONNEES AUX DIRECTIONS REGIONALES LORS DU DECOUPAGE.

a) sur la cartographie.

- utiliser les tableaux d'assemblage des planches cadastrales. Il faut commander les planches cadastrales et les tableaux d'assemblage quand la DR ne les possède pas.
- il faut récupérer systématiquement les plans associés aux M6.
- dans les petites communes rurales, la cartographie est plus faible, les plans sont mauvais. Les M6 listent les logements de manière séquentielle.
- plans : CICN ne donne pas le bâti. Il faut recourir aux plans du cadastre avec n°s de voies.

b) sur l'utilité des déplacements terrain.

- il importe, d'une manière générale, de viser à concentrer les déplacements.
- en zone urbanisée, les M6 sont en général suffisamment précis (comportant nom et n° des voies, ou, au minimum, sens de la numérotation et pair/impair) pour rendre inutile le déplacement terrain.
- en zone très urbanisée (Paris..), il se peut qu'un immeuble entier ou une fraction d'immeuble (cas des grandes tours) soit un secteur. Dans ce cas, le M6 ne suffit plus, les aires seront des étages, voire des fractions d'étages.
- éviter le déplacement terrain pour la délimitation des secteurs ; ceci implique d'être le plus clair et complet possible dans la description des secteurs avant le tirage.
- en urbain, une aire peut être délimitée par des tronçons de voirie. Pour respecter la connexité, si les plans ne donnent pas le sens de la numérotation, le déplacement terrain peut être nécessaire. Mais dans les

villes plus petites, cela peut aider au repérage : mairie, église, hydrographie.

- en urbain, le terrain peut seulement être utile pour vérifier la contiguïté des aires, ou préciser le positionnement des bâtiments, ou confirmer les limites d'un secteur tiré.
- en rural, compte tenu de la faible qualité des plans, le déplacement peut être utile. Notamment, il peut être utile d'aller sur le terrain pour vérifier où sont les maisons quand c'est de l'habitat épars.
- quand les M6 ne permettent pas d'identifier clairement les logements sur les plans (notamment, s'il n'y a pas de n° sur les plans), les déplacements terrain peuvent être nécessaires. Il peut y avoir aussi des n°s sur le plan qu'on ne retrouve pas dans le M6 : immeuble construit depuis le RP.
- quand des n°s d'immeuble figurent sur la planche cadastrale mais pas sur le M6 : entrée de magasin, immeuble désaffecté, n° introuvable, terrain vague, le terrain peut confirmer. Cependant, les écarts M6 (situation 1999) / terrain (2000) sont possibles et l'objet de l'exercice n'est pas de leur donner des explications dans tous les cas.

c) sur le découpage.

- à toute étape, le découpage est exhaustif.
- les logements à prendre en compte dans la composition des secteurs ou des aires et leur dénombrement sont exclusivement ceux du RP.
- les secteurs ont intérêt à s'appuyer sur les districts, c'est plus facile. Et quand on réagrège des districts, commencer par le plus petit. Inversement, il vaudra mieux, parfois, constituer un secteur dans la fourchette haute de la taille, pour éviter d'avoir à couper artificiellement des districts.
- semi-urbain : délimitation des secteurs = routes, limites communales. Ces dernières peuvent ne pas être matérialisées, mais, si elles le sont, il peut être utile de le signaler comme aide aux enquêteurs qui feront l'enquête ultérieurement. On a intérêt à avoir un plan d'ensemble.
- attention aux cas où une limite communale passe à l'intérieur d'un pâté de maison.
- s'il y a eu une légère erreur sur le calcul de la taille d'un secteur et même si celle-ci a été identifiée après tirage du secteur, **il importe absolument de**

conserver la taille erronée dans la feuille de tirage car c'est elle qui sert pour calculer le poids de sondage.

