

N°71-526-X au catalogue

# Méthodologie de l'Enquête sur la population active du Canada



## Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Service à la clientèle, Division de la statistique du travail, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : 1-866-873-8788, courriel : travail@statcan.ca).

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.ca](http://www.statcan.ca). Vous pouvez également communiquer avec nous par courriel à [infostats@statcan.ca](mailto:infostats@statcan.ca) ou par téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

### Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369

Appels locaux ou internationaux :

Service de renseignements	1-613-951-8116
Télécopieur	1-613-951-0581

### Programme des services de dépôt

Service de renseignements	1-800-635-7943
Télécopieur	1-800-565-7757

## Renseignements pour accéder au produit

Le produit n° 71-526-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à [www.statcan.ca](http://www.statcan.ca) et de choisir la rubrique « Publications ».

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui sont observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.ca](http://www.statcan.ca) sous « À propos de nous » > « Offrir des services aux Canadiens ».

# Méthodologie de l'Enquête sur la population active du Canada

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2008

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juin 2008

N° 71-526-X au catalogue

ISBN 978-0-662-08294-1

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 71-526-X).

---

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

# Signes conventionnels

---

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0<sup>s</sup> valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- <sup>p</sup> provisoire
- <sup>r</sup> révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- <sup>E</sup> à utiliser avec prudence
- F trop peu fiable pour être publié

## Remerciements

Nous remercions sincèrement les nombreuses personnes qui ont contribué de différentes façons à ce document.

Un remerciement tout particulier aux auteurs principaux des divers chapitres : Jean-François Beaumont, René Boyer, Guy Laflamme, Danielle Lebrasseur, John Lindeyer et Claude Turmelle. Ces personnes ont travaillé à différents aspects du remaniement de 2004 de la méthodologie de l'Enquête sur la population active, faisant en sorte qu'elles étaient les personnes les mieux placées pour écrire sur le sujet.

De surcroît, bien d'autres personnes à la Division des enquêtes auprès des ménages, à la Division de la statistique du travail et ailleurs à Statistique Canada ont contribué à la planification, à l'examen, à la vérification, à la révision, à la traduction et à la production de ce manuel. Nous les remercions tous très sincèrement.

# Table des matières

Chapitre 1: Introduction et aperçu général de l'enquête .....	6
Chapitre 2: Plan de sondage.....	9
Chapitre 3: Création et tenue à jour de la base de sondage .....	26
Chapitre 4: Collecte.....	45
Chapitre 5: Traitement et imputation .....	47
Chapitre 6: Pondération et estimation .....	52
Chapitre 7: Estimation de la variance .....	61
Chapitre 8: Qualité des données.....	64
Chapitre 9: Base de sondage de l'EPA pour d'autres enquêtes.....	74
Bibliographie .....	76
Annexe A.1 Glossaire .....	79
Annexe A.2 Abréviations .....	84
Annexe B Caractéristiques de la base de sondage et du plan de sondage .....	85
Annexe C Plan de sondage de l'Enquête sur la population active - depuis 2005 .....	95
Annexe D Exemples de cartes d'UPE (formule F01).....	96
Annexe E Cartes provinciales.....	105

## Chapitre 1 Introduction et aperçu général de l'enquête

### Introduction

La présente publication se veut un ouvrage de référence sur la méthodologie de l'Enquête sur la population active (EPA). L'EPA couvre l'ensemble du territoire canadien. Toutefois, cette publication se concentre sur la méthodologie utilisée pour les dix provinces. Elle présente les changements introduits lors du récent remaniement du plan de sondage et couvre toutes les étapes de l'enquête.

Un document distinct intitulé *Guide de l'Enquête sur la population active* (disponible sur Internet à l'adresse [www.statcan.ca](http://www.statcan.ca)) sert de complément au présent rapport en mettant l'accent sur les concepts, les définitions et sur les données produites dans le cadre de l'EPA.

### 1.1 Historique

L'EPA a été créée après la Seconde Guerre mondiale pour répondre à un besoin urgent de données fiables et actuelles sur le marché du travail reflétant le passage d'une économie de guerre à une économie de paix. L'enquête a été conçue dans le but de produire des estimations sur l'emploi et le chômage tant à l'échelle régionale que nationale.

Effectuée sur une base trimestrielle à son origine en 1945, l'EPA est devenue une enquête mensuelle en 1952. En 1960, le Comité interministériel de la statistique du chômage recommandait que l'EPA devienne l'instrument officiel de mesure du chômage au Canada. Après l'adoption de cette recommandation, la demande de données a augmenté, les utilisateurs voulant disposer d'une plus vaste gamme de statistiques sur le marché du travail, notamment des données régionales plus détaillées. L'éventail des estimations produites de cette enquête s'est élargi considérablement au fil des ans, et présente aujourd'hui un portrait détaillé du marché du travail canadien.

### 1.2 Concepts et produits de l'EPA

L'EPA est la source officielle d'estimations mensuelles touchant l'emploi total (du travail rémunéré et du travail autonome, du travail à temps plein et du travail à temps partiel) et le chômage. Parmi les principaux indicateurs mensuels publiés, signalons le taux de chômage, le taux d'emploi et le taux d'activité. Il s'agit d'une des principales sources d'information sur les caractéristiques individuelles de la population en âge de travailler (notamment l'âge, l'état matrimonial, le niveau d'instruction et la situation familiale).

Les estimations de l'emploi sont ventilées à divers niveaux, dont le secteur d'activité, la profession, la durée de l'emploi, le nombre d'heures habituellement travaillées et le nombre d'heures réellement travaillées. Certaines des questions posées permettent d'étudier une grande variété de sujets d'actualité comme le travail à temps partiel non choisi, le cumul d'emplois et l'absentéisme au travail.

Les estimations sur le chômage sont produites par secteur d'activité et par profession, ainsi que selon la durée du chômage, le genre de travail recherché et l'activité avant la recherche de travail. Il existe également des données sur l'activité récente sur le marché du travail des personnes actuellement inactives. Le Guide de l'Enquête sur la population active donne une description complète du contenu du questionnaire de l'EPA.

En plus des estimations nationales et provinciales, l'EPA produit également des données pour des régions infra-provinciales, comme les régions économiques de l'assurance-emploi (REAE) et les régions métropolitaines de recensement (RMR). Ces dernières années, on a également totalisé, à l'aide de techniques spéciales d'estimation, des indicateurs normalisés du marché du travail pour de petites régions, comme les divisions de recensement (DR) et les centres d'emploi du Canada. Le gouvernement fédéral et les gouvernements provinciaux se servent des données de l'EPA pour la répartition des ressources financières et autres entre les diverses juridictions politiques et administratives.

Les estimations normalisées de l'EPA paraissent chaque mois dans la publication *Information - population active* (numéro 71-001-X au catalogue et disponible sur Internet). On peut également accéder à une variété de données propres au marché du travail par l'intermédiaire de CANSIM, la base de données et le système d'extraction électronique de Statistique Canada. Cette base, qui contient plus de mille séries chronologiques, est mise à jour mensuellement avec les données de l'EPA.

L'EPA est également à la source du CD-ROM *Revue chronologique de la population active* (numéro 71F0004X au catalogue), qui contient des données détaillées sous forme de séries transversales et chronologiques de 1976 à l'année en cours. L'enquête peut produire beaucoup plus d'information que ce qui est publié périodiquement. Des totalisations spéciales sont produites en recouvrement des coûts. Pour plus de

détails concernant les produits et services rattachés à l'enquête, veuillez consulter la section 8 du Guide de l'Enquête sur la population active.

### **1.3 Aperçu général de l'enquête et structure du document**

Dans les provinces, l'EPA couvre 98 % de la population. Les résidents des réserves indiennes, des terres publiques et de certaines régions éloignées sont exclus du champ de l'enquête. Il en est de même pour les pensionnaires d'un établissement institutionnel et les membres à temps plein des Forces canadiennes. Ces groupes sont exclus de la population cible, car la collecte de leurs données soulèverait des problèmes opérationnels. À titre d'exemple, plusieurs membres des Forces canadiennes vivent dans des lieux inaccessibles aux interviewers de l'EPA (notamment à bord de navires de guerre ou dans des camps militaires et des casernes). Il serait donc difficile d'interviewer ces personnes. Les pensionnaires sont également exclus, car la plupart d'entre eux sont incapables de participer au marché du travail.

L'enquête utilise un plan de sondage à deux degrés<sup>1</sup>. Au premier degré, on sélectionne un échantillon de régions géographiques. Ces régions sont appelées : unités primaires d'échantillonnage (UPE). Dans chacune des UPE sélectionnées, on choisit un échantillon de logements au second degré. Les logements sélectionnés demeurent dans l'échantillon pour une période de six mois. Chaque mois, les logements ayant déjà collaboré à l'enquête pendant six mois, soit le sixième des logements échantillonnés, sont remplacés par d'autres logements. On obtient ainsi un chevauchement de cinq répondants sur six entre deux mois consécutifs. Cette approche permet d'améliorer la qualité des estimations du changement entre deux mois. Elle permet également de contrôler le fardeau de réponse imposé à chaque logement. Les chapitres 2 et 3 fournissent plus de détails au sujet du plan de sondage.

La collecte des données de l'EPA a lieu pendant la semaine suivant la semaine de référence. Normalement, la semaine de référence comprend le quinzième jour du mois. En 2004, environ 51 000 ménages répondaient chaque mois au questionnaire de l'EPA. On notera au passage que l'interview de l'EPA ne requiert que huit minutes en moyenne. La collecte des données est effectuée grâce à un système d'interview assistée par ordinateur. Afin de réduire les déplacements, et ainsi les coûts de collecte, on tente d'effectuer un nombre

important d'interviews au téléphone. Les détails concernant la stratégie de collecte sont présentés au chapitre 4.

Dans les jours suivant la collecte, on effectue le traitement, l'imputation, la pondération et on dérive des indicateurs de la qualité. Ces sujets sont traités aux chapitres 5, 6, et 8. Malgré le volume d'information à traiter chaque mois, Statistique Canada publie les estimations de l'EPA 13 jours seulement après la fin de la semaine d'interview.

Il est important de souligner que la base de sondage de l'EPA est également utilisée par la plupart des enquêtes sociales menées par Statistique Canada. Ce point est couvert au chapitre 9. Les annexes, dont chacune traite d'un sujet particulier, et les autres documents d'enquête sont inclus à la fin de la publication.

### **1.4 Les changements introduits en novembre 2004**

Après chaque recensement décennal de la population, on remanie le plan de sondage de l'EPA pour tenir compte de l'évolution des caractéristiques de la population et des besoins des utilisateurs de données. Le remaniement permet également de mettre à jour l'information géographique requise pour mener l'enquête. Contrairement au remaniement de 1994, le questionnaire et l'application de collecte n'ont pas été modifiés lors de ce remaniement.

On a procédé au remaniement de 2004 dans un cadre budgétaire serré. Par conséquent, certaines études effectuées lors des précédents remaniements n'ont pas pu être mises à jour. Nous avons dû supposer que les conclusions obtenues lors des précédents remaniements sont toujours valides. Cette contrainte budgétaire signifiait également que le remaniement devait être mis en place sans devoir changer les systèmes informatiques existants. Finalement, une partie du remaniement a dû être financée par l'entremise d'une réduction de la taille de l'échantillon de l'enquête sur une période de trois ans. La stratégie de réduction de la taille de l'échantillon est détaillée à la section 2.5.3.

Le nouveau plan de sondage a été introduit progressivement à partir du mois de novembre 2004. Afin de réduire les coûts d'exploitation de l'enquête, deux changements majeurs à la méthodologie ont été introduits à cette occasion.

Avant novembre 2004, la première des six interviews était faite en personne. Aux fins de réduction des coûts de collecte, la première interview est dorénavant

---

1. Un plan à trois degrés est utilisé à de rares occasions pour traiter les centres urbains isolés de moyenne taille.

effectuée par téléphone pour certains logements des centres urbains. La stratégie implantée est décrite dans le chapitre 4.

Le second changement vise à réduire le coût de transition de l'ancien plan de sondage vers le nouveau. Une part importante du budget d'un remaniement est consacrée à l'obtention d'une liste d'adresses pour chacune des UPE sélectionnées selon le nouveau plan. En vue de réduire les coûts de cette opération de listage<sup>2</sup>, et améliorer la couverture de la base de sondage, nous avons utilisé le Registre des adresses (RA). Le RA est une base de données contenant l'adresse des logements des centres urbains. Les détails concernant ce changement sont donnés au chapitre 3.

En plus de ces deux changements majeurs, d'autres améliorations ont été apportées au plan de sondage. Par le passé, la base de données géographiques de Statistique Canada ne couvrait pas l'ensemble du territoire des dix provinces. Cette base est requise pour établir les frontières des UPE. Par conséquent, avant le remaniement de 2004, une partie des UPE étaient définies pour répondre aux besoins spécifiques de l'enquête. Pour le reste du territoire, nous devons

utiliser les secteurs de dénombrement de 1991<sup>3</sup> comme UPE. La qualité de la base de données géographiques s'est grandement améliorée depuis le remaniement de 1994. Pour la première fois en 2004, nous avons pu définir les frontières des UPE pour l'ensemble du territoire des dix provinces. La stratégie de formation des UPE est présentée à la section 2.6.4. On a également conçu une application permettant de tracer les cartes routières requises pour effectuer le travail de collecte pour chaque UPE sélectionnée. La base de données géographique est la pierre angulaire de cette application.

Afin de mieux contrôler la répartition de l'échantillon, et par conséquent les coûts de collecte, nous avons mis en place une stratégie particulière pour les régions ayant un coût de collecte élevé (voir section 2.5.2 et Chen, Lindeyer et Laflamme 2004). Nous avons également introduit des méthodes permettant de cibler la population des immigrants dans les grands centres et la population Autochtones dans les quatre provinces de l'ouest (voir section 2.6.4.2). En vue de réduire les coûts de maintenance de la base de sondage, le nouveau plan de sondage ne contient plus une base de sondage des appartements. Finalement, la sélection de l'échantillon dans les régions rurales à faible densité de population est dorénavant effectuée avec un plan à deux degrés plutôt que le plan à trois degrés utilisé par le passé.

---

2. Une opération de listage consiste à dresser la liste des adresses résidentielles dans l'UPE. Une carte est fournie à l'intervieweur pour préciser les limites de l'UPE.

---

3. Petite région composée d'un ou plusieurs pâtés de maisons voisins, utilisée par Statistique Canada pour la livraison des questionnaires aux ménages et aux logements (collecte du recensement). L'ensemble du Canada est divisé en secteurs de dénombrement.

## Chapitre 2 Plan de sondage

### Introduction

Le plan de sondage décrit toutes les étapes à accomplir pour sélectionner un échantillon de personnes. Il vise à améliorer la qualité des estimations produites tout en contrôlant les coûts. Diverses stratégies sont mises en place pour atteindre cet objectif.

Une part importante du budget d'une enquête est consacrée à l'activité de collecte. Par ailleurs, le plan de sondage a peu d'effet sur les autres postes budgétaires d'une enquête (par exemple, le traitement et la diffusion). Dans la pratique, le plan de sondage tente donc de réduire les coûts de la collecte tout en maximisant la qualité des données.

La section 2.1 propose un survol du plan de sondage de l'EPA. La section 2.2 présente quelques notions de base de la théorie des sondages. Ces notions seront utilisées tout au long de ce chapitre. Le lecteur connaissant déjà ces principes peut passer directement aux sections suivantes. La section 2.3 motive la stratégie d'utilisation du RA. Le reste du chapitre est consacré à quatre techniques permettant d'améliorer l'efficacité du plan de sondage.

### 2.1 Description du plan de sondage

Il est impossible de prendre contact avec l'ensemble des personnes âgées de 15 ans et plus chaque mois pour établir leur statut d'emploi pour produire les estimations requises. De plus, aucune source administrative ne permet de produire ces estimations. Par conséquent, nous devons les produire à partir d'un échantillon de personnes.

Nous ne pouvons sélectionner directement un échantillon de personnes sur lesquelles enquêter, car nous ne disposons pas d'une liste complète des personnes habitant les dix provinces. Même si cette liste existait, les personnes ainsi sélectionnées seraient éparpillées sur l'ensemble du territoire. Les coûts de déplacement seraient alors exorbitants lorsque l'intervieweur doit prendre contact, en personne, avec certains répondants.

Dans l'optique de réduire les coûts de déplacement, l'échantillon de personnes est tiré au moyen de deux processus de sélection consécutifs. Ce mode de sélection porte le nom d'échantillonnage à deux degrés<sup>4</sup>. Au premier degré, on sélectionne un échantillon de régions géographiques. Ces régions sont appelées : unités

primaires d'échantillonnage (UPE). Pour chacune des UPE sélectionnées, on établit la liste des logements. On choisit un échantillon de logements au second degré de ces listes. Tous les habitants des logements sélectionnés au second degré font partie de l'échantillon de personnes de l'EPA. Ce mode de sélection permet de réduire la dispersion géographique des personnes échantillonnées. Il évite également d'avoir à créer une liste de toutes les adresses dans les dix provinces.

En plus des estimations mensuelles décrites à la section 1.2, l'EPA produit également des estimations du changement entre deux périodes de référence données. En vue d'améliorer la qualité de ces estimations du changement, il est préférable d'accroître le chevauchement entre les échantillons de ces deux périodes. On ne peut augmenter ce chevauchement qu'en conservant les mêmes logements dans l'échantillon durant plusieurs mois. Par ailleurs, l'accroissement du chevauchement s'accompagne d'une augmentation du fardeau imposé aux répondants, car ils doivent participer à l'enquête à plusieurs reprises. À son tour, cette augmentation induit du fardeau pourrait entraîner une augmentation du taux de non-réponse. En fin de compte, l'accroissement du chevauchement a donc un effet négatif sur le fardeau de réponse. Il est donc nécessaire d'établir un compromis entre la qualité des estimations du changement et le fardeau imposé aux répondants. On notera au passage que les coûts de collecte sont également influencés par l'ampleur du chevauchement. En effet, il est plus coûteux d'obtenir une réponse au cours du premier mois qu'au cours des mois suivants. Un plus grand chevauchement entraîne donc une réduction du coût d'exploitation de l'enquête.

En tenant compte de ces facteurs, il a été décidé que chaque logement doit demeurer dans l'échantillon de l'EPA pour une période de six mois consécutifs. Sujet à cette contrainte, le chevauchement maximum de l'échantillon entre deux mois consécutifs est de cinq sixièmes. Il est donc nécessaire de remplacer le sixième de l'échantillon de logements chaque mois. Pour établir cette stratégie, l'échantillon de logements de l'EPA est scindé en six panels ou groupes de renouvellement. Chacun de ces groupes est représentatif de la population observée. Les logements appartenant au premier groupe de renouvellement sont contactés pour la première fois en janvier. Ces logements demeureront dans l'échantillon jusqu'au mois de juin inclusivement. Au mois de juillet, tous les logements du groupe un sont remplacés par de nouveaux logements. Le second groupe

4. Une plan à trois degrés est utilisé à de rares reprises pour traiter les centres urbains isolés de moyenne taille. Cette approche est présentée à la section 2.6.4.

comprend les logements enquêtés de février à juillet inclusivement et ainsi de suite pour les autres groupes de renouvellement. Les détails concernant la rotation de l'échantillon de logements sont présentés à la section 2.7.2.

La gestion de l'échantillon de l'EPA par les groupes de renouvellement fournit une méthode simple pour sélectionner les échantillons pour d'autres enquêtes de Statistique Canada. Comme chaque groupe de renouvellement est représentatif de la population, on peut constituer l'échantillon pour une autre enquête en regroupant les logements d'un nombre approprié de groupes de renouvellement. Les détails concernant l'utilisation de la base de sondage de l'EPA pour sélectionner les échantillons des d'autres enquêtes auprès des ménages sont présentés au chapitre 9.

On notera au passage que le chevauchement des échantillons mensuels de l'EPA ouvre la porte à des méthodes de traitement et d'estimation plus performantes. Les méthodes présentement utilisées par l'EPA sont décrites aux chapitres 5 et 6.

## 2.2 Quelques notions de base de la théorie des sondages

Cette section présente quelques notions requises pour comprendre la description du plan de sondage fournie dans ce chapitre. Un survol conceptuel de la théorie des sondages est disponible dans Satin et Shastry (1992). Pour plus de détails concernant cette théorie, le lecteur est prié de consulter un des nombreux livres traitant de ce sujet (par exemple Cochran (1977) ou Särndal, Swensson et Wretman (1992)).

Les estimations de l'EPA sont produites à partir d'un échantillon probabiliste, c'est-à-dire un échantillon dans lequel les personnes à enquêter sont choisies aléatoirement. Il est important de remarquer que l'estimation produite à partir d'un échantillon est différente de celle que l'on obtiendrait si l'ensemble de la population était interviewé. Il s'agit de l'erreur d'échantillonnage. De même, l'estimation produite à partir d'un échantillon donné est différente de celle qui serait produite avec un autre échantillon. En plus des erreurs d'échantillonnage, les résultats d'une enquête sont également entachés d'erreurs non échantillonnelles. Le chapitre 8 présente les procédures mises en place pour contrôler ce type d'erreur.

Deux importantes mesures de l'erreur sont le biais et la variance d'échantillonnage. Ces deux concepts peuvent être définis de façon intuitive. Supposons qu'il est possible de sélectionner plusieurs échantillons selon

un plan de sondage donné. Pour chacun de ces échantillons, on produit une estimation de la caractéristique d'intérêt (par exemple, le nombre de chômeurs, le revenu personnel moyen, *etc.*). L'estimation est biaisée si la moyenne des estimations produites à partir de tous les échantillons possibles est différente de la valeur que l'on obtiendrait en enquêtant sur toutes les personnes de la population. De plus, la variance entre ces diverses estimations est la variance due à l'échantillonnage.

Il y a plusieurs sources de biais. À titre d'exemple, l'imperfection de la base de sondage, la méthode utilisée pour produire l'estimation, et la non-réponse des personnes peuvent entraîner des biais dans les estimations. Cette composante de l'erreur est difficilement mesurable dans la pratique. De plus, elle a peu d'effet sur la définition du plan de sondage. Par conséquent, nous n'avons pas tenu compte du biais potentiel lors de la mise en place des quatre premières étapes du plan de sondage qui sont documentées dans ce chapitre.

La variance d'échantillonnage mesure la dispersion entre les estimations produites à partir de tous les échantillons possibles. Plus la variance d'échantillonnage est petite, meilleure est la précision de l'estimation. Une estimation de la variance d'échantillonnage peut toutefois être obtenue à partir d'un seul échantillon probabiliste.

D'autres mesures sont dérivées de la variance d'échantillonnage. Une de ces mesures est l'erreur type. Cette mesure est obtenue en prenant la racine carrée de la variance d'échantillonnage. Elle est principalement utilisée pour obtenir un intervalle de confiance afin d'effectuer un test statistique. Une autre mesure est le coefficient de variation (c.v.). Le c.v. est une mesure relative de la qualité d'une estimation. Par définition, le c.v. est l'erreur type divisée par l'estimation. Encore une fois, plus le c.v. est petit, meilleure est la précision de l'estimation. Une troisième mesure est l'effet de plan. Cette dernière est une mesure relative qui permet de comparer l'efficacité d'un plan de sondage à un autre. Plus de détails sur ces mesures sont donnés dans le chapitre 8.

La variance d'échantillonnage est une mesure de l'efficacité du plan. Lors de la conception du plan de sondage, nous avons tenté de réduire la variance d'échantillonnage. D'un autre point de vue, un plan de sondage plus efficace, c'est-à-dire entraînant une variance d'échantillonnage plus petite, permet de réduire la taille de l'échantillon par rapport à un autre plan

moins efficace tout en maintenant la qualité des estimations.

Plusieurs facteurs influencent la variance d'échantillonnage d'une estimation. Les plus importants sont : le nombre de personnes dans la population, le nombre de personnes dans l'échantillon, le plan de sondage utilisé pour tirer l'échantillon, le taux de réponse et l'homogénéité de la caractéristique d'intérêt dans la population. La répartition de l'échantillon (voir section 2.5) assure que le nombre de logements sélectionnés est suffisant pour produire les diverses estimations de l'enquête avec une qualité adéquate. La stratification (voir section 2.6) permet d'améliorer l'efficacité du plan de sondage en regroupant des logements similaires. Dans le cadre d'une enquête à deux degrés comme l'EPA, l'homogénéité de la caractéristique d'intérêt à l'intérieur de l'UPE et dans le logement influence également la variance d'échantillonnage. Plus les logements d'une même UPE sont homogènes, moins le plan de sondage est efficace. La même logique s'applique à l'homogénéité à l'intérieur des logements.

Pour sa part, la forme des UPE permet de réduire les coûts de collecte. La formation des UPE est présentée à la section 2.4. Finalement, certaines méthodes de sélection de l'échantillon permettent d'améliorer l'efficacité du plan de sondage. La section 2.7 décrit la méthode de sélection utilisée par l'EPA.

En contrôlant chacun de ces paramètres, il est possible de définir un plan de sondage efficace, c'est-à-dire ayant une petite variance d'échantillonnage pour une taille d'échantillon donnée et un coût d'exploitation fixe.

### **2.3 Le Registre des adresses et son effet sur le plan de sondage**

Le Registre des adresses (RA) est une base de données contenant l'adresse des logements des centres urbains (voir chapitre 3.1 pour une description du RA). Lors du remaniement, nous avons considéré la possibilité d'utiliser le RA afin de sélectionner un échantillon de logements directement de cette liste. En milieu urbain, cette approche aurait pu être plus économique et plus efficace que le plan de sondage à deux degrés traditionnellement utilisé par l'EPA. Toutefois, elle se heurte à trois problèmes majeurs.

Dans un premier temps, la qualité du RA n'est pas uniforme géographiquement. Il aurait été difficile de délimiter les régions où l'on aurait pu sélectionner un échantillon de logements directement du RA. De plus, au moment du remaniement du plan de sondage,

processus de mise à jour du RA était toujours incertain<sup>5</sup>. Dans ce contexte, il aurait été difficile de maintenir la qualité de la base de sondage au fil du temps. Finalement, la sélection de logements directement du RA aurait requis d'importantes modifications aux systèmes informatiques existants. Toutefois, le cadre budgétaire du remaniement ne permettait pas d'apporter des changements majeurs aux systèmes informatiques existants. Pour ces raisons, il a été décidé que la sélection de logements directement du RA n'était pas souhaitable pour ce remaniement.

Toutefois, il demeurerait souhaitable d'utiliser le RA afin de réduire les coûts de collecte et améliorer la couverture de l'enquête. Le compromis consiste à utiliser le RA comme un outil pour améliorer l'efficacité de l'opération de listage des adresses dans les UPE sélectionnées. Pour chaque UPE sélectionnée, on établit la qualité des informations disponibles sur le RA. En fonction de cette mesure de qualité, on choisit une méthode appropriée d'utilisation du registre. Cette stratégie est décrite au chapitre 3.

L'utilisation du RA permet de réduire les coûts de listage. Elle ne réduit toutefois pas les autres composantes des coûts de la collecte (coûts de déplacement pour effectuer une interview, temps requis pour mener une interview, *etc.*) On notera au passage que l'utilisation du RA ne requiert aucune modification aux procédures traditionnellement utilisées pour mettre à jour la base de sondage. Dans ce contexte, on profite de bénéfices associés à l'utilisation du RA sans perturber de façon majeure le travail effectué sur le terrain.

Selon cette stratégie, l'utilisation du RA a peu d'effet sur l'élaboration des autres composantes du plan de sondage. L'utilisation du RA a un effet direct sur la structure de coût de la collecte et, par conséquent, sur la stratégie de formation des UPE. Toutefois, on n'a pas tenu compte de cet effet lors du remaniement.

### **2.4 Formation des UPE**

La première étape du plan de sondage consiste à déterminer quelles sont les UPE qui seront utilisées pour la sélection de l'échantillon de premier degré. Une fois déterminé, l'ensemble des UPE formera la base de sondage aréolaire de l'enquête. Nous verrons à la section 2.6 qu'il est possible de structurer la base de sondage afin d'accroître l'efficacité du plan.

---

5. Le RA est principalement utilisé pour la collecte du Recensement de la population. Un processus adapté aux besoins du recensement est mis en place pour effectuer la mise à jour du RA.

Il existe des unités géographiques normalisées définies dans la Classification géographique type. Afin de réduire les coûts du remaniement, nous avons évalué la possibilité d'utiliser une de ces unités pour définir les UPE. Malheureusement, aucune des unités normalisées ne répondait aux besoins de l'enquête. Certaines étaient trop petites alors que d'autres étaient trop grosses. De plus, il aurait été difficile de corriger ces problèmes. Pour ces raisons, nous avons dû former les UPE pour l'ensemble du territoire des dix provinces. On notera au passage qu'il a été possible de former les UPE pour l'ensemble du territoire puisque la base de données de la géographie couvre maintenant tout le territoire. Cette solution de rechange n'était pas envisageable lors du remaniement précédent.

La taille et la forme des UPE ont un effet direct sur l'efficacité du plan de sondage et sur les coûts de collecte. Du point de vue de l'efficacité du plan, il est préférable de choisir peu de logements à partir de petites UPE. En portant cet argument à l'extrême, l'idéal serait de former des UPE ne contenant qu'un seul logement et de sélectionner ce logement lorsque l'UPE correspondante est choisie. Cette situation est similaire à la sélection des logements directement d'une liste de logements. Cet idéal n'est malheureusement pas atteignable avec une base aréolaire.

La taille des UPE et le nombre de ménages sélectionnés par UPE ont un effet sur plusieurs facettes à la fois. Ces effets touchent souvent plusieurs composantes du plan. À titre d'exemple, lorsque l'on réduit le nombre de logements sélectionnés par UPE, il est nécessaire de sélectionner plus d'UPE pour atteindre la taille d'échantillon désirée. Dans ce cas, les logements sélectionnés seront en général plus loin les uns des autres, ce qui augmentera les coûts de déplacement pour les interviews faites en personne. De même, il sera nécessaire de lister un plus grand nombre d'UPE. Par ailleurs, la taille de l'UPE a un impact également sur les coûts de listage. Plus l'UPE est grande, plus il sera difficile et coûteux pour l'intervieweur d'établir la liste des logements. En contrepartie, une grosse UPE pourra demeurer dans l'échantillon plus longtemps avant d'être remplacée par une autre. Le coût du listage d'une grosse UPE est ainsi amorti sur une plus longue période.

Pour établir la taille idéale des UPE et le nombre optimal de logements à sélectionner par UPE, deux éléments sont requis. D'abord, on doit disposer d'un outil permettant d'évaluer la variance d'échantillonnage découlant de différents scénarios. Cet outil peut être construit à partir des données du recensement. On doit également disposer d'un modèle relativement précis des

coûts de collecte. Ce modèle doit permettre d'estimer les coûts pour différents scénarios de taille d'UPE et de nombre de logements sélectionnés par UPE. Pour construire ce modèle, on doit disposer d'information détaillée concernant les coûts. Avec l'introduction du premier contact par téléphone (voir section 4.3) et l'utilisation du RA, il devenait pratiquement impossible de construire un modèle de coûts valable. Dans ce contexte, nous n'avons pu réévaluer la taille idéale des UPE ni le nombre optimal de logements à choisir provenant des UPE échantillonnées. Par conséquent, les conclusions obtenues lors du dernier remaniement ont été reconduites.

La taille des UPE a été établie à 200 logements occupés. D'autres critères ont été établis pour encadrer leur formation. Afin de réduire les coûts de collecte, il est préférable de former des UPE compactes. Une UPE est compacte si le ratio du périmètre sur la surface est petit. Dans la pratique, nous désirons éviter les UPE de forme allongée afin de réduire la distance à parcourir lors de la collecte. De plus, pour réduire le nombre de kilomètres à parcourir, il est souhaitable que toutes les routes de l'UPE soient accessibles sans devoir quitter les frontières de l'unité (par exemple, les UPE ne devraient pas chevaucher une autoroute ou une rivière). Il est également souhaitable que les UPE respectent les limites des régions pour lesquelles on désire produire des estimations.

Les UPE ont été formées avec le Système automatisé de regroupement des territoires (SARTE). Ce système a été initialement conçu par la Division de la géographie de Statistique Canada pour répondre aux besoins du recensement. Les besoins de l'EPA étant similaires à ceux du recensement, l'expertise acquise avec ce dernier pouvait être mise à profit pour l'EPA.

Les UPE de l'EPA sont formées en regroupant des îlots de recensement<sup>6</sup>. Pour ce faire, SARTE commence avec une solution initiale et tente d'améliorer cette solution en déplaçant des îlots de recensement d'une UPE à une autre. Chaque solution est comparée à la précédente à l'aide d'une mesure combinant les divers critères énoncés plus haut. Si cette mesure est inférieure à la mesure de la solution précédente, elle est conservée. SARTE explore des permutations jusqu'à ce qu'il soit impossible d'améliorer la solution trouvée ou que le temps alloué à la recherche soit épuisé.

---

6. Territoire équivalant à un pâté de maisons dont les côtés sont délimités par des rues formant des intersections. Ces territoires couvrent l'ensemble du Canada.

Les UPE formées par SARTE ne peuvent être directement utilisées par l'EPA. En effet, selon certaines circonstances, ces UPE comptent parfois trop ou pas assez de logements. Il est nécessaire de combiner les UPE de petites tailles. Ces UPE de petites tailles sont causées par des contraintes opérationnelles de SARTE. À l'autre extrême, certains îlots de recensement comptent beaucoup plus de 200 logements. Les UPE comportant ces gros îlots de recensement seront, par définition, trop grosses. Ces grosses UPE ont été scindées en exploitant l'information du RA. Finalement, une étape de validation manuelle a été effectuée pour un échantillon d'UPE. Certaines corrections ont été apportées aux limites des UPE à la suite de cette étape.

Nous avons également dû effectuer un traitement particulier aux UPE des régions éloignées. Ces régions sont caractérisées par de grandes surfaces non habitées. Dans le passé, une unité particulière de la Classification géographique type était utilisée pour former les UPE de ces régions. Cette unité n'existe plus selon la classification de 2001. Il a donc été nécessaire de mettre en place une stratégie particulière pour ces régions.

Lorsque laissé à lui-même, SARTE veille à ce que chaque parcelle de territoire est assignée à une UPE. Dans les régions éloignées, cette approche crée des UPE gigantesques couvrant plusieurs centaines de kilomètres carrés. L'utilisation de ces UPE gigantesques aurait généré d'énormes coûts de collecte et frustré les intervieweurs qui auraient dû parcourir plusieurs centaines de kilomètres pour contacter les répondants. Pour former des UPE couvrant de plus petites surfaces, les UPE formées par SARTE ont été modifiées afin d'exclure les parcelles de territoire ayant une très faible densité de population. Après avoir exclu ces parcelles, les UPE étaient redéfinies afin de ne couvrir que les parcelles de terre ayant une densité de population relativement élevée.

Une fois les UPE formées, une analyse détaillée a été effectuée afin de repérer les UPE dont le coût de collecte serait très élevé. Cette analyse a permis d'ajuster à la marge les stratégies de répartition et de stratification afin de contrôler les coûts de collecte dans ces UPE. Ces stratégies sont décrites aux sections 2.5.2 et 2.6 respectivement.

## 2.5 Répartition de l'échantillon

Comme expliqué à la section 2.2, le nombre de logements échantillonnés a un effet direct sur la qualité des estimations qui pourront être produites par l'enquête. Comme l'EPA est utilisée pour produire des estimations à divers niveaux géographiques (Canada,

provinces, régions économiques, *etc.*) et pour plusieurs variables, il est nécessaire d'atteindre un compromis valable pour toutes ces estimations lors de la répartition de l'échantillon.

Cette recherche d'un compromis n'est pas facile, car elle repose sur des besoins diamétralement opposés. À titre d'exemple, la répartition optimale pour produire des estimations à l'échelle nationale serait approximativement proportionnelle à la population de chaque province. Toutefois, une telle répartition produirait des estimations de piètre qualité pour les provinces ayant de faibles populations et des estimations d'excellente qualité pour les provinces ayant le plus de citoyens. Cette approche n'est donc pas retenue. Par ailleurs, une répartition assurant une qualité uniforme pour chaque province produirait de moins bonnes estimations à l'échelle nationale. Encore une fois, cette solution n'est pas appropriée.

La répartition de l'échantillon précise le nombre de logements à sélectionner dans chacun des niveaux géographiques. Cette répartition est déterminée pour s'assurer que l'échantillon pourra produire des estimations rencontrant les objectifs de qualité préétablis. Cette étape est déterminante, car les étapes suivantes en dépendent. Elle est également importante puisqu'elle assure une utilisation efficace des ressources de l'enquête. Un nombre trop important de logements alloué pour une région donnée produira des estimations de qualité supérieure à celle établie dans les objectifs de l'enquête. Toutefois, ce surplus de logements échantillonnés dans une région est obtenu au détriment des autres régions puisque l'enveloppe budgétaire de l'enquête est fixe. La répartition de l'échantillon est faite de façon à satisfaire le mieux possible les besoins tout en visant un plan de sondage efficace.

On notera au passage que la répartition de l'échantillon de l'EPA vise à satisfaire les objectifs de cette enquête. Toutefois, l'échantillon de l'EPA sert souvent de point de départ à d'autres enquêtes, dites supplémentaires, de Statistique Canada. Dans ce contexte, il est fort probable que la répartition de l'échantillon de l'EPA ne soit pas appropriée pour ces autres enquêtes. Il est possible de contourner ce problème. L'utilisation de la base de sondage et de l'échantillon de l'EPA par d'autres enquêtes est traitée au chapitre 9.

La répartition de l'échantillon de l'EPA est effectuée en plusieurs étapes. Les premières étapes déterminent la répartition de base de l'échantillon. Ces étapes sont décrites à la section 2.5.1. La section 2.5.2 décrit les

ajustements faits à la répartition de base afin de mieux cibler les ressources de l'enquête. La méthode utilisée pour mettre en place la réduction de la taille de l'échantillon requise pour financer les activités du remaniement est présentée à la section 2.5.3. Le tableau B.4 (annexe B) présente la répartition de l'échantillon de l'EPA selon diverses unités géographiques.

### 2.5.1 Répartition de base

Des objectifs de qualité de l'EPA sont établis pour les provinces, les régions économiques (RÉ) et les régions économiques de l'assurance-emploi (REAE). Le but de la répartition est de veiller à ce que l'échantillon satisfasse ces objectifs.

Pour les provinces, le c.v. de l'estimation provinciale mensuelle du nombre de chômeurs devrait être inférieur à 7%. Dans le cas des RÉ, le c.v. de l'estimation mensuelle du nombre de chômeurs, par moyenne mobile de trois mois, devrait être inférieur à 25%. Pour les REAE, la cible pour l'estimation par moyenne mobile de trois mois du nombre de chômeurs est de 15%<sup>7</sup>. Puisque la plupart des régions métropolitaines de recensement (RMR) sont également des REAE, fixer un objectif pour les REAE garantit également la qualité des estimations pour les RMR.

Au moment de la rédaction de ce document, deux organismes financent l'échantillon de l'EPA. D'une part, Statistique Canada finance un échantillon de 36 000 ménages afin de répondre aux deux premiers objectifs cités au paragraphe précédent. La deuxième source de financement provient de Ressources humaines et Développement social Canada (RHDSC). Ce financement vise à garantir la qualité des estimations produites pour les REAE. La répartition de l'échantillon repose sur l'hypothèse que l'enveloppe budgétaire de Statistique Canada pour l'enquête est assurée sur une longue période de temps. Par contre, le financement reçu de RHDSC pourrait fluctuer dans le temps.

Dans l'éventualité d'une réduction de financement de RHDSC, l'échantillon financé par Statistique Canada doit assurer la qualité des estimations aux niveaux national, provincial et par RÉ. Pour ce faire, la répartition de ces deux parties de l'échantillon doit être faite de façon séquentielle : répartition de l'échantillon financé par Statistique Canada et répartition de l'échantillon financé par RHDSC pour améliorer la qualité des estimations du nombre de chômeurs par REAE.

---

7. Ces objectifs sont les mêmes que ceux utilisés lors du précédent remaniement.

#### 2.5.1.1 Répartition de l'échantillon financé par Statistique Canada

La première étape consiste à effectuer la répartition de l'échantillon financé par Statistique Canada aux dix provinces. Diverses stratégies avaient été considérées pour effectuer cette étape. En fin de compte, le comité directeur du remaniement a statué que la répartition provinciale utilisée avant le remaniement devait être reconduite dans le nouveau plan de sondage. On notera au passage que les stratégies considérées auraient entraîné peu de changement à cette répartition provinciale.

La seconde étape effectue la répartition de l'échantillon provincial pour chaque RÉ la composant. Certaines RÉ sont de petites tailles. La taille d'échantillon requise pour produire des estimations fiables pour ces petites RÉ serait trop élevée. Par conséquent, quatre petites RÉ ont été combinées avec leur voisine et l'objectif de qualité est appliqué à l'union de ces RÉ. Cette situation se produit pour de petites RÉ dans le nord des provinces de Québec, du Manitoba, de la Saskatchewan et de la Colombie-Britannique.

La répartition de l'échantillon provincial en RÉ doit établir un équilibre entre la qualité des estimations provinciales et la qualité des estimations pour chaque RÉ la composant. Après avoir considéré plusieurs solutions de rechange, il a été décidé que l'estimation provinciale a préséance sur l'estimation par RÉ. Dans le but d'optimiser la qualité des estimations provinciales, nous avons, dans un premier temps, réparti l'échantillon provincial de façon proportionnelle au nombre de ménages dans chaque RÉ. Avec ce premier jet, la qualité des estimations pour certaines RÉ n'aurait pas satisfait les objectifs de qualité de l'enquête. Des ajustements mineurs ont donc été faits pour corriger le tir.

Pour effectuer ces ajustements, il est nécessaire de pouvoir prédire la qualité des estimations du nombre de chômeurs par moyenne mobile de trois mois pour chaque RÉ avec une taille d'échantillon donnée. Cette prédiction est fondée sur un modèle de l'estimation du c.v. en fonction de la taille de l'échantillon, d'un taux de réponse anticipé, d'une estimation du nombre de chômeurs de même qu'une mesure de l'efficacité du plan de sondage pour chaque RÉ. La mesure de l'efficacité du plan utilisée dans ce modèle est l'effet de plan. Ce modèle est fondé sur les données des cinq dernières années de l'EPA. Implicitement, on suppose ainsi que le nouveau plan de sondage se comportera de façon similaire à l'ancien.

À l'aide du modèle, il est possible de quantifier l'effet d'un changement à la répartition de départ sur le c.v. des RÉ. En plus des objectifs déjà stipulés, une contrainte additionnelle est imposée pour chaque RÉ. On désire avoir au moins 200 ménages échantillonnés par mois dans chaque RÉ. Cette contrainte permet qu'un nombre suffisant de ménages soit enquêté dans chaque RÉ et donne ainsi une protection contre les erreurs découlant du modèle utilisé pour prédire les c.v.

La répartition de l'échantillon financé par Statistique Canada dans chaque RÉ est obtenue en changeant le moins possible la répartition de départ pour que : a) un minimum de 200 ménages soient échantillonnés chaque mois dans les RÉ et b) le c.v. de l'estimation mensuelle du nombre de chômeurs obtenue par moyenne mobile de trois mois soit inférieur à 25 %. La solution à ce problème est obtenue à l'aide d'un algorithme de programmation non linéaire.

#### **2.5.1.2 Répartition de l'échantillon financé par RHDSC**

Afin de satisfaire simultanément les trois objectifs de l'enquête, on doit contrôler la taille de l'échantillon alloué à chaque RÉ et à chaque REAE. Pour répondre à ces besoins, il est nécessaire d'effectuer la répartition de l'échantillon aux intersections des RÉ et REAE. Pour la même raison, la stratification de l'enquête est également effectuée séparément dans chaque intersection des RÉ et REAE (voir section 2.6).

Certaines intersections RÉ/REAE sont de petites tailles. Il était impossible d'établir un plan de sondage efficace pour ces petites intersections. Par conséquent, les petites intersections RÉ/REAE ont été combinées avec une intersection voisine. Un principe de base a guidé la combinaison des petites intersections. Dans la mesure du possible, les petites intersections, une fois combinées, devaient respecter les frontières des REAE. Cette approche donne implicitement plus d'importance aux estimations selon la REAE qu'à celles selon la RÉ, car elle améliore l'efficacité du plan de sondage pour les REAE. À la suite de cet exercice de combinaison d'intersections, l'ensemble du territoire des dix provinces est subdivisé en 140 intersections.

Avant de procéder à la répartition de l'échantillon financé par RHDSC, il est nécessaire d'effectuer la répartition de l'échantillon financé par Statistique Canada à une RÉ aux intersections touchant cette RÉ. Cette première répartition est effectuée de façon proportionnelle à la taille de chaque intersection. À partir de cette répartition dans les intersections, on

dérive la répartition de l'échantillon financé par Statistique Canada dans chaque REAE.

La stratégie de répartition de l'échantillon financé par RHDSC repose sur quatre critères :

- Le c.v. de l'estimation du nombre de chômeurs par moyenne mobile de trois mois doit être plus petit que 15 % pour chaque REAE.
- La taille minimale de l'échantillon pour chaque REAE est fixée à 500.
- Les estimations produites par l'EPA sont utilisées pour établir les paramètres d'admissibilité et la durée des prestations au programme de l'assurance-emploi. Dans ce contexte, la qualité des estimations produites pour chaque REAE doit être similaire d'une REAE à l'autre.
- La part de l'échantillon financé par RHDSC pour chaque province doit être similaire à la répartition utilisée avant le remaniement.

Cette dernière contrainte a été introduite pour deux raisons. D'abord, au moment du remaniement, RHDSC amorçait des travaux pour mettre à jour les frontières des REAE. Il aurait donc été prématuré de perturber la répartition provinciale de l'échantillon financé par RHDSC alors que d'autres modifications seront probablement requises sous peu. Cette contrainte offre également une protection contre de potentielles erreurs dans le modèle prédisant l'efficacité du nouveau plan de sondage.

Encore une fois, la programmation non linéaire est utilisée pour résoudre ce problème. Après avoir complété la répartition de l'échantillon financé par la RHDSC, la taille de l'échantillon allouée à chaque REAE est finalement répartie aux intersections RÉ/REAE. Cette dernière étape est de nouveau faite de façon proportionnelle à la taille de chaque intersection.

Ensuite, on a comparé cette nouvelle répartition à celle utilisée avant le remaniement afin de repérer de potentielles erreurs du modèle utilisé pour prédire l'efficacité du nouveau plan. Quelques ajustements ont été effectués lors de cette validation.

La répartition produit deux paramètres qui seront utilisés par les étapes suivantes : la fraction de sondage inverse (FSI) et le nombre de répondants requis pour chaque intersection. La fraction de sondage inverse sera utilisée pour déterminer la taille des strates (voir section 2.6) et lors du processus de sélection (voir section 2.7). Le nombre de répondants requis dans chaque

intersection RÉ/REAE sera, pour sa part, utilisé pour stabiliser la taille de l'échantillon au fil du temps.

### 2.5.2 Les ajustements à la répartition de départ

Comme expliqué à la fin de la section 2.4, le coût pour effectuer la collecte dans certaines UPE est parfois très élevé par rapport à l'effet de ces UPE sur les estimations produites par l'enquête. Deux méthodes ont été utilisées pour identifier ces UPE. Dans un premier temps, nous avons tiré profit de l'expertise détenue par les bureaux régionaux pour identifier un premier ensemble d'UPE. Nous avons également utilisé le taux de vacance<sup>8</sup> observé lors du dernier recensement pour identifier d'autres UPE à fort coût de collecte.

La sélection d'UPE à fort taux de vacance augmente les coûts de collecte, car en plus de faire sa tâche habituelle, l'interviewer doit vérifier si le logement échantillonné est vacant ou non pendant une période de six mois. Le coût de collecte pour un ménage répondant dans une UPE à fort taux de vacance est donc anormalement élevé. On notera au passage que les coûts de listage des UPE à fort taux de vacance sont également élevés. D'un point de vue coûts-bénéfices, il est donc souhaitable de contrôler le nombre de logements sélectionnés dans les UPE à fort taux de vacance.

Les UPE ayant des coûts de collecte élevés ont été assignés à deux groupes. Le premier groupe comprend les UPE ayant un coût de collecte exorbitant. Comme expliqué plus tôt, ces UPE ont été identifiées par les bureaux régionaux. Peu d'UPE répondent aux critères sévères appliqués pour former ce groupe. Pour améliorer le rendement de l'effort de collecte, ces UPE ont simplement été exclues de la base de sondage de l'enquête. L'exclusion de personnes appartenant à la population cible de l'enquête introduit automatiquement un biais dans les estimations de l'enquête. Toutefois, les UPE exclues totalisent moins de 1 % de la population canadienne. Leur exclusion ne peut donc causer un biais significatif dans les estimations. Le tableau B.1 (annexe B) présente la ventilation du nombre de ménages exclus de la base de sondage pour chaque province.

Les UPE ayant un fort taux de vacance sont assignées au second groupe. Afin de réduire les coûts de collecte, on a réduit le nombre de ménages à sélectionner de ces UPE. Pour ce faire, les UPE appartenant au second groupe ont été assignées à des strates spéciales (voir section 2.6). On peut ainsi leur appliquer un traitement

particulier sans pour autant affecter les strates environnantes.

L'exclusion de quelques UPE de même que la réduction de la taille de l'échantillon dans les UPE appartenant au second groupe perturbent légèrement la répartition de l'échantillon. En effet, si rien n'est fait, la réduction de la taille de l'échantillon pour ces UPE entraînera une taille d'échantillon plus petite que celle déterminée par la répartition. Pour pallier ce problème, on a augmenté la fraction de sondage applicable aux UPE à coût de collecte normal dans les intersections contenant des UPE à fort coût de collecte.

Un dernier ajustement a été apporté à la répartition de l'échantillon. Après chaque recensement, Statistique Canada passe en revue la liste des RMR. À la suite du Recensement de 2001, Statistique Canada prévoyait la formation de six nouvelles RMR pour le Recensement de 2006. Puisque l'EPA produit des estimations pour chaque RMR, il fallait s'assurer que la taille de l'échantillon tiré dans chacune de ces nouvelles RMR serait suffisante pour produire des estimations de qualité. Pour ce faire, on a simplement récupéré quelques logements qui étaient alloués à l'échantillon des régions voisines pour les assigner à une de ces six RMR. Cet ajustement garantit que nous aurons assez de répondants pour produire des estimations de qualité pour ces RMR. Il sera toutefois insuffisant si ces RMR sont promues au rang des REAE.

### 2.5.3 Réduction de la taille de l'échantillon

Comme expliqué à la section 1.4, le remaniement de l'EPA a été effectué avec des ressources financières limitées. Pour mener à bien le remaniement, il a été nécessaire de réduire la taille de l'échantillon de l'enquête de 3 % sur une période de trois ans. L'argent ainsi économisé lors de l'activité de collecte a permis de financer les activités du remaniement.

Les systèmes informatiques utilisés par l'EPA pour gérer l'échantillon au fil du temps sont complexes. Il aurait été difficile d'appliquer une réduction spécifique à chaque région et augmenter la taille de l'échantillon dans chacune de ces régions à la fin de la période de trois ans. Par conséquent, nous avons choisi une option simple pour permettre la réduction de la taille de l'échantillon. La réduction de la taille de l'échantillon a été mise en oeuvre de façon uniforme dans l'ensemble des régions. Pour que cette réduction soit mise en oeuvre, nous avons tiré profit de la méthode de stabilisation de la taille de l'échantillon. Cette méthode est décrite à la section 3.4. La taille de l'échantillon ciblé par la stabilisation a été réduite de 3 % sans modifier les

8. Voir le glossaire (annexe A) pour une définition du taux de vacance.

fractions de sondage établies lors de la répartition de l'échantillon. À la fin de la période de trois ans, il suffira de modifier les cibles de la stabilisation afin de ramener la taille de l'échantillon à son niveau normal.

Des simulations ont montré que l'effet de la réduction de la taille de l'échantillon sur les principales estimations de l'enquête sera mineur. Par contre, cette réduction peut avoir un effet plus important sur des estimations touchant des sous-populations. Afin de minimiser ce risque, il est prévu que la taille de l'échantillon sera ramenée à son niveau normal dès avril 2008.

## 2.6 Stratification

Dans le but d'améliorer l'efficacité du plan de sondage, il est préférable de former des strates. Une strate est un groupe d'unités d'échantillonnage. Dans le cas de l'EPA, les strates regroupent les UPE. La stratification consiste à assigner chaque UPE à une seule strate. Une fois la stratification terminée, on peut créer une base de sondage contenant toutes les UPE et leur strate correspondante.

La sélection de l'échantillon est faite de façon indépendante dans chaque strate (voir section 2.7). La stratification améliorera l'efficacité du plan si les UPE assignées à une même strate sont homogènes, c'est-à-dire si les ménages appartenant à la strate ont des caractéristiques similaires. La stratification offre d'autres avantages.

Puisque la sélection est effectuée de façon indépendante dans chaque strate, on peut utiliser une méthode de sélection spécifique et plus appropriée pour chacune. De plus, l'indépendance des strates permet d'appliquer un plan de sondage adapté aux caractéristiques d'une région. À titre d'exemple, pour certains centres urbains isolés sur le territoire, il sera bénéfique d'utiliser un plan de sondage à trois degrés afin de réduire les coûts de collecte. Pour les autres régions, un plan à deux degrés sera plus efficace.

La stratification permet également la mise à jour du plan de sondage pour un ensemble de strates si l'on observe que la population, et ses caractéristiques, ont grandement évolué au fil du temps depuis le recensement décennal. Finalement, on peut tenir compte de contraintes opérationnelles spécifiques à certaines régions pour faciliter le travail de collecte.

### 2.6.1 Changements apportés lors de ce remaniement

Depuis les années 1960, l'EPA utilisait une base d'appartements. Cette base regroupait tous les immeubles d'au moins cinq étages et comptant 30 logements

dans les grands centres urbains. Son utilisation avait deux principaux avantages. Dans un premier temps, elle permettait de mieux contrôler l'effet de la construction de nouveaux immeubles à appartements sur la taille de l'échantillon. Par ricochet, cette base minimisait l'effet de ces constructions sur les probabilités de sélection des UPE. On supposait également que les ménages résidant dans ces immeubles avaient des caractéristiques différentes des autres ménages.

Une étude du Recensement de 2001 a révélé que les ménages demeurant dans les immeubles d'appartements n'étaient pas différents des autres ménages. De plus, entre 1994 et 2004, la croissance de la base d'appartements a été plutôt modeste. Par conséquent, peu de nouveaux immeubles ont été sélectionnés pendant cette période. Toutefois, les coûts associés à la mise en place et la gestion de cette base étaient assez élevés. En s'appuyant sur ces résultats, nous avons conclu que les bénéfices de la base d'appartements n'étaient pas assez importants pour en justifier l'implantation. Nous l'avons donc éliminée du plan de sondage. En 1994, on avait défini des strates d'immeubles à faibles revenus à partir de cette base. L'élimination de la base d'appartements a également entraîné l'élimination de la strate des immeubles à faibles revenus<sup>9</sup>.

Dans le passé, on utilisait un plan de sondage à trois degrés dans certaines régions rurales à faible densité de population. Grâce aux nouveaux outils géographiques, nous avons constaté que les UPE de ces strates couvraient parfois un immense territoire et qu'elles n'étaient pas toujours contiguës. Nous avons également observé que la densité de population de ces régions n'était pas significativement inférieure à celles des régions pour lesquelles nous utilisons un plan à deux degrés. Pour ces raisons, nous avons abandonné le plan de sondage à trois degrés dans les zones rurales. Depuis 2004, nous utilisons donc un plan à trois degrés seulement pour traiter les zones urbaines isolées.

Afin de réduire les coûts de collecte, nous avons également introduit des strates regroupant des UPE ayant un fort taux de vacance. La collecte dans ces UPE est à la fois coûteuse et source de frustration pour les intervieweurs. En isolant des UPE dans des strates spécifiques, on peut réduire la taille de l'échantillon sélectionné dans ces régions et ainsi réduire les coûts de

---

9. Nous avons rapidement considéré la possibilité de construire des strates de ménages à faibles revenus en regroupant des UPE ayant un fort taux de ménages à faibles revenus. Cette approche a été mise de côté, car ces strates auraient été peu stables dans le temps.

collecte. D'autres strates ciblant des populations rares ont également été créées afin de répondre aux besoins des utilisateurs. La méthodologie appliquée pour définir ces strates spéciales est décrite à la section 2.6.4.

### 2.6.2 Stratification de base

Dans la mesure du possible, il est préférable que les strates respectent les régions géographiques pour lesquelles on désire produire des estimations fiables. Comme nous l'avons vu à la section 2.5, des objectifs de qualité sont précisés pour les RÉ et les REAE. Par conséquent, les intersections des RÉ et des REAE forment le premier niveau de stratification de la base de sondage de l'EPA. Comme mentionné à la section 2.5.1.2, certaines de ces intersections sont trop petites et doivent être combinées avec une intersection voisine. Après le regroupement, les intersections sont de nouveau mises à contribution pour la stratification<sup>10</sup>.

La stratification est déterminée en plusieurs étapes. À l'intérieur de chacune de ces intersections, lorsque nécessaire, on identifie les UPE à assigner aux strates spéciales (voir section 2.6.4). Par la suite, les UPE restantes sont stratifiées de façon géographique ou optimale (voir section 2.6.5).

### 2.6.3 Taille des strates

La méthodologie de sélection et renouvellement de l'EPA ajoute une contrainte à la taille des strates. Comme expliqué plus tôt, on effectue le renouvellement d'un sixième de l'échantillon chaque mois. Pour mettre cette approche en place, il est préférable de sélectionner six UPE, ou parfois 12, dans chaque strate. De plus, pour améliorer l'efficacité du plan de sondage, des études menées lors du précédent remaniement ont montré qu'il est préférable de sélectionner 10 ménages par UPE dans les strates rurales, huit dans les secteurs urbains et six pour les strates couvrant les RMR de Montréal, Toronto et Vancouver. En sélectionnant plus de ménages par UPE dans les secteurs ruraux, on espère amortir les coûts de déplacement sur un plus grand nombre d'unités. À l'autre extrême, la sélection de six

ménages par UPE des trois plus grands RMR du pays permet d'augmenter le nombre d'UPE requis dans l'échantillon pour ces régions. La réduction du nombre de ménages requis par UPE dans les trois grands RMR devrait donc entraîner une diminution de l'effet de plan. Finalement, cette réduction permet d'accroître le nombre de strates à former. Des strates plus nombreuses, et donc plus petites, devraient entraîner une augmentation de l'homogénéité de celles-ci ce qui devrait également améliorer l'efficacité du plan de sondage.

En combinant ces contraintes, il s'avère que le nombre de logements à regrouper dans chaque strate est :

$$M_h = FSI * 6 * m_h^* \quad (1)$$

où

$M_h$  : nombre de ménages à regrouper dans chaque strate d'une région;

$FSI$  : fraction de sondage inverse telle qu'établie lors de la répartition de l'échantillon. Comme nous l'avons vu à la section 2.5, cette fraction de sondage est la même pour toutes les strates dans une intersection RÉ/REAE;

$m_h^*$  : nombre de ménages à sélectionner par UPE choisie au premier degré. Comme expliqué au précédent paragraphe, ce nombre varie selon la densité de la population pour la région (rurale, urbaine, trois plus grands RMR).

En divisant le nombre de ménages dans une région par cette cible, on peut déterminer le nombre de strates à former dans chaque région. Le résultat de cette division n'est pas un nombre entier. Par conséquent, il est nécessaire de l'arrondir. On notera au passage qu'avec cette contrainte, les strates d'une intersection RÉ/REAE ont toutes approximativement la même taille.

### 2.6.4 Strates spéciales

Les strates spéciales peuvent être regroupées en deux catégories : celles qui sont définies pour traiter les zones géographiques isolées ou ayant une faible densité de population; et celles qui sont formées pour cibler des populations particulières. La première catégorie permet de définir un plan de sondage adapté à la géographie du territoire canadien. Elle comprend les strates pour les régions éloignées, les strates pour les régions à haut taux de vacance et les strates à trois degrés de sélection pour les zones urbaines isolées. La seconde catégorie permet

10. L'Enquête sur la Santé dans les collectivités canadiennes sélectionne une part importante de son échantillon de la base aréolaire élaborée pour l'EPA. Cette enquête produit des estimations selon la région socio-sanitaire. Nous avons donc considéré la possibilité de tenir compte des régions sociosanitaires pour définir les strates de l'EPA. Cette approche a été rejetée, car les frontières de ces régions changent au fil du temps. De plus, le croisement des régions sociosanitaires avec les RÉ et les REAE aurait entraîné un nombre important de petites intersections. Il aurait donc été impossible de former des strates qui respectent chacune de ces classifications.

de mieux cibler certaines populations d'intérêt pour les analystes utilisant les données de l'EPA. Elle inclut les strates à forte proportion d'Autochtones, les strates à forte proportion d'immigrants et les strates à forte proportion de ménages à hauts revenus. Afin d'alléger le texte, on utilisera dorénavant les termes strates d'Autochtones, strates d'immigrants et strates à haut revenu pour désigner ces strates. Remarquons, qu'il s'agit ici d'un usage quelque peu abusif, car ces strates ne comptent pas que des Autochtones, des immigrants ou des ménages à hauts revenus.

#### **2.6.4.1 Des strates adaptées aux particularités du territoire canadien**

Une part importante du territoire canadien est habitée par une petite portion de la population. Les coûts de collecte dans les régions à faible densité de population sont très élevés. Par ailleurs, l'effet de ces régions sur les estimations principales de l'EPA est relativement faible. Il est donc nécessaire d'élaborer des approches bien adaptées pour ces régions afin d'allouer efficacement nos ressources limitées.

Pour satisfaire cet objectif, on définit d'abord des strates de régions éloignées. Ces strates regroupent les parties du territoire canadien ayant la plus faible densité de population. Une fois ces régions consignées dans des strates particulières, il est possible de mieux contrôler la taille de l'échantillon sélectionné dans ces strates. On peut ainsi mieux contrôler la répartition de nos ressources. Les limites de ces régions sont essentiellement les mêmes que lors du précédent remaniement. Quelques corrections ont été apportées à ces limites afin d'exclure de ces strates les régions dont la densité de la population a augmenté depuis 1991. À la suite de ce travail, une strate de régions éloignées a été formée pour la partie nord de toutes les provinces sauf pour l'Île-du-Prince-Édouard, la Nouvelle-Écosse et le Nouveau-Brunswick. Ces strates sont formées de territoires contigus.

Comme nous l'avons vu plus tôt, les coûts de collecte sont anormalement élevés lorsque le taux de vacance d'une UPE est élevé. La densité de population de ces UPE est également faible et elles sont parfois situées près d'un centre urbain. Cette caractéristique les distingue des UPE des régions éloignées. La strate à haut taux de vacance regroupe ces UPE. Contrairement aux strates des régions éloignées, les strates à haut taux de vacance ne couvrent pas un territoire contigu.

Le dernier type de strate de cette catégorie apporte une solution aux contraintes opérationnelles associées à la collecte dans ces centres urbains de moyenne taille,

mais isolés géographiquement. Ces centres sont trop petits pour combler la tâche d'un intervieweur pendant une longue période. Ils sont toutefois trop isolés des autres centres urbains pour qu'un intervieweur résidant hors de ses limites puisse se déplacer pour y effectuer la collecte. Pour réduire les coûts de déplacement et les coûts de formation associés à l'embauche d'un nouvel intervieweur, il est donc préférable d'utiliser un plan de sondage à trois degrés pour couvrir ces centres urbains.

Au premier degré, on choisit un centre urbain parmi tous les centres regroupés dans une strate<sup>11</sup>. Le territoire du centre urbain sélectionné est scindé en parcelles de terrain (Unités secondaires d'échantillonnage (USE)). L'approche appliquée pour définir les UPE est de nouveau utilisée pour définir ces USE. Toutes les USE d'une UPE sont assignées au même groupe de renouvellement. Ainsi, un intervieweur pourra prendre contact avec tous les nouveaux ménages introduits dans l'échantillon pour l'UPE dans un même mois. Puisque la première interview dans un ménage est souvent faite en personne, cette approche devrait réduire les déplacements à effectuer par l'intervieweur. Au second degré, on choisit un échantillon d'USE. Au troisième et dernier degré, on sélectionne un échantillon de logements dans les USE sélectionnées.

Cette approche garantit qu'un centre urbain échantillonné demeurera dans l'échantillon pour une longue période de temps. Elle réduit également les coûts de collecte en concentrant l'échantillon dans un nombre réduit de ces petits centres urbains. Des strates à trois degrés ont été formées dans les provinces de Québec, de l'Ontario, de l'Alberta et de la Colombie-Britannique.

Le tableau B.2 (annexe B) présente le nombre de ménages dans les strates spéciales de la première catégorie.

#### **2.6.4.2 Des strates pour cibler certaines sous-populations**

Les sous-populations ciblées par ces strates spéciales sont relativement rares. Une base de sondage aréolaire n'est pas un outil très efficace pour cibler des sous-populations lorsque celles-ci ne sont pas très concentrées géographiquement.

Les membres de ces sous-populations qui nous intéressent n'habitent pas tous les mêmes quartiers. De plus, dans un quartier regroupant plusieurs membres d'une sous-population donnée, une part importante de ménages ne comptent aucun membre de cette sous-

---

11. Ces centres urbains correspondent aux limites des subdivisions de recensement.

population. Par conséquent, le pourcentage de ces sous-populations dans les strates spéciales demeure relativement faible. Il est important de noter que ce pourcentage est toutefois bien supérieur à celui observé sur l'ensemble du territoire canadien. Dans le cadre du plan de sondage utilisé par l'EPA, elles fournissent le seul outil disponible, bien qu'imparfait, pour cibler ces sous-populations.

Trois critères déterminent l'efficacité des strates spéciales. Comme nous l'avons vu au paragraphe précédent, le premier de ces critères est la prévalence. Le second critère est la proportion de la population ciblée résidant dans des UPE assignées aux strates spéciales. Ces deux premiers critères vont de paire : ils réussissent ensemble ou échouent ensemble. Par exemple, une strate d'Autochtones ayant un pourcentage de ménages Autochtones de 60 % serait peu efficace si elle ne couvre qu'un pour cent de la population des Autochtones. L'inverse est également vrai. Le troisième critère est la stabilité des strates spéciales au fil du temps. La mise en place de strates spéciales de très bonne qualité au départ est inutile si ces strates se détériorent très rapidement après quelques mois.

Un dernier facteur primordial a été considéré pour élaborer la stratégie de formation des strates spéciales : leur effet sur les estimations produites pour la population totale. L'obtention de bonnes strates spéciales ne peut se justifier si leur introduction résulte en une diminution majeure de la qualité des estimations principales de l'EPA. Afin de trouver un compromis valable, nous avons utilisé les données du Recensement de 2001 pour mesurer l'effet de divers scénarios sur la prévalence, la proportion de la population cible couverte par les strates spéciales et l'effet sur les estimations principales de l'enquête. Les données du Recensement de 1996 ont été mises à contribution pour évaluer la stabilité dans le temps de chaque scénario. Les lignes directrices de la stratégie de formation des strates spéciales découlent de cette étude.

La première ligne directrice détermine que la formation des strates se fait en fonction de la prévalence de caractéristiques particulières. À titre d'exemple, la proportion d'immigrants à l'Île-du-Prince-Édouard est très faible. Dans ce contexte, il serait inutile de tenter de former une strate d'immigrants pour cette province. Selon ce critère, la formation de strates d'Autochtones n'est donc possible qu'au Manitoba, en Saskatchewan, en Alberta et en Colombie-Britannique. De même, des strates d'immigrants sont requises pour les RMR de Montréal, d'Ottawa, de Toronto, de Calgary et de

Vancouver. Finalement, on peut former des strates de hauts revenus dans les plus importantes RMR du pays.

La seconde ligne directrice précise une limite au nombre de strates spéciales qui peuvent être formées. Cette contrainte garantit que les strates spéciales n'auront pas un effet négatif majeur sur les estimations principales de l'EPA. L'étude menée avec les données des recensements de 1996 et de 2001 a montré que chaque catégorie de strates spéciales ne devait pas couvrir plus de 8 % de la population d'une région.

À partir de ces lignes directrices, les strates spéciales ont été formées de façon séquentielle. Pour chaque catégorie, les strates spéciales sont formées en identifiant les UPE ayant la plus forte prévalence de la sous-population d'intérêt. En raison de leur construction, ces strates ne sont donc pas contiguës.

Nous avons d'abord formé les strates de hauts revenus dans les plus importantes RMR. Pour ce faire, dans une RMR donnée, les UPE étaient d'abord classées en ordre décroissant selon la proportion de ménages ayant des revenus supérieurs à 125 000 \$ selon le Recensement de 2001<sup>12</sup>. Les UPE du haut de cette liste étaient assignées à une strate de hauts revenus jusqu'à ce que cette strate soit de la taille requise (voir section 2.6.3). Si la limite de 8 % n'était pas atteinte, une autre strate de hauts revenus était formée pour ce même RMR. Une fois le travail complété pour ce RMR, on passait au RMR suivant.

La même approche a été appliquée pour former les strates d'immigrants et d'Autochtones. Pour former les strates d'immigrants, les UPE étaient classées par ordre décroissant selon la proportion de ménages comptant au moins un immigrant selon le Recensement de 2001. À Montréal, Toronto et Vancouver, un ménage était classé immigrant si au moins un de ses membres avait immigré au pays lors de cinq dernières années. Pour Ottawa et Calgary, la règle était fondée sur l'entrée au pays dans les dix dernières années puisque ces RMR comptent moins d'immigrants.

Pour former les strates d'Autochtones, nous avons dû modifier légèrement cette stratégie de base. Les strates des hauts revenus et d'immigrants respectent les limites des RMR. Toutefois, une portion importante des Autochtones demeure hors des limites des RMR. De

---

12. Nous avons considéré la possibilité de mesurer le pourcentage en se basant sur les données fiscales plutôt que sur le Recensement de 2001. Nous avons opté pour l'utilisation du recensement, car il était impossible d'assigner correctement chaque déclaration fiscale à une UPE.

plus, certaines intersections RÉ/REAE étaient trop petites pour former une strate d'Autochtones bien que plusieurs UPE dans ces intersections comptaient une forte proportion de ménages d'Autochtones. Pour pallier ce problème, les strates d'Autochtones ne respectent pas les limites de chaque intersection RÉ/REAE, mais plutôt les limites des REAE. Finalement, les UPE déjà assignées à une strate de région éloignée, une strate à fort taux de vacance ou une strate à trois degrés ne pouvaient être assignées à une strate Autochtones<sup>13</sup>.

Le tableau B.3 (annexe B) fournit le nombre de ménages regroupés dans les strates spéciales, la prévalence de la population ciblée et la proportion de la sous-population comprise par les strates spéciales.

### 2.6.5 La stratification des UPE restantes

La stratégie utilisée pour définir plusieurs types de strates spéciales a été décrite dans la section précédente. Ces strates ne comprennent toutefois qu'une petite part du territoire canadien. Cette section décrit l'approche implantée pour stratifier les UPE normales, c'est-à-dire non classées dans une strate spéciale.

Avec le temps la demande d'information pour les subdivisions de recensement (SDR) a crû fortement. Par conséquent, nous avons tenté de former des strates respectant les limites des SDR. Grâce à cette approche, on stabilise la taille de l'échantillon sélectionné dans la SDR. On améliore ainsi la qualité des estimations pour cette SDR. Plusieurs SDR sont relativement petites. Il est donc impossible de former une strate à l'intérieur de ces SDR, car l'EPA impose une contrainte sur le nombre de ménages à regrouper dans chaque strate. L'équation (1) de la section 2.6.3 fournit la taille ciblée pour chaque strate. De ce fait, cette équation précise également la taille minimale d'une SDR pour qu'on puisse former une strate, ou plusieurs, respectant ses frontières. Règle générale, nous avons pu former des strates dans les SDR comptant au moins 20 000 ménages.

Dans la mesure du possible, nous avons également tenté de former des strates respectant les zones urbaines et les zones rurales. Cette approche se justifie par trois raisons : les strates rurales comptent plus de ménages que les strates urbaines (voir équation (1) de la section 2.6.3); les personnes habitant dans les zones rurales ont des caractéristiques différentes de celles habitant les zones urbaines; une stratification respectant ces zones

nous permet de mettre en place des stratégies de collecte plus appropriées.

Ces trois règles de base servent de canevas à la stratification des UPE restantes. La stratégie de stratification définit ces règles opérationnelles en une séquence d'étapes à accomplir. À chaque étape, on doit déterminer le nombre de strates à former. On doit également décider comment traiter les petites zones résultant de l'intersection des limites des RMR, des RÉ, des REAE, et des zones urbaines ou rurales.