- éviter les aires à cheval sur 2 immeubles : un bâtiment entier est préférable.
Idem : des étages entiers, si possible.
- quand une aire est délimitée entre deux portions de voirie dont l'une est numérotée et pas l'autre (cette dernière ne comportant pas d'accès de logements), cela peut poser problème par la suite pour définir l'affectation d'un nouveau logement à l'aire ou aux aires voisines. Dans ce cas, c'est le plan qui fera foi ; il doit donc être aussi précis que possible.
- les zones vides à l'intérieur d'un secteur doivent être attribuées à une aire ; en rural, un champ doit être affecté à une aire donnée.
- en semi-urbain : il est plus facile de délimiter des aires par des rues entières, " du tant au tant " ou " du n° tant à l'angle de la voie X "... Il faut pouvoir identifier les coins de rue sur le M6.
- si le M6 ne comporte pas les n°s de voie mais que la voie soit entièrement dans l'aire, ça ne pose pas de problème, sinon il faut éventuellement recourir aux FL.
- attention aux descriptions du type : " 20 premiers logements en partant de la gauche "... Préciser plutôt l'orientation ou les points de repère : " quand on est dos à l'église, c'est ... ".
- quand on a un district à découper, c'est plus simple : grâce aux M6 et aux DIC, on arrive facilement à mettre les logements sur des voies.
- en urbain, une aire peut être définie par un tronçon de voie, mais en rural, il faut une ligne fermée.
- dans le dossier de découpage, le nombre total de logements suffit, pas besoin du détail par type (principal, vacant...).

5.5 Constitution du dossier de découpage.

Les descriptions du secteur et des aires découpées sont rassemblées dans un *dossier de découpage* qui adopte une forme normalisée (cf. encadré).

Cet état descriptif comprend :

- un exemplaire du plan de l'unité échantillonnée.
- Sur la 1^{ère} page, la nature et le nom de l'unité à découper. Exemple : dossier 1835, IRIS 0703 de la commune de Courbevoie (92).
- une description précise des limites du secteur sélectionné. Dans la description des secteurs, préciser “ *partie* de district ” si le secteur n'est pas composé de la totalité d'un district donné.
- le plan cadastral ; les limites des secteurs et des aires à l'intérieur du secteur sélectionné doivent être surlignées et les voies grisées. La numérotation littérale des aires doit être indiquée. Les plans doivent être orientés (indication du Nord, même approximative).
- une description précise des limites des aires en précisant leur composition dans les parties denses où les aires ne sont pas des portions de surface.
- une liste et un dénombrement de tous les logements au RP composant chacune des aires, avec leur identifiant RP (qui permettra ultérieurement l'appariement avec les données de l'exploitation principale). Quand un immeuble entier fait partie d'une aire, il n'y a pas lieu d'en distinguer les logements.

Cet état est ensuite envoyé à la Direction Générale, ainsi que la feuille de tirage du secteur sélectionné et, éventuellement, de la grande zone.

ENCADRE
PRESENTATION DES RESULTATS
(DOSSIER DE DECOUPAGE).

Les résultats du découpage doivent être concrétisés par la constitution de dossiers dont le format sera identique. *Cette uniformisation est indispensable pour une exploitation correcte du travail.*

1. Descriptif.

On créera un dossier par unité échantillonnée. Celui-ci contiendra donc in fine les descriptions de toutes les aires (entre 6 et 10) qui seront découpées à partir de cette unité. Ce dossier sera repéré par *l'identifiant Emploi de l'unité échantillonnée* (qui accompagnera les aires découpées tout au long de leur phase d'utilisation, cf. annexe 3) ; on complétera cette information par le nom de la commune, le code du département d'appartenance de cette unité, ainsi que son identifiant RP (numéro d'IRIS, ou de district) quand celle-ci n'est pas une commune. Dans le cas où l'unité échantillonnée est composée de plusieurs communes ou IRIS, on les fera toutes figurer.

Ce dossier comportera 1, 2 ou 3 parties :

- une partie découpage en grandes zones (pour les unités de plus de 800 logements)
- une partie découpage en secteurs (pour les unités de plus de 240 logements)
- une partie découpage en aires (pour toutes les unités)

En tête de chacune de ces parties, figureront les données d'identification définies plus haut.

2. Parties découpage en grandes zones et en secteurs.

Les parties découpage en grandes zones et découpage en secteurs suivent le même modèle.