Puisqu'on désire produire des estimations de qualité pour les RMR, il est préférable que les limites des strates respectent les frontières des RMR. Nous avons vu plus tôt que les limites des RMR respectent généralement les limites des REAE. À de rares occasions, ces frontières ne sont pas exactement les mêmes. Pour ces RMR, il est impossible de former des strates respectant à la fois les RMR et les intersections RÉ/REAE. Une décision concernant chacun de ces cas a été prise afin de minimiser l'effet de l'ajustement à la marge d'une de ces frontières sur les estimations des RMR et des RÉ et REAE concernées. La même situation se présente lorsque l'on désire former des strates respectant les zones urbaines et rurales. Dans certains cas, la zone urbaine est trop petite pour qu'une strate soit formée à l'intérieur de celle-ci. Dans ce cas, il est nécessaire de combiner ces zones à une zone urbaine voisine ou une zone rurale. Encore une fois, chaque cas a été évalué individuellement. Les petites zones rurales sont traitées de la même façon.

Après avoir réglé ces quelques problèmes de frontières, on effectue d'abord la stratification des RMR. Chaque RMR est traitée séparément. Dans une RMR donnée, on identifie d'abord les SDR suffisamment grandes pour former des strates respectant leurs frontières. L'équation (1) (section 2.6.3) permet d'identifier ces SDR. Parmi les SDR suffisamment grandes, on vérifie ensuite s'il est nécessaire de former plus d'une strate. Pour ce faire, il suffit de diviser le nombre de ménages dans la SDR par la taille de strate ciblée. Comme ce quotient n'est pas un nombre entier, on procède à un arrondissement. Si la SDR est suffisamment grande pour former au moins 10 strates, on forme d'abord des super strates. Ces super strates scindent le territoire de la SDR en territoires compacts comptant un nombre similaire de ménages. Cette approche assure une meilleure répartition géographique des UPE sélectionnées. Finalement, dans chaque SDR comptant plus d'une strate et chaque super strate, on effectue une stratification optimale.

---

13. Cette contrainte n'est pas requise pour les strates d'immigrants et de hauts revenus, car les RMR ne sont pas touchées par ces strates spéciales.

Cette stratification tente de réduire la variance d'échantillonnage de plusieurs variables d'intérêt en regroupant des UPE ayant des caractéristiques similaires. La liste de ces variables d'intérêt, 29 au total, est identique à celle du dernier remaniement. La liste de ces variables est disponible à la fin de l'annexe B.

L'algorithme implanté pour effectuer cette stratification repose sur une méthode élaborée par Friedman et Rubin (1967) et modifiée par Drew et coll. (1985). Ces modifications ont pour objet de l'adapter au contexte des sondages avec probabilités inégales et pour produire des strates de tailles similaires. L'algorithme permet d'accorder plus d'importance à certaines variables. Pour notre application, la même importance est accordée à chacune d'elles. La seule exception à cette règle est le revenu du ménage à laquelle on a accordé trois fois plus d'importance, car le revenu est corrélé avec plusieurs variables.

L'algorithme utilise une approche itérative. À partir d'une stratification de départ respectant les contraintes, il effectue l'échange d'une UPE entre deux strates et vérifie si cette nouvelle stratification réduit la variance. Si l'échange entraîne une réduction, cette nouvelle solution remplace la précédente. Si l'échange résulte en une augmentation de la variance, la stratification précédente est conservée. Ce processus d'échange est répété jusqu'à ce qu'aucun échange n'entraîne de réduction de la variance. Une fois complété, on utilise d'autres stratifications de départ différentes et on répète le processus. La stratification associée à la plus petite variance parmi toutes celles considérées est conservée. Il est possible de former des strates compactes et contiguës avec cette méthode. Ces contraintes supplémentaires limitent grandement les combinaisons d'UPE qui peuvent être considérées. Elles produisent donc des solutions ayant une plus grande variance d'échantillonnage. Pour cette raison, comme lors du dernier remaniement, les strates formées lors de la stratification optimale ne sont pas contiguës ni compactes.

Une fois la stratification des grandes SDR terminée, on passe au traitement des zones urbaines. La stratégie appliquée pour les gros SDR est de nouveau utilisée. Dans chaque zone urbaine, on détermine d'abord le nombre de strates requises. Lorsque plus de 10 strates sont requises, on procède à la formation de super strates. On effectue ensuite une stratification optimale par zone urbaine ou par super strate pour former les strates finales. Les mêmes étapes sont accomplies pour la zone rurale.

Cette approche est également appliquée pour stratifier les UPE situées hors des RMR. Dans chaque intersection RÉ/REAE, on identifie d'abord les SDR suffisamment grandes pour contenir au moins une strate. Le reste du territoire est par la suite stratifié de façon optimale selon l'appartenance à la zone urbaine ou à la zone rurale. On notera au passage que la zone urbaine hors des RMR est morcelée. En effet, elle est composée de petites villes dispersées sur le territoire de l'intersection et entourées de la zone rurale. Par conséquent, les UPE regroupées dans une même strate peuvent être situées à plusieurs kilomètres de distance les unes des autres. Ce morcellement de la strate urbaine n'entraîne pas une augmentation des coûts de collecte puisqu'on utilise le même taux de sondage pour les strates rurales entourant les strates urbaines. Dans ce contexte, une UPE sélectionnée dans une strate urbaine sera souvent près d'une autre UPE sélectionnée. Comme expliqué plus tôt, les centres urbains isolés géographiquement sont inclus dans des strates à trois degrés de sélection.

## 2.7 Sélection de l'échantillon

Une fois la stratification terminée, toutes les pièces sont en place pour procéder à la sélection de l'échantillon. Cette section fournit une description conceptuelle de la méthode de sélection et de renouvellement utilisée par l'EPA. Des détails supplémentaires concernant le traitement de la croissance et la maintenance de la base de sondage sont fournis au chapitre 3.

Lorsque l'on utilise un plan à deux degrés, la théorie des sondages stipule qu'il est préférable de sélectionner les UPE avec une probabilité proportionnelle à leur taille quand cette mesure de taille est également corrélée aux estimations d'intérêt. Cette condition est satisfaite pour l'EPA. À titre d'exemple, le nombre de personnes travaillant dans une UPE est fortement corrélé au nombre de personnes demeurant dans l'UPE. Par cette raison, la sélection des UPE de l'EPA est effectuée avec une probabilité proportionnelle à leur taille. La mesure de taille que nous utilisons pour calculer la probabilité de sélection de chaque UPE est le nombre de ménages dans l'UPE selon le recensement de 2001<sup>14</sup>.

La sélection de l'échantillon s'effectue de façon indépendante d'une strate à l'autre. Il est donc possible d'adopter une méthode de sélection adaptée à chaque strate. Il existe plusieurs méthodes pour sélectionner un échantillon d'UPE avec probabilité proportionnelle à

---

14. Dans la pratique, on dérive plutôt une mesure de taille à partir du nombre de ménages. Cette mesure se nomme l'inverse de la fraction de sondage pour chaque UPE. Les détails concernant ce calcul sont fournis à la section 2.7.1.

leur taille. L'EPA en utilise deux et une de celles-ci est appliquée dans la grande majorité des strates. Ces deux méthodes sont présentées à la section 2.7.1.

Comme expliqué à la section 2.1, on remplace le sixième de l'échantillon chaque mois. Pour simplifier le processus de renouvellement de l'échantillon, il est donc préférable de sélectionner six UPE, ou un multiple de six, dans chaque strate. Ainsi, il suffit de remplacer les logements sélectionnés dans une de ces six chaque mois pour effectuer le renouvellement. Nous reviendrons sur la méthode de renouvellement de l'échantillon à la section 2.7.2.

Pour déterminer le nombre d'UPE à sélectionner dans la strate, on calcule d'abord le nombre de ménages à enquêter dans la strate selon la répartition de l'échantillon. Pour ce faire, il suffit de diviser le nombre de ménages dans la strate par la fraction de sondage inverse (FSI) déterminée lors de la répartition de l'échantillon. On divise ensuite ce quotient par le nombre idéal de ménages à enquêter par UPE sélectionnée. Comme nous l'avons vu à la section 2.6.3, ce nombre idéal est de six ménages par UPE dans les strates des RMR de Montréal, Toronto et Vancouver; huit dans les strates urbaines hors de ces trois RMR et dix dans les strates rurales. Si le résultat de la seconde division est plus près de six que de douze, on sélectionnera six UPE dans la strate. Sinon, douze UPE seront sélectionnées. Lors de la stratification, la taille des strates était déterminée selon l'hypothèse que six UPE seraient sélectionnées (voir section 2.6.3). Par conséquent, la seconde division nous conduira à la conclusion qu'il est nécessaire de sélectionner six UPE dans l'immense majorité des strates. Dans ce qui suit, les explications seront fournies pour la sélection de six UPE. La même approche s'applique lorsque l'on désire sélectionner douze UPE.

### 2.7.1 Sélection des UPE et du premier échantillon de logements

Deux méthodes sont utilisées par l'EPA pour sélectionner l'échantillon d'UPE : la méthode de Rao-Hartley-Cochran (RHC) et la méthode par sélection systématique avec probabilité proportionnelle à la taille et ordre aléatoire. La méthode de RHC est appliquée pour la plupart des strates, car elle ouvre la porte à la mise à jour des probabilités de sélection lorsqu'on observe une forte croissance pour certaines UPE. La méthode décrite dans Keyfitz (1951) peut être combinée avec la méthode RHC pour effectuer la mise à jour des probabilités tout en maximisant le chevauchement des UPE choisies avant et après la mise à jour.

Dans ce qui suit, on présente de façon sommaire la méthode RHC. Pour plus de détails, le lecteur peut consulter l'article de Rao, Hartley et Cochran (1962). La méthode par sélection systématique avec probabilité proportionnelle à la taille et ordre aléatoire est peu utilisée par l'EPA, elle n'est donc pas décrite dans ce document. Les principes décrits pour la méthode RHC s'appliquent également à cette dernière. Le lecteur désirant plus d'information sur cette méthode peut consulter Cochran (1977).

Pour sélectionner six UPE dans une strate, il faut d'abord distribuer l'ensemble des UPE de la strate dans six groupes contenant chacun le même nombre d'UPE. Chacun de ces groupes est associé à un groupe de renouvellement (voir section 2.1). Dans un second temps, il suffit de sélectionner une UPE par groupe avec probabilité proportionnelle à la taille dans le groupe. Cette idée peut se résumer à l'aide de l'équation suivante :

$$\pi_{hij} = \frac{M_{hij}}{\sum_{j \in hi} M_{hij}} \quad (2)$$

où

$M_{hij}$  : le nombre de ménages dans l'UPE  $j$  du groupe  $i$  de la strate  $h$  selon le Recensement de 2001.

$\sum_{j \in hi} M_{hij}$  : est la somme sur toutes les UPE appartenant au groupe  $i$  de la strate  $h$ .

$\pi_{hij}$  : est la probabilité de sélection de l'UPE  $j$  dans le groupe  $i$  de la strate  $h$

Au second degré, on désire sélectionner un échantillon de ménages dans les UPE choisies. Afin de simplifier ce processus, l'EPA sélectionne les ménages par échantillonnage systématique. Cette méthode est préconisée, car, elle est simple à implanter, elle assure une bonne répartition des ménages choisis dans l'UPE et elle permet de traiter facilement l'ajout de nouveaux logements dans l'UPE. Pour effectuer un tirage par échantillonnage systématique, il est nécessaire de déterminer le pas de sondage. Ce pas de sondage correspond à la fraction de sondage inverse (FSI) de l'UPE. Cette dernière est établie en fonction du nombre de ménages dans l'UPE selon le Recensement de 2001 et la FSI déterminée lors de la répartition de l'échantillon. On l'obtient avec l'équation suivante :

$$FSI_{hij} = \left( \frac{M_{hij}}{\sum_{j \in hi} M_{hij}} \right) FSI_h \quad (3)$$

où

$FSI_{hij}$  : fraction de sondage inverse dans l'UPE  $j$  du groupe  $i$  de la strate  $h$ .

$FSI_h$  : fraction de sondage inverse de la strate  $h$  établie lors de la répartition de l'échantillon<sup>15</sup>.

Comme la  $FSI_h$  est constante pour toutes les UPE d'un groupe,  $FSI_{hij}$  est proportionnelle au nombre de ménages dans chaque UPE. Par conséquent, il est possible de sélectionner un échantillon avec probabilité proportionnelle à la taille en utilisant ces  $FSI_{hij}$  comme mesure de taille. Le système de sélection de l'EPA est configuré pour utiliser des fractions de sondage inverses entières. Le résultat de l'équation (3) est donc arrondi à l'entier inférieur ou supérieur afin que  $\sum_{j \in hi} FSI_{hij} = FSI_{hi} = FSI_h, \forall i \in h$ .

À la suite de cet arrondissement, il existe deux interprétations plus intuitives pour  $FSI_{hij}$ . Selon la première,  $FSI_{hij}$  est le nombre d'échantillons distincts disponibles dans l'UPE. Dans le jargon de l'EPA, ce concept s'appelle le nombre de départs aléatoires. La fraction de sondage inverse de l'UPE est également le pas de sondage à utiliser si l'UPE correspondante est sélectionnée au premier degré. En appliquant ce pas de sondage, on sélectionnera le nombre approprié de ménages dans l'UPE pour atteindre la cible prévue pour le groupe<sup>16</sup>.

En résumé, l'équation (2), après arrondissement contrôlé, fournit la mesure de taille à utiliser pour sélectionner un échantillon avec probabilité proportionnelle à la taille. La probabilité de sélection du premier degré associée à chaque UPE est donc :

$$\pi_{hij}^* = \frac{FSI_{hij}}{\sum_{j \in hi} FSI_{hij}} = \frac{FSI_{hij}}{FSI_h}. \quad (4)$$

De plus, la probabilité de sélection au second degré lorsque l'UPE  $j$  du groupe  $i$  de la strate  $h$  est choisie est  $1/FSI_{hij}$ . Par conséquent, la probabilité de sélection du ménage  $k$  situé dans l'UPE  $j$  du groupe  $i$  de la strate  $h$  est :

$$\pi_{hij}^* = \frac{FSI_{hij}}{\sum_{j \in hi} FSI_{hij}} \times \frac{1}{FSI_{hij}} = \frac{1}{FSI_h}. \quad (5)$$

La probabilité de sélection des ménages appartenant à une strate est la même pour tous les ménages. Le plan de

15. Comme expliqué à la section 2.5, on utilise la même FSI pour toutes les strates appartenant à une intersection RÉ/REAE.

16. Cette cible correspond au nombre de ménages dans le groupe (selon le Recensement de 2001) divisé par la fraction de sondage de la strate.

sondage de l'EPA est donc autopondéré<sup>17</sup>. Finalement, on notera que  $\pi_{hij}^* \approx \pi_{hij}$  car

$$\pi_{hij}^* = \frac{FSI_{hij}}{\sum_{j \in hi} FSI_{hij}} \approx \frac{M_{hij}}{\sum_{j \in hi} M_{hij}} = \pi_{hij}. \quad (6)$$

La différence entre ces deux probabilités résulte de l'arrondissement effectué sur les  $FSI_{hij}$ .

Dans la pratique, pour sélectionner une UPE dans un groupe, on ordonne d'abord les UPE du groupe de façon aléatoire. On tire ensuite un nombre aléatoire entier suivant une loi uniforme. Ce nombre aléatoire  $U$  est tel que  $U \in [1, FSI_h]$ . Il a deux fonctions. Il permet d'abord d'identifier la première UPE sélectionnée. Cette UPE est la première pour laquelle la somme cumulée des  $FSI_{hij}$  est inférieure ou égale à  $U$  (c'est-à-dire  $\sum_{j \leq k} FSI_{hij} \leq U$  où l'indice  $j$  suit l'ordre aléatoire).

Il précise également le nombre de départs aléatoires à utiliser dans cette UPE  $k$  avant de passer à l'UPE suivante. Le nombre de départs à utiliser dans la première UPE est  $D_k = (\sum_{j \leq k} FSI_{hij}) - U + 1$ . On sélectionne finalement un second nombre aléatoire entier  $U_k \in [1, FSI_{hik}]$ <sup>18</sup>. Ce nombre indique le premier départ aléatoire à utiliser pour sélectionner l'échantillon de logements pour l'UPE  $k$ . Ces logements demeureront dans l'échantillon pour une période de six mois.

Gray (1973) et Alexander et coll. (1982) utilisent deux approches différentes pour démontrer que cette méthode produit un échantillon respectant les probabilités de sélection spécifiées. Laflamme (2003) illustre le processus de sélection de l'échantillon à l'aide d'un diagramme.

## 2.7.2 Renouvellement de l'échantillon

La section 2.7.1 décrit comment on sélectionne le premier échantillon de logements dans chacun des groupes formés avec la méthode RHC. Après une période de six mois, il est nécessaire de remplacer cet échantillon par de nouveaux logements. En poursuivant avec l'exemple donné à la fin de la précédente section,

17. Un plan de sondage est autopondéré si tous les individus d'une région ont le même poids de sondage. Dans le cas de l'EPA, tous les individus appartenant à la même intersection RÉ/REAE ont le même poids de sondage.

18. Ce second nombre aléatoire a deux fonctions. Il tient compte que la taille de l'échantillon associé aux derniers départs aléatoires est parfois plus petite que celle des premiers départs. On espère ainsi stabiliser la taille de l'échantillon global au fil du temps. Il prépare également le terrain pour appliquer la règle du nombre minimum de départs à utiliser.

le premier échantillon correspondait au départ aléatoire  $U_k$  de l'UPE  $k$ .

Si le nombre de départs aléatoires à utiliser de l'UPE  $k$  est un ( $D_k = 1$ ), le second échantillon de logements correspond au départ  $U_{k+1}$  de l'UPE  $k + 1$  où  $U_{k+1} \in [1, FSI_{hi(k+1)}]$ . Dans le cas contraire, le second échantillon correspond au départ  $U_k + 1$  de l'UPE  $k$ . Si  $U_k + 1 > FSI_k$ , on passe au départ 1 de l'UPE  $k$ . De façon générale, avec cette méthode, l'UPE  $k$  demeure dans l'échantillon pendant  $D_k$  périodes de six mois. Lorsqu'il est nécessaire de remplacer les logements enquêtés, il suffit de passer au départ aléatoire suivant. Après  $D_k$  périodes, on passe au départ  $U_{k+1}$  de l'UPE  $k + 1$ . Cette UPE demeurera dans l'échantillon jusqu'à ce que tous ses départs aléatoires aient été utilisés. Il en va de même pour les UPE qui entreront dans l'échantillon à une date ultérieure.

Cette méthode produit les résultats escomptés : les probabilités de sélection sont toujours respectées au fil du temps. Malheureusement, elle souffre d'un inconvénient de taille. Comme nous l'avons vu, la première UPE demeure dans l'échantillon un nombre aléatoire de périodes. Dans certains cas, la première UPE sélectionnée demeure dans l'échantillon pour peu de périodes. Ce renouvellement rapide de la première UPE sélectionnée causerait une utilisation inefficace de nos ressources limitées. En effet, l'introduction d'une UPE dans l'échantillon requiert un investissement de notre part : préparation du matériel, listage et parfois l'embauche et la formation d'un intervieweur. Pour être efficace, il serait donc préférable d'amortir cet investissement en évitant, dans la mesure du possible, un renouvellement trop rapide de la première UPE.

Pour pallier ce problème, l'EPA a conçu une correction qui permet d'augmenter le nombre de départs

aléatoires à utiliser de la première UPE sans introduire de biais dans les probabilités de sélection. Lorsque  $D_k$  est trop petit, selon un critère prédéterminé, on l'augmente afin de conserver cette UPE plus longtemps dans l'échantillon. Dans ce cas, il est nécessaire de réduire d'autant le nombre de départs à utiliser pour l'UPE  $k + 1$  afin de ne pas biaiser les probabilités de sélection. Certaines contraintes sont requises pour s'assurer que l'augmentation du nombre de départs associés à la première UPE ne réduira pas trop le nombre de départs à enquêter de la seconde UPE. Gray (1973) démontre que cette approche ne biaise pas les probabilités de sélection. Laflamme (2003) fournit des explications concernant ces contraintes.

Cette méthode est appliquée de façon indépendante dans chaque groupe créé par la méthode RHC. Le renouvellement des échantillons n'est toutefois pas fait au même moment. Le renouvellement dans un groupe RHC associé au groupe de renouvellement 1 est fait en janvier et juillet de chaque année. Celui pour les groupes RHC associés au groupe de renouvellement 2 est fait en février et août, etc. En appliquant cette méthode, il est possible de déterminer au lendemain du remaniement la liste des départs qui seront dans l'échantillon de l'EPA pour chaque mois au cours des dix prochaines années.

Il est important de noter que la méthode de renouvellement décrite dans cette section, y compris l'ajustement effectué au nombre de départs associé à la première UPE sélectionnée du groupe, s'applique également lorsque que l'échantillon est sélectionné avec la méthode par sélection systématique avec probabilité proportionnelle à la taille et ordre aléatoire. Gray (1973) en fournit la preuve.

## Chapitre 3 Création et tenue à jour de la base de sondage

### 3.1 Utilisation du registre des adresses

#### 3.1.1 Introduction

Comme défini au chapitre précédent, l'EPA comporte un plan d'échantillonnage à deux degrés avec une base aréolaire au premier degré. Il est donc nécessaire d'obtenir une liste complète des logements au sein de chaque unité primaire d'échantillonnage (UPE) choisie pour sélectionner l'échantillon du second degré. Par le passé, la liste des adresses de logements servant à constituer une base de sondage du second degré était obtenue grâce à un exercice de listage effectué sur le terrain pour chaque UPE sélectionnée. Pour éviter cet exercice coûteux, les adresses sont maintenant obtenues à partir d'une nouvelle source, soit le Registre des adresses (RA), qui est une liste comportant près de 90 % des logements du Canada. Le RA avait le potentiel de réduire les coûts et les délais associés à la création de la base de sondage du second degré et peut-être même d'améliorer la qualité.

La disponibilité d'une liste d'adresses a permis l'élaboration d'un plan d'échantillonnage à un degré plus simple et plus efficace que le plan d'échantillonnage à deux degrés utilisé auparavant dans le cadre de l'EPA. Toutefois, en raison des contraintes de temps et des restrictions budgétaires, on a décidé de maintenir le plan d'échantillonnage à deux degrés pour le moment. Par conséquent, tel qu'il est décrit dans le présent chapitre, une stratégie pour l'utilisation la plus efficace possible du RA dans un contexte de plan d'échantillonnage à deux degrés a été élaborée et mise en œuvre.

#### 3.1.2 Registre des adresses (RA)

Le RA a initialement été créé pour le Recensement de la population du Canada de 1991 et avait pour objectif d'améliorer la couverture du recensement. Il a été créé à l'aide de plusieurs fichiers administratifs comme les fichiers des factures de téléphone et les fichiers des permis de bâtir. Après le recensement, le RA a été mis à jour à l'aide de la liste d'adresses créée durant le processus de dénombrement du recensement. Depuis, le processus est demeuré inchangé : le RA est mis à jour à l'aide de fichiers administratifs avant que le recensement n'ait lieu et il est de nouveau mis à jour ensuite grâce aux renseignements recueillis dans le cadre du recensement. Pour obtenir une description plus détaillée de la façon dont le RA est créé et tenu à jour, veuillez consulter Swain, Drew, Lafrance et Lance (1992).

À l'origine, le RA était conçu pour fournir une liste d'adresses des collectivités dont la population était supérieure à 50 000 habitants et pour assurer sa tenue à jour. La couverture du RA a été élargie après le Recensement de 2001 de manière à inclure les régions moins peuplées. Parce que les listes d'adresses des grandes collectivités sont tenues à jour depuis longtemps, le RA de ces collectivités est plus exact. Au fil du temps, le RA des régions moins peuplées devrait lui aussi se préciser.

Puisque le recensement n'est effectué qu'aux cinq ans, le RA peut se détériorer entre les mises à jour. Bien que la plupart des changements soient causés par la construction de nouveaux logements, d'autres changements résultent de la conversion d'entreprises en logements privés (et vice versa), de la conversion de logements unifamiliaux en appartements, de la démolition de logements, etc.

Malgré ces deux questions de couverture, le taux de couverture global du RA a pu être estimé, à partir d'études postcensitaires, à environ 96 % dans les régions couvertes. D'un autre côté, on savait que le taux de couverture variait en fonction des régions, ce qui est un important facteur à considérer lorsqu'on conçoit une méthodologie pour utiliser le RA dans le contexte de l'EPA.

En 2005, le RA comportait environ 13 millions d'adresses. Le Recensement de 2001 a permis d'indiquer que la majorité de ces adresses étaient des logements résidentiels valides. On a constaté après le Recensement de 2001 que certaines de ces adresses, qui dataient d'avant le Recensement de 2001, étaient invalides. Elles ont quand même été gardées. D'autres adresses du RA sont obtenues grâce aux mises à jour apportées aux fichiers administratifs (en vue du Recensement de 2006).

Pour qu'un logement résidentiel figure dans le RA, une adresse de voirie standard valide doit lui être attribué. Par conséquent, nous pouvons nous attendre à un sous-dénombrement dans les régions rurales où certains logements résidentiels ne possèdent pas une adresse de voirie valide.

Les logements collectifs représentent une autre catégorie de logements disponibles par l'entremise du processus d'extraction du RA, mais ils ne font cependant pas partie du RA. Une liste complète des logements collectifs est dressée au cours du recensement et sert à élaborer le fichier de listage des logements

collectifs (FLLC). Cette liste n'est pas mise à jour au cours du processus régulier de mise à jour du RA et demeure plutôt inchangée entre deux recensements. Les logements collectifs font partie de la population cible de l'EPA et devraient donc être couverts par la base de sondage.

Quand le RA permet de créer une liste d'adresses résidentielles pour une UPE en particulier, les logements collectifs figurant dans le FLLC qui peuvent être associés à la région du RA correspondante sont ajoutés à la liste. La section 3.3 explique plus en détail la définition des logements collectifs et la façon dont ils sont traités.

### 3.1.3 Registre des adresses (RA) et Base nationale de données géographiques (BNDG)

Afin d'utiliser le RA selon un plan d'échantillonnage à deux degrés (tel qu'il sera ultérieurement décrit à la section 3.1.5), nous devons être en mesure d'assigner des adresses à une UPE en particulier. Il existe deux façons d'effectuer cette étape. La première consiste à lier le RA à la Base nationale de données géographiques (BNDG), qui contient le réseau routier de tout le pays, ainsi que les tranches d'adresses pour la plupart des sections de rue, aussi connues sous le nom de côtés d'îlot. Comme c'est le cas avec le RA, l'exactitude de la BNDG peut varier d'une région à une autre. Pour toutes les adresses liées à la BNDG, nous savons, grâce aux tranches d'adresses, sur quel côté d'îlot elles tombent, ce qui permet de les assigner à une UPE en particulier. On les appelle les adresses « structurées ». En ce qui a trait aux adresses qui ne sont pas liées à la BNDG, mais qui étaient valides lors du dernier recensement, nous pouvons avoir recours à un autre lien géographique appelé îlot de recensement, qui correspond à l'îlot urbain auquel les adresses appartenaient en 2001, selon le recenseur. Puisque les UPE avaient comme point de départ des îlots de recensement (comme il a été expliqué au chapitre précédent), nous pouvons également assigner ces adresses à une UPE en particulier. Puisque nous ne savons pas exactement sur quel côté d'îlot ces adresses tombent, nous disons de ces enregistrements qu'ils sont « non structurés ». Au moment du remaniement, environ 86 % de toutes les adresses du RA étaient « structurées ».

L'îlot urbain de 2001 est disponible pour tous les enregistrements valides de 2001, mais pour assurer la cohérence avec la carte de l'UPE qui est préparée à partir de la BNDG, il a été décidé d'utiliser d'abord le lien « côté d'îlot » lorsqu'il était disponible. Il est

important de prendre note que tout enregistrement de croissance, ainsi que toute adresse consignée avant 2001 mais jugée invalide en 2001, ne pouvant être géocodé comme côté d'îlot ne peut être assigné à une UPE en particulier (aucun îlot urbain n'existait pour ces adresses en 2001) et ne peut donc pas être utilisé dans le cadre de l'EPA<sup>19</sup>.

### 3.1.4 Algorithme de mise en séquence

Pour chaque UPE sélectionnée, une liste d'adresses est préparée et vérifiée sur le terrain. Les adresses doivent être classées selon un ordre précis pour faciliter et optimiser l'exercice de listage. Un algorithme a été élaboré de façon à énumérer les côtés d'îlot selon un ordre permettant de couvrir l'UPE en entier tout en minimisant la distance totale que doit parcourir l'intervieweur lorsqu'il procède à la vérification de la liste des adresses. Cet algorithme utilise l'information géographique contenue dans l'UPE, tirée de la BNDG. Puisque l'ordre est défini à partir des côtés d'îlot, il est impossible d'utiliser cet algorithme pour placer les enregistrements non structurés au bon endroit sur la liste. Par conséquent, ils sont ajoutés à la fin de la liste et triés selon le nom de la rue et une adresse de voirie. Pour obtenir plus de renseignements à propos de l'algorithme de mise en séquence et prendre connaissance de la discussion au sujet de ses forces et de ses lacunes, consultez Laflamme et Dochitoui (2005).

La qualité de la liste d'adresses d'une UPE donnée dépend de la qualité du RA, de la qualité de la BNDG et de la qualité du FLLC.

### 3.1.5 Utilisation du RA

Comme il a été mentionné précédemment, en raison des contraintes de temps et des restrictions budgétaires, la possibilité d'utiliser le RA comme base de liste dans un plan d'échantillonnage à un degré a vite été écartée à l'étape de l'élaboration. On a donc plutôt étudié les différentes façons de l'utiliser dans le plan d'échantillonnage à deux degrés traditionnel de l'EPA. Différentes possibilités ont été envisagées et, en fin de compte, la plus prometteuse est la suivante :

1. Selon l'estimation de la qualité du RA de chaque UPE, les UPE sont divisées en trois groupes :

---

19. Une nouvelle étape est présentement mise en œuvre pour assigner à une UPE en particulier un enregistrement de croissance qui n'est pas géocodé comme côté d'îlot, mais qui figure dans une rue contenue dans les limites de l'UPE.

- a. dans le groupe RA 1, où l'on s'attend à ce que la qualité du RA soit excellente, aucun listage préalable n'est effectué sur le terrain et l'échantillon initial des logements est directement sélectionné à partir de la liste du RA correspondant à une UPE;
  - b. dans le groupe RA 2, où l'on s'attend à ce que la qualité du RA soit bonne, une liste préliminaire est créée à partir du RA, puis elle est vérifiée et mise à jour par l'intervieweur sur le terrain durant l'exercice de listage. Le parcours déterminé par l'algorithme de mise en séquence figure sur la carte de l'intervieweur et doit être respecté;
  - c. dans le groupe RA 3, où le RA est de piètre qualité ou n'existe pas, on ne tente pas d'utiliser le RA et un listage traditionnel sur le terrain est effectué. Le parcours déterminé par l'algorithme de mise en séquence figure sur la carte et doit être respecté, comme dans le cas du groupe RA 2.
2. La liste des adresses de chaque UPE, peu importe son groupe RA, est tenue à jour (c'est-à-dire mise à jour) sur le terrain au moins une fois par année, comme cela se faisait auparavant. À l'avenir, si le RA commence à être mis à jour infra-annuellement, il sera possible d'éviter cet exercice sur le terrain dans la plupart des cas.

Cette stratégie vise à utiliser le RA dans la mesure du possible tout en tenant compte que la qualité du RA varie en fonction des régions. Cette stratégie nécessite également qu'une mesure de la qualité du RA soit établie à l'échelle de l'UPE, ce qui ne sera pas facile.

Un essai sur le terrain a été effectué à l'automne 2003 pour évaluer la stratégie proposée et inciter l'élaboration de règles pour assigner des UPE aux groupes RA. Cet essai a démontré, en termes de sous-dénombrement total, que l'utilisation d'un RA non vérifié en tant que base de sondage (comme pour le groupe RA 1) générerait des résultats semblables à ceux obtenus par la méthode traditionnelle de listage. D'un autre côté, le surdénombrement total serait beaucoup plus élevé. Heureusement, après avoir vérifié la liste RA (comme dans le groupe RA 2), le sous-dénombrement n'est plus aussi problématique et est même moins marqué qu'avec la méthode traditionnelle. De plus, le surdénombrement diminue considérablement même s'il semble demeurer un peu plus élevé que le surdénombrement obtenu lorsque la méthode traditionnelle était utilisée. À la lumière de ces résultats, la stratégie proposée a été

acceptée et mise en œuvre tel que décrit à la section suivante. Pour en apprendre davantage sur les essais sur le terrain, consultez Thompson et Turmelle (2004).

Avant que les extractions du RA ne puissent servir sur le terrain aux fins de listage (groupe RA 2) ou de sélection de l'échantillon (groupe RA 1), un certain traitement des données est nécessaire. Il s'agit d'éliminer les enregistrements en double, de transférer certaines adresses d'une UPE à une autre, d'éliminer des unités dans les logements collectifs et de reformater les listes pour qu'elles soient compatibles avec les systèmes de l'EPA. Consultez Gouzi et coll. (2004) pour obtenir une description complète des modifications apportées au RA et des interventions manuelles effectuées.

### 3.1.6 Stratégie de répartition dans les groupes RA

Le principal enjeu de la mise en œuvre de la stratégie de répartition dans les groupes RA consistait à élaborer une méthode visant à assigner adéquatement des UPE aux trois groupes RA. Deux objectifs se faisaient concurrence : réduire le nombre d'UPE nécessitant une vérification sur le terrain (ou maximiser le groupe RA 1) et maximiser la qualité de la base de sondage obtenue. La qualité du RA pour l'UPE, la qualité de l'ordre du listage et le rendement des intervieweurs qui vérifient l'UPE auront des répercussions sur la qualité de la base de sondage. Une recommandation proposée par les intervieweurs ayant participé à l'essai sur le terrain était de minimiser le nombre d'adresses non structurées sur les listes envoyées pour vérification. Les adresses non structurées ne sont pas ordonnées adéquatement (elles figurent à la fin de la liste). S'il y en a plusieurs sur une liste, beaucoup de déplacements motorisés ou à pied seront nécessaires pour les vérifier, rendant ainsi l'exercice de listage plutôt fastidieux et susceptible de laisser passer des erreurs. Toutes ces préoccupations ont été prises en considération lors de l'élaboration de la stratégie.

La stratégie de répartition initiale de l'UPE a été élaborée à l'aide des résultats obtenus lors de l'essai sur le terrain ainsi que des résultats du Test du recensement de 2004. À partir de ces deux tests, un ensemble de caractéristiques importantes ont été définies, puis ont servi à établir une série de règles permettant d'assigner les UPE aux groupes RA. Les principales caractéristiques sont :

- le nombre de logements non structurés dans l'UPE et dans quelle mesure ils sont dispersés (cent maisons unifamiliales sur quinze rues

différentes causent davantage de problèmes qu'un gros immeuble d'appartements);

- la couverture du RA après le Recensement de 2001;
- le nombre d'îlots de recensement dans l'UPE (puisque l'algorithme de mise en séquence ne fonctionne pas toujours bien lorsqu'il y a un grand nombre d'îlots de recensement);
- le nombre d'enregistrements de croissance ajoutés depuis le dernier recensement;
- le nombre d'immeubles à logements multiples de deux ou trois logements (cette catégorie de logements est connue pour contenir un sur-dénombrement considérable et devrait donc être vérifiée).

Voici les critères ayant servi à classer les UPE dans les trois groupes RA :

- Groupe RA 3

S'il y a (a) trop de logements non structurés ou qu'ils sont trop dispersés **ou** (b) que la couverture du RA en 2001 était inférieure à 90 % **ou** (c) que le nombre d'îlots de recensement dans l'UPE était supérieur à 20, il faut alors classer l'UPE dans le groupe RA 3. L'utilisation du RA pour ces UPE a été jugée inadéquate.

- Groupe RA 1

Par ailleurs, si (a) la couverture du RA en 2001 variait entre 97,75 % et 103 % **et** (b) que le nombre d'enregistrements non structurés est très faible ou qu'ils sont regroupés **et** (c) que le nombre d'immeubles à logements multiples de deux ou trois logements est assez petit **et** (d) que le nombre d'enregistrements ajoutés depuis le dernier recensement est très restreint **et** (e) que le nombre de logements collectifs est inférieur à cinq (avec au plus un logement collectif non structuré), il faut alors classer l'UPE dans le groupe RA 1. Il s'agit de classer dans le groupe RA 1 uniquement les UPE stables à faible risque.

- Groupe RA 2

Autrement, il faut classer l'UPE dans le groupe RA 2.

Ces règles ont été appliquées à presque toutes les UPE. Une intervention manuelle n'a été nécessaire que pour quelques UPE problématiques. En fin de compte, la répartition initiale a permis d'assigner environ 39 % des UPE échantillonnées au groupe RA 1, 24 % au

groupe RA 2 et 37 % au groupe RA 3. C'est cette répartition qui a été utilisée dans la production quand les exercices de listage ont débuté à la fin de l'été 2004.

Après quelques semaines de listage, nous avons reçu certains commentaires plutôt négatifs de la part des intervieweurs sur le terrain. Une plainte fréquente était que le réseau routier présenté sur la carte de l'UPE était parfois désuet, particulièrement dans les secteurs en forte expansion. Une autre plainte concernait l'algorithme de mise en séquence. En effet, le parcours n'avait parfois pas de sens aux yeux de l'intervieweur ou ne semblait pas du tout idéal. Ces deux problèmes rendaient l'exercice de listage déroutant et pénible, ce qui avait pour conséquence d'augmenter le temps accordé et les sommes consacrées au listage des UPE problématiques, et également d'accroître les frustrations du personnel sur le terrain à l'égard de la nouvelle technologie.

Puisque rien ne pouvait être fait à ce moment pour améliorer la qualité de présentation du réseau routier ou l'optimisation de la mise en séquence, le problème a été résolu en modifiant la méthodologie sur le terrain. Pour les cas extrêmement problématiques, il a été décidé que les intervieweurs pourraient procéder au listage en utilisant la manière traditionnelle, c'est-à-dire :

- pour les UPE du groupe RA 3, ne pas tenir compte du parcours prédéterminé qui figure sur la carte et se servir de son expérience et de son jugement pour trouver le parcours idéal;
- pour les UPE du groupe RA 2, ne pas tenir compte de la liste RA et le parcours prédéterminé et recommencer du début, comme pour les UPE problématiques du groupe RA 3. Il est simple de ne pas tenir compte de la liste RA dans la liste de départ puisque tous les logements sur la liste peuvent simplement être supprimés.

Puisque le nouvel échantillon de l'EPA a été introduit graduellement, un groupe de renouvellement à la fois et sur une période de six mois, l'exercice de listage a également été fait avec un groupe de renouvellement à la fois. Nous avons ainsi eu l'occasion de revoir les règles servant à assigner les UPE aux groupes RA en fonction des nouveaux renseignements obtenus du premier groupe de rotation (pour les autres groupes de rotation). Nous avons minutieusement examiné plusieurs UPE problématiques qui avaient déjà été répertoriées et avons défini certaines caractéristiques communes qui n'avaient pas encore été utilisées, mais qui pouvaient aider à déterminer les UPE problématiques. Les principales caractéristiques qui semblaient

être liées aux UPE problématiques étaient : un important accroissement (ou une importante réduction) de la longueur du réseau routier et l'ajout de nouvelles rues, et ce, depuis le dernier recensement. Nous nous sommes servis de ces caractéristiques pour adapter les règles et avons également resserré certaines des règles préliminaires. Finalement, nous avons transféré plus de trois cents UPE échantillonnées des groupes RA 1 et 2 au groupe RA 3. Thompson et Rodrique (2005) propose une description complète du processus d'élaboration des règles ainsi qu'une description du travail manuel ayant dû être fait pour assigner certaines UPE.

La répartition finale des UPE dans les groupes RA, à l'échelle de la population et à l'échelle de l'échantillon, est présentée aux tableaux 3.1 et 3.2.

La répartition parmi les groupes RA varie considérablement d'une province à une autre. Le RA n'est pas aussi utile dans les provinces de l'Atlantique (les quatre premières lignes) que dans les autres provinces, particulièrement en Ontario, au Manitoba et en Colombie-Britannique, où 40 % des UPE échantillonnées font partie du groupe RA 1.

Cette répartition deviendra dynamique au fil du temps. Au fur et à mesure que les UPE sont introduites dans l'échantillon puis en sont éliminées et que la qualité du RA évolue (particulièrement après le Recensement de 2006), la répartition des groupes RA changera probablement aussi.

## 3.2 Cartographie

### 3.2.1 Introduction

Le plan de l'EPA tire parti du fait que, depuis 2001, les données sur le pays tout entier sont accessibles dans un Système d'information géographique (SIG). Les 10 provinces ont été divisées en UPE à l'aide d'un Système automatisé de regroupement des territoires (SARTE) modifié, comme il est décrit à la section 2.4. Une fois effectuée la délimitation des UPE, y compris les interventions manuelles, un certain nombre de nouvelles caractéristiques ont été incorporées pour produire des cartes des UPE.

- Les limites des UPE ont été ajoutées à l'entrepôt des données géographiques de la Division de la géographie, rendant ainsi les limites de l'EPA accessibles aux applications logicielles du SIG.

**Tableau 3.1 Répartition des UPE dans les groupes Registre des adresses**

Province	Groupe 1		Groupe 2		Groupe 3		Total nombre
	nombre	%	nombre	%	nombre	%	
Terre-Neuve et Labrador	100	10	164	17	729	73	993
Île-du-Prince-Édouard	48	17	45	16	193	67	286
Nouvelle-Écosse	357	21	324	19	1 002	60	1 683
Nouveau-Brunswick	252	18	200	14	964	68	1 416
Québec	5 824	40	3 366	23	5 465	37	14 655
Ontario	9 276	46	4 560	23	6 131	31	19 967
Manitoba	1 108	52	237	11	771	37	2 116
Saskatchewan	618	31	269	13	1 121	56	2 008
Alberta	2 250	42	1 129	21	1 973	37	5 352
Colombie-Britannique	2 945	40	2 269	31	2 122	29	7 336
Total	22 778	41	12 563	22	20 471	37	55 812

Voir annexe A.2 pour les abréviations.

**Tableau 3.2 Répartition des UPE échantillonnées dans les groupes Registre des adresses**

Province	Groupe 1		Groupe 2		Groupe 3		Total nombre
	nombre	%	nombre	%	nombre	%	
Terre-Neuve et Labrador	27	12	38	17	157	71	222
Île-du-Prince-Édouard	25	20	21	17	80	63	126
Nouvelle-Écosse	88	26	59	17	197	57	344
Nouveau-Brunswick	58	18	45	14	217	68	320
Québec	429	33	314	23	553	43	1 296
Ontario	935	46	459	22	644	32	2 038
Manitoba	215	51	49	11	162	38	426
Saskatchewan	151	35	80	18	207	47	438
Alberta	213	38	125	23	220	39	558
Colombie-Britannique	333	41	242	30	232	29	807
Total	2,474	38	1,432	22	2,669	40	6,575

Voir annexe A.2 pour les abréviations.

- Comme il est expliqué à la prochaine section, la Division de la géographie a réalisé suffisamment de cartes de différentes tailles et de médaillons automatisés pour pouvoir dresser la carte de chaque UPE et indiquer chaque petit détail.
- Le Registre des adresses (RA), avec le concours de la Division de la géographie, a établi l'ordre des îlots, des côtés d'îlot dans les îlots et des adresses dans les côtés d'îlot pour chaque UPE couverte par le RA.
- L'ordre des numéros d'îlot et le point de départ optimal aux fins de listage dans chaque îlot ont été ajoutés aux couches successives d'information des cartes.
- La Division de la géographie a développé une application de cartographie pour créer et imprimer des cartes des UPE et fureter dans celles-ci.

La stratification et la sélection initiale de l'UPE ont généré plus de 6 500 UPE pour la mise en place de ce plan d'échantillonnage. Une fois le plan d'échantillonnage mis en place, le remplacement d'une UPE génère de 30 à 60 UPE supplémentaires par mois. Le logiciel du SIG, soit l'Arc/Info, peut être utilisé pour explorer des cartes et les imprimer dans une grande variété de formats et de tailles. Sur le terrain, on appelle ces cartes des *schémas de la grappe F01*. Les intervieweurs s'en servent pour lister les logements faisant partie des UPE.

Les étapes d'identification, de cartographie, d'impression, de vérification et de mise à jour de ces UPE, et de transmission aux bureaux régionaux sont décrites ci-après.

### 3.2.2 Logiciel cartographique de l'EPA

La Division de la géographie a créé une application Arc/Info appelée le logiciel cartographique de l'EPA pour générer des *schémas de la grappe F01* dans un format Arc/Info MXD et en Adobe PDF pour chacune des UPE sélectionnées. Les fichiers Arc/Info MXD permettent la manipulation manuelle de chacune des UPE, au besoin. Les données de sortie du logiciel Adobe PDF sont utilisées aux fins d'impression et d'affichage. Le logiciel cartographique de l'EPA utilise des données géographiques, y compris le réseau routier, la géographie du recensement, différents traits physiques et les données sur les limites de l'UPE dans le cadre de l'EPA, données qui sont enregistrées sur un serveur central dans la Division de la géographie. Pour obtenir des renseignements supplémentaires, veuillez consulter Cillis (2004a).

La carte principale de chaque UPE peut être générée dans l'un des trois formats suivants : 11 po x 17 po, 17 po x 22 po ou 22 po x 34 po. Pour toutes les régions qui ne sont pas assez détaillées sur la carte principale, des médaillons sont générés sur les cartes de taille 11 po x 17 po. La plupart des UPE n'ont pas de médaillon, mais certaines UPE en régions rurales qui couvrent de grandes étendues de terre peuvent avoir 10 médaillons et plus. Il y a naturellement une certaine influence réciproque entre le nombre de médaillons et la taille de la carte. En effet, les grandes cartes n'ont pas besoin d'autant de médaillons. En moyenne, chaque UPE requiert près de 0,6 médaillon, ce qui génère plus de 10 000 cartes pour les 6 500 UPE.

Une base de données contenant les UPE, les tailles de cartes et les coordonnées des médaillons détermine l'ensemble des cartes à produire pour chaque UPE en cours. Le tableau de l'UPE contient tous les paramètres d'échantillonnage qui doivent être imprimés dans la légende de la carte. Un indicateur peut être réglé pour permettre le traitement par lots d'un ensemble d'UPE d'un mois au complet. Le tableau de l'UPE est mis à jour tous les mois en remplaçant tous les nouveaux ensembles d'UPE.

Une interface pour l'utilisateur pilotée par menus offre une variété de fonctions (voir Cillis, 2004b). En résumé :

- Les fichiers MXD peuvent être édités;
- Les documents en format PDF peuvent être imprimés;
- Les données saisies dans les tableaux pour une UPE peuvent être modifiées pour une taille de carte donnée, et des médaillons peuvent être créés ou supprimés;
- La création de nouveaux fichiers MXD ou PDF peut être contrôlée par la date d'introduction, pour les nouvelles UPE, ou individuellement, par UPE.

Des fonctions supplémentaires peuvent être ajoutées au logiciel cartographique de l'EPA, notamment un indicateur pour désactiver automatiquement les points de départ et les numéros d'îlot si la mise en ordre des îlots n'est pas nécessaire, et l'utilisation de couleurs plutôt qu'une échelle de gris.

### 3.2.3 Système de conception de l'échantillon

Pour compléter le logiciel cartographique de l'EPA, des améliorations ont été apportées au Système de conception de l'échantillon. Un système de contrôle des

cartes des UPE gère les milliers de cartes devant être traitées. Un système d'impression automatisé permet le traitement par lots d'importants volumes d'imprimés PDF à partir de multiples imprimantes. Un ensemble d'interfaces assure le transfert de données listées, tirées du Registre des adresses, au système de listage des échantillons. Une copie de travail de chaque carte est générée aux fins de la vérification de la qualité des données du RA, cette étape de la vérification est décrite à la section 3.1.

### 3.2.4 Vérification des cartes

La vérification administrative du *schéma de la grappe FOI* doit être effectuée pour ajuster la taille de la carte, supprimer ou créer des médaillons, ajouter les noms de rue manquants, indiquer les numéros de voirie, le cas échéant, inclure des descriptions quand de grandes UPE sont divisées, enlever des numéros d'îlot et des points de départ, le cas échéant, et inscrire des remarques relatives à la vérification de la qualité des données du RA.

La taille des cartes et les limites des médaillons sont générées automatiquement, mais elles ne sont peut-être pas toujours pertinentes. Une limite de sous-îlot est attribuée à beaucoup de logements collectifs et d'immeubles d'appartements, ce qui d'habitude génère un médaillon. Les sous-îlots sont un artéfact du Programme des îlots de recensement qui tente d'isoler les logements spéciaux. L'EPA n'est pas concernée par ces sous-îlots collectifs et le médaillon est généralement supprimé, à moins qu'il ne forme une limite de l'UPE. Le nombre de médaillons peut souvent être réduit simplement en augmentant la taille de la carte. Une fois que la taille de la carte et les médaillons sont réglés, la carte peut alors être recrée à l'aide du logiciel cartographique de l'EPA. Un aperçu des étapes détaillées de la vérification administrative est donné par Lindeyer (2004).

Dans certaines situations, les numéros d'îlot et les points de départ sont éliminés de la carte. En régions rurales, le nombre d'îlots est parfois excessif, ce qui rend plus difficile la recherche de l'ordre approprié des îlots par le programme de mise en ordre automatisé. Il est arrivé dans de rares cas que la géographie de l'UPE altère l'ordre des numéros d'îlot. Il n'était possible de supprimer des numéros d'îlot et des points de départ que dans le groupe RA 3 puisque les deux autres groupes nécessitaient un ordre visible pour être liés à la liste du Registre des adresses.

Les noms de rue sont indiqués sur le schéma à l'aide des procédures par défaut de l'application Arc/Info. Il

arrive parfois qu'une rue importante n'ait pas de nom. À l'occasion, d'autres sources sont utilisées pour ajouter le nom manuellement.

Les numéros de voirie ne sont généralement pas utilisés sur les cartes. Dans le cadre de l'EPA, procéder au listage sans l'aide des numéros de voirie est une tradition, en partie parce que cette façon de faire favorise un examen plus approfondi de la région et en partie en raison de la mauvaise qualité des numéros de voirie dans certaines régions. Les numéros de voirie sont ajoutés manuellement, au besoin, pour indiquer une limite imaginaire qui traverse une rue. Les limites imaginaires supposent qu'il n'existe aucun trait physique pour délimiter l'UPE. Dans un tel cas, le numéro de voirie sert à délimiter l'UPE. Cependant, dans bon nombre de cas, le numéro de voirie n'est pas disponible. L'intervieweur doit l'estimer en fonction de la distance parcourue ou il peut l'estimer parce qu'il connaît la limite municipale, par exemple.

La vérification de la qualité des données du Registre des adresses permet de détecter les situations irrégulières qui pourraient nécessiter l'ajout de remarques supplémentaires sur la carte. Pour certains grands immeubles d'appartements, la description du contenu est inscrite sur le diagramme en tant qu'inclusions ou exclusions, selon le cas. Comme il est indiqué dans la section ci-après, lorsque de grandes UPE sont divisées, des annotations sont également nécessaires.

Dans des cas exceptionnels, l'UPE de forme irrégulière générée par la conciliation au dernier réseau routier doit être modifiée manuellement pour obtenir une limite facilement repérable et un compte sur les ménages semblable au plan original de l'UPE.

### 3.2.5 Division des grandes UPE

En formant des UPE, on tentait d'obtenir des secteurs de 200 ménages. En règle générale, le programme SARTe a généré une ou plusieurs UPE affichant un grand nombre de ménages dans chaque unité de travail. Quand on travaille avec des îlots de recensement comme unité d'immeuble pour la formation des UPE, certains îlots sont gros au début. Le nombre dépasse parfois 450 ménages; il est alors considéré comme étant trop élevé pour une seule UPE. De plus, les mises à jour du Registre des adresses génèrent un certain nombre d'UPE dont la croissance est considérable; le nombre de ménages est alors mis à jour. Certaines UPE dépassaient également la limite de 450 ménages. Pour cette raison, l'unité délimitée par le SARTe est appelée le géocode\_EPA. Ce n'est qu'après la division que les unités prennent le nom d'UPE. Les grands

géocodes\_EPA sont divisés en deux UPE ou plus avant que l'échantillon ne soit choisi. La plupart des géocodes\_EPA ne sont associés qu'à une UPE. Seulement 1,5 % des géocodes\_EPA ont besoin d'être divisés. Trois sortes de division sont générées comme suit. Les deux premières sortes de division fournissent un nouveau compte sur les ménages pour chaque composante d'UPE.

La division la plus simple, c'est lorsqu'une adresse de voirie, en général un immeuble d'appartements, comporte, par hasard, environ la moitié des logements. Une UPE n'est constituée que de l'adresse de l'appartement, tandis que l'autre UPE se compose de ce qui reste du géocode\_EPA, à l'exception de l'adresse de l'appartement. Il est à noter que le logiciel cartographique de l'EPA automatisé ne pouvait que dresser la carte du géocode\_EPA original. Le *schéma de la grappe F01* est annoté manuellement en vue de clarifier quelle UPE doit être listée. Il est rare d'utiliser plus d'une adresse d'appartement pour diviser le géocode\_EPA en trois UPE. Au total, 436 géocodes\_EPA sont divisés par adresse de voirie avant qu'un échantillon ne soit choisi; 24 d'entre eux sont divisés en trois UPE.

La deuxième division la plus simple utilise une rue sur toute sa longueur et s'appelle une division d'un côté d'îlot. Une UPE se compose de tous les logements ayant une adresse sur cette rue et l'autre, de tous les logements n'ayant pas d'adresse sur cette rue. Pour certaines UPE, la rue offrait une multitude d'intersections, forçant ainsi le tracé d'une limite irrégulière sur la carte. Les annotations manuelles indiquaient clairement toutes les adresses sur la rue. Quelque 260 géocodes\_EPA ont été divisés par côté d'îlot. Ils ont tous été divisés en deux UPE.

Le dernier mode de division ne peut être généré de manière automatisée comme c'est le cas pour les deux premiers modes de division. Le compte total des ménages est divisé également par le nombre de divisions, sans savoir s'il est ou non possible de générer une telle division. En choisissant la première de ces divisions, une division manuelle des îlots ou des côtés d'îlot est effectuée et conservée pour pouvoir ultérieurement introduire les autres UPE dans le même géocode\_EPA. Bien que ces divisions « virtuelles » aient comme objectif premier la division égale des ménages, une divergence pouvant atteindre 20 % est permise au sein des procédures administratives. Seuls 111 géocodes\_EPA doivent être divisés virtuellement, mais 12 d'entre eux doivent avoir entre 5 et 14 UPE chacun.

L'annotation manuelle de la description des divisions sur les cartes des UPE représente une importante mesure de la qualité lorsqu'on veut s'assurer que la région appropriée est listée.

### 3.2.6 Envoi aux bureaux régionaux

Il est nécessaire d'établir un calendrier serré des envois afin de transmettre à temps tous les *schémas de la grappe F01* en vue de l'exercice de listage. La section suivante offre une description du listage et de la tenue à jour du listage. Tous les groupes RA sont envoyés ensemble bien que les UPE du groupe RA 1 ne soient pas requis avant que la sélection des logements ne soit faite.

En ce qui concerne l'important afflux de cartes lié au plan de 2004, les envois aux bureaux régionaux ont été effectués entre juin et novembre 2004 et le listage a débuté en juillet de la même année. Les cartes de plus de 6 500 UPE ont été dressées et plus de 30 000 cartes individuelles ont été produites, y compris des médaillons et des copies. Deux copies ont été envoyées, une à l'intention du gestionnaire régional de programme et l'autre pour l'intervieweur. Vu le nombre de remplacements d'UPE plus faible entre les remaniements, une troisième copie est générée à l'intention de l'intervieweur principal.

### 3.2.7 Rétroaction des bureaux régionaux

Puisque l'introduction de la liste du RA au sein des opérations sur le terrain a nécessité que les tâches de l'intervieweur soient modifiées de façon substantielle, beaucoup de commentaires ont été demandés et reçus. Malgré les essais préliminaires sur les listes générées par le RA, un certain nombre de recommandations ont été reçues et mises en œuvre.

L'attribution des UPE aux groupes RA a été perfectionnée à la lumière des difficultés rencontrées par certains intervieweurs relativement à l'ordre des îlots suggéré sur la carte. En ce qui a trait aux UPE du groupe RA 2, l'intervieweur est autorisé à supprimer tous les logements listés dans le RA si l'ordre porte complètement à confusion. Dans ce cas, l'ordre des îlots sur la carte peut également être ignoré.

La région des Prairies a remarqué que les désignations canton, rang et méridien ne figuraient pas sur ses cartes. Plusieurs gestionnaires ont demandé d'avoir des cartes en couleurs et des courbes de niveau pour les aider à trouver les repères terrestres. Des couleurs ont été ajoutées aux conditions du logiciel cartographique de l'EPA.

La qualité des données sur les réseaux routiers est devenue un problème dans beaucoup de régions. D'anciennes routes et d'anciens chemins de fer qui n'existent plus depuis de nombreuses années figurent toujours dans la base de données géographiques. D'autres routes sont absentes de la base de données bien qu'elles ne soient pas très récentes.

Le SARTe fonctionnait à partir d'un instantané du réseau routier tiré du Recensement de 2001. Ce réseau routier a été mis à jour au début de 2004, ce qui nécessite un effort de conciliation lorsqu'un changement dans les rues affecte également la limite d'une UPE. Environ une douzaine d'UPE ont une forme irrégulière ou ne sont pas contiguës en raison des imperfections de cette conciliation.

Des mises à jour futures du réseau routier nécessiteront également une conciliation qui pourrait modifier les cartes actives.

### **3.2.8 Registre des adresses et mises à jour du réseau routier**

Il est prévu d'améliorer la liste régulièrement en se conformant au plus récent Registre des adresses. Le RA a été utilisé pour le Recensement de 2006; les données géographiques du Recensement de 2006 sont donc utilisées, y compris les îlots requis pour la délimitation des UPE. Une nouvelle conciliation est nécessaire pour mettre à jour les limites actuelles des UPE en fonction de la nouvelle géographie. Environ 400 UPE présentent des différences considérables entre l'ancienne et la nouvelle géographie si la limite doit respecter les nouveaux îlots. La liste de quelques d'UPE actives doit peut-être être mise à jour pour refléter les changements apportés à la région devant faire l'objet d'un listage.

Ce n'est que lorsque la conciliation sera terminée que les étapes pour l'utilisation du plus récent RA pourront être entreprises. La mise en ordre des îlots dans une UPE, des côtés d'îlot dans un îlot et des logements du RA dans un côté d'îlot doit être refaite. Une nouvelle extraction est ensuite effectuée, suivie d'une analyse subséquente pour l'identification du groupe RA. La taille de la carte ne changera probablement pas, mais il pourrait être nécessaire d'obtenir de nouvelles coordonnées pour les médaillons en raison d'un nouveau réseau routier dans certains secteurs. Toute mise à jour du RA pourrait passer par ces étapes.

## **3.3 Listage et tenue de liste**

### **3.3.1 Introduction**

Le listage consiste à localiser et à saisir tous les logements inclus dans les limites de l'UPE, comme on

peut les voir dans le *Schéma de la grappe F01*. L'interprétation correcte des détails de la carte relativement aux traits physiques des lieux s'avère primordiale. L'intervieweur sur le terrain saisit les adresses ou les descriptions des logements avec l'application de listage de l'EPA installée sur son ordinateur portable. Le manuel de l'intervieweur aux IPAO EPA-100 (2006) donne des instructions sur le listage et sur l'utilisation de l'application de listage. Le manuel a nécessité des mises à jour importantes dictées par l'utilisation du Registre des adresses (RA).

La liste des logements est non limitative. Durant la vie de l'UPE, l'intervieweur peut y ajouter régulièrement de nouveaux logements, à mesure que la population augmente. Si le caractère non limitatif de la liste permet une sélection additionnelle de logements sur le terrain, il peut également exiger un traitement particulier si la croissance devient un fardeau pour l'intervieweur.

Le listage s'effectue différemment selon le groupe RA, tel qu'il est précisé à la section 3.1. Cette section porte sur le traitement par groupe RA, les différences entre le listage initial et la tenue de liste, de même que sur le traitement des logements collectifs, les choix possibles dans les zones où il y a croissance de logements et le programme en place pour contrôler la qualité du listage.

### **3.3.2 Traitement par groupe RA**

Les UPE du groupe RA 3 n'ont aucun logement inscrit au RA. L'intervieweur commence avec un listage vierge et doit localiser tous les logements composant l'UPE.

La plupart des logements des UPE du groupe RA 2 proviennent d'une liste produite à partir du RA. Dans une vérification de liste, on doit confirmer que ces logements appartiennent à l'UPE, on supprime les logements inscrits par erreur, on ajoute les logements sur le terrain qui n'apparaissent pas sur la liste et on corrige les adresses inexactes, en particulier quand le nombre d'unités est vague ou fictif.

Presque tous les logements des UPE du groupe RA 1 proviennent d'une liste produite à partir du RA. Contrairement aux UPE du groupe 2, celles du groupe 1 sont exemptées de la vérification du listage avant la sélection des logements, *c'est-à-dire* que la liste du RA sert à la sélection initiale des logements sans profiter des mises à jour sur place. Une fois les interviews commencées, les mises à jour du listage devraient permettre de corriger toute divergence entre la liste du RA et les logements réels inclus dans l'UPE.

Le groupe RA dicte uniquement la méthode de listage initiale. Une fois la liste vérifiée sur le terrain, le groupe RA original n'a plus d'importance. Présentement il n'y a aucune rétroaction au Registre des adresses pour quelque changement apporté à une liste produite à partir du RA. Diverses études permettraient de traiter différemment les différences constatées.

### 3.3.3 Listage de départ

Le listage de départ consiste à dresser une liste des logements pour la première sélection d'échantillon dans une UPE. La création d'une liste d'un groupe RA 1 transmise directement pour la sélection de l'échantillon fait partie du processus de listage de départ, bien qu'il n'y ait aucune vérification sur place.

Les UPE attribuées au groupe RA 3 font l'objet d'un listage de départ sur le terrain. Les UPE attribuées au groupe RA 2 sont transmises sur le terrain pour la vérification du listage de départ. On ne peut faire aucune sélection d'échantillon avant que l'intervieweur sur place ne retourne la liste mise à jour pour ajout à la base de données du bureau central.

L'application de listage de l'EPA traite le listage de départ et la tenue de liste de façon très différente. Au cours du listage de départ, les intervieweurs sur place peuvent ajouter, supprimer et déplacer très facilement les logements dans la liste. Les lignes d'adresses sont renumérotées à chacun des changements apportés à la liste. À cette étape, aucun logement n'affiche de données de sélection; par conséquent, le renumérotage n'a aucun effet sur l'historique de la sélection de l'échantillon.

On trouvera de plus amples détails dans un guide d'autoapprentissage (EPA Listage initial pour les grappes du registre des adresses 2004) créé afin d'aider les intervieweurs à passer à la nouvelle méthode de listage des UPE attribuées au groupe RA 2. On a intégré le contenu du guide d'autoapprentissage au Manuel de l'intervieweur aux IPAO EPA-100 (2006). Il faudra consulter le manuel pour prendre connaissance de tous les détails sur le listage.

### 3.3.4 Traitement des logements collectifs

Le listage des logements collectifs n'est pas aussi évident que celui des logements privés. En règle générale, le personnel du bureau central indique le nombre voulu de logements à inscrire sur la liste pour un logement collectif. En raison des effectifs limités et du volume important des UPE soumis chaque mois au listage durant le remaniement, on a confié aux intervieweurs sur le terrain la responsabilité de décider

eux-mêmes du nombre de logements à inscrire sur la liste pour chaque logement collectif, selon un ensemble de critères préétablis (voir ci-dessous). Les recommandations par type de logement collectif devraient s'appliquer dans 99 % des cas.

Deux critères principaux régissent le listage des logements collectifs. Selon le premier, les pensionnaires des établissements ne font pas partie de la population visée dans l'EPA. Il faut d'abord déterminer si l'établissement compte des résidents qui s'y trouvent contre leur gré (p. ex. prisons) ou à cause d'une infirmité (par exemple maisons de repos). En règle générale, on n'inscrirait sur la liste que la résidence du propriétaire et s'il y a lieu, la résidence du personnel. Les nouveaux types de logements collectifs pour aînés proposent un mélange complexe d'appartements simples, de chambres avec repas fournis, de soins infirmiers à plein temps et des soins palliatifs. Dans les cas plus complexes, il faudra s'en remettre au bureau central pour déterminer la bonne procédure de listage.

Le second grand critère a trait à la probabilité de joindre un répondant ou plus n'ayant aucun lieu de résidence habituel ailleurs. L'enquête tient compte des motels ayant des locataires à long terme, des maisons de pension, des résidences du personnel des divers terrains de camping, des auberges et des hôpitaux. Quand plusieurs chambreurs partagent des aires communes, on peut indiquer trois chambres ou plus sur une ligne de listage.

### 3.3.5 Tenue de liste

Après l'échantillonnage d'une UPE jusqu'au niveau du logement, la tenue de liste consiste à effectuer régulièrement des mises à jour de la liste d'adresses. Avant d'obtenir le premier échantillon de logements, on aura créé et possiblement vérifié sur le terrain une liste initiale. Après l'échantillonnage, les règles pour les mises à jour du listage deviennent plus strictes afin de maintenir le dossier historique de la sélection de l'échantillon. On ne peut pas supprimer un logement. On ne peut que le désactiver en indiquant les motifs de la désactivation dans la description du logement. On peut déplacer les logements, mais la ligne de listage n'est pas mise à jour. On tient plutôt une séquence d'impression distincte afin d'enregistrer l'ordre de listage privilégié par l'intervieweur. Afin de garantir que l'intervieweur apporte un soin particulier à la tenue d'une liste, l'application de listage nécessite l'utilisation de touches additionnelles pour lancer et confirmer les mises à jour du listage.

Comme la liste des logements collectifs de l'UPE est ouverte, on y fait des ajouts à mesure qu'il se construit de nouveaux logements. On sélectionne une partie des nouveaux logements au moyen de l'application de listage pendant la semaine des interviews, en utilisant le taux d'échantillonnage de l'UPE. Les logements sélectionnés par les intervieweurs (LSI) génèrent des cas à traiter pour l'intervieweur utilisant l'application IPAO.

On crée deux formes de LSI pendant la tenue de liste. D'abord, les nouvelles structures ajoutées en fin de liste sont échantillonnées au moyen du taux d'échantillonnage inverse et à partir de la prochaine ligne d'interviews provenant de la dernière sélection de l'échantillon dans l'UPE. Une fois qu'on a sélectionné un logement, on passe à la prochaine ligne d'interviews en appliquant le taux d'échantillonnage inverse.

On désigne la seconde forme de LSI par le terme « multiples ». Au cours de l'interview dans un logement sélectionné, l'intervieweur peut déterminer que la structure contient des logements distincts, généralement des unités de sous-sol ou d'étage pas facilement visibles de la rue. Si la liste des logements n'indique pas ces unités additionnelles sous forme de lignes distinctes, on peut avoir raté l'occasion de sélectionner ces logements pendant la vie de l'UPE. Afin de compenser les occasions ratées ainsi et dans d'autres cas semblables non résolus, on sélectionne toutes les unités omises avec le logement d'origine. Puis, on les ajoute à la liste comme multiples du logement sélectionné à l'origine et l'application génère des cas pour chaque multiple.

On procède habituellement à la tenue de liste une ou deux fois par année pour chaque UPE, selon son potentiel de croissance et la disponibilité de l'intervieweur. Afin de réduire les coûts, on procède, en règle générale, à la tenue de liste pendant le mois initial de l'UPE, le mois où s'effectue le renouvellement des logements et une communication personnelle s'impose pour au moins un certain nombre des logements nouvellement sélectionnés.

### **Tenue de liste initiale du groupe RA 1**

On doit apporter un soin particulier à la tenue de liste du groupe RA 1. La tenue de liste initiale de ces UPE correspond aussi à la première fois qu'un intervieweur a accès à la liste produite à partir du RA aux fins de mise à jour.

En règle générale, on s'efforce de terminer les interviews tôt dans la première semaine d'interviews. À tout le moins, il faut aussi vérifier pendant le listage les logements sélectionnés des UPE attribués au groupe RA 1. Sinon, on risque de ne pas déceler les erreurs de

listage qui ont une incidence sur la sélection de l'échantillon. Les adresses comptant plusieurs unités nécessitent une attention particulière. Il faut vérifier la structure entière afin de déterminer s'il y a eu ajout d'une unité inexistante ou plus à partir du RA.

On trouvera de plus amples détails dans le guide d'autoapprentissage (EPA Manuel de l'intervieweur des grappes du groupe RA 1 2004) préparé à l'intention de tous les intervieweurs sur le terrain pour les aider à tenir compte des exigences particulières de la tenue de liste dans le nouveau plan, en particulier pour le traitement des UPE du groupe RA 1. On a mis à jour le Manuel de l'intervieweur aux IPAO EPA-100 (2006) afin d'y inclure ces nouvelles caractéristiques.

### **3.3.6 Traitement des zones en croissance**

Comme les listes de logements de l'UPE sont ouvertes, une croissance extrême est possible. La capacité de l'intervieweur de tenir des listes élargies a ses limites, entre autres : la limite physique de 999 listages de l'application de listage, le coût associé à la tenue de listes élargies et une contrainte temporelle pour procéder aux interviews face à l'afflux important d'échantillons nouveaux que produit une liste élargie. Pour désigner les méthodes servant à réduire le nombre d'interviews à effectuer, on parle du sous-échantillonnage de l'UPE.

#### **UPE fictives**

Si le seul problème posé par la croissance d'une liste des logements d'une UPE se résume à l'atteinte de la limite physique de 999 listages, on génère une UPE fictive afin d'intégrer les logements qui restent. D'ordinaire, la valeur de la grappe dans le numéro d'identification de l'échantillon passe de la valeur existante à 999. Au besoin, on générera une seconde UPE fictive, dont la valeur sera de 998.

D'habitude, l'UPE en cause se trouve dans une zone où l'on trouve un fort pourcentage de logements vacants. Tous les logements figurent sur la liste et le nombre d'interviews se situe dans la norme pour une UPE. Aucun sous-échantillonnage de l'UPE n'est nécessaire et pourtant, la tenue d'une liste élargie pourrait encore causer des problèmes étant donné le grand nombre de logements vacants.

La modification de l'application afin de permettre plus de 999 listages exigerait un nouveau changement dans le numéro d'identification de l'échantillon des UPE de l'EPA et des changements subséquents dans chacun des systèmes qui utilisent ce numéro d'identification de l'échantillon.

Dans les faits, le nombre 999 ne représente pas la limite véritable. Afin de conserver un nombre égal de logements, quel que soit le point de départ dans la sélection des logements, on a fixé la limite du listage à un multiple du taux d'échantillonnage inverse des logements ( $I_c$ ) :

$$\text{Dernière liste} = \text{int}(999/I_c) * I_c.$$

Cette limite permet de bien distinguer les échantillons sélectionnés dans les listes de l'UPE d'origine et de l'UPE fictive. L'UPE fictive respecte les règles habituelles touchant la croissance continue et les logements sélectionnés par l'intervieweur.

### **Sous-échantillonnage de l'UPE**

En fonction de la rétroaction provenant du terrain, les UPE à forte croissance peuvent avoir une incidence sur la capacité de l'intervieweur de terminer une tâche ayant beaucoup de nouvelles adresses de logements. Le cas échéant, on fait un sous-échantillonnage de l'UPE afin d'alléger le fardeau. Il y a deux formes de sous-échantillonnage : la modification simple du taux d'échantillonnage et la formation d'une sélection d'échantillon secondaire.

On a recours à la modification simple du taux d'échantillonnage quand la croissance est modérée. Le potentiel de construction additionnelle est limité et le pourcentage d'augmentation par rapport au chiffre prévu pour le ménage varie de 100 % à 200 %. Si la croissance est inférieure à 100 %, elle ne suffit pas à justifier le sous-échantillonnage. Si la croissance est supérieure à 200 %, on devrait songer à la formation d'unités secondaires et à un sous-échantillonnage des unités, afin d'ajouter une autre étape de sélection de l'échantillon.

Le plus souvent, il suffit de réduire le taux d'échantillonnage d'un facteur de deux pour réduire le travail de l'intervieweur. On modifie le sous-poids de l'UPE afin de représenter la procédure utilisée pour sélectionner le sous-échantillon, tel qu'il est indiqué à la section 6.2.2.

### **Formation de sous-grappes**

La formation de sous-grappes en tant que forme d'unités secondaires nécessite une croissance importante dans l'UPE d'au moins 200 % supérieure à sa taille prévue, un réseau routier utilisable permettant de former au moins quatre unités secondaires et une augmentation importante de la charge de travail de l'intervieweur.