Elles comportent une feuille " descriptif " décrivant ces zones ou secteurs :

- tout d'abord le ou les districts RP sur lesquels ils sont situés (s'il s'agit d'une fraction de district, on ajoutera entre parenthèse la mention " partie "), puis une description des limites de la zone (limites naturelles, voies, et, éventuellement dans l'urbain, numéros de rue ou étages).

- au bas de cette feuille figure un tableau de tirage. On indiquera pour chaque zone ou secteur le nombre total de logements. On calculera, dans une troisième colonne, le pour-millième cumulé arrondi de ce nombre total de logements. On choisira la

zone ou le secteur disposant du plus petit pour-millième supérieur ou égal au nombre aléatoire fourni pour chaque unité échantillonnée (ce numéro sera le même pour le choix de la zone et du secteur). Cette zone ou ce secteur seront grisés. Un fichier Quattro sera mis à la disposition des DR pour faciliter cette procédure.

Après la feuille “ descriptif ” sera annexé un plan A3 montrant les différentes zones ou secteurs. Ce plan pourra être omis dans le seul cas où la zone ou le secteur n’est composé que d’un immeuble (il n’est pas question en effet d’établir ici de plans détaillés d’immeubles).

3. Partie découpage en aires.

La partie découpage en aires est constituée d’un descriptif général du secteur choisi, qui comporte, outre les données d’identification communes à tout le dossier, la liste des districts couverts par le secteur (ceci permettra de remonter facilement aux bordereaux M6), une description précise du contenu et des limites du secteur, le nombre de logements et le nombre d’aires découpées.

Viennent ensuite une feuille par aire (à chacune est attribuée une lettre de A à J) indiquant le contenu et les limites de cette aire, le nombre de logements au moment du RP et la liste de ces logements. Ces informations sont très importantes car ce sont celles-ci (ainsi que les informations complémentaires qu’on pourra récupérer des données du RP à partir des identifiants saisis) qui seront mises à la disposition des enquêteurs quand ils seront sur le terrain.

Ces données devront être saisies sous Quattro, dans un fichier dont le nom sera l’identifiant de l’unité échantillonnée. La saisie se fera, à partir d’un fichier vierge fourni, dans des zones bien délimitées afin que ces données puissent être récupérées sans trop de manipulation. Ce fichier est constitué d’un onglet pour le descriptif du secteur, suivi d’un onglet par aire découpée. Pour lister les logements, on indiquera les identifiants de districts et d’immeubles (il n’est pas utile de répéter les numéros de communes et de district d’une ligne sur l’autre quand ceux-ci sont identiques). On ne précisera les numéros de logements (à raison de un par ligne) que dans le cas où seule une fraction de l’immeuble est prise dans l’aire.

Dans cette partie « découpage en aires » figurera également un plan détaillé du secteur, au format A3 (sauf là aussi si cette aire n’est composée que d’un immeuble).

4. Envoi et stockage.

Ce dossier devra être conservé par les DEM. Toutes les 2 semaines, et par paquet, il sera envoyé une copie de ces dossiers à la DG, ainsi que les fichiers Quattro de découpage des aires par courrier électronique.

6. Achèvement de la constitution du fichier d'aires.

6.1 Préparation du fichier des logements de l'échantillon.

A partir de la réception et de la validation des fichiers de découpage à la DG, on dispose d'un fichier QUATTRO contenant une page de descriptif "littéraire" de l'aire et la liste des identifiants des logements RP qu'elle contient.

Une première étape consiste à fabriquer un fichier hiérarchisé comprenant :

- le niveau " dossier " (repéré par son n° de dossier)
- le niveau " aire " (repéré par une lettre)
- le niveau " logement " (repéré par l'identifiant RP figurant dans le descriptif du contenu de l'aire).

6.2 Validation de la cohérence et appariement avec les données RP.

L'appariement avec le RP permettra de vérifier que les logements supposés appartenir aux aires découpées existent bien et ont été repérés correctement : pas d'erreur sur les n°s de districts ou sur les identifiants logements tels qu'ils figurent sur le dossier de découpage. En cas d'incohérence, le dossier devra être retraité en DR.

De plus, cet appariement permettra de récupérer des variables RP disponibles à partir de la FL, essentiellement en vue du redressement de la non-réponse.