L'intervieweur doit présenter une ventilation détaillée des chiffres des nouveaux logements par côté d'îlot, incluant toutes les nouvelles rues construites dans cette

zone en croissance. Le personnel du bureau central se sert des chiffres des nouveaux logements et des nouvelles rues pour définir quatre sous-grappes ou plus, en s'efforçant de nouveau de générer des unités d'échantillonnage d'environ 200 ménages.

On sélectionne deux des unités secondaires, pour créer deux nouvelles unités à inscrire dans la liste sur le terrain. Le taux global de sous-échantillonnage pour l'UPE se situe généralement entre deux et trois. Au-delà de trois, on risque davantage que les valeurs aberrantes aient une incidence marquée sur les estimations de certaines variables, tandis qu'en dessous de deux, on n'obtient aucune réduction sensible de la charge de travail de l'intervieweur. On corrige le poids de l'UPE pour les deux sous-grappes afin de compenser le sous-échantillonnage. Veuillez consulter la section 6.2.2 pour de plus amples détails.

Les unités de sous-échantillonnage sont cartographiées en dehors de l'opération habituelle de cartographie et soumises au listage. La sélection des logements ne peut commencer qu'une fois le listage des unités de sous-échantillonnage terminé. D'ordinaire, il faut procéder rapidement au listage des sous-grappes puisque le listage initial de l'UPE d'origine est déjà bien avancé dans le calendrier des activités lorsqu'on détermine la croissance.

### **Mise à jour de la strate**

À l'occasion, une UPE connaît une croissance extrême. Même le sous-échantillonnage de l'UPE ne parvient pas à réduire la charge de travail de l'intervieweur, ce qui peut avoir une incidence sur les estimations du coefficient de variation (c.v.) Il est alors préférable de remanier la strate. D'ordinaire, d'autres UPE de la strate connaîtront une croissance importante. Ces UPE ne sont pas toutes sélectionnées et la strate risque de ne plus contenir un échantillon représentatif de la population de la strate, d'où la nécessité de mettre la strate à jour.

On doit effectuer un dénombrement total des logements pour toutes les UPE de la strate. Un Registre des adresses plus récent, un nouveau recensement ou à défaut, un dénombrement sur le terrain s'impose. En outre, si la croissance d'une UPE est importante, on doit entreprendre une ventilation détaillée du chiffre des logements par côté d'îlot pour former des unités secondaires d'échantillonnage, tel qu'il est indiqué plus haut.

On met alors en œuvre le programme de mise à jour de la strate à partir des chiffres des nouveaux logements,

y compris les nouvelles unités secondaires. Ce programme, basé sur Keyfitz (1951), tel que modifié par Drew, Choudhry et Gray (1978), conserve le plus grand nombre possible des UPE sélectionnées au moment de la mise à jour. Les UPE nouvellement sélectionnées nécessitent un listage, mais généralement, le nouvel échantillon conserve de 60 à 70 % des grappes déjà listées. La mise en place du nouvel échantillon s'étend sur une période de six mois. On applique un coefficient de pondération provisoire à toutes les UPE de la strate jusqu'à la mise en place complète. Ce coefficient de pondération tient compte des données nouvelles tirées du dernier recensement des logements, absentes de l'échantillon actif.

Dès le début du plan actuel, il a fallu procéder à la mise à jour d'une strate à Calgary.

### 3.3.7 Contrôleur de rendement de l'UPE

Un certain nombre d'UPE suscitent des difficultés au moment de la cartographie et du listage. L'interprétation de la carte peut donner des listes erronées. Le contrôleur de rendement de l'UPE signale les UPE comptant trop ou trop peu de ménages. La présence de trop peu de ménages semble indiquer la présence d'une zone où il y a des démolitions, des inexactitudes des limites indiquées dans la création de l'UPE ou de difficultés à détecter ces limites sur le terrain. La présence de trop de ménages indique habituellement des zones en croissance, mais peut aussi indiquer des fractionnements de grandes UPE qui avaient été mal définies sur la carte ou mal interprétées sur le terrain. On trouvera une explication des fractionnements d'UPE aux sections 2.4 et 3.2.

Le suivi sur place vise à justifier ou à corriger les divergences. On produit des cartes explicatives ou des révisions si les limites originales ne sont pas claires. Les détails additionnels sur le réseau routier le plus récent, y compris les nouveaux numéros de voirie, peuvent aussi aider à cerner le groupe exact de logements à inscrire sur la liste.

À l'occasion, des corrections s'imposent à l'extraction du RA, pas tellement pour faire coïncider la sélection actuelle d'UPE avec le groupe voulu de logements, mais plutôt pour offrir des commentaires au sujet des UPE avoisinantes également touchées par les changements.

### 3.4 Processus de sélection des logements

La sélection des logements est un processus qui vient après la sélection des unités primaires d'échantillonnage (UPE) dans chaque strate et qui comprend : l'attribution

des numéros de renouvellement aux UPE, la création d'un système de renouvellement des UPE, non seulement pour effectuer une rotation des logements dans les UPE, mais également pour remplacer les UPE elles-mêmes, l'échantillonnage systématique des logements et la stabilisation de la taille de l'échantillon total. Nous décrivons d'abord le numéro de renouvellement.

#### Numéro de renouvellement

Chaque mois, une partie de l'échantillon de l'EPA est remplacée. Le remplacement des unités d'échantillonnage, soit leur renouvellement, s'effectue à chaque étape du plan d'échantillonnage à plusieurs degrés. L'unité ultime de la sélection, le logement, est remplacée tous les six mois, tandis que les unités plus globales demeurent plus longtemps dans l'échantillon. On a fixé à six mois la période de renouvellement des logements, car cette durée représentait un compromis entre le coût du renouvellement et l'augmentation du taux de non-réponse qui pourrait se produire si l'on demandait aux répondants de participer à l'enquête plus longtemps.

Pour veiller à ce que les intervieweurs aient une charge de travail uniforme et réduire au minimum le biais associé au nombre de mois pendant lesquels un logement a été sondé, on a adopté un mode de renouvellement selon lequel un sixième des logements est remplacé chaque mois. Ce mode de renouvellement est mis en œuvre en associant à chaque UPE un numéro de renouvellement entre un et six. Ce numéro détermine le mois durant lequel le renouvellement des logements (leur arrivée dans l'échantillon) a lieu. Si le numéro de renouvellement est un, les logements dans l'UPE sont donc remplacés en janvier. Puisque les logements sont actifs pour une période de six mois, le renouvellement 1 s'effectue également en juillet pour remplacer les logements sélectionnés en janvier. À ce stade, les logements de janvier sont considérés comme ayant été « remplacés ». De manière semblable pour le renouvellement 2, les logements sont remplacés en février et en août; pour le renouvellement 3, ils sont remplacés en mars et en septembre; pour le renouvellement 4, ils sont remplacés en avril et en octobre; pour le renouvellement 5, ils sont remplacés en mai et en novembre et, finalement, pour le renouvellement 6, ils sont remplacés en juin et en décembre.

*Hors renouvellement* : Dans certaines situations, une UPE doit être remplacée, mais elle est ensuite introduite dans l'enquête au cours d'un mois différent de la date habituelle d'introduction pour ce renouvellement. On parle alors d'introduction hors renouvellement. La date d'introduction est plus tard que d'habitude, mais la date

d'expiration demeure la même; moins de six interviews sont alors effectuées dans le premier échantillon de l'UPE. On effectue des introductions hors renouvellement chaque fois qu'un nouvel échantillon doit être introduit plus rapidement qu'il est prévu selon la date habituelle d'introduction.

### **Attribution des numéros de renouvellement**

Les numéros de renouvellement sont attribués de façon à répartir également l'échantillon prévu dans l'ensemble des six panels de renouvellement. L'échantillon prévu est le rendement de toutes les UPE échantillonnées en fonction du compte d'origine sur les ménages utilisés pour créer la base de l'EPA. Une répartition égale est requise à l'échelle de la strate, de la région économique de l'assurance-emploi (REAE) et de la province. L'atteinte de cet objectif suppose ce qui suit :

- La charge de travail par strate est stable au fil du temps étant donné qu'un nombre à peu près égal de logements est remplacé chaque mois.
- L'échantillon est composé d'un nombre égal de ménages qui en font partie depuis un, deux, trois, quatre, cinq ou six mois, ce qui atténue l'effet sur les estimations des mois passés dans l'échantillon.
- L'échantillon est bien divisé en six parties représentatives de même taille, dont on peut se servir lorsque l'on souhaite obtenir des sous-échantillons à partir de la base de l'EPA.

Pour commencer, on attribue aux UPE d'une strate un numéro de renouvellement pour équilibrer l'échantillon de logements total prévu dans la strate. Dans la plupart des secteurs, chaque strate comporte six ou douze sélections, de sorte que chaque groupe de renouvellement peut avoir le même nombre de sélections. Dans quelques strates, il pourrait y avoir de une à cinq UPE. Dans la mesure du possible, on combine ces strates pour plus ou moins équilibrer la répartition. Souvent, les strates éloignées comportent moins de six sélections et font l'objet d'un traitement à part – leur rendement réel est, en tout cas, très incertain.

L'attribution des numéros de renouvellement s'effectue de façon indépendante au sein de chaque REAE. Dans une REAE, les strates sont traitées au fur et à mesure qu'elles sont créées. En effet, chaque strate dépend de l'échantillon cumulatif prévu obtenu à la suite du renouvellement de la strate précédente. L'attribution commence par un tableau hiérarchisé des échantillons

prévus par groupe de renouvellement. Pour chaque strate figurant dans la liste, les échantillons prévus pour chaque UPE sont classés du plus petit au plus grand. Un compte cumulatif des échantillons prévus par groupe de renouvellement, tel qu'ils ont été recueillis dans la strate précédente pour cette REAE, est classé de manière semblable, c'est-à-dire du plus petit au plus grand.

Les numéros de renouvellement sont attribués en appariant l'UPE comportant le plus petit échantillon prévu dans la strate au renouvellement affichant le compte cumulatif le plus élevé. En même temps, l'UPE comportant le deuxième plus petit échantillon est appariée au renouvellement affichant le deuxième compte cumulatif le plus élevé, et ainsi de suite pour les six renouvellements. S'il existe 12 sélections dans la strate, elles sont traitées en deux lots séquentiels de six. Après avoir traité chaque REAE, la variation dans l'échantillon prévu ne devrait pas être plus importante que la variation dans une strate donnée. Dès le début du traitement de chaque REAE, un ordre aléatoire est attribué aux numéros de renouvellement pour assurer une répartition la plus égale possible à l'échelle provinciale et nationale. Soulignons qu'il s'agit d'échantillons prévus en fonction des comptes d'origine. Dans certains cas, l'échantillon réel s'écartera considérablement de cette valeur prévue.

La méthode susmentionnée s'applique à la méthode des groupes aléatoires décrite à la section 2.7. Dans le cadre de cette méthode, on attribue un renouvellement à chaque groupe et chaque sélection d'UPE génère un échantillon de taille prévue qui diffère d'un groupe à un autre. Pour certaines strates qui font appel à une probabilité aléatoire proportionnelle à la taille, la taille prévue de l'échantillon est la même pour toutes les UPE sélectionnées. Un ordre aléatoire des six numéros de renouvellement possibles est attribué à l'échantillon systématique des UPE.

Dans les strates à trois degrés d'échantillonnage, un numéro de renouvellement est attribué à chaque étape de sélection de l'UPE. Les étapes subséquentes ont toutes le même numéro de renouvellement.

### **Renouvellement des UPE**

Les méthodes de sélection des UPE sont décrites à la section 2.7. Tous les mois, un petit sous-ensemble d'UPE sélectionnées atteignent leur durée de vie prédéterminée dans les échantillons de l'EPA et doivent être remplacées. Des méthodes de renouvellement des UPE légèrement différentes sont utilisées selon le type de plan de chaque strate.

*Secteurs où l'on applique la méthode des groupes aléatoires.* Dans la méthode des groupes aléatoires, chaque groupe se compose d'un sous-ensemble aléatoire des UPE de chaque strate. Les UPE dans le sous-ensemble sont classées aléatoirement et une UPE est choisie selon une probabilité proportionnelle à sa taille. La taille de l'UPE est liée au nombre de ménages déterminé lors du dernier recensement. Le nombre aléatoire servant à sélectionner l'UPE *initiale* détermine également le nombre d'échantillons systématiques à prélever dans cette UPE; c'est ce qu'on appelle la vie d'une UPE. Une fois que le nombre requis d'échantillons a été interviewé, l'UPE est remplacée.

Il est nécessaire d'attribuer à chaque première UPE choisie un nombre aléatoire de périodes pendant lesquelles elle restera dans l'échantillon pour maintenir la probabilité initiale de sélection des UPE au fil du temps. Si, par exemple, on gardait les premières unités choisies jusqu'à épuisement de l'échantillon, c'est-à-dire jusqu'à ce que tous les échantillons systématiques de logements aient été prélevés, l'échantillon finirait par devenir biaisé, car les grandes UPE y seraient sur-représentées. En outre, puisque beaucoup d'UPE dans une strate sont de taille semblable, un grand nombre d'entre elles s'épuiseront en même temps, ce qui alourdira les opérations de listage sur le terrain.

On procède au renouvellement en passant à l'UPE suivante dans la liste randomisée des UPE du groupe. En atteignant le bas de la liste, on revient à la première UPE de la liste. Finalement, on revient à l'UPE sélectionnée initialement, mais dans presque toutes les strates, le plan de l'EPA sera remplacé avant que cela ne se produise.

Le premier remplacement d'une UPE sélectionnée initialement s'effectue dans l'échantillon jusqu'à ce qu'il n'y ait plus de choix de logements possible dans l'UPE, et ce, conformément à la règle de la durée de vie minimale expliquée à la section 2.7. Le prolongement de la durée de vie de la sélection initiale d'une UPE s'annule en réduisant la durée de vie de la sélection de la première UPE de remplacement. Conséquemment, la deuxième unité, c'est-à-dire l'UPE de remplacement, demeure dans l'échantillon jusqu'au mois de sortie normalement fixé sans correction en fonction de la durée de vie minimale de l'UPE initiale. Le remplacement ultérieur des UPE au sein de l'échantillon peut s'effectuer jusqu'à ce que tous les échantillons de logements possibles aient été échantillonnés.

Pour chaque UPE, choisie ou non, on génère aléatoirement un autre nombre entre 1 et la fraction de

sondage inverse (FSI). Ce nombre détermine le départ aléatoire pour l'échantillonnage systématique des logements dans l'UPE. L'échantillonnage systématique des logements produit généralement un logement de plus dans la première sélection par rapport à la dernière. La randomisation du point de départ initial permet d'éviter la réduction graduelle de la taille de l'échantillon qui s'ensuit à mesure que la grappe vieillit. La valeur du point de départ augmente avec chaque possibilité d'échantillonnage dans l'UPE et revient à 001 seulement après avoir atteint la valeur maximale qui équivaut à la FSI.

*Secteurs où l'on applique la méthode d'échantillonnage systématique avec classement aléatoire et PPT.* Pour le renouvellement des unités d'échantillonnage, on procède plus ou moins de la manière décrite dans les paragraphes précédents. Plutôt que de choisir une UPE par groupe aléatoire, une ou plusieurs UPE peuvent être choisies à partir d'une liste de la strate complète classée aléatoirement. Avec  $n$  correspondant au nombre d'UPE à sélectionner et  $I_s$  correspondant à la FSI de la strate, un nombre aléatoire  $r$  est sélectionné de 1 à  $n * I_s$ . Les UPE sont initialement sélectionnées à l'aide d'une séquence de points de départ  $r + k * I_s$  lorsque  $k = 0$  à  $n - 1$ . Ces points de départ déterminent la période de conservation de chaque UPE sélectionnée *initialement*. Comme précédemment, un nombre aléatoire distinct de 1 à la FSI de l'UPE est choisi pour commencer la sélection des logements.

Un exemple illustrera cette méthode. Voici une liste de cinq UPE classées aléatoirement. Pour la FSI de la strate  $I_s = 84$ ,  $n = 2$ .

**Tableau 3.3 Échantillonnage systématique avec classement aléatoire et PPT de l'EPA**

UPE	Comptes sur les ménages	FSI <sub>u</sub>	FSI	CM	S	Vie
3	153	24,637	25	25	...	...
4	241	38,808	39	64	61	4
1	224	36,071	36	100	...	...
5	218	35,105	35	135	...	...
2	207	33,333	33	168	145	24

Note: CM est cumulatif, S est le départ aléatoire. Voir l'annexe A.2 pour les autres abréviations.

Les comptes sur les ménages tirés du recensement sont additionnés pour donner un résultat de 1 043, puis ils sont divisés par  $I_s * n = 168$  pour générer une taille de l'échantillon prévue de 6 208 par UPE sélectionnée pour cet exemple. Cette taille prévue de l'échantillon est divisée par le compte total des ménages de chaque UPE pour donner les FSI<sub>u</sub> non arrondies. Les FSI<sub>u</sub> sont ensuite arrondies et rajustées pour donner la colonne FSI

de l'UPE. Il est nécessaire de rajuster les arrondissements pour que la somme des FSI demeure égale à  $I_s * n$ . La colonne FSI est la mesure de la taille réelle dans la procédure d'échantillonnage PPT. La colonne CM cumule les FSI de rangée en rangée. La sélection de l'UPE provient d'un départ choisi au hasard  $r$  entre 1 et 168, disons que  $r = 61$  dans notre exemple. En examinant la colonne CM, choisissez la rangée  $i$ , où  $CM_{i-1} < 61$  et  $61 \leq CM_i$ . La rangée 2 satisfait à ces exigences. Pour la deuxième sélection,  $n = 2$ , où  $r + k * I_s = 61 + 1 * 84 = 145$ , et la rangée 5 est sélectionnée, comme le montre la colonne S. On peut voir que la condition requise pour pouvoir choisir plus d'une UPE est que chaque FSI soit inférieure à la FSI de la strate. Cette condition s'applique facilement si les comptes sur les ménages sont plus ou moins égaux dans toutes les UPE de la strate. La durée de vie de la sélection d'une UPE est équivalente à  $FSI_i - (S_i - CM_{i-1}) + 1$ . La première UPE demeure dans l'échantillon pour  $39 - (61 - 25) + 1 = 4$  possibilités ou départs. La seconde UPE demeure dans l'échantillon pour 24 départs.

Une autre façon d'y penser est de considérer la liste des UPE classées aléatoirement comme une liste de classement aléatoire partielle de tous les échantillons possibles dans ces UPE, chaque échantillon correspondant à un point de départ pour la sélection des logements. Elle n'est que partiellement aléatoire, car tous les points de départ possibles pour une même UPE sont regroupés. Nous sélectionnons ensuite un échantillon systématique de points de départ dans cette liste.

*Secteurs où l'on applique la méthode d'échantillonnage à trois degrés.* Quelques strates urbaines font l'objet d'un échantillonnage à trois degrés. Le même mode de remplacement des échantillons s'applique à chaque étape de l'échantillonnage. La sélection des logements dans l'unité secondaire se poursuit jusqu'à ce que la période de conservation de l'unité secondaire se soit écoulée et que la sélection passe à la prochaine unité secondaire. Les unités secondaires dans l'UPE sont renouvelées jusqu'à ce que la période de conservation de l'UPE se soit écoulée et que l'UPE soit remplacée. Dans la nouvelle UPE, on choisit des unités secondaires pour commencer une nouvelle séquence. Les numéros de renouvellement sont attribués au niveau de l'UPE pour minimiser les coûts. Ces strates ont des caractéristiques semblables à celles des régions éloignées. Puisque les régions éloignées ne sont visitées qu'une fois tous les six mois, il vaut mieux que toute l'UPE ait le même mois de renouvellement.

## Sélection des logements

La sélection des logements est effectuée chaque mois, un renouvellement à la fois, en fonction du mois de chaque renouvellement. L'échantillonnage systématique des logements exige la FSI de l'UPE et un point de départ. (Dans un échantillonnage à trois degrés, l'unité secondaire de la FSI correspondra à l'unité pertinente pour la sélection des logements.)

Les logements sont sélectionnés à partir de l'UPE à l'aide d'une liste d'adresses générée par le Registre des adresses ou du listage effectué sur le terrain dans le cadre de l'UPE. Les UPE de remplacement doivent être complétées et ajoutées à la base de données de listage avant la première sélection de logements, sinon on parle alors d'un listage en retard. L'échantillon systématique est généré en sélectionnant les listes qui commencent par le numéro de ligne équivalant au départ  $r$  et en sélectionnant les lignes supplémentaires par la suite en ajoutant la FSI de l'UPE  $I$ . Nous sélectionnons  $l = r + k * I$ ,  $k = 0, 1, 2, \dots$  jusqu'à ce que le numéro de la ligne dépasse le nombre de lignes disponibles dans la liste.

La première sélection de logements dans une UPE s'effectue avec le départ aléatoire choisi à l'étape de sélection de l'UPE. Pour chaque sélection ultérieure, on ajoute 1 au départ aléatoire pour l'UPE, jusqu'à ce que la durée de vie de l'UPE ait atteint son maximum. Si la valeur augmentée dépasse la FSI de l'UPE, le départ redevient 1.

En ce qui a trait aux UPE sélectionnées initialement, il n'est pas nécessaire que chaque logement soit sélectionné avant le remplacement de l'UPE.

## Stabilisation de la taille de l'échantillon

Le listage des UPE sur le terrain constitue une liste ouverte. Les ajouts à la liste génèrent un nouvel échantillon conformément à la fraction de sondage de l'UPE. Le programme de stabilisation a été institué lors du remaniement du plan de 1974 afin de restreindre la croissance des échantillons de l'EPA. Dans le plan actuel, la réduction de la taille de l'échantillon de 3 % nécessaire au financement du remaniement a été mise en œuvre en partie en ayant recours au programme de stabilisation.

Le nombre de logements nouvellement échantillonnés (nouvelles unités) sélectionnés chaque mois est restreint à une valeur établie d'avance appelée allocation de base. Tous les logements qui dépassent cette limite sont retirés aléatoirement. On attribue un poids de stabilisation aux logements restants pour compenser.

Les allocations de base et les échantillons retirés sont calculés en fonction du secteur et du renouvellement. Un seul numéro de renouvellement est attribué à l'échantillon mensuel de l'EPA, selon le mois de renouvellement. Le calcul mensuel des poids de stabilisation suppose qu'un poids différent est accordé à chaque numéro de renouvellement.

Chaque secteur commence avec une Région économique de l'assurance-emploi (REAE). Des régions distinctes sont créées dans chaque REAE pour la strate spéciale qui est peut-être présente, y compris des régions éloignées, Autochtones, à revenu élevé et à taux élevé de vacance. Certaines REAE, lorsqu'elles sont exprimées en région physique, sont équivalentes à la région métropolitaine de recensement (RMR) et, par conséquent, les secteurs de stabilisation respectent les limites de la RMR.

Les régions ne sont pas toutes soumises à la stabilisation. Les strates des régions éloignées et des régions à taux élevé de vacance, les échantillons des territoires du Nord et quelques régions Autochtones dont les échantillons prévus sont très petits ne sont pas inclus dans le programme de stabilisation. Leur échantillon mensuel réel est trop variable pour justifier une stabilisation.

Le retrait des logements est effectué conformément à l'algorithme suivant. Les UPE hors renouvellement et les UPE dont la croissance est forte sont définies d'abord.

Les UPE hors renouvellement sont celles qui sont introduites au cours d'un mois et qui ne sont pas imposées par le numéro de renouvellement. Elles sont entièrement exclues de la stabilisation. Ces UPE appartiennent vraiment à un mois de renouvellement différent, un mois d'échantillonnage différent.

Comme il est expliqué à la section 3.3, les UPE à forte croissance sont sous-échantillonnées. On attribue un poids à ces UPE pour compenser le sous-échantillonnage. Comme il est expliqué ci-après, les UPE sous-échantillonnées sont incluses dans le compte des logements pour le secteur de stabilisation, mais elles ne peuvent être retirées. Les poids de stabilisation ne sont pas appliqués aux UPE sous-échantillonnées afin d'éviter toute augmentation du poids de l'UPE déjà présente. Le poids des niveaux d'UPE est décrit plus en détail au chapitre 6.2.2.

Les UPE qui n'ont pas été stabilisées ont un poids de stabilisation par défaut de 1. Le poids de stabilisation

appliqué aux UPE qui restent est calculé pour chaque région (et renouvellement) à l'aide de la formule :

$$w = (N - C) / (B - C)$$

où :

- w correspond au poids de stabilisation;
- N correspond au compte initial des logements sélectionnés;
- B correspond à l'allocation de base;
- C représente le compte des logements sélectionnés dans les UPE sous-échantillonnées. D'habitude, C égale à zéro.

Le nombre de logements à retirer est  $N - B$ . Le retrait est systématiquement appliqué aux logements  $N - C$  à l'aide d'un point de départ aléatoire dans chaque secteur de stabilisation.

Si  $N$  est inférieur à  $B$ , aucune stabilisation n'est alors appliquée à cette région pour le mois en question. Autrement dit, le poids de stabilisation prend une valeur par défaut de 1. Nous ne pouvons accroître la taille de l'échantillon lorsque l'allocation de base  $B$  est supérieure à la taille réelle de l'échantillon. Cette situation se produit quand de nouvelles UPE ne sont pas listées à temps pour l'échantillonnage ou que la population d'une région diminue au fil du temps. En outre, la variabilité par renouvellement augmente au fil du temps et certains renouvellements généreront des échantillons plus petits qu'il était prévu.

Les logements sélectionnés sur le terrain en raison de la croissance de la grappe n'ont aucune chance d'être inclus dans le programme de stabilisation. Ces logements sont exclus du poids de stabilisation. Quand des logements sont sélectionnés parce que des immeubles à logements multiples ont été répertoriés comme des résidences individuelles, tous les logements pourraient alors être sélectionnés lors de l'échantillonnage de la résidence principale. Il est possible d'accorder un poids de stabilisation à de tels « immeubles à logements multiples », leur attribuant en réalité le poids de la résidence principale.

D'autres enquêtes ont recours au programme de stabilisation. Par exemple, dans le cadre de l'Enquête sur les dépenses des ménages, un ensemble indépendant de ménages est choisi à partir de l'ensemble d'UPE actives de l'EPA. Les renouvellements complets sont généralement réservés à cette fin. La stabilisation réduit le rendement de l'échantillon à une valeur totale établie.

## Établissement des allocations de base

Le plan de l'EPA est élaboré en fonction de la répartition des ménages dans plus de 1 000 strates. Le processus de sélection des logements est établi en fonction des logements listés dans l'ensemble des UPE sélectionnées dans ces strates. Afin de fixer ces sélections de logements à l'ensemble de répartitions des ménages requis, les allocations de base sont calculées comme une estimation du nombre de logements qui, finalement, produiront le compte sur les ménages requis. Les allocations de base sont calculées mensuellement. Les facteurs qui ont une incidence sur ce calcul comprennent la répartition de l'échantillon dans les ménages, le déséquilibre du renouvellement, le cas échéant, la conversion aux logements et la présence de secteurs non stabilisés.

*Répartition de l'échantillon* : Le point de départ du calcul des allocations de base est l'échantillon attribué à partir du plan de la base de l'EPA. Pour chaque secteur de stabilisation, il s'agit du nombre prévu de logements occupés (répondants et non-répondants) dans le cadre de l'enquête.

*Déséquilibre du renouvellement* : En raison du processus d'attribution des numéros de renouvellement expliqué précédemment, le nombre de ménages prévus peut ne pas être exactement identique pour les six renouvellements. Les différences sont rarement de plus de 1 %. L'utilisation d'échantillons prévus par renouvellement suppose que les poids de stabilisation sont plus uniformes parmi les renouvellements. Au fur et à mesure que le plan prend de l'âge, le mérite relatif de ce déséquilibre perd de l'importance et une simple division par six est utilisée.

*Conversion aux logements* : L'étape d'échantillonnage finale de l'EPA est fondée sur le listage des logements de chaque UPE sélectionnée six semaines avant de déterminer si le logement correspond ou non à un ménage. La taille souhaitée d'un échantillon de ménage doit être convertie en une allocation de base des logements pour chaque secteur de stabilisation. Les principaux facteurs à prendre en considération lors de cette conversion comprennent le taux de vacance, les nouveaux logements sélectionnés sur le terrain et les logements qui sont désactivés durant la tenue à jour de la liste.

La différence principale entre le nombre de ménages et le nombre de logements est l'ensemble des réponses comprises dans les catégories « vacant », « saisonnier », « en construction », « commercial » et « résidents non admissibles ».

De tels logements hors du champ de l'enquête représentent de 5 % à 20 % du nombre total de logements, selon le secteur de stabilisation. Pour chaque secteur de stabilisation, le taux de vacance est estimé en fonction des plus récents résultats d'enquête. Au début du plan de sondage, une estimation plus appropriée est obtenue à partir des enregistrements de recensement en effectuant un petit rajustement pour les différences connues entre les logements recensés et les logements de l'EPA. Le calcul précis des allocations de base adéquates pour un nouveau plan de sondage peut prendre un an.

Il arrive parfois qu'une nouvelle UPE ne soit pas listée à temps pour l'échantillonnage. La liste est achevée durant la semaine d'enquête et elle génère des échantillons sur le terrain appelés logements sélectionnés par les intervieweurs. Durant la semaine d'enquête, l'intervieweur peut découvrir d'autres logements dans d'autres UPE qui génèrent également des logements sélectionnés par les intervieweurs. Ces derniers ajouts ne sont pas disponibles pour le programme de stabilisation. Par conséquent, l'allocation de base est réduite pour permettre l'intégration de ces ménages supplémentaires après l'échantillonnage. Ces logements supplémentaires représentent en moyenne de 1 % à 2 % de la taille initiale d'un échantillon. L'estimation des logements sélectionnés par les intervieweurs est calculée en fonction des plus récents résultats d'enquête, mais ne peut dépasser 4 % pour un secteur de stabilisation donné afin d'éviter les valeurs aberrantes.

Les erreurs de listage, les immeubles incendiés et les roulottes déplacées peuvent réduire l'échantillon de nouveaux logements d'environ 1 % durant la semaine d'enquête. Les allocations de base sont augmentées pour compenser cette perte. L'estimation des logements désactivés est calculée en fonction des plus récents résultats d'enquête, mais ne peut dépasser 2 % pour un secteur de stabilisation donné afin d'éviter les valeurs aberrantes.

Les trois facteurs précédents subissent les conséquences du fait que le rendement passé n'est pas toujours un bon indicateur des valeurs futures. La taille finale des échantillons dans les ménages variera légèrement tous les mois comparativement à la répartition requise de l'échantillon. Le facteur le plus important est le taux de vacance. Il peut avoir une composante saisonnière et changer de façon imprévisible avec chaque UPE de remplacement.

*Secteurs non stabilisés* : La taille de l'échantillon dans les secteurs non stabilisés représente environ 1,3 % de

l'échantillon total. La taille des échantillons dans les secteurs non stabilisés peut augmenter, tandis que la taille des échantillons dans les secteurs stabilisés est maintenue relativement constante. Heureusement, les secteurs non stabilisés sont des régions surtout rurales et éloignées dont la perspective de croissance est faible. Il n'existe aucun contreponds associé aux secteurs stabilisés pour la croissance potentielle des secteurs non stabilisés. En général, les problèmes liés au listage et aux déplacements vers les régions plus éloignées contribuent à maintenir la taille des échantillons plus petite qu'elle était prévue.

### **Nouveau poids de stabilisation**

Le poids de stabilisation utilisé pour compenser les logements retirés de l'échantillon est calculé de nouveau une fois les retraits effectués. Les strates d'un secteur de stabilisation n'ont pas toutes la même FSI de strate. De grandes différences entre les strates peuvent survenir, ce qui pourrait avoir une incidence sur le poids de stabilisation applicable à toutes les UPE de la région. Le poids est recalculé comme si le retrait était effectué séparément pour chaque ensemble de strates affichant la même FSI au sein de la région. Ces petits ensembles de strates ne peuvent être utilisés directement comme secteurs de stabilisation, car la taille des échantillons sera trop petite pour produire une estimation viable des valeurs d'allocation de base.

## Chapitre 4 Collecte

### Introduction

Les données servant à produire les estimations mensuelles touchant l'emploi total et le chômage sont obtenues sur une base mensuelle en communiquant avec les ménages ciblés par l'enquête (voir la partie 1.3 pour plus de détails sur la couverture).

Le calendrier de l'enquête ainsi que les méthodes de collecte utilisées sont décrits dans les sections suivantes.

#### 4.1 Calendrier de l'enquête

La collecte des données est effectuée par une équipe d'environ 1 500 intervieweurs qui relèvent des bureaux régionaux.

Le cycle de l'enquête pour un mois donné commence aussitôt que le traitement des données du mois précédent se termine. Les activités de collecte sont gérées suivant un échéancier serré établi en fonction des autres processus de l'enquête.

La collecte des données aux fins de l'EPA a lieu tous les mois durant la semaine qui suit la semaine de référence de l'EPA, qui est habituellement la semaine où tombe le 15 du mois. Les interviews débutent le dimanche de la semaine de collecte et se poursuivent généralement jusqu'au mardi de la semaine suivante. En de rares occasions, il arrive que la collecte soit prolongée d'une journée.

Les données recueillies par les intervieweurs sont transmises au bureau central pour le traitement.

#### 4.2 Méthode de collecte

Aux fins de la collecte, on distingue deux types de ménages dans l'EPA. Un sixième de l'échantillon mensuel est constitué de ménages qui en sont à leur premier mois dans l'enquête et qu'on appelle les « naissances » ou nouvelles unités. Le reste, soit cinq sixièmes, est constitué de ménages qui en sont entre leurs deuxième et sixième mois dans l'enquête et qu'on appelle les « subséquents ». Il est à noter que, en cas de non-réponse ou d'un changement complet des occupants dans le ménage, les procédures de collecte pour les nouvelles unités s'appliquent aux ménages le premier mois qu'ils sont enquêtés.

Les interviews de l'EPA sont réalisées à l'aide de deux méthodes de collecte : les interviews sur place assistées par ordinateur (IPAO) et les interviews téléphoniques assistées par ordinateur (ITAO). Historiquement, l'IPAO était utilisée pour les ménages qui en étaient à leur premier mois dans l'enquête et les

intervieweurs se rendaient sur les lieux pour mener l'interview. Les interviews subséquentes sont normalement menées à l'aide de l'ITAO, à partir des centres d'appels. Si les membres du ménage le demandent, les interviews subséquentes se déroulent en personne par la méthode IPAO.

Auparavant, les interviews subséquentes étaient effectuées par téléphone par les intervieweurs IPAO. Ce qui signifie que le même intervieweur effectuait tant l'interview avec la nouvelle unité que les interviews subséquentes pour un ménage donné. Vers le milieu de l'année 2000, une application centralisée de l'ITAO a été introduite pour les interviews subséquentes, ce qui signifiait que les intervieweurs pour l'interview subséquente étaient différents de ceux qui avaient fait la première interview bien que dans plusieurs cas, le même intervieweur menait toutes les interviews subséquentes pour un ménage. De mars à septembre 2005, un ordonnanceur d'appels ITAO a été introduit graduellement. L'ordonnanceur d'appels sélectionne automatiquement des cas pour l'intervieweur à partir d'un centre de répartition, ce qui fait qu'un ménage peut maintenant être contacté par n'importe quel intervieweur.

Un changement important a eu lieu en 2004; les nouvelles unités des régions urbaines pour lesquelles des numéros de téléphone sont disponibles sont maintenant enquêtées au moyen du premier contact par téléphone (PCT). Ces nouvelles unités sont contactées par la méthode ITAO par des intervieweurs du lieu de travail ITAO. Cette approche a été introduite afin de réduire les coûts de collecte associés à une première interview en personne.

Lors de l'interview des nouvelles unités, l'intervieweur doit recueillir l'information de tous les membres du ménage sélectionné. Pour les interviews subséquentes, l'intervieweur vérifie la liste des membres du ménage, puis recueille les renseignements relatifs à la population active pour le mois en cours. Pour les personnes âgées de 70 ans et plus, on réduit le fardeau imposé au répondant en réutilisant les réponses fournies à l'interview initiale pour les mois subséquents.

Les réponses des membres du ménage sont habituellement obtenues par un membre bien informé du ménage. Cette façon de procéder est appelée « réponse par procuration » et est mise en pratique, car il serait trop coûteux en temps et trop dispendieux de faire plusieurs visites ou appels pour obtenir l'information directement de chaque membre du ménage. Environ

65 % des renseignements obtenus à l'EPA le sont de cette manière.