6.3 Tirage des aires retenues dans l'échantillon.

Si le secteur comporte exactement 6 aires, elles feront toutes partie de l'échantillon. Sinon, 6 d'entre elles seront tirées au sort par la Direction Générale. Dans les deux cas, l'ordre d'introduction dans l'échantillon sera également déterminé de manière aléatoire⁵³ : il constituera le n° de série (allant de 1 à 6)⁵⁴, qui remplacera la désignation provisoire des aires sous forme littéraire.

Ce n° détermine le rang d'introduction au cours des périodes successives (par bloc de 6 trimestres). Les aires non retenues dans l'échantillon principal sont également numérotées de manière définitive en séquence (7, 8 ...) pour servir éventuellement de réserve pour des tests.

⁵³ Pour la numérotation définitive des aires et leur tirage, la Direction Générale attribuera aléatoirement le n°1 à l'une des lettres, puis il y aura numérotation en séquence dans l'ordre des lettres avec retour en A lorsqu'on a épuisé la liste et sélection des aires numérotées 1 à 6.

⁵⁴ Celui-ci sera enregistré dans l'identifiant de l'aire sous le code SA.

6.4 Affectation de différents numéros.

On affecte à chaque unité échantillonnée (= dossier) :

- le *n° définitif de sous-échantillon* (le *n°* provisoire étant constitué par le rang). Les aires de rang 7 (devant constituer une réserve) seront redistribués parmi les *n°*s 1 à 6 pour pouvoir être affectés à un sous-échantillon donné en cas d'activation. Pour les unités des autres rangs, l'idée initiale était de les réaffecter entre les différents sous-échantillons, en satisfaisant si possible des conditions d'équilibrage sur certaines variables RP, issues de l'exploitation principale, et en tenant compte de la stratification.

Néanmoins, seule une partie des unités ayant été traitées par les Directions régionales au démarrage de l'enquête en vraie grandeur, cette solution optimale n'a pu être mise en oeuvre. On a donc réaffecté dans un premier temps les dossiers des rangs 1 et 2 entre les *n°*s de sous-échantillon 1 et 2, en gommant les petits écarts de cardinalité dus à des arrondis. L'opération est prévue pour être refaite ultérieurement avec les unités des rangs suivants.

- le *n° de semaine* : on doit partager l'ensemble des dossiers en 13 parties égales correspondant aux 13 semaines de référence du trimestre. Compte tenu, si possible avec des conditions d'équilibrage. Ce numéro doit bien entendu être indépendant du précédent. Le problème pratique évoqué ci-dessus (date de disponibilité des différentes unités) empêche de réaliser un tirage équilibré nécessitant des données sur l'ensemble des aires prises individuellement (même si le *n°* de semaine est attribué à l'ensemble des aires issues d'un même dossier) : un simple tirage aléatoire sera donc réalisé.

6.5 Affectation du poids d'échantillonnage.

a) Le poids initial d'échantillonnage (servant pour une estimation fondée sur l'échantillon d'un trimestre) est obtenu comme inverse de la probabilité de sondage, calculée comme produit :

- de la probabilité de tirage de l'unité de 1ère phase
- de la probabilité de tirage du secteur (voire de la grande zone, le cas échéant)
- de la probabilité de tirage de l'aire au sein des aires découpées.

Sur ce dernier point, deux méthodes étaient envisageables : soit tirer l'aire de manière équiprobable, soit proportionnellement à sa taille (sachant qu'il peut y avoir des fluctuations autour de la norme de 20, des tolérances exceptionnelles ayant été accordées jusqu'à 27, notamment).

L'inconvénient de la 2ème méthode est que les aires d'un même secteur, interrogées successivement, n'ont pas le même poids, alors que cette condition est assurée par la 1ère. *C'est la première méthode qui a donc été adoptée.*

On note que, en utilisant cette 1ère méthode, le poids d'échantillonnage résultant ne fait pas apparaître explicitement la taille de l'aire mais le ratio [taille du secteur / nombre d'aires découpées], qui doit peu fluctuer autour de 20 d'un secteur à l'autre (s'il n'y a pas, du moins, d'aires systématiquement plus grosses ou plus petites que la moyenne au sein d'un même secteur..).