Afin de maintenir élevés les taux de réponse à l'EPA, différents types de lettres sont envoyés aux répondants des logements sélectionnés. Par exemple, lorsqu'un ménage est sélectionné pour la première fois, une lettre d'introduction ainsi qu'une brochure explicative lui sont envoyées avant la première interview. Des lettres à remettre en cas de refus sont également envoyées afin de convaincre le ménage hésitant de participer à l'enquête.

### **4.3 Premier contact par téléphone**

L'approche du premier contact par téléphone (PCT) a été introduite en novembre 2004. Cette approche permet d'effectuer certains premiers contacts par téléphone auprès des nouvelles unités à partir des centres d'appel ITAO.

Le PCT est utilisé pour les ménages sélectionnés dans les régions urbaines seulement. Les listes administratives pour ces secteurs sont mises à jour régulièrement et les adresses ont tendance à avoir une forme normalisée, ce qui permet un meilleur appariement. Lorsque la majorité des contacts sont effectués par téléphone, les coûts de collecte sont considérablement réduits.

Chaque mois, environ 9 000 nouveaux ménages sont introduits dans l'échantillon. Parmi ceux-ci, environ 6 000 sont sélectionnées dans les grandes zones urbaines. Toutes les naissances des zones urbaines seront

soumises à un appariement avec les fichiers administratifs des numéros de téléphone. Un numéro de téléphone est trouvé pour environ 60 % à 70 % d'entre elles, bien que ce pourcentage peut varier. En 2007 environ 4 000 naissances ont fait partie de l'approche du PCT.

Certaines procédures ont été mises en place pour permettre qu'au moins une tentative soit faite pour contacter chaque ménage durant les deux premiers jours de collecte, soit le dimanche ou le lundi.

La composante de contact du questionnaire est utilisée pour permettre la vérification de l'adresse du ménage. Cette vérification est primordiale pour s'assurer que le ménage contacté est celui qui occupe le logement sélectionné. Si le ménage contacté n'est pas le bon, d'autres sources sont utilisées pour trouver un numéro de téléphone valide pour le logement. Si on ne peut trouver un numéro valide, le cas est transféré à un intervieweur sur le terrain qui se rendra en personne à l'adresse indiquée.

Si aucun contact n'est établi pendant deux mois consécutifs, le cas sera transféré à un intervieweur sur le terrain lorsqu'il commencera son troisième mois dans l'enquête. Les logements peuvent également être transférés aux intervieweurs sur le terrain durant la collecte si le répondant demande une interview en personne.

## Chapitre 5 Traitement et imputation

### Introduction

Il existe deux types de non-réponses dans l'Enquête sur la population active (EPA) : la non-réponse des personnes et la non-réponse pour l'item. La non-réponse des personnes se produit s'il n'est pas possible d'obtenir le moindre renseignement d'une personne pour l'enquête en raison d'un refus de répondre ou d'une impossibilité de joindre un membre du ménage concerné. La non-réponse pour l'item se produit lorsqu'il n'est pas possible d'obtenir de renseignements sur un ou plus des items du questionnaire. Actuellement, dans le cadre de la stratégie de traitement des non-réponses, différentes méthodes sont utilisées pour gérer les deux cas possibles. En ce qui concerne la non-réponse pour l'item, trois méthodes d'imputation sont utilisées : l'imputation par enregistrement donneur ou « hot deck » (HD), l'imputation historique et l'imputation par déduction. Pour le traitement de la non-réponse de la personne, on utilise, soit la méthode d'imputation par hot deck, soit un ajustement de poids pour la non-réponse en fonction de l'historique du ménage contenant la personne.

Pour ce qui est de l'imputation HD pour l'EPA, les valeurs manquantes sont remplacées par les valeurs correspondantes d'un donneur sélectionné aléatoirement dans la même classe d'imputation. Les classes d'imputation sont définies en utilisant des variables disponibles pour les receveurs et pour les donneurs potentiels. Une innovation dans la stratégie de l'imputation HD utilisée par l'EPA est d'utiliser les valeurs des mois précédents (peut-être imputées) de certaines variables dans la définition des classes d'imputation. Cette innovation, quelques fois appelée l'imputation hot deck longitudinale, a été mise en œuvre en janvier 2005 et est fondée sur les recherches de Bocci et Beaumont (2004).

Plusieurs aspects reliés au plan de sondage de l'EPA et aux procédures de collecte ont été pris en considération lors du développement de la stratégie pour le traitement des non-réponses. Ils sont résumés ci-dessous :

- L'échantillon est sélectionné selon un plan d'échantillonnage stratifié à plusieurs degrés et est divisé en six groupes de renouvellement. Par définition, chaque groupe de renouvellement est représentatif d'une population entière.

- Les logements sélectionnés font partie de l'échantillon pendant six mois consécutifs (le terme ménages sélectionnés doit être compris comme des ménages occupant actuellement les logements sélectionnés). Chaque groupe de renouvellement comprend des ménages (logements) qui font partie de l'échantillon depuis le même nombre de mois. Pour un mois donné, il y a un groupe de renouvellement naissant comprenant les ménages qui sont dans l'échantillon depuis seulement un mois et cinq autres groupes de renouvellement faisant partie de l'échantillon depuis une période allant de deux à six mois. En conséquence, les ménages échantillonnés sont les mêmes d'un mois à l'autre pour cinq des six groupes de renouvellement.
- Les principales variables d'intérêt sont des données catégoriques, liées à la situation d'activité sur le marché du travail des membres des ménages. Les variables liées à la rémunération sont secondaires et elles sont recueillies uniquement le premier mois au cours duquel une personne fait partie de l'échantillon ou si elle change d'emploi. Les données liées à la rémunération ne sont pas recueillies pour les autres mois et sont simplement reportées au mois suivant en vue d'alléger le fardeau des répondants.
- La collecte est principalement effectuée par téléphone, sauf pour le groupe de renouvellement naissant pour lequel les interviews sont dans une large mesure réalisées sur place. La collecte est principalement effectuée par téléphone, sauf pour le groupe de renouvellement naissant pour lequel les interviews sont dans une large mesure réalisées sur place.

Le reste du chapitre décrit plus en détail les étapes du traitement et de l'imputation. D'autres renseignements sur la méthodologie d'imputation de l'EPA se trouvent dans les rapports de Lorenz (1996) et de Bocci et Beaumont (2004). Le premier décrit les particularités du système d'imputation par hot deck. L'autre se concentre sur la méthode d'imputation hot deck longitudinale.

### 5.1 Étapes générales de traitement des données

Le traitement des données, de la collecte à la pondération, peut être divisé en six grandes étapes :

- 1) Réception des données des bureaux régionaux et première phase de vérification
- 2) Codage des industries et des professions
- 3) Vérification de la cohérence
- 4) Imputation HD
- 5) Dérivation des variables
- 6) Pondération et production d'un fichier de microdonnées épurées (fichier TABS)

Au cours de la première phase de vérification, on procède à l'imputation par déduction et, dans une moindre mesure, à de l'imputation historique. L'imputation par déduction est appliquée lorsqu'une valeur manquante peut logiquement être déduite des réponses données aux autres variables. L'imputation historique consiste à transférer des données passées au mois en cours. En même temps sont mis en place différents indicateurs comme ceux déterminant le type de non-réponse ou la nécessité d'imputation par hot deck. Les vérifications de la cohérence sont appliquées au cours de l'étape 3. Les dossiers qui échouent aux vérifications de la cohérence sont inspectés à la main afin que les incohérences soient retirées. Les variables socio-démographiques sont toutes imputées avant que ne soient effectuées les imputations HD. Ces variables servent à former des classes d'imputation.

## 5.2 Imputation lors du prétraitement

Avant de pouvoir effectuer l'imputation réelle des valeurs manquantes en utilisant la méthode HD, il faut suivre quelques étapes de prétraitement des données. Tout d'abord, les personnes sont divisées en trois groupes A, B et C. Le groupe A comprend des donneurs potentiels. Il s'agit de l'ensemble de personnes dont les données enregistrées ne contiennent aucune valeur manquante et ces valeurs sont cohérentes entre elles. Le groupe B est formé de toutes les personnes qui n'appartiennent pas au groupe A et qui ne présentent aucune valeur manquante et ces valeurs sont cohérentes après la première phase de vérification (au cours de laquelle l'imputation par déduction ou de l'imputation historique ont été effectuées). Ces personnes ont un dossier complet, mais ne peuvent être des donneurs potentiels. Les autres personnes, les receveurs, forment le groupe C et nécessitent l'imputation par hot deck. Au cours d'une deuxième étape de prétraitement, toutes les données concernant la rémunération sont converties sur une base horaire. Cela permet de s'assurer que la valeur imputée pour la rémunération est cohérente avec la valeur concernant le nombre d'heures consignées pour un receveur.

La dernière étape de prétraitement constitue l'identification des valeurs aberrantes relatives à la rémunération. Les valeurs qui sont, soit extrêmement élevées, soit extrêmement basses sont réputées suspectes et c'est pourquoi elles sont donc considérées comme manquantes et sont imputées. Pour les personnes dont la rémunération est très forte ou très faible, sans pour autant tomber dans les extrêmes, les valeurs enregistrées sont conservées, mais elles ne peuvent pas être des donneurs potentiels. Ces personnes sont d'emblée attribuées au groupe B. La méthode utilisée pour la détection des valeurs aberrantes est la méthode du quartile. Fondamentalement, la méthode permet de repérer les rémunérations qui sont beaucoup plus importantes ou beaucoup plus faibles que la valeur médiane. Les enregistrements contenant des valeurs au-dessus d'un certain seuil ou trop éloignées de la valeur médiane sont considérés comme aberrants. Différents seuils sont utilisés pour définir les valeurs aberrantes présumées incorrectes et celles qui empêchent des personnes d'être des donneurs potentiels. Le rapport de Lorenz (1996) offre davantage de détails sur la détection des valeurs aberrantes dans l'EPA.

## 5.3 Imputation en cas de non-réponse des items

Une fois que toutes les étapes de prétraitement ont été menées à bien, les valeurs manquantes peuvent être imputées. L'imputation par hot deck aléatoire au sein des différentes classes est utilisée pour imputer les valeurs manquantes. La procédure est appliquée de telle sorte que les données d'un receveur après imputation répondent aux règles de vérification de la cohérence et de la validité (les variables qui nécessitent des valeurs autres qu'un blanc pour un receveur donné doivent être imputées en utilisant des valeurs autres qu'un blanc). Dans une classe d'imputation donnée, chaque receveur est imputé en sélectionnant des donneurs par un simple échantillonnage aléatoire sans remplacement jusqu'à ce qu'un donneur répondant aux règles de vérification soit trouvé.

Les classes d'imputation initiales sont formées en croisant les variables catégoriques classées par ordre d'importance :

- 1) TPATH (12 catégories)
- 2) LMLFS3 (3 catégories)
- 3) COW (3 catégories)
- 4) OCC4 (4 catégories)
- 5) PROV (10 catégories)
- 6) AGE3P3 (5 catégories)
- 7) ABQ1 (2 catégories)

- 8) IMM (3 catégories)
- 9) LMLFS7 (7 catégories)
- 10) LMINDG (20 catégories)
- 11) MULTJOB (2 catégories)
- 12) AGECP1 (5 catégories)
- 13) SEX (2 catégories)
- 14) OCC10 (10 catégories)
- 15) AGECP2 (8 catégories)
- 16) STUD (2 catégories)
- 17) EDUC (2 catégories)
- 18) DWELRENT (2 catégories).

Dans le cadre des études empiriques, on a estimé que ces variables étaient classées selon un ordre d'importance décroissant pour expliquer les variables relatives à la population active. Une description plus détaillée des catégories de variables se trouve dans la section 5.5.

Notez que les variables LMSFL3, LMSLFS7 et LMINDG se rapportent aux valeurs du mois précédent.

L'utilisation de la variable TPATH est quelque peu complexe. En vue d'expliquer son rôle, il faut tout d'abord remarquer la situation d'activité de la variable. LFSSTAT peut accepter une des sept valeurs correspondant aux sept premières valeurs de la variable TPATH. Une de ces sept valeurs LFSSTAT est affectée à chaque donneur. Pour les receveurs, il est possible que nous ne connaissions pas la valeur de LFSSTAT. Cependant, nous disposons de suffisamment de renseignements pour exclure quelques-unes des sept valeurs possibles. Le rôle de la variable TPATH est de veiller à ce que seules les valeurs valides soient imputées aux receveurs. Pour ce faire, une seule valeur de TPATH est attribuée à chaque receveur et chaque donneur est reproduit par son nombre de valeurs TPATH valides. Ces donneurs reproduits sont retirés à la fin de l'étape de l'imputation. Par exemple, supposons que pour un donneur, LFSSTAT = 2. Alors, il a comme valeurs valides TPATH = (2, 8, 10). Il est ainsi reproduit trois fois; chaque répétition est attribuée à une de ces trois valeurs TPATH.

L'imputation est effectuée au sein de toutes les classes qui contiennent un nombre suffisant de donneurs. Deux conditions permettent de déterminer si une classe dispose d'un nombre suffisant de donneurs :

- (i) Le nombre de donneurs doit être plus important que le nombre de receveurs de cette classe.
- (ii) Chaque classe doit comprendre un minimum de trois donneurs.

Si l'une de ces conditions n'est pas satisfaite, l'imputation est de nouveau effectuée en retirant la variable la moins importante (DWELRENT) lorsque sont formées les classes d'imputation. Si après cette deuxième série d'imputation, il reste encore des receveurs n'ayant pas été imputés en raison de classes qui ne répondent pas aux deux conditions précitées, l'imputation est effectuée à une troisième reprise, en retirant la deuxième variable la moins importante (EDUC). Ce processus qui consiste à enlever une variable afin d'effectuer l'imputation continue jusqu'à ce que tous les receveurs soient imputés ou jusqu'à ce que seules les cinq premières variables – TPATH, LMLFS3, COW, OCC4 et PROV – restent. Tous les receveurs pas encore imputés sont envoyés pour l'imputation du dossier complet, de telle sorte que toutes les variables reliées à l'activité du receveur, y compris celles qui étaient enregistrées, sont remplacées par celles d'un donneur sélectionné aléatoirement – voir la section 5.4.

Dans une classe d'imputation donnée répondant aux deux conditions ci-dessus, chaque receveur est imputé en sélectionnant d'abord un donneur, de telle sorte que les règles de vérification de la validité sont satisfaites. Si aucun donneur ne peut être trouvé, alors l'imputation du dossier complet est effectuée. Si un donneur convenable peut être trouvé (c'est-à-dire un qui répond aux règles de vérification de la validité après imputation des valeurs manquantes du receveur), les valeurs manquantes du receveur sont remplacées par les valeurs correspondantes du donneur et l'on procède à l'examen des règles de vérification de la cohérence. Lorsque toutes les règles de vérification sont satisfaites, le processus d'imputation pour le receveur est achevé. Dans le cas contraire, on procède à une deuxième tentative avec un autre donneur convenable (c'est-à-dire répondant aux vérifications de la validité) et à l'examen des règles de vérification de la cohérence. Lorsque toutes les règles de vérification sont satisfaites après la deuxième tentative, le processus d'imputation pour le receveur est achevé. Dans le cas contraire, le dossier complet est soumis à l'imputation avec les valeurs du dernier donneur utilisé.

Au cours du processus d'imputation, des indicateurs de sortie sont inscrits au fichier de sortie. Ils sont utiles pour produire différents tableaux qui peuvent servir à surveiller et à améliorer le processus d'imputation. Le rapport de Lorenz (1996) fournit des explications détaillées sur les spécificités de la stratégie d'imputation indiquée ci-dessus.

## 5.4 Imputation en cas de non-réponse des personnes

La non-réponse des personnes se produit lorsqu'une personne dans un ménage est non-répondant ou quand tout le ménage est non-répondant. Si un ménage est non-répondant, mais qu'il a été répondant le mois précédent, alors l'ensemble des variables reliées à l'activité de toutes les personnes appartenant au ménage sont imputées. Si le ménage était également non-répondant le mois précédent alors on procède à un ajustement de poids pour la non-réponse.

Avant janvier 2005, on utilisait l'imputation transversale par hot deck et l'imputation historique afin de traiter la non-réponse des personnes. L'imputation historique de renseignements passés était utilisée pour les variables sociodémographiques. Elle servait également à toutes les autres variables si le ménage non-répondant avait répondu au cours du mois précédent. Si la personne non-répondante n'avait pas répondu au cours du mois précédent, mais l'avait fait par le passé, l'imputation transversale par hot deck était alors préférée à l'imputation historique.

Le principal problème de l'imputation historique est qu'elle sous-estime les changements qui surviennent d'un mois à l'autre. Par ailleurs, l'imputation transversale par hot deck a tendance à surestimer les changements qui surviennent d'un mois à l'autre. Afin de pallier ces problèmes, de nouvelles variables longitudinales par hot deck sont utilisées dans la définition des classes de l'imputation hot deck (voir Bocci et Beaumont 2004). Cette méthode a été mise en œuvre en janvier 2005 afin de traiter les non-réponses des personnes lorsque des renseignements historiques sont disponibles. Avant l'imputation hot deck, toutes les variables sociodémographiques restent imputées par imputation historique dans la mesure où il n'est pas prévu qu'elles changent de manière significative d'un mois à l'autre.

Les variables utilisées pour former l'imputation par hot deck pour la non-réponse des personnes sont données ci-dessous. Il s'agit des variables suivantes, classées par ordre d'importance :

- 1) PROV (10 catégories)
- 2) AGE GP1 (5 catégories)
- 3) LMLFS7 (7 catégories)
- 4) SEX (2 catégories)
- 5) ABQ1 (2 catégories)
- 6) IMM (3 catégories)
- 7) REAE (Région économique de l'assurance-emploi – 55 catégories)

- 8) EDUC (2 catégories).

Il faut remarquer que les donneurs et les receveurs pour l'imputation longitudinale par hot deck sont des personnes, même si nous nous intéressons à la non-réponse des ménages. De plus, les règles de vérification de la validité et de la cohérence perdent toute pertinence dans la mesure où est effectuée une imputation du dossier complet (le donneur répond déjà à toutes les règles). Enfin, il faut noter que les valeurs manquantes des receveurs sont imputées en utilisant les valeurs du donneur du mois en cours même si les classes d'imputation sont fondées sur les valeurs du mois précédent.

## 5.5 Définition des variables utilisées pour déterminer les classes d'imputation

Classement par âge (AGE GP1 et AGE GP3 sont des groupements plus larges d'AGE GP2).

AGE GP1	AGE GP2	AGE GP3	
1	1	1	15 à 19
2	2	2	20 à 24
3	3	2	25 à 29
3	4	2	30 à 34
3	5	2	35 à 44
4	6	2	45 à 54
4	7	2	55 à 64
5	8	3	65+

Classement par profession (OCC4 est un groupement plus large d'OCC10).

OCC4	OCC10	
01	01	Direction, gestion, sciences naturelles, sciences sociales, religion, enseignement, médecine et artistique
02	02	Administratif
02	03	Ventes
02	04	Services
03	05	Agriculture, pêche, forêts et exploitation minière
03	06	Traitement, usinage et fabrication
03	07	Construction
03	08	Transports
03	09	Manutention, artisanat, autres
04	10	N'a jamais travaillé ou dernier emploi occupé il y a plus d'un an ou incapacité permanente

### EDUC

- |   |  |
|---|--|
| 0 | La personne n'a pas de diplôme d'études secondaires. |
| 1 | La personne a un diplôme d'études secondaires.       |

### COW - catégorie de travailleurs

- |   |                                   |
|---|-----------------------------------|
| 1 | Employé rémunéré                  |
| 2 | Travailleur autonome              |
| 3 | Travailleur familial non rémunéré |

STUD		Identité autochtone – Indien de l'Amérique du nord, Métis ou Inuit.
0	Pas étudiant à temps-plein	
1	Étudiant à temps-plein	
DWELRENT		ABQ1
1	Propriétaire d'un logement	1 Oui
2	Locataire d'un logement	2 Non
PROV		Situation vis-à-vis de l'activité du dernier mois. LMLFS3 est un groupement plus large de LMLFS7
10	Terre-Neuve et Labrador	LMLFS7
11	Île-du-Prince-Édouard	1 Employé et au travail
12	Nouvelle-Écosse	2 Employé et en congé
13	Nouveau-Brunswick	3 Chômeur, mis à pied temporairement
24	Québec	4 Chômeur à la recherche d'un emploi
35	Ontario	5 Chômeur, commencera un emploi
46	Manitoba	6 Inactif
47	Saskatchewan	7 Incapacité permanente
48	Alberta	
59	Colombie-Britannique	
SEX		LMLFS3
M	Homme	1 Employé (LMLFS7=1,2)
F	Femme	2 Chômeur (LMLFS7=3,4,5)
		3 Inactif (LMLFS7=6,7)
TPATH		Groupe d'industrie du dernier mois.
1	Employé et au travail (LFSSTAT=1)	LMINDG
2	Employé et en congé (LFSSTAT=2)	1 Agriculture, foresterie, pêche et chasse
3	Mis à pied temporairement (LFSSTAT=3)	2 Extraction minière et extraction de pétrole et de gaz
4	Chômeur à la recherche d'un emploi (LFSSTAT=4)	3 Services publics
5	Chômeur, commencera un emploi (LFSSTAT=5)	4 Construction
6	Inactif (LFSSTAT=6)	5 Fabrication
7	Incapacité permanente (LFSSTAT=7)	6 Commerce de gros
8	LFSSTAT valeurs 2 à 6	7 Commerce de détail
9	LFSSTAT valeurs 3 à 6	8 Transport et entreposage
10	LFSSTAT valeurs 2, 4, 5 ou 6	9 Information et industries culturelles
11	LFSSTAT valeurs 4, 5 ou 6	10 Finance et assurance
12	LFSSTAT valeurs 5 ou 6	11 Immobilier et location
MULTJOB - Avez-vous plus d'un emploi ou plus d'une entreprise la semaine dernière?		12 Services professionnels, scientifiques et techniques
1	Oui	13 Gestion de sociétés et d'entreprises
2	Non ou pas de réponse	14 Services administratifs et de soutien, services de gestion des déchets et d'assainissement
IMM - pays de naissance		15 Services d'enseignement
1	Canada	16 Soins de santé et assistance sociale
2	États-Unis	17 Arts, divertissements et loisirs
3	Autre	18 Hébergement et services de restauration
		19 Autres services (sauf administrations publiques)
		20 Administrations publiques

## Chapitre 6 Pondération et estimation

### Introduction

L'estimation est le processus d'enquête qui établit une approximation des paramètres de population inconnus à partir des données de l'échantillon, en combinaison possible avec des données auxiliaires provenant d'autres sources. Les résultats de l'estimation servent à formuler des inférences sur ces paramètres inconnus, c'est-à-dire à tirer des conclusions sur les caractéristiques de la population totale à partir d'un seul échantillon de cette population. Les paramètres de population d'intérêt comprennent notamment les chiffres, les moyennes et les ratios de population, de même que leurs moyennes sur plusieurs mois d'enquête. Une estimation<sup>20</sup>  $\hat{\theta}$  représente donc une approximation d'un paramètre inconnu de population  $\theta$  et la différence  $\hat{\theta} - \theta$  entre ces deux quantités porte le nom d'erreur d'enquête totale. Il existe deux principaux types d'erreur d'enquête totale : l'erreur d'échantillonnage et les erreurs non dues à l'échantillonnage. L'erreur d'échantillonnage est attribuable au fait qu'on calcule les estimations à partir d'un seul échantillon de l'ensemble de la population, tandis que les erreurs non dues à l'échantillonnage sont attribuables à d'autres causes, comme une base de sondage imparfaite, des erreurs de mesure ou les non-réponses. Dans l'Enquête sur la population active (EPA), on corrige l'erreur d'échantillonnage et une partie de l'erreur due à la non-réponse du ménage en attribuant un poids d'estimation, appelé poids final, à chaque personne échantillonnée sur laquelle on a des données, qu'elles soient imputées ou non. Pour simplifier les propos du présent chapitre, on présume qu'il n'y a aucune erreur non due à l'échantillonnage autre que l'erreur de non-réponse des ménages, même si on souligne brièvement qu'il est possible de corriger partiellement les imperfections de la base de sondage en ayant recours aux poids finaux.

Le principe fondamental de la pondération consiste à pondérer le poids de chaque personne par l'inverse de sa probabilité d'inclusion dans l'échantillon. Cela garantit que les estimations sont sans biais ou presque, dans la mesure où l'on s'attend, pour tous les échantillons possibles, à ce que l'erreur d'enquête soit égale à zéro ou presque. Pour évaluer la qualité d'une estimation et obtenir des inférences valides, on calcule habituellement des mesures de précision, comme le coefficient de variation estimé. On détermine le coefficient de

variation par la formule  $c.v.(\hat{\theta}) = \sqrt{V(\hat{\theta})}/\hat{\theta}$ , où  $V(\hat{\theta})$  correspond à la variance des estimations pour tous les échantillons possibles. Puisqu'on ne sélectionne, en pratique, qu'un seul échantillon, la variance  $V(\hat{\theta})$  est inconnue. Cependant, il est possible de l'estimer en n'utilisant que cet échantillon (voir le chapitre 7), dans le but d'obtenir les mesures de précision voulues.

La façon de sélectionner l'échantillon a une incidence sur la probabilité d'inclusion des ménages et, par conséquent, sur leur poids final. L'ensemble du processus de sélection peut se diviser en trois étapes principales : (i) sélection de l'échantillon stratifié initial à plusieurs degrés, (ii) ajustements de l'échantillon pour tenir compte des grappes en croissance et (iii) ajustements de l'échantillon visant à conserver la taille de l'échantillon au fil du temps. La section 6.2 traite de ces trois étapes et de leur incidence sur les probabilités d'inclusion. Après la sélection d'échantillon, il est possible de calculer un poids appelé le poids déterminé par le plan d'échantillonnage. On considère souvent ce dernier comme étant le nombre de fois où l'on doit reproduire chaque unité échantillonnée afin de représenter la population cible. Pour quantité de raisons, dont les refus ou l'impossibilité de communiquer avec certains des ménages échantillonnés, le nombre de ménages sur lesquels on recueille de l'information est inférieur au nombre de ménages échantillonnés. Cette réduction de la taille de l'échantillon occasionne une erreur de non-réponse du ménage. Pour compenser pour la non-réponse du ménage, on a recours à de l'imputation (voir le chapitre 5) et à des ajustements de poids. L'ajustement de poids pour la non-réponse consiste à ajuster le poids déterminé par le plan d'échantillonnage pour chaque ménage répondant, au moyen d'un facteur de correction de non-réponse. Le principe de base consiste à choisir un modèle qui convient aux probabilités de réponse inconnues, puis à calculer les facteurs de correction de non-réponse selon l'inverse des probabilités de réponse estimées. Le poids résultant de l'ajustement pour la non-réponse est appelé le sous-poids. La section 6.3 décrit de quelle manière le sous-poids est obtenu. Enfin, on a souvent recours au calage dans les enquêtes pour obtenir les poids finaux. Le calage consiste essentiellement à obtenir des poids finaux qui se rapprochent le plus possible des sous-poids, tout en respectant certaines contraintes précises. Pour l'EPA, on choisit les contraintes afin (i) de garantir la conformité avec les estimations externes de la population, (ii) de tenir compte, dans une certaine

20. Par souci de simplicité, on ne fait pas de distinction entre l'estimateur et l'estimation.

mesure, de la sous-couverture et (iii) d'améliorer l'efficacité des estimations. En outre, l'EPA utilise un estimateur de calage composite depuis janvier 2000 afin d'améliorer l'efficacité des estimations, en particulier celle des estimations du changement. Le calage et le calage composite sont décrits à la section 6.4 avec la méthode de pondération intégrée, qui garantit un poids final commun pour chacune des personnes du ménage. Dans la prochaine section, on définira certains concepts de base et on présentera la notation.

### 6.1 Concepts de base et notation

Pour amorcer la discussion sur la pondération et l'estimation, il faut présumer au départ que le paramètre de population d'intérêt est le chiffre de population

$$t_y = \sum_{k \in P} y_k;$$

où  $P$  indique la population du mois courant et  $y$  constitue une variable d'intérêt. Par exemple,  $y_k$  pourrait être une variable binaire indiquant si une personne donnée  $k$  de la population occupe un emploi ( $y_k = 1$ ) ou non ( $y_k = 0$ ). Le chiffre de population  $t_y$  représenterait alors le nombre de personnes occupant un emploi au sein de la population  $P$ . À partir de cette population, on sélectionne un échantillon de ménages selon un plan d'échantillonnage stratifié à plusieurs degrés et on recueille l'information de l'EPA sur chaque personne issue des ménages sélectionnés. En l'absence de non-réponse des ménages, on pourrait estimer le chiffre de population  $t_y$  au moyen de la formule,

$$\hat{t}_y^D = \sum_{k \in s} w_k^D y_k,$$

où  $s$  représente l'échantillon de tous les membres des ménages sélectionnés et  $w_k^D$  correspond au poids déterminé par le plan d'échantillonnage pour la personne  $k$ . Puisqu'il y a non-réponse des ménages, on n'obtient pas la variable  $y$  pour toutes les personnes de l'échantillon  $s$  et on ne peut pas calculer l'estimation selon le poids déterminé par le plan d'échantillonnage  $\hat{t}_y^D$ . Dans ce cas, le chiffre de population  $t_y$  est plutôt estimé par la formule

$$\hat{t}_y^{\text{NA}} = \sum_{k \in s_r} w_k^{\text{NA}} y_k,$$

où  $s_r$  correspond au sous-ensemble de toutes les personnes de  $s$  appartenant à un ménage répondant (ou imputé) et  $w_k^{\text{NA}}$  correspond au poids corrigé de non-réponse pour la personne  $k$ , lequel est appelé le sous-poids de la personne  $k$ .

Comme il est mentionné dans l'introduction, on utilise le calage dans l'EPA par souci de conformité, afin de tenir compte du sous-dénombrement et

d'améliorer l'efficacité. Le calage permet d'en arriver à l'estimation

$$\hat{t}_y^C = \sum_{k \in s_r} w_k^C y_k,$$

où  $w_k^C$  correspond au poids de calage pour la personne  $k$ . Enfin, pour obtenir des estimations plus efficaces dans l'EPA, on utilise le calage composite. Il permet d'en arriver à l'estimation

$$\hat{t}_y^{\text{CC}} = \sum_{k \in s_r} w_k^{\text{CC}} y_k,$$

où  $w_k^{\text{CC}}$  correspond au poids de calage composite pour la personne  $k$ , lequel est souvent appelé le poids final de la personne  $k$ . Les étapes de la pondération sont décrites plus en détail aux sections 6.2, 6.3 et 6.4.

Dans bien des cas, l'intérêt ne réside pas dans l'estimation d'un chiffre de population, mais plutôt d'un taux de population

$$r_{y_1, y_2} = \frac{\sum_{k \in P} y_{1k}}{\sum_{k \in P} y_{2k}},$$

où  $y_1$  et  $y_2$  représentent deux variables d'intérêt. Par exemple,  $y_{1k}$  pourrait représenter une variable binaire indiquant si une personne donnée  $k$  de la population est en chômage ( $y_{1k} = 1$ ) ou non ( $y_{1k} = 0$ ), et  $y_{2k}$  pourrait représenter une variable binaire indiquant si la personne  $k$  fait partie de la population active ( $y_{2k} = 1$ ) ou non ( $y_{2k} = 0$ ). Le taux de population  $r_{y_1, y_2}$  représente alors le taux de chômage au sein de la population. Il est possible de l'estimer à partir des poids finaux  $w_k^{\text{CC}}$  en utilisant la formule

$$\hat{r}_{y_1, y_2}^{\text{CC}} = \frac{\sum_{k \in s_r} w_k^{\text{CC}} y_{1k}}{\sum_{k \in s_r} w_k^{\text{CC}} y_{2k}}.$$

Enfin, il vaut parfois la peine d'estimer la moyenne d'un paramètre de population sur plus d'un mois, selon la formule suivante

$$\theta = \sum_{t=1}^T \frac{\theta_t}{T},$$

où  $\theta_t$  correspond au paramètre de population au mois  $t$  et  $T$  correspond au nombre de mois retenus dans la définition de la moyenne ci-dessus. Par exemple,  $\theta_t$  pourrait correspondre au taux de chômage ou au nombre de personnes ayant un emploi au mois  $t$ . Le paramètre  $\theta$  est appelé un paramètre de moyenne mobile de  $T$ -mois. Il est possible de l'estimer au moyen de la formule

$$\hat{\theta}^{\text{CC}} = \sum_{t=1}^T \frac{\hat{\theta}_t^{\text{CC}}}{T},$$

où  $\hat{\theta}_t^{CC}$  correspond à une estimation de  $\theta_t$  obtenue en utilisant les poids finaux du mois  $t$ . Dans l'EPA, on produit, pour chaque mois de l'enquête, des estimations de moyenne mobile de trois mois pour les taux de chômage de chaque Région économique de l'assurance-emploi (REAE), à partir des trois derniers mois. De telles estimations sont plus stables que des estimations mensuelles, mais leur interprétation est différente étant donné qu'elles évaluent un paramètre de population différent.

## 6.2 Poids déterminé par le plan d'échantillonnage

En principe, le poids déterminé par le plan d'échantillonnage d'une personne  $k$  est égal à l'inverse de sa probabilité de sélection,  $\pi_k^D$ , dans l'échantillon  $s$ , c'est-à-dire,  $w_k^D = 1/\pi_k^D$ . S'il n'y avait aucune erreur de non-réponse ou aucune autre erreur non due à l'échantillonnage, cela garantirait que les estimations obtenues selon le poids déterminé par le plan d'échantillonnage sont sans biais ou du moins, à peu près sans biais. Comme chacune des personnes d'un ménage sélectionné est incluse dans l'échantillon, calculer la probabilité de sélection d'une personne donnée équivaut à calculer la probabilité de sélection de son ménage. Divers mécanismes aléatoires interviennent dans la sélection des ménages. Chaque mécanisme a une incidence sur la probabilité de sélection obtenue. Dans le reste du présent chapitre, on décrit ces mécanismes ainsi que leur incidence sur le poids déterminé par le plan d'échantillonnage.

### 6.2.1 Poids de base

Au moment de la conception de l'enquête, on constitue les strates en regroupant les unités géographiques. Le chapitre 2 présente une explication détaillée de la stratification. Dans chaque strate, on sélectionne un échantillon des ménages (ou plus précisément, des logements) au moyen d'un plan d'échantillonnage à plusieurs degrés. Dans la plupart des strates, on utilise un plan d'échantillonnage à deux degrés, et l'accent sera mis sur ces strates afin d'expliquer de quelle façon on obtient le poids de base. De plus, aucune distinction ne sera faite entre les ménages et les logements dans le reste du présent chapitre, et le terme « ménages » sera employé pour décrire aussi bien les uns que les autres.

On divise chaque strate  $h$  en  $N_{1h}$  unités primaires d'échantillonnage (UPE). Au premier degré d'échantillonnage, on en sélectionne  $n_{1h}$ , la probabilité étant proportionnelle à la quantité  $R_{hj}^*$ , où l'indice  $hj$  indique l'UPE  $j$  dans la strate  $h$ . La probabilité de sélection de

premier degré de l'UPE  $j$  dans la strate  $h$  correspond donc à la formule

$$\pi_{1hj} = \frac{n_{1h} R_{hj}^*}{\sum_{j \in P_h} R_{hj}^*},$$

où  $P_h$  correspond à la population des UPE dans la strate  $h$ . La quantité  $R_{hj}^*$  est égale au ratio  $R_{hj} = \tilde{N}_{hj} / \tilde{n}_{2hj}$  arrondi au nombre entier, où  $\tilde{N}_{hj}$  correspond au nombre approximatif de ménages dans l'UPE  $j$  de la strate  $h$  provenant du Recensement de 2001 et  $\tilde{n}_{2hj}$  correspond au nombre de ménages dont la sélection est initialement prévue dans l'UPE  $j$  de la strate  $h$ . À noter que  $R_{hj}^*$  correspond au nombre de renouvellements dans l'UPE  $hj$  et s'apparente étroitement à la fraction de sondage inverse (FSI) pour l'UPE définie dans l'équation (3) à la section 2.7. Pour l'instant, on présume que  $\tilde{n}_{2hj}$  est constant dans la strate, donc  $\tilde{n}_{2hj} \equiv \tilde{n}_{2h}$ . Cette hypothèse est abandonnée plus loin, lorsqu'il est question de pondération avec la méthode de sélection Rao-Hartley-Cochran (RHC). Sélectionner des UPE avec une probabilité proportionnelle à  $R_{hj}^*$  équivaut essentiellement à sélectionner des UPE avec une probabilité proportionnelle à  $R_{hj}$  si ces ratios ne sont pas trop petits; cela équivaut donc à sélectionner les UPE avec une probabilité proportionnelle à  $\tilde{N}_{hj}$  puisque  $\tilde{n}_{2hj}$  est une constante dans la strate.

Dans chaque UPE sélectionnée  $j$ , on choisit systématiquement un échantillon de ménages selon l'intervalle d'échantillonnage fixe  $R_{hj}^*$ . Au deuxième degré d'échantillonnage, la probabilité de sélection du ménage  $i$  dans l'UPE  $j$  de la strate  $h$  correspond donc à  $\pi_{2hji} = 1/R_{hj}^*$ . Il faut noter que le nombre réel de ménages sélectionnés dans l'UPE  $j$  de la strate  $h$ ,  $n_{2hj} \approx N_{hj} \pi_{2hji} \approx N_{hj} / R_{hj}$ , sera probablement différent du nombre prévu initialement,  $\tilde{n}_{2hj} = \tilde{N}_{hj} / R_{hj}$ , car  $\tilde{N}_{hj}$  est une mesure approximative (provenant du Recensement de 2001) du nombre réel de ménages dans l'UPE  $j$  de la strate  $h$ ,  $N_{hj}$ .

La probabilité globale de sélection du ménage  $i$  dans la strate  $h$  correspond à la formule

$$\pi_{hi}^B = \pi_{1hj} \pi_{2hji} = \frac{n_{1h}}{\sum_{j \in P_h} R_{hj}^*},$$

qui représente une constante dans chaque strate, c'est-à-dire,  $\pi_{hi}^B \equiv \pi_h^B$ , pour tous les ménages  $i$  dans la strate  $h$ . On appelle aussi cette probabilité de sélection globale la probabilité de sélection de base. Le taux d'échantillonnage initialement prévu dans la strate  $h$  correspond à  $f_h = n_h / \tilde{N}_h$ , où l'on définit  $n_h$  pendant la répartition

du plan de l'enquête et  $\tilde{N}_h = \sum_{j \in P_h} \tilde{N}_{hj}$ . À noter que l'échantillon initial prévu par UPE correspond à  $\tilde{n}_{2h} = n_h / n_{1h}$ . Dans l'EPA, on souhaite respecter le taux d'échantillonnage initial, déterminé après la répartition de l'échantillon aux strates pour que  $\pi_h^B = f_h$ . Pour satisfaire à cette exigence, on arrondit  $R_{hj}$  aux nombres entiers  $R_{hj}^*$  afin d'obtenir  $\sum_{j \in P_h} R_{hj}^* = n_{1h} / f_h = \tilde{N}_h / \tilde{n}_{2h} = \sum_{j \in P_h} R_{hj}$ . Puisque  $\sum_{j \in P_h} R_{hj}^*$  doit être un nombre entier, on arrondit  $n_{1h} / f_h$  pour respecter l'égalité  $\sum_{j \in P_h} R_{hj}^* = n_{1h} / f_h$ .

Dans la méthode RHC, on regroupe d'abord les UPE dans chaque strate de manière aléatoire en  $n_{1h}$  groupes. Puis, on considère chaque groupe  $g$  comme une strate distincte et on sélectionne une seule UPE dans chaque groupe, selon la méthodologie décrite plus haut. Avec la méthode RHC, on choisit le nombre de ménages qu'on prévoit initialement sélectionner dans un groupe  $g$  donné de la strate  $h$ ,  $\tilde{n}_{2hg}$ , de sorte que le taux global d'échantillonnage dans chacun des  $n_{1h}$  groupes de la strate  $h$  soit égal au taux d'échantillonnage initial  $f_h$ . Par conséquent,  $\tilde{n}_{2hg} = \tilde{N}_{hg} f_h$ , où  $\tilde{N}_{hg}$  correspond au nombre approximatif de ménages du groupe  $g$  de la strate  $h$  selon le Recensement de 2001, et il se peut fort bien que  $\tilde{n}_{2hg}$  ne soit pas parfaitement identique pour deux groupes différents dans la strate  $h$ .

Comme on recueille l'information désirée pour chacune des personnes admissibles au sein d'un ménage sélectionné, la probabilité de sélection de base d'une personne  $k$  dans la strate  $h$  correspond à  $\pi_{hk}^B = \pi_h^B = f_h$  et son poids de base correspond à

$$w_{hk}^B = 1 / \pi_h^B = f_h^{-1}.$$

Dans le cas d'un plan d'échantillonnage avec un poids de base constant dans chaque strate, on parle d'autopondération dans les strates.

Le poids de base serait égal au poids déterminé par le plan d'échantillonnage si le plan d'échantillonnage et la population demeuraient inchangés. Cependant, comme les UPE sont appelées à croître au fil du temps et que le taux de l'échantillonnage systématique est fixe, la taille de l'échantillon augmenterait continuellement (tout comme le coût de la collecte des données). On constaterait aussi, au fil du temps, de grandes variations de la charge de travail chez les intervieweurs et entre les intervieweurs. Pour éviter cette situation, on a recours à deux méthodes d'échantillonnage pour contrôler la taille de l'échantillon : le sous-échantillonnage de l'UPE et la stabilisation de l'échantillon. Ces méthodes modifient la probabilité de sélection de base des ménages (et des personnes). Il devient donc nécessaire d'ajuster le poids

de base afin de contrebalancer ces méthodes d'échantillonnage. Cette question est examinée de plus près aux sections 6.2.2 et 6.2.3.

## 6.2.2 Poids de la grappe

Une grappe correspond à une UPE dans les strates d'un plan d'échantillonnage à deux degrés et à une avant-dernière unité dans les autres strates. Dans les zones urbaines, le nombre de logements dans une grappe peut s'accroître considérablement au fil du temps à cause de la construction. Compte tenu du taux d'échantillonnage fixe (ou le pas d'échantillonnage) dans chaque grappe, la taille de la tâche de l'intervieweur risque alors de s'accroître grandement. Cela peut avoir des répercussions sur la qualité de son travail et sur sa capacité de mener à bien sa tâche. Si la croissance dans une grappe dépasse 100 %, mais qu'elle n'est pas trop extrême, on peut sous-échantillonner la grappe au hasard en utilisant la méthode I ou la méthode II décrite ci-dessous. Ces méthodes de sous-échantillonnage modifient la probabilité de sélection des ménages. On modifie donc le poids de base  $w_{hk}^B$  au moyen d'un facteur de correction de grappe  $a_{hk}^P$  afin d'obtenir le poids de la grappe

$$w_{hk}^P = w_{hk}^B a_{hk}^P.$$

Malheureusement, on perd la caractéristique d'autopondération en procédant au sous-échantillonnage. On trouvera de plus amples détails sur les méthodes I et II dans Kennedy (1998). Lorsque la croissance est extrême, le sous-échantillonnage risque de ne pas être pratique et on procède à la mise à jour de la strate, comme le décrit la méthode III ci-dessous.

### Méthode I : Sous-échantillonnage de grappe

Cette méthode est la plus simple et la plus courante de toutes les méthodes de sous-échantillonnage. On y a recours lorsqu'une grappe doit être sous-échantillonnée à cause de sa croissance et que ni la méthode II ni la méthode III ci-dessous ne conviennent. On modifie le taux d'échantillonnage par grappes afin de réduire le nombre de ménages sélectionnés, tout en évitant l'échantillonnage des ménages déjà sélectionnés. On multiplie le poids de base des ménages interviewés par ce facteur. En raison des valeurs aberrantes que peuvent donner les enquêtes spéciales effectuées au moyen de la base de sondage de l'EPA, la valeur maximale du facteur de correction de grappe correspond à 3. Avant d'utiliser cette méthode, on doit aussi s'assurer que la croissance suffit à justifier un facteur d'au moins 2.

## Méthode II : Formation de sous-grappes

Lorsque la croissance dans une grappe dépasse 200 % et que les tracés des rues sont bien définis, on divise la grappe de croissance en plusieurs sous-grappes. On prend d'abord un échantillon des petites sous-grappes, puis on retient un échantillon des ménages dans chaque sous-grappe sélectionnée. Cette méthode équivaut à ajouter un autre degré d'échantillonnage dans les grappes de croissance. Elle ne modifie pas la probabilité de sélection des grappes, mais plutôt la probabilité de sélection des ménages dans les grappes de croissance.

## Méthode III : Mise à jour de strates

Lorsque la croissance est extrême, même la formation de sous-grappes peut s'avérer insuffisante et une mise à jour de strate s'impose, comme l'indique la section 3.3.6. On doit avoir le dernier recensement des logements dans toutes les grappes de la strate et on forme de nouvelles grappes en subdivisant des grappes existantes dans la base de sondage à partir de ce nouveau recensement. On effectue une mise à jour de l'échantillon de la strate, selon la méthode Keyfitz (1951), modifiée par Drew, Choudhry et Gray (1978), en conservant le plus grand nombre possible des UPE sélectionnées à l'origine. Le nouvel échantillon s'étend sur une période de six mois. On applique un coefficient de pondération provisoire à toutes les UPE de la strate jusqu'à la mise en place complète. Ce coefficient de pondération tient compte des données nouvelles tirées du dernier recensement des logements, absentes de l'échantillon actif.

### 6.2.3 Poids de stabilisation

Au dernier degré d'échantillonnage, on effectue un échantillonnage systématique à un taux fixe. Comme on emploie toujours le même taux de sondage, la croissance de la population et, par conséquent, du nombre de ménages, entraînerait un élargissement progressif de l'échantillon et une augmentation des coûts de sondage si aucune stabilisation de l'échantillon n'était effectuée. La stabilisation de l'échantillon consiste à sélectionner de façon aléatoire des ménages de l'échantillon afin de garder la taille de ce dernier à son niveau prévu. On effectue cette sélection aléatoire en échantillonnant systématiquement dans chaque secteur de stabilisation et indépendamment d'un secteur de stabilisation à un autre. Par définition, un secteur de stabilisation rassemble tous les ménages qui appartiennent à la même REAE et au même groupe de renouvellement. L'ensemble de personnes appartenant aux ménages qui continuent de faire partie de l'échantillon après la stabilisation est noté par la valeur  $s$  à la section 6.1.

La stabilisation de l'échantillon modifie la probabilité de sélection des ménages. On modifie donc le poids de la grappe  $w_{hk}^P = w_{hk}^B a_{hk}^P$  au moyen d'un facteur de correction par stabilisation  $a_{hk}^S$  afin d'obtenir le poids de stabilisation  $w_{hk}^S = w_{hk}^B a_{hk}^P a_{hk}^S$ . Par définition, le poids déterminé par le plan d'échantillonnage d'une personne  $k$  dans la strate  $h$ ,  $w_{hk}^D$ , est égal à son poids de stabilisation  $w_{hk}^S$ , c'est-à-dire,

$$w_{hk}^D \equiv w_{hk}^S = w_{hk}^B a_{hk}^P a_{hk}^S.$$

On calcule séparément le facteur de correction par stabilisation  $a_{hk}^S$  dans les sous-secteurs. Par définition, un sous-secteur réunit toutes les strates d'un secteur de stabilisation ayant un taux d'échantillonnage commun. La pondération de stabilisation s'écarte légèrement du principe de pondération par l'inverse de la probabilité de sélection, car on l'applique dans les sous-secteurs plutôt que dans les secteurs de stabilisation. On appelle souvent cette méthode de pondération stratification a posteriori, les strates a posteriori correspondant aux sous-secteurs.

Pour simplifier, supposons un secteur de stabilisation où tous les ménages présentent une probabilité de sélection de base de 1 sur 200 au moment de la conception de l'enquête et un facteur de correction de grappe de 1. Dans cet exemple, le secteur de stabilisation n'est pas réparti en sous-secteurs. Si la taille prévue de l'échantillon du secteur de stabilisation est de 300 ménages au moment de la conception de l'enquête et si les taux d'échantillonnage utilisés produisent en fait 350 ménages, il faut retirer aléatoirement 50 ménages du secteur de stabilisation. La probabilité de sélection des ménages passe donc de 1 sur 200 à 3 sur 700 (c'est-à-dire  $1/200$  multiplié par  $300/350$ ). En multipliant le poids de base de 200 par le facteur  $350/300$ , on obtient le poids de stabilisation  $700/3$ .

Les ménages qui présentent une des deux caractéristiques suivantes sont exclus de la stabilisation de l'échantillon et de la pondération de stabilisation :

- Les ménages qui appartiennent à une grappe sous-échantillonnée avec la méthode I ou II décrite à la section 6.2.2;
- Les ménages habitant un logement de construction récente qui a été ajouté à la liste des grappes et qu'on ne peut retirer de l'échantillon (logement sélectionné par l'intervieweur).

Comme ces ménages n'ont pas la chance d'être retirés de l'échantillon, ils sont aussi exclus de la pondération de stabilisation.

### 6.3 Sous-poids

Les ménages de l'échantillon sélectionné  $s$  ne sont pas tous interviewés, à cause des refus et d'autres facteurs qui rendent impossible la communication avec certains ménages. On traite d'abord une partie de cette non-réponse des ménages au moyen d'une méthode d'imputation longitudinale (voir le chapitre 5). On traite ensuite les ménages non répondants qui restent en les retirant et on corrige le poids déterminé par le plan d'échantillonnage des ménages répondants, en incluant ceux qui ont été imputés, au moyen d'un facteur de correction de non-réponse. Comme le souligne l'introduction, le principe de base consiste à choisir un modèle qui convient aux probabilités de réponse inconnues, puis à calculer les facteurs de correction de non-réponse selon l'inverse des probabilités de réponse estimées.

Dans l'EPA, le modèle de non-réponse utilisé est le modèle uniforme de non-réponse à l'intérieur des catégories. Avec ce modèle, on présume que tous les ménages appartenant à une catégorie donnée de non-réponse  $c$  présentent la même probabilité de réponse  $p_c$ . La probabilité de réponse estimée  $\hat{p}_c$  correspond simplement au taux de réponse des ménages dans la catégorie  $c$  selon le poids déterminé par le plan d'échantillonnage. Le facteur de correction de non-réponse pour une personne  $k$  appartenant à un ménage répondant dans la catégorie  $c$  correspond à  $a_{ck}^{NA} = 1/\hat{p}_c$  et le poids corrigé de non-réponse ou le sous-poids correspond à la formule

$$w_{ck}^{NA} = w_{ck}^B a_{ck}^P a_{ck}^S a_{ck}^{NA} = w_{ck}^D a_{ck}^{NA}.$$

Chaque personne d'un ménage répondant donné a droit au même facteur de correction de non-réponse et donc, au même sous-poids.

La clé pour réduire le biais de non-réponse consiste à cerner les catégories de non-réponse qui expliquent bien le mécanisme de non-réponse inconnue et dont l'élaboration justifie l'hypothèse d'une probabilité de réponse constante à l'intérieur des catégories. Sur le plan de l'efficacité, il est aussi souhaitable que les catégories de non-réponse soient les plus homogènes possible par rapport aux principales variables d'intérêt, autrement dit, la formation des catégories devrait faire en sorte que les répondants d'une catégorie donnée s'apparentent aux non-répondants au chapitre des variables d'intérêt principales. Ainsi, les variables servant à élaborer les catégories devraient apporter une explication au mécanisme de non-réponse et aux variables d'intérêt principales.

Dans l'EPA, chaque strate Autochtone ou de revenu élevé forme une catégorie distincte de non-réponse. On obtient les autres catégories par le croisement des variables REAE, TYPE et RENOUVELLEMENT (sauf les ménages qui appartiennent à une catégorie Autochtone ou de revenu élevé). La variable TYPE compte cinq catégories et indique le type de strate auquel le ménage appartient : isolée, rurale, urbaine à trois degrés d'échantillonnage, urbaine hors d'une région métropolitaine de recensement (RMR) et urbaine RMR. La variable RENOUVELLEMENT correspond à un des six groupes de renouvellement. Il faut noter que les catégories de non-réponse ne se chevauchent pas et que, réunies, elles couvrent l'ensemble de la population. On procède aussi à un regroupement des catégories lorsqu'un facteur de correction de non-réponse est supérieur à deux dans une catégorie donnée. On regroupe alors cette catégorie avec une autre qui a été choisie pour que le facteur de correction de non-réponse de la catégorie combinée soit inférieur à deux. La catégorie choisie doit appartenir à la même province, au même type de strate et au même groupe de renouvellement que la catégorie à regrouper, autrement dit, le regroupement s'applique à l'ensemble des REAE. On regroupe les catégories de non-réponse afin d'éviter les facteurs importants de correction de non-réponse, puisqu'ils ont tendance à accroître la variabilité des estimations.

On a songé à adopter une nouvelle méthodologie de correction du poids de non-réponse, décrite en plus grand détail par Alavi et Beaumont (2004). Avec cette nouvelle méthodologie, on élabore les catégories à partir de l'information tirée de la collecte des données, comme le nombre de tentatives de communication avec un ménage et le moment de la dernière tentative, en plus de l'information sur le plan de sondage. Une étude empirique a démontré que le nombre de tentatives constituait la variable la plus importante pour expliquer la non-réponse. De plus, on a corrélé cette information avec l'emploi et le chômage. Par conséquent, cette information semble très utile pour contrebalancer la non-réponse. Pour la formation des catégories de non-réponse, on a proposé la méthode de pointage (voir Little, 1986). Cette méthode consiste (i) à modéliser et estimer les probabilités de réponse en utilisant la régression logistique et (ii) à former des catégories homogènes par rapport à ces probabilités estimées. Contrairement à la méthode courante, la méthode de pointage proposée permet de contrôler facilement le nombre de catégories et le nombre minimal de ménages répondant dans chaque catégorie.

## 6.4 Poids final

La présente section décrit la manière de calculer les poids finaux qui permettent d'obtenir les estimations officielles. Le calage composite et la méthode de pondération intégrée constituent les éléments clés pour obtenir les poids finaux. La dernière méthode sert à garantir un poids final commun pour chacune des personnes du ménage. Il est d'abord question de calage à la section 6.4.1, puis du calage composite à la section 6.4.2. On décrit la méthode de pondération intégrée à la section 6.4.3. Enfin, à la section 6.4.4, on indique de quelle façon le système de pondération traite les poids finaux négatifs.

### 6.4.1 Calage

Le calage est une technique qui permet d'attribuer à toutes les personnes  $k \in s_r$ , des poids  $w_k^C$ , qui se rapprochent le plus possible des sous-poids  $w_k^{NA}$ , selon une certaine fonction de distance et de manière que les estimations selon le poids de calage pour un vecteur de variables auxiliaires  $\mathbf{x}$ ,  $\hat{\mathbf{t}}_x^C = \sum_{k \in s_r} w_k^C \mathbf{x}_k$ , soient identiques au vecteur de chiffres de population connus,  $\mathbf{t}_x = \sum_{k \in P} \mathbf{x}_k$ . Dans l'EPA, ces chiffres de population connus, souvent appelés totaux de contrôle, sont en fait des estimations du recensement (projections démographiques), pour le mois courant, du nombre de personnes de 15 ans et plus dans les régions économiques (RÉ) et les RMR/régions de recensement (RR), ainsi que du nombre de personnes compris dans les 24 groupes d'âge-sexe par province. On emploie des totaux de contrôle additionnels afin de garantir que le nombre estimatif de personnes de 15 ans et plus est le même pour chaque groupe de renouvellement. Pour procéder au calage, il faut connaître le vecteur  $\mathbf{x}$  pour chaque personne  $k \in s_r$ . Dans le cas de l'EPA, cela signifie qu'il faut savoir à quel groupe d'âge-sexe appartient chaque personne  $k \in s_r$ , de même que sa RÉ et sa RMR/RR.

Plus précisément, on obtient les poids de calage  $w_k^C$  dans l'EPA en minimisant la fonction de distance

$$\sum_{k \in s_r} \frac{(w_k^C - w_k^{NA})^2}{w_k^{NA}}$$

selon la contrainte de calage  $\sum_{k \in s_r} w_k^C \mathbf{x}_k = \mathbf{t}_x$ . On pourrait aussi employer d'autres fonctions de distance (voir Deville et Särndal 1992). Cette minimisation permet d'en arriver aux poids de calage  $w_k^C = w_k^{NA} g_k^C$ , où l'on obtient le facteur de correction par calage  $g_k^C$  au moyen de la formule

$$g_k^C = 1 + \mathbf{x}_k' \left( \sum_{k \in s_r} w_k^{NA} \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \left( \mathbf{t}_x - \sum_{k \in s_r} w_k^{NA} \mathbf{x}_k \right).$$

Le poids de calage obtenu  $w_k^C$  peut aussi être perçu comme un poids de régression. Dans l'EPA, on inclut (implicitement) un point d'intersection distinct pour chaque province dans le vecteur  $\mathbf{x}$  de manière que le nombre total de personnes de chaque province soit implicitement inclus dans le vecteur des totaux de contrôle  $\mathbf{t}_x$ . Cela s'explique par le fait que les 24 groupes d'âge-sexe couvrent toute la population de la province. On peut démontrer que l'inclusion (implicite) du point d'intersection dans le vecteur des variables auxiliaires permet d'en arriver à un facteur de correction par calage qui le réduit à la formule

$$g_k^C = \mathbf{x}_k' \left( \sum_{k \in s_r} w_k^{NA} \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{t}_x.$$

Cette expression simplifiée du facteur de correction par calage est utilisée dans le système d'estimation de la variance.

Comme le souligne l'introduction, on utilise le calage dans l'EPA pour les trois raisons suivantes : (i) garantir la conformité avec les estimations de recensement estimées et avec toutes les enquêtes qui utilisent ces estimations de recensement, (ii) tenir compte, dans une certaine mesure de la sous-couverture et (iii) améliorer l'efficacité des estimations. Afin de tenir compte de la sous-couverture et d'améliorer l'efficacité des estimations, on doit faire la corrélation entre les variables auxiliaires servant au calage et les variables d'intérêt principales. Une façon d'atteindre cet objectif consiste à choisir les variables auxiliaires en modélisant les variables d'intérêt. Ainsi, un modèle approprié permet de constater facilement que l'emploi ou le chômage est lié à l'âge et au sexe d'une personne.

### 6.4.2 Calage composite

Le calage composite (ou l'estimateur composite par régression) est essentiellement identique au calage, sauf que certains totaux de contrôle sont des estimations du mois précédent et que les variables auxiliaires associées à ces totaux de contrôle estimatifs ne sont pas connues pour toutes les personnes  $k \in s_r$  et sont donc imputées. Ces totaux de contrôle et ces variables auxiliaires sont appelés respectivement totaux de contrôle composites et variables auxiliaires composites. Il existe 25 variables auxiliaires composites pour chaque province et elles sont toutes définies par rapport au mois précédent (voir la liste complète à la section 6.5).

Définissons maintenant le vecteur des variables auxiliaires composites pour l'unité  $k$  selon  $\mathbf{z}_{t-1,k}$  et le vecteur correspondant des totaux de contrôle estimatifs selon  $\hat{\mathbf{t}}_z$ . Puisque  $\mathbf{z}_{t-1,k}$  est défini pour le mois précédent (mois  $t-1$ ), on doit calculer les totaux de contrôle estimatifs  $\hat{\mathbf{t}}_z$  à partir des données du mois précédent. Malheureusement, on ne connaît pas le vecteur des variables auxiliaires composites  $\mathbf{z}_{t-1}$  pour les personnes appartenant au groupe de renouvellement naissant, puisqu'elles n'ont pas été interviewées au cours du mois précédent. Pour contourner ce problème, on comble les valeurs manquantes par imputation en combinant deux méthodes d'imputation.

La première méthode, l'imputation par la moyenne, permet d'obtenir le vecteur modifié :

$$\mathbf{z}_{\bullet,k}^{(1)} = \begin{cases} \mathbf{z}_{t-1,k} & \text{si } k \in s_r - s_r^b \\ \hat{\mathbf{t}}_z / N_{15+} & \text{si } k \in s_r^b \end{cases},$$

où  $s_r^b$  correspond au sous-ensemble de personnes  $k \in s_r$  appartenant au groupe de renouvellement naissant et  $N_{15+}$  correspond au nombre de personnes de 15 ans et plus pour la province. Dans une étude empirique précédente, on a pu démontrer que cette méthode d'imputation permettait d'évaluer de façon efficace les paramètres de population définis pour le mois courant  $t$ .

Avec la seconde méthode, on utilise le vecteur modifié :

$$\mathbf{z}_{\bullet,k}^{(2)} = \begin{cases} \mathbf{z}_{t-1,k} + (\delta_k^{-1} - 1)(\mathbf{z}_{t-1,k} - \mathbf{z}_{tk}) & \text{si } k \in s_r - s_r^b \\ \mathbf{z}_{tk} & \text{si } k \in s_r^b \end{cases},$$

où  $\mathbf{z}_{tk}$  correspond au vecteur  $\mathbf{z}_{t-1,k}$  défini pour le mois courant  $t$  et  $\delta_k$  correspond à la probabilité  $k \in s_r - s_r^b$  étant donné  $k \in s_r$ . Dans l'EPA,  $\delta_k = 5/6$ , étant donné  $k \in s_r$ , et cette valeur est remplacée dans l'équation précédente par l'estimation  $\hat{\delta}_k = \sum_{k \in s_r - s_r^b} w_k^{NA} / \sum_{k \in s_r} w_k^{NA}$ . Essentiellement, il s'agit d'effectuer une imputation rétrospective (imputation à partir des valeurs du mois courant pour produire les valeurs du mois précédent) afin d'imputer  $\mathbf{z}_{t-1}$  pour le groupe de renouvellement naissant, car on sait qu'il existe une forte corrélation d'un mois à l'autre pour les variables auxiliaires composites. Cependant, les valeurs de  $\mathbf{z}_{t-1}$  hors des groupes de renouvellement naissant sont modifiées étant donné que l'imputation rétrospective élimine tout changement pour les personnes appartenant au groupe de renouvellement naissant. On détermine la correction hors du groupe de renouvellement naissant de manière à préserver l'absence de biais asymptotique qui caractérise les estimations. Dans une étude empirique précédente, on a constaté que cette méthode

d'imputation (qui permet d'en arriver à  $\mathbf{z}_{\bullet,k}^{(2)}$ ) permettait d'évaluer de façon efficace les paramètres de population définis comme des différences entre deux mois successifs.

En fait, on n'utilise ni  $\mathbf{z}_{\bullet,k}^{(1)}$  ni  $\mathbf{z}_{\bullet,k}^{(2)}$  dans l'enquête. Les variables auxiliaires composites sont plutôt définies à partir de la formule

$$\mathbf{z}_{\bullet,k} = (1 - \alpha)\mathbf{z}_{\bullet,k}^{(1)} + \alpha\mathbf{z}_{\bullet,k}^{(2)},$$

où  $\alpha$  est une constante de réglage choisie qui équivaut à 2/3. On parvient ainsi à un compromis entre les deux méthodes d'imputation. Chen et Liu (2002) ont effectué une étude sur le choix de  $\alpha$ . Bocci et Beaumont (2005) ont également étudié d'autres méthodes d'imputation qui font appel à l'imputation calée.

On obtient les poids de calage composite de l'EPA  $w_k^{CC}$  en minimisant la fonction de distance donnée à la section 6.4.1, laquelle a permis d'obtenir les poids de calage  $w_k^C$ , sauf que la contrainte  $\sum_{k \in s_r} w_k^C \mathbf{x}_k = \mathbf{t}_x$  est remplacée par la contrainte

$$\sum_{k \in s_r} w_k^{CC} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{z}_{\bullet,k} \end{pmatrix} = \begin{pmatrix} \mathbf{t}_x \\ \hat{\mathbf{t}}_z \end{pmatrix}.$$

À noter que les poids de calage composite  $w_k^{CC}$  sont toujours calés en fonction des totaux de contrôle habituels  $\mathbf{t}_x$ . On les obtient par la formule  $w_k^{CC} = w_k^{NA} g_k^{CC}$ , où le facteur de correction par calage composite  $g_k^{CC}$  présente la même forme que  $g_k^C$ , sauf que  $\mathbf{x}_k$  et  $\mathbf{t}_x$  sont remplacés par  $(\mathbf{x}'_k, \mathbf{z}'_{\bullet,k})'$  et  $(\mathbf{t}'_x, \hat{\mathbf{t}}'_z)$  respectivement. Vous pouvez trouver d'autres renseignements sur le calage composite de l'EPA dans les rapports de Singh, Kennedy et Wu (2001), de Fuller et Rao (2001) et de Gambino, Kennedy et Singh (2001). Gambino, Kennedy et Singh (2001) traitent aussi des questions liées aux personnes manquantes ou inadmissibles pour le mois précédent, hors des groupes de renouvellement naissants. On impute les valeurs manquantes au moyen de la méthode d'imputation par hot deck aléatoire et on applique la formule  $\mathbf{z}_{\bullet,k} = \mathbf{0}$  aux personnes inadmissibles pour le mois précédent. Il s'agit de définir  $\mathbf{z}_{\bullet,k}$  pour que  $\sum_{k \in s_r} w_k^{NA} \mathbf{z}_{\bullet,k}$  demeure, tout comme  $\hat{\mathbf{t}}_z$ , une estimation du vecteur inconnu des totaux de contrôle  $\mathbf{t}_z$ , laquelle est définie pour le mois précédent. On traite de la façon habituelle les valeurs pour les personnes manquantes ou inadmissibles pour le mois courant.

On a recours au calage composite afin d'améliorer l'efficacité des estimations. On obtient un accroissement important de l'efficacité des estimations pour  $\mathbf{z}_t$  s'il y a une forte corrélation d'un mois à l'autre entre  $\mathbf{z}_t$  et  $\mathbf{z}_{t-1}$ . Cet accroissement est attribuable au

chevauchement de l'échantillon de l'EPA. D'un côté, on y gagne en efficacité, car le calage composite utilise l'information provenant du groupe de renouvellement final pour le mois précédent. D'un autre côté, on y perd en efficacité en raison des valeurs manquantes dans le groupe de renouvellement naissant et de l'imputation de  $\mathbf{z}_{t-1}$ . Dans l'ensemble, on a constaté de façon empirique que le calage composite profite à l'EPA.

#### 6.4.3 Méthode de pondération intégrée

Étant donné que certaines variables auxiliaires et toutes les variables auxiliaires composites sont définies à l'échelle de la personne, les poids de calage  $w_k^C$  et les poids de calage composite  $w_k^{CC}$  ne sont pas constants dans le ménage, contrairement aux sous-poids  $w_k^{NA}$ . Cela ne pose aucun problème tant qu'on cherche à évaluer les paramètres de population liés aux personnes, comme le nombre total de personnes ayant un emploi au sein de la population. Dans l'EPA, on cherche parfois à évaluer les paramètres de population liés aux ménages, quoique dans une moindre mesure. Par exemple, on pourrait chercher à évaluer le nombre total de ménages ayant une certaine caractéristique en commun, notamment la présence d'au moins un membre occupant un emploi. Il existe plus d'une possibilité de pondération pour ces paramètres de population.

Pour éviter de produire deux ensembles de poids finaux, on a adopté dans l'EPA la méthode de pondération intégrée qui permet d'obtenir un seul ensemble de poids finaux pour les paramètres de population liés tant aux personnes qu'aux ménages; voir Lemaître et Dufour (1987). Avec cette méthode, le poids de calage composite final est constant pour tous les membres du ménage. Il s'agit de remplacer  $\mathbf{x}_k$  et  $\mathbf{z}_{\bullet k}$  pour une personne donnée  $k$  par la moyenne de  $\mathbf{x}$  et  $\mathbf{z}_{\bullet}$  pour tous les membres du même ménage, puis de calculer les poids de calage ou les poids de calage composite conformément à la section 6.4.1 ou 6.4.2. Cela garantit un poids final commun pour tous les membres du ménage. Cette contrainte additionnelle sur les poids finaux devrait réduire l'efficacité des estimations. Cependant, Pandey, Alavi et Beaumont (2003) ont constaté de façon empirique que la réduction de l'efficacité est faible dans le contexte de l'EPA.

#### 6.4.4 Traitement des poids finaux négatifs

Parfois, on obtient des poids finaux négatifs. On répète alors le calage composite, sauf qu'on remplace les sous-poids par les poids finaux lorsqu'ils sont positifs et qu'on les conserve tels quels lorsque les poids finaux sont négatifs. S'il subsiste encore des poids finaux négatifs après ce second calage composite, on

attribue à ces poids négatifs une valeur égale à 1 et on accepte que la contrainte de calage composite ne soit pas respectée. Cela se produit rarement.

### 6.5 Variables auxiliaires composites, définies pour la province

Un astérisque (\*) indique qu'il n'est pas nécessaire de préciser la variable auxiliaire, car on peut la déduire à partir d'autres variables auxiliaires.

#### Caractéristiques de la population active du mois précédent (non ventilées)

Personnes occupées, 15+  
Chômeurs(euses), 15+  
\* *Inactifs*, 15+

#### Caractéristiques de la population active du mois précédent par groupes d'âge-sexe

Hommes occupés, 15 à 24  
Chômeurs, 15 à 24  
\* *Hommes inactifs*, 15 à 24

Hommes occupés, 25+  
Chômeurs, 25+  
\* *Hommes inactifs*, 25+

Femmes occupées, 15 à 24  
Chômeuses, 15 à 24  
\* *Femmes inactives*, 15 à 24

\* *Femmes occupées*, 25+  
\* *Chômeuses*, 25+  
\* *Femmes inactives*, 25+

#### Emploi pour le mois précédent par secteur d'activité

Personnes occupées - agriculture, 15+  
Personnes occupées - construction, 15+  
Personnes occupées - information, culture et loisirs, 15+  
Personnes occupées - services publics, 15+  
Personnes occupées - secteur manufacturier, 15+  
Personnes occupées - ressources naturelles, 15+  
Personnes occupées - transport et entreposage, 15+  
Personnes occupées - finance, assurance et immobilier, 15+  
Personnes occupées - services professionnels, scientifiques et techniques, 15+  
Personnes occupées - gestion d'entreprises, services administratifs et autres services de soutien, 15+  
Personnes occupées - secteur d'enseignement, 15+  
Personnes occupées - soins de santé et assistance sociale, 15+  
Personnes occupées - hébergement et services de restauration, 15+  
Personnes occupées - administrations publiques, 15+  
Personnes occupées - commerce, 15+  
\* *Personnes occupées dans d'autres services*, 15+

#### Emploi pour le mois précédent par catégorie de travailleur

Personnes occupées, employés du secteur public, 15+  
Personnes occupées, employés du secteur privé, 15+  
\* *Personnes occupées, secteur privé, travail autonome*, 15+

## Chapitre 7 Estimation de la variance

### Introduction

Dans une enquête fondée sur un échantillon probabiliste, comme l'EPA, les inférences statistiques doivent tenir compte de l'erreur d'échantillonnage. La variance mesure la précision d'un estimateur. En raison de la complexité de la méthode d'estimation et du plan d'échantillonnage, il se peut qu'on ne puisse disposer d'une forme explicite de l'estimateur de variance. Il est toutefois possible d'obtenir une estimation de la variance à partir des données tirées de l'échantillon.

L'EPA fait appel à la méthode du jackknife pour évaluer la variabilité d'échantillonnage. La section 7.1 décrit l'application de la méthode du jackknife dans le contexte des échantillons transversaux et des passages répétés de l'EPA, tandis que la section 7.2 présente les étapes principales de l'estimation de la variance dans le cadre de l'EPA. La section 7.3 porte sur le système informatique développé à cette fin.

### 7.1 La méthode du jackknife

L'EPA fait appel à la méthode du jackknife pour estimer la variance d'échantillonnage de l'estimateur composite par régression (ou le calage composite), utilisée pour chacune des dix provinces et de l'estimateur de régression généralisée, utilisé pour chacun des trois territoires. La méthode du jackknife est une méthode de rééchantillonnage qui nécessite que les strates et les répliques soient définies aux fins de l'estimation de la variance. Dans le plan de sondage de l'EPA de 2004, les strates de variance et les répliques pour toutes les provinces et dans bien des cas, pour les trois territoires aussi, sont respectivement identiques aux strates du plan et aux unités primaires d'échantillonnage (UPE). Cela concorde avec le point de vue de Särndal, Swensson et Wretman (1992), observation 11.5.2, : « Pour l'échantillonnage à plusieurs degrés, on emploie habituellement la technique jackknife au niveau de l'UPE ». En raison de la sélection d'une seule UPE dans quelques strates pour les territoires, on exige parfois le regroupement des strates afin d'avoir au moins deux répliques par strate de variance. Lorsqu'on regroupe deux strates ou plus, chaque strate originale à UPE simple devient alors une réplique aux fins de l'estimation de la variance.