On obtient finalement la formule suivante pour la probabilité d'inclusion d'une aire constituée à l'intérieur du secteur j de l'unité i appartenant à la strate h :

$$\Pi_{jih} = k_h \frac{T_i}{\sum_{j \in P_h} T_j} * \frac{S_{ij}}{T_i} * \frac{1}{n_{ij}},$$

où :

k_h est le nombre d'unités de 1ère phase tirées dans la strate h

T_i est la taille de l'unité échantillonnée i

P_h est la population des unités de 1ère phase dans la strate h

S_{ij} est la taille du secteur j tiré dans l'unité i

n_{ij} est le nombre d'aires découpées à l'intérieur du secteur j de l'unité i .

Cette formule permet de calculer, strate par strate, le poids de sondage affecté aux différentes aires et, par voie de conséquence, à tous les logements de chaque aire. On constate, pour une valeur moyenne de 560 environ, des écarts par région (allant de 282 à 819) et par catégorie de commune (668 pour les communes rurales), qui s'expliquent par les contraintes de précision régionale qui ont conduit à augmenter la taille de l'échantillon dans certaines petites régions (cf. supra).

b) une fois l'enquête réalisée au cours d'un trimestre, les poids d'échantillonnage seront remplacés par des poids finaux, différenciés cette fois-ci par logement. Ces poids seront calculés à partir d'une *procédure de correction de la non-réponse* et du calage sur un certain nombre de totaux ou de structures relatifs aux logements (données du RP 1999) ainsi qu'une *structure d'âge courante* (pyramide des âges du trimestre en cours ou décalée...). On obtiendra ainsi un deuxième jeu de poids qui servira pour l'estimation de données trimestrielles (en niveau moyen).

La mise en oeuvre de ces procédures de calage nécessitera de vérifier quel type de données on peut récupérer, avec quelle périodicité, quels délais et quelle qualité (régionalisation possible, apport ultérieur des données issues du RRP...).

Enfin, des procédures spécifiques devront être mises en oeuvre pour définir les poids relatifs à des estimations de natures différentes : variations trimestrielles, niveaux annuels moyens etc...

6.6 Disjonction avec les autres échantillons.

Il n'a pas été édicté de règle générale visant à faire en sorte que des logements faisant partie des aires des séries antérieures (notamment, la série 1990) soient exclus du futur échantillon Emploi. Le découpage en aires s'est donc fait sans tenir compte du précédent échantillon Emploi.

Néanmoins, dans certains cas, des configurations problématiques peuvent être examinées et faire l'objet d'exceptions (cas de logements sollicités actuellement par le dispositif léger et devant rentrer prochainement dans le nouvel échantillon, par exemple...).

En ce qui concerne l'échantillon-maître, la disjonction avec l'échantillon Emploi doit se faire au fur et à mesure de la constitution des aires, en excluant de l'échantillon-maître des logements figurant déjà dans l'échantillon Emploi : les DR n'ont donc pas non plus à tenir compte, pour le découpage des aires Emploi, de la localisation des unités primaires rurales de l'échantillon-maître qui leur ont été livrées.

6.7 Constitution du dossier d'aires.

Le " dossier d'aires " remis aux enquêteurs comprendra :

- l'identifiant (n° de série de 1 à 6 et n° du dossier), le plan et le descriptif " littéraire " de l'aire (repris du dossier de découpage)
- une feuille par logement RP avec :
 - * l'identifiant RP
 - * les variables récupérées du RP
 - * l'adresse précise et le nom de l'occupant RP du logement, issus de la FL.

Ce document permettra à l'enquêteur de faire son ratissage, au vu des adresses qu'on lui donne, et de vérifier les noms (ce qui permet l'envoi ultérieur d'une lettre-avis). Pour chaque logement neuf qu'il rencontrera, il devra remplir une feuille vierge analogue.

ANNEXE 2 : TIRAGE DES SECTEURS ET GRANDES ZONES.

La Direction Générale a remis à la Direction régionale, au moment de la livraison des unités échantillonnées, un *nombre aléatoire (associé à chaque unité)*. On calculera le cumul de la part relative, en nombre de logements, de chaque secteur (préalablement numéroté), par rapport à l'ensemble de l'unité échantillonnée (ou de la grande zone) et le nombre aléatoire permettra de désigner le secteur échantillonné. Ceci pourra être mis en oeuvre grâce à une application QUATTRO dont le modèle est ci-dessous.

Secteur	Nombre de logements	% cumulé du total
2	130	501
3	145	743
4	154	1000
	599	(aléa=245)

Instructions	
1. Appuyer sur le bouton Effacer	<input type="button" value="Effacer"/>
2. Remplir la liste des secteurs et leur nombre de logements	
3. Renseigner le nombre aléatoire	<input type="text" value="245"/>
4. Appuyer sur le bouton Tirage	<input type="button" value="Tirage"/>

La mise en oeuvre de ce tirage sera récapitulée sur une *feuille de tirage* qui sera envoyée à la DG, comme élément du dossier de découpage.

Nota :

1. S'il y a lieu, le même nombre aléatoire et la même procédure seront utilisés pour le tirage des grandes zones.

2. Pour des raisons de commodité, un nombre aléatoire sera fourni même pour des unités pour lesquelles il n'y aurait pas lieu, a priori, de réaliser de tirage de secteur (unités petites). Dans ce cas, ce nombre aléatoire sera tout simplement inutilisé.

ANNEXE 3 : ELEMENTS D'IDENTIFICATION DES UNITES ECHANTILLONNEES.

Les éléments d'identification de l'unité ou de ses subdivisions seront les suivants :

1. Numéro d'identification Emploi (*n° d'ordre séquentiel national*, de 1 à ..., sur 4 positions) pour repérer de manière unique toutes les unités échantillonnées. Ce numéro accompagnera les aires issues d'une même unité échantillonnée tout au long de la vie de l'échantillon. Pour des raisons techniques, ce numéro ne sera pas nécessairement séquentiel au sein de chaque région.

Nota : ce numéro figure en tête de l'unité à découper, sous l'appellation " *dossier* ".

2. Type d'unité (commune, regroupement de communes, IRIS, regroupement d'IRIS, district) : code en clair.

3. Identifiants géographiques :

* région (2 positions)

* code DEPCOM (5 positions) : code de la commune à laquelle appartient l'unité ou de la commune correspondant à une subdivision de l'unité, si celle-ci est un regroupement de communes.

* n° d'IRIS (4 positions) : n° de l'unité " IRIS " sélectionnée ou n° de l'IRIS correspondant à une subdivision de l'unité, si celle-ci est un regroupement d'IRIS.

* n° de district (4 positions) : n° de l'unité " district " sélectionnée ou n° du district correspondant à une subdivision de l'unité (commune ou IRIS).

* nom de la commune de rattachement de l'unité (cf. remarques sur le code DEPCOM).

4. Critères de stratification :

* code TU99 : tranche de taille d'unité urbaine (1 position).

* code ZE : zone d'emploi au sens 1990 de la commune à laquelle appartiennent l'unité ou ses subdivisions (4 positions).

5. Taille : *nombre total de logements* au RP 1999 de l'unité ou de ses subdivisions (4 positions).

6. Nombre aléatoire : nombre entier de 1 à 1000 (qui sert au tirage manuel des secteurs découpés à l'intérieur de chaque unité ou au tirage des grandes zones).

7. Rang de traitement : nombre de 1 à 7, indiquant, dans chaque strate, l'ordre dans lequel les DR devront avoir achevé le traitement des unités à découper. Les unités de rang 7 donneront naissance aux aires de réserve.

8. Caractéristiques dans l'échantillon Emploi.

* Numéro de sous-échantillon. Il prend les valeurs de 1 à 6, distinguées selon le trimestre d'introduction dans le nouvel échantillon.

* Numéro de semaine. Il prend les valeurs de 1 à 13 et donne le n° de la semaine de référence sur laquelle doit porter l'interrogation au sein de chaque trimestre.

9. Caractéristique de l'aire dans l'échantillon : numéro de série, de 1 à 6 (valeurs supérieures pour les aires de réserve). Il indique l'ordre d'utilisation des aires au sein d'un même secteur pendant la période de vie de l'échantillon.