Pour chaque province et chaque territoire désignées par la lettre  $p$ , pour la strate de variance  $ph$  et une réplique donnée  $phi$ , on obtient une réplique jackknife  $\hat{Y}_{p(phi)}$  d'un estimateur de total mensuel  $\hat{Y}_p$  comme il est indiqué ci-dessous. On répète la méthode d'estimation à

l'échelle de la province ou du territoire, après avoir omis les dossiers d'échantillons pour la réplique  $phi$  et multiplié les sous-poids des dossiers restants dans la strate de variance  $ph$  par le facteur  $A_{ph}/(A_{ph}-1)$ , où  $A_{ph}$  correspond au nombre de répliques dans la strate  $ph$ , afin de tenir compte des dossiers omis. De même, à l'échelle du Canada, on obtient une réplique jackknife  $\hat{Y}_{C(phi)}$  d'un estimateur du total mensuel  $\hat{Y}_C$  en répétant la procédure de calage à l'échelle du Canada, après avoir omis les dossiers d'échantillons pour la réplique  $phi$  et corrigé comme auparavant les sous-poids des dossiers restants dans la strate de variance  $ph$ .

On obtient un estimateur de la variance d'échantillonnage d'une estimation du total mensuel pour une province ou un territoire  $\hat{Y}_p$  au moyen de la formule

$$\hat{V}(\hat{Y}_p) = \sum_{h=1}^{H_p} [(A_{ph}-1)/A_{ph}] \sum_{i=1}^{A_{ph}} [\hat{Y}_{p(phi)} - \hat{Y}_p]^2, \quad (1)$$

tandis qu'on obtient un estimateur de la variance d'échantillonnage d'une estimation à l'échelle du Canada  $\hat{Y}_C$  au moyen de la formule

$$\hat{V}(\hat{Y}_C) = \sum_p \sum_{h=1}^{H_p} [(A_{ph}-1)/A_{ph}] \sum_{i=1}^{A_{ph}} [\hat{Y}_{C(phi)} - \hat{Y}_C]^2, \quad (2)$$

où  $H_p$  correspond au nombre de strates de variance pour la province ou le territoire  $p$ . La variance des totaux mensuels est additive pour les provinces, puisque l'échantillonnage se fait de façon indépendante d'une province ou d'un territoire à l'autre et des méthodes d'estimation s'appliquent individuellement à chaque province ou territoire, de manière que  $\hat{Y}_{C(phi)} - \hat{Y}_C = \hat{Y}_{p(phi)} - \hat{Y}_p$ ; toutefois, ce n'est pas le cas pour quantité d'autres statistiques telles que les taux.

Singh, Kennedy et Wu (2001) énoncent les conditions nécessaires qui garantissent la validité de la méthode du jackknife pour les enquêtes transversales ou répétées. Les estimations pour les répliques doivent avoir une moyenne et une variance identiques et la sélection des répliques doit prévoir ou présumer un remplacement. Si l'on sélectionne les répliques sans remplacement, l'estimateur de la variance jackknife devient conservateur quand la covariance entre les répliques est négative. Dans le cas des enquêtes répétées, si les répliques sont communes (ou reliées) au fil du temps, la méthode du jackknife doit en tenir compte.

Comme il est indiqué à la section 3.4, le renouvellement des unités d'échantillonnage survient à chaque étape du plan échantillonnage à plusieurs degrés.

Des ménages attribués à une UPE sont supprimés de l'échantillon tous les six mois, tandis que les UPE comme telles sont supprimées après quelques années. On attribue aux UPE, que le renouvellement vient ajouter à l'échantillon, le même numéro de réplique qu'aux UPE qu'elles remplacent. Ainsi, on peut créer des vecteurs répétés, représentant plusieurs mois de données, pour évaluer, par exemple, la variance d'une moyenne mobile de trois mois.

Le nombre de strates de variance et les répliques peuvent différer d'un mois à l'autre. Cela peut se produire parce qu'une zone est devenue inaccessible en raison du mauvais temps, de l'ajout ou du retrait des UPE à la suite d'une augmentation ou d'une diminution de la taille de l'échantillon ou de la présence d'une UPE dont l'échantillon est plutôt petit et qui n'affiche aucune donnée réelle ou imputée sur les répondants depuis des mois (tous les logements sont vacants, saisonniers, en construction ou occupés par des personnes qui ne doivent pas être interviewées). L'ensemble de vecteurs répétés, p. ex., au cours d'une période de six mois, réunit les six ensembles mensuels de répliques. Il est donc possible, si les ensembles mensuels de répliques ne sont pas identiques, que la réunion contienne des strates de variance ou des répliques qui n'étaient pas présentes dans un mois donné. L'estimation obtenue pour la province viendrait alors remplacer les répliques manquantes.

D'après la description précédente de l'estimateur de la variance jackknife, chaque strate de variance doit contenir au moins deux répliques. Toutefois, une seule UPE est sélectionnée à partir de quelques strates isolées du plan de l'EPA de 2004, quand le coût de collecte est important et quand le nombre total d'UPE dans chaque strate est petit. Afin de satisfaire à l'exigence de sélectionner au moins deux répliques dans chaque strate, on peut opter pour une des deux stratégies suivantes : (i) regrouper les strates d'une UPE par province et voir chaque strate comme une réplique ou (ii) présumer qu'on sélectionne deux UPE dans chaque strate, mais que la seconde UPE représente une pseudo-population et ne produit aucune estimation. Une étude de simulation à partir des données du Recensement de la population de 2001 a démontré qu'il y a peu de différence entre les deux stratégies pour les estimations et les estimations de la variance et on a donc adopté la seconde stratégie, qui est déjà intégrée au système d'estimation de la variance de l'EPA.

Pourvu que les répliques soient communes ou reliées au fil du temps, la méthode du jackknife s'applique à

une fonction linéaire ou non linéaire des totaux mensuels, telle qu'un taux mensuel, une différence des taux d'un mois à l'autre ou une moyenne annuelle. Il est possible de calculer l'estimation de la variance jackknife d'une différence des totaux d'un mois à l'autre pour une caractéristique donnée  $Y$ , par exemple, à partir de l'équation (1) après avoir remplacé  $\hat{Y}_R$  et  $\hat{Y}_{p(phi)}$ , respectivement, par  $\hat{Y}_{p1} - \hat{Y}_{p2}$  et  $\hat{Y}_{p1(phi)} - \hat{Y}_{p2(phi)}$ , où les indices numériques renvoient au mois et  $phi$  renvoie aux répliques au fil du temps. De façon plus générale, on obtient l'estimation de la variance d'une fonction linéaire ou non linéaire  $g$  des totaux mensuels pour une province ou un territoire  $p$  à partir de l'équation (1) après avoir remplacé  $\hat{Y}_p$  et  $\hat{Y}_{p(phi)}$  par  $g(\hat{Y}_{p1}, \hat{Y}_{p2}, \dots, \hat{Y}_{pM})$  et  $g(\hat{Y}_{p1(phi)}, \hat{Y}_{p2(phi)}, \dots, \hat{Y}_{pM(phi)})$ , respectivement, alors que  $\hat{Y}_{pm}$  représente un vecteur des estimations pour le mois  $m$  et la province ou le territoire  $p$ , et  $\hat{Y}_{pm(phi)}$  représente un vecteur connexe de répliques jackknife. L'équation (2) sert, de façon semblable, à estimer la variance d'une fonction linéaire ou non linéaire des totaux mensuels à l'échelle du Canada. L'équation (2) se prête bien à l'estimation de la variance d'échantillonnage des taux pour plus d'une province, par exemple les taux de chômage à l'échelle canadienne, que l'on évalue en tant que ratios mixtes dans l'EPA.

Comme méthode de rééchantillonnage pour l'estimation de la variance, la méthode du jackknife nécessite beaucoup de calculs, mais s'avère simple à appliquer. Dans le contexte de l'EPA, cette méthode demeure plutôt simple à utiliser. Rao (2005) et Särndal, Swensson et Wretman (1992) comparent les méthodes courantes de rééchantillonnage.

## 7.2 Estimation de la variance pour l'estimation composite par régression et l'estimateur par régression généralisée

Afin d'évaluer la variabilité d'échantillonnage d'un estimateur, on devrait, en théorie, répéter chacune des étapes menant au calcul des poids finaux pour chaque réplique jackknife. Dans le cas de l'EPA toutefois, on répète uniquement la méthode de calage. On ne calcule qu'une seule fois les poids de grappe, de stabilisation et de correction de non-réponse, qui permettent d'obtenir les sous-poids utilisés pour le calage. Les sections 6.4.1 et 6.4.2 décrivent respectivement le calage pour l'estimateur par régression généralisée et le calage pour l'estimation composite par régression. Comme nous le verrons au paragraphe suivant, l'estimation de la variance pour l'estimateur composite par régression, qui utilise les données de deux mois consécutifs, s'avère

plus complexe que pour l'estimateur par régression généralisée.

On obtient les poids finaux composites pour la province en trois étapes :

- i) À partir du fichier de totalisation du mois précédent, en utilisant les poids finaux composites, obtenir les estimations des caractéristiques clés du marché du travail du mois précédent, corrigées afin de concorder avec les chiffres de population du mois courant. Cette étape permet d'obtenir les totaux de contrôle composite par régression.
- ii) Fusionner les dossiers des fichiers de totalisation du mois courant et du mois précédent pour les cinq groupes de renouvellement communs aux deux fichiers, afin d'imputer les variables secondaires composites pour les dossiers manquants dans le fichier du mois précédent. Joindre le fichier obtenu aux dossiers du groupe de renouvellement naissant du fichier de totalisation du mois courant. On imputera les variables secondaires composites manquantes pour le renouvellement à partir de la moyenne, tel qu'il est indiqué à la section 6.4.2.
- iii) Calibrer les sous-poids du mois courant à partir du fichier obtenu à l'étape ii) jusqu'aux totaux de contrôle composites par régression et démographiques du mois courant en utilisant la procédure de calage par régression généralisée. Cette étape permet d'obtenir les poids finaux composites du mois courant.

Pour l'estimation de la variance, on répète les étapes i) et iii) dans le cadre du processus de calage pour chaque réplique jackknife, tandis qu'on procède une seule fois à l'étape ii) pour la province. On peut également signaler qu'on ne tient pas compte de la présence de poids négatifs pendant l'estimation de la variance, tel qu'il est mentionné à la section 6.4.4.

Comme il est indiqué au chapitre précédent, les répliques dans les strates de variance sont communes ou reliées au fil du temps afin d'évaluer, par exemple, la variance d'une moyenne annuelle. Pour évaluer la variance de l'estimateur composite pendant la période d'introduction progressive de l'échantillon, lorsqu'on présente un nouveau plan, on a relié les répliques des plans d'échantillonnage de 1984 et 1994, de même que celles des plans de 1994 et 2004. Pour l'appariement, on a utilisé un fichier de chevauchement de grappes entre des plans d'échantillonnage consécutifs. Même si un tel appariement entre des plans d'échantillonnage consécutifs ne peut être qu'approximatif, il propose une transition acceptable des estimations de la variance pendant la période d'introduction progressive de l'échantillon.

### **7.3 Le système d'estimations de la variance de l'EPA**

On a conçu le système d'estimation de la variance de l'EPA afin d'atteindre les objectifs suivants : (en utilisant la méthode du jackknife), (i) obtenir les estimations et les estimations de la variance des totaux mensuels des fonctions linéaires ou non linéaires définies par l'utilisateur, par exemple une moyenne mobile de trois mois d'emploi total par province ou une différence des taux de chômage dans deux domaines au cours du même mois et (ii) évaluer les effets des totaux mensuels du plan d'échantillonnage, pour les méthodes d'estimation composite par régression et les méthodes par régression généralisée. On peut actuellement obtenir les estimations de la variance pour les plans d'échantillonnage de 1984, 1994 et 2004, ainsi que pour la période d'introduction progressive de l'échantillon entre des plans consécutifs. Le programme s'applique à l'EPA de même qu'aux autres enquêtes stratifiées à plusieurs degrés qui utilisent le même type d'estimateurs.

## Chapitre 8 Qualité des données

### Introduction

Les estimations de l'EPA, comme celles produites à l'aide de toute enquête-échantillon, peuvent comporter des erreurs d'échantillonnage et des erreurs non dues à l'échantillonnage. Conséquemment, pour interpréter correctement les estimations de cette enquête, il faut une connaissance de leur qualité.

Dans une enquête par sondage, des inférences sont faites au sujet de la population visée à partir des données recueillies auprès d'une partie seulement de cette population. Les résultats sont probablement différents de ceux qu'on obtiendrait si on menait un recensement complet de cette population dans les mêmes conditions. L'erreur causée par l'étendue des conclusions fondées sur un échantillon seulement à toute la population est appelée erreur d'échantillonnage. Au nombre des facteurs qui contribuent aux erreurs d'échantillonnage, on trouve : la taille de l'échantillon, la variabilité des caractéristiques étudiées, le plan d'échantillonnage et la méthode d'estimation.

L'erreur non due à l'échantillonnage, comme son nom l'indique, n'a rien à voir avec le processus d'échantillonnage et se produit dans un recensement aussi bien que dans une enquête par sondage. Ce type d'erreur peut survenir à n'importe quelle étape d'une enquête (planification, conception, collecte des données, codage, saisie, vérification, estimation, analyse et diffusion des données) et est principalement attribuable à des erreurs humaines. On peut également associer l'erreur non due à l'échantillonnage à d'autres types d'erreurs, par exemple à des erreurs dans les sources d'information, les méthodes utilisées pour obtenir des projections de population, des erreurs d'ajustements saisonniers, etc.

Pour assurer et contrôler la qualité de ses données, l'EPA s'est dotée d'un programme élaboré pour mesurer la qualité des données. Une gamme d'indicateurs de qualité sont produits sur une base régulière et analysés avec soin. En présence de valeurs inhabituelles, les responsables de l'EPA sont immédiatement avisés de faire les corrections nécessaires le plus rapidement possible. Par ailleurs, certains indicateurs sont contrôlés d'une façon moins régulière puisque leur rôle est de déceler des tendances ou des effets à long terme. À titre d'exemple, certains indicateurs permettent de mesurer les conséquences de certains changements d'ordre opérationnel alors que d'autres mesurent l'effet de modifications mineures au plan de sondage. Ces

renseignements à long terme au sujet de la fiabilité des données peuvent servir à apporter des changements susceptibles d'améliorer la qualité générale des résultats et d'aider les analystes et les utilisateurs de données, tant à l'interne qu'à l'externe, dans leur travail. Dans les lignes qui suivent, les indicateurs de qualité produits pour l'EPA sont présentés sous deux rubriques : erreur d'échantillonnage et erreurs non dues à l'échantillonnage.

### 8.1 Erreurs d'échantillonnage

Les répercussions de l'erreur d'échantillonnage sur les estimations de l'enquête sont fonction de plusieurs facteurs (voir section 2.2 pour une définition de l'erreur d'échantillonnage). Le plus évident est la taille de l'échantillon. Tout autre facteur étant constant, l'erreur d'échantillonnage diminue généralement au fur et à mesure que la taille de l'échantillon augmente. Outre la taille d'échantillon, l'erreur d'échantillonnage dépend de facteurs tels que la variabilité de la population, la méthode d'estimation et le plan de sondage.

Pour un échantillon d'une taille donnée, l'erreur d'échantillonnage est liée à diverses caractéristiques du plan de sondage comme la méthode de stratification utilisée, la répartition de l'échantillon, le choix des unités d'échantillonnage et la méthode de sélection employée à chaque degré d'échantillonnage pour un plan à plusieurs degrés.

De même, pour un plan de sondage donné, la méthode d'estimation utilisée joue un rôle important. L'EPA a apporté une modification majeure à la méthode d'estimation. Cette nouvelle méthode, l'estimation composite (voir chapitre 6), réduit l'erreur d'échantillonnage de façon significative.

Enfin, l'erreur d'échantillonnage varie d'une variable à l'autre puisque le degré de variabilité diffère également d'une variable à une autre. Ces erreurs sont généralement plus grandes pour les caractéristiques qui sont relativement rares. Elles sont également plus importantes lorsque la caractéristique d'intérêt n'est pas répartie uniformément dans la population. Ainsi, bien qu'elles se fondent sur le même échantillon, les estimations relatives au chômage comportent généralement une erreur d'échantillonnage plus élevée que les données relatives à l'emploi.

L'une des principales caractéristiques d'un échantillon probabiliste, comme celui utilisé dans le cadre de l'EPA, est que la variance d'échantillonnage peut être

estimée au moyen de l'échantillon. Le chapitre 7 décrit la méthode présentement utilisée pour estimer la variance des estimations produites par l'EPA.

Trois mesures sont dérivées de la variance d'échantillonnage : l'erreur type, le c.v. et l'effet de plan (voir section 2.2). L'écart type peut servir à calculer l'intervalle de confiance associée à une estimation. L'intervalle de confiance est construit autour de l'estimation obtenue et sa largeur dépend de l'écart type.

Pour mettre en lumière les liens entre les différentes mesures de précision, prenons l'exemple suivant. En mars 2005, le taux de chômage de la population canadienne âgée de 15 ans et plus était de 7,4 % et l'estimation de l'écart-type correspondant était de 0,0013. Par conséquent, le coefficient de variation est 1,8 %. (0,0013/0,074). L'intervalle de confiance de 95 % se situe entre 7,14 % et 7,66 %, soit  $0,074 \pm 0,0026$ . Cela signifie que si le processus de sélection était répété un grand nombre de fois, 95 % des intervalles de confiance contiendraient la valeur que l'on obtiendrait à l'aide d'un recensement.

Compte tenu de leur stabilité, les c.v. inclus dans la publication mensuelle de l'EPA ne sont pas mis à jour chaque mois. On fournit plutôt une estimation du coefficient de variation correspondant à la moyenne des c.v. de l'année précédente. Ces estimations sont mises à jour deux fois par année (janvier à juin et juillet à décembre). Le tableau suivant fournit les c.v. observés pour les estimations mensuelles de l'emploi et du chômage.

**Tableau 8.1 Coefficients de variation (c.v.) mensuels observés, 2005**

Province	Employés		Chômeurs
	%		
Terre-Neuve et Labrador	1,6	5,7	
Île-du-Prince-Édouard	1,4	7,5	
Nouvelle-Écosse	0,9	5,6	
Nouveau-Brunswick	1,0	5,5	
Québec	0,6	4,0	
Ontario	0,4	3,2	
Manitoba	0,7	7,0	
Saskatchewan	0,8	7,4	
Alberta	0,6	5,3	
Colombie-Britannique	0,7	5,3	
Canada	0,3	1,9	

Grâce aux données recueillies dans le cadre de l'EPA, il est possible de produire des milliers d'estimations relatives aux caractéristiques de la population. Il est également possible d'obtenir des estimations du changement d'un mois à l'autre, des estimations annuelles, des estimations nationales, provinciales et infraprovinciales. En raison des contraintes d'espace

dans les publications courantes et spéciales, il n'est pas possible d'inclure les c.v. de toutes les estimations d'enquête publiées. Toutefois, il existe des tableaux qui présentent les c.v. approximatifs pour différents groupes d'estimations. Ces tableaux sont disponibles dans la Revue chronologique de la population active (71F0004X au catalogue de Statistique Canada).

Les estimations du changement d'un mois à l'autre ont pris une grande importance au fil du temps. À cet égard, le communiqué mensuel de l'EPA indique maintenant les écarts-types (ET) relatifs aux variations à l'échelle provinciale et nationale pour les employés et les chômeurs. Ces chiffres sont donnés pour la période de 2005 dans le tableau ci-dessous.

**Tableau 8.2 Écart-type de la variation d'un mois à l'autre Employés et chômeurs**

Province	Employés	Chômeurs
	milliers	
Terre-Neuve et Labrador	3	2
Île-du-Prince-Édouard	1	1
Nouvelle-Écosse	4	3
Nouveau-Brunswick	3	2
Québec	18	14
Ontario	20	15
Manitoba	4	3
Saskatchewan	3	2
Alberta	9	6
Colombie-Britannique	12	9
Canada	32	24

Par ailleurs, on peut utiliser l'effet du plan en tant qu'indice de la détérioration du plan d'échantillonnage avec le temps. Dans le cadre de l'EPA, on calcule deux types d'effet du plan, chacun dépendant des données utilisées pour l'établir. On détermine l'*effet du plan non ajusté* à partir de poids non calés, c'est-à-dire sans ajustement tenant compte des totaux de population. On calcule l'*effet du plan ajusté* au moyen de poids finaux. Par conséquent, l'effet du plan non ajusté montre l'efficacité du plan d'échantillonnage, tandis que l'effet du plan ajusté donne une évaluation générale de la stratégie adoptée en combinant toutes les caractéristiques du plan de sondage (stratification, échantillonnage à plusieurs degrés, stratification a posteriori et estimation). Plus l'effet est faible, plus le plan est efficace en ce qui concerne la variance de l'échantillonnage. L'observation de l'effet du plan permet donc de mesurer les changements qualitatifs subis par le plan avec le temps. Soulignons que les effets du plan non ajusté (plan d'échantillonnage) sont généralement plus importants que les effets du plan ajusté (plan de sondage) fondés sur les pondérations finales, puisqu'ils ne profitent pas

du gain de précision qu'apporte la stratification a posteriori.

Dans le cadre de l'EPA, on utilise l'effet du plan non ajusté, conjointement à d'autres renseignements, pour déterminer les régions où le plan de sondage a perdu une part importante de son efficacité au fil du temps. Dans certains cas, il sera nécessaire d'effectuer un mini remaniement dans ces régions afin de pallier ce problème. Le tableau suivant présente quelques valeurs représentatives des effets du plan ajustés et non ajustés pour les caractéristiques emploi et chômage à l'échelle nationale et provinciale, selon les données d'enquête de janvier à juin 2004.

**Tableau 8.3 Effets du plan - employés et chômeurs, 2004**

Province	Employés		Chômeurs	
	Ajusté	Non-ajusté	Ajusté	Non-ajusté
Terre-Neuve et Labrador	0,68	2,44	1,10	1,23
Île-du-Prince-Édouard	0,42	1,86	1,27	1,21
Nouvelle-Écosse	0,40	2,81	1,07	1,06
Nouveau-Brunswick	0,50	3,06	1,28	1,35
Québec	0,35	2,77	1,02	1,11
Ontario	0,32	6,21	1,03	1,19
Manitoba	0,23	3,73	0,97	1,03
Saskatchewan	0,31	3,94	0,97	0,93
Alberta	0,28	9,56	1,03	1,26
Colombie-Britannique	0,33	3,08	1,01	1,10
Canada	0,33	4,87	1,03	1,16

## 8.2 Erreurs non dues à l'échantillonnage

Les erreurs non dues à l'échantillonnage peuvent survenir à toutes les étapes d'une enquête et sont causées en général par des erreurs humaines telles des erreurs d'inattention, de mauvaise compréhension ou d'interprétation. L'effet sur les estimations peut se manifester sous la forme de biais et/ou de variabilité accrue dans les estimations. L'effet net de la variance non due à l'erreur d'échantillonnage peut être négligeable si le nombre d'observations est élevé ou s'il s'agit de grands secteurs. Par contre, son effet peut être élevé lorsqu'il s'agit de petits secteurs ou lorsque les caractéristiques à l'étude sont rares ou rattachées à des questions délicates.

Par ailleurs, l'effet net du biais non dû à l'échantillonnage a tendance à être cumulatif. Il peut être attribuable à la formation ou à l'attitude de l'intervieweur, à une mauvaise conception du questionnaire ou à la méthode d'imputation utilisée pour pallier la non-réponse. L'ensemble de ces facteurs peut contribuer à provoquer l'accumulation des erreurs dans une direction plutôt qu'une autre.

La variance et/ou le biais non dus à l'échantillonnage peuvent provenir de différentes sources. Dans ce qui suit, on s'intéresse tout d'abord à la couverture, à la

non-réponse, à la vacance, à la réponse, au traitement et aux activités sur le terrain.

### 8.2.1 Erreur de couverture

Les erreurs de couverture peuvent se produire à plusieurs étapes de l'enquête : pendant la conception de la base de sondage, lors de l'identification des logements et/ ou des personnes à inclure dans l'enquête ou lors de la collecte et du traitement des données. En ce qui concerne l'EPA, l'indicateur utilisé pour mesurer l'erreur de couverture s'appelle le *taux de glissement*. Par définition, ce taux représente l'écart entre les estimations de la taille de la population produites à partir des poids avant calage et les plus récentes estimations démographiques du recensement. Les estimations démographiques utilisées dans la détermination du taux de glissement peuvent également comporter des erreurs, et ces erreurs sont en fait un des facteurs qui contribuent au glissement. Dans le cadre de l'EPA, on observe du sous-dénombrement; il se traduit par un taux de glissement positif. Pour réduire au maximum le biais qui en résulte, on ajuste le poids associé à chaque répondant en fonction de totaux de contrôle provenant de sources indépendantes (voir chapitre 6).

L'omission de logements ou de personnes de la population cible, c'est-à-dire la présence de sous-couverture à l'EPA, peut introduire des erreurs non dues à l'échantillonnage. Par logement, on entend toute construction habitable répondant à certains critères. Les personnes qui vivent dans un logement composent le ménage. Il se peut qu'un logement occupé ne soit pas inscrit dans la liste des UPE pour diverses raisons : l'omission lors de l'établissement de la liste, immeuble en construction durant la dernière vérification, erreurs dans les délimitations de la grappe ou encore classifié vacant par erreur. Il est également possible que des personnes soient oubliées dans un ménage, soit parce que le répondant ne révèle pas leur présence ou encore qu'on leur a attribué un lieu de résidence habituel ailleurs que dans le ménage échantillonné. Les étudiants sont souvent oubliés puisqu'ils résident ailleurs durant leurs études, quoique leur résidence habituelle soit dans l'échantillon. Des erreurs peuvent donc se glisser dans les estimations de l'enquête, si les caractéristiques des personnes non incluses dans l'enquête diffèrent de celles des personnes incluses. Par exemple, si l'enquête n'atteint pas une partie de la population qui est jeune et grandement mobile, qui affiche des taux de chômage plus élevés que la population du même âge dans l'enquête, alors le glissement biaise les estimations du chômage à la baisse. Finalement, comme mentionné

précédemment, les estimations de population ont également un rôle à jouer en ce qui concerne le glissement.

D'autres facteurs pouvant contribuer au glissement de l'EPA ont été identifiés. Par exemple, la population s'accroît entre les remaniements, généralement dans des endroits spécifiques et non pas de manière uniforme. L'échantillon peut surestimer ou sous-estimer cette croissance ou en rendre compte de façon précise. Autre exemple, l'ajustement pour pallier la non-réponse (voir chapitre 5) peut également influencer le glissement. En effet, si les ménages non-répondants comptent moins de membres et s'ils sont représentés dans l'échantillon par des ménages de grande taille, alors il peut y avoir un effet sur le taux de glissement.

**Tableau 8.4 Taux de glissement moyen (%) - Canada par groupes d'âge et Provinces, 2005**

Canada	%
Tous âges	10,3
âges 15 à 19	8,3
âges 20 à 24	18,3
âges 25 à 29	19,8
âges 30 à 39	14,6
âges 40 à 54	8,8
âges 55+	4,7
Terre-Neuve et Labrador	8,4
Île-du-Prince-Édouard	7,0
Nouvelle-Écosse	8,0
Nouveau-Brunswick	9,0
Québec	7,9
Ontario	10,8
Manitoba	6,1
Saskatchewan	7,1
Alberta	13,2
Colombie-Britannique	14,2

Tous les mois, les taux de glissement sont analysés en détail. Ils sont produits mensuellement pour les régions, et à l'échelle nationale (en excluant les territoires) et provinciale, pour douze groupes d'âge-sexe (15-19, 20-24, 25-29, 30-39, 40-54, 55+). Ils sont aussi produits pour chaque territoire, mais sans décomposition pour les groupes d'âge-sexe. Dans le cadre du dernier remaniement de l'EPA, on a produit des séries révisées des taux de glissement. Ces nouvelles séries utilisent les estimations démographiques fondées sur le Recensement de 2001. Le tableau suivant contient les taux de glissement moyens pour l'année civile 2005.

Finalement, on produit périodiquement des estimations du nombre de ménages selon la taille du ménage. Ces estimations fournissent un autre point de vue du glissement.

Tous ces indicateurs permettent de déceler des problèmes potentiels relativement à la couverture de l'échantillon et de réagir en conséquence. Pour y

remédier ou ralentir sa progression, on peut par exemple songer à créer des exercices à l'intention des intervieweurs pour accroître leurs connaissances des règles de composition du ménage, à distribuer un bulletin qui explique ce qu'est le glissement ou le concept de logement multiple ou encore à établir un programme de relistage d'un certain nombre d'UPE jugées en pleine expansion.

Le glissement sera toujours surveillé minutieusement puisqu'il peut induire un biais dans les estimations. De plus, en dépit de l'application d'une méthode d'estimation pour corriger le glissement, on peut s'attendre à ce que subsiste un certain biais dans l'estimation, autre que le biais habituel d'estimation, puisque les caractéristiques des personnes et des logements omis peuvent être différentes de celles des personnes qui sont comprises dans l'échantillon.

### 8.2.2 Non-réponse

Chaque mois, durant la semaine d'enquête, les intervieweurs s'affairent à déterminer quels sont les logements sélectionnés qui contiennent des personnes admissibles à l'enquête. Les logements identifiés non admissibles pour le mois d'enquête le sont pour les raisons suivantes :

- logement hors du champ d'enquête, c'est-à-dire un logement occupé par des personnes ne faisant pas partie de la population cible, par exemple, des membres des Forces armées canadiennes ;
- logement vacant : logement non occupé, logement saisonnier ou logement en construction;
- logement non existant : logement démoli, logement transformé en local d'affaires, maison mobile démenagée ou encore logement abandonné ou inscrit par erreur.

Lorsqu'on a déterminé qu'un logement est admissible à l'enquête, il n'est pas toujours possible de réaliser l'interview, et ce, pour les raisons suivantes :

- non-réponse du ménage : personne à la maison, absence temporaire, interview impossible (mauvais temps, circonstances inhabituelles dans le ménage, etc.), problèmes d'ordre technique ou refus.

L'importance du biais attribuable à la non-réponse est habituellement inconnue, mais on sait qu'elle est directement liée aux différences de caractéristiques entre les groupes d'unités répondantes et les groupes d'unités non-répondantes. Comme l'effet de ce biais s'accroît

avec l'augmentation du taux de non-réponse, on essaie de maintenir le taux de réponse à un niveau aussi élevé que possible durant les activités de collecte.

Depuis 1993, l'EPA se soumet aux normes et lignes directrices de Statistique Canada pour la déclaration des taux de non-réponse. Tous les mois, les taux de non-réponse pondérés et non pondérés sont acheminés à la banque de données centrale sur la non-réponse de Statistique Canada, dont le mandat est de compiler les données longitudinales pour plusieurs enquêtes régulières. Cette base de données exige des taux de non-réponse à l'étape de la collecte et à l'étape de l'estimation.

Le tableau suivant présente les taux de non réponse moyens pour l'année 2005 ainsi que le minimum et le maximum atteints durant cette année. À l'EPA, le maximum pour la non réponse est normalement atteint au mois de juillet, étant donné le haut pourcentage de personnes qui ne sont pas à la maison, et le minimum au mois d'octobre.

**Tableau 8.5 Taux de non-réponse (non pondéré) Canada et provinces, 2005**

Province	Moyenne	Maximum	Minimum
	%		
Terre-Neuve et Labrador	4,2	5,4	3,0
Île-du-Prince-Édouard	3,5	4,8	2,4
Nouvelle-Écosse	6,3	7,3	4,6
Nouveau-Brunswick	4,6	5,4	3,1
Québec	5,4	6,6	3,7
Ontario	4,8	5,7	3,7
Manitoba	3,6	5,4	2,1
Saskatchewan	3,6	4,6	2,4
Alberta	4,9	6,3	3,1
Colombie-Britannique	5,7	6,7	4,5
Canada	4,9	5,5	3,8

Tous les mois, l'EPA produit les taux de non-réponse ventilés selon la cause de la non-réponse (un simple refus, aucun contact, personnes temporairement absentes, problème technique ou autre raison). Ces taux sont analysés avec soin afin de déceler les causes majeures de la non-réponse. On peut ainsi rapidement identifier des problèmes et apporter les correctifs requis.

Depuis 1999, plusieurs facteurs ont perturbé la série du taux de non-réponse à l'EPA. Voir EPA (2005) pour plus de détails. Tout d'abord, l'introduction d'un nouveau système d'Interview assistée par ordinateur (IAO) à l'automne 1999 dans les bureaux régionaux a généré des difficultés techniques causant une hausse de la non-réponse durant le début de son implantation. Le système d'interview téléphonique assistée par ordinateur (ITAO) qui a été introduit en septembre 2000 pour les interviews subséquentes a également eu un effet sur la

série de non-réponse. Au cours de la même période, l'arrivée de nouveaux ordinateurs portatifs a amélioré le rendement de l'application. Par le fait même, la série s'est établie à son niveau habituel, c'est-à-dire, celui juste avant l'introduction du nouveau système IAO. Ensuite, le nouveau plan d'échantillonnage introduit en novembre 2004 a également eu un effet. Il a nécessité l'embauche de nouveaux intervieweurs qui ont tendance à obtenir des taux de non-réponse légèrement plus élevés durant leurs six premiers mois à l'EPA. Finalement, durant la même période, l'introduction de la méthodologie du premier contact par téléphone (PCT) pour les nouveaux logements entrant dans l'échantillon a généré un nouveau type de non-réponse pour les cas transférés au terrain. Pour une revue historique des enjeux de la non-réponse à l'EPA, se reporter à l'article de Sheridan, Drew et Allard (1996).

Depuis l'implantation du PCT (voir chapitre 4), environ le tiers des nouveaux logements sont contactés chaque mois par téléphone pour répondre à l'EPA. Les gestionnaires de l'EPA étaient préoccupés par les effets possibles de ces changements sur les refus. Par conséquent, la sous-section de la qualité portait une attention particulière à cette composante en produisant plusieurs tableaux qui sont analysés de façon minutieuse. On note un taux de non-réponse plus élevé pour les cas qui ont un premier contact par téléphone comparé à ceux qui ont eu un premier contact par téléphone et une visite personnelle. Toutefois, ce taux de non-réponse plus élevé semble être principalement causé par l'augmentation des cas sans contact pour les ménages faisant partie du groupe de PCT.

Les taux de refus pour l'EPA sont habituellement très bas. Les taux mensuels canadiens varient entre 1 et 2 %. Les taux de refus à l'échelle provinciale sont ordinairement du même ordre de grandeur, mais peuvent descendre aussi bas que 0,5 % ou monter aussi haut que 3 %. Le système de collecte permet d'obtenir plus de détails concernant le motif du refus. Il est ainsi possible de surveiller au fil du temps des changements d'attitude de la part des répondants face à l'enquête.

### 8.2.3 Logements vacants

Les logements identifiés correctement comme étant vacants ou inexistantes n'introduisent aucun biais dans les estimations de l'EPA. Par contre, la variance de l'estimation s'en trouve plus élevée puisque l'échantillon compte un nombre moins élevé de ménages. Les intervieweurs de l'EPA retournent visiter les logements vacants tous les mois afin d'interviewer les personnes ciblées par l'enquête qui peuvent s'y être installées

depuis l'enquête précédente. Les logements inexistantes sont tout simplement retirés de la base de sondage. Une attention particulière doit être apportée à la détermination des logements qui sont vacants puisque ceux-ci influencent directement deux autres indicateurs. En effet, si un logement est codé vacant alors que les occupants sont temporairement absents, le taux de non-réponse produit pour l'EPA est quelque peu sous-estimé. Par ailleurs, le taux de glissement s'en trouve surestimé puisque ce logement mal codé aurait dû être considéré lors de la détermination de ce taux. Les intervieweurs se doivent donc de faire un travail très minutieux pour déterminer si un logement est vacant, et par conséquent, hors du champ de l'enquête, ou tout simplement occupé par un ménage temporairement absent et donc dans le champ de l'enquête. À l'EPA, un programme intitulé Programme de vérification des logements vacants a été mis sur pied pour obtenir de l'information sur cette erreur.

Le tableau suivant présente les taux de vacance moyens, les valeurs minimales et maximales pour l'année 2005 à l'échelle provinciale et nationale.

**Tableau 8.6 Taux de vacance (non pondéré) Canada et provinces, 2005**

Province	Moyenne	Maximum	Minimum
	%		
Terre-Neuve et Labrador	15,4	14,9	16,4
Île-du-Prince-Édouard	20,5	18,6	23,0
Nouvelle-Écosse	16,8	15,2	18,7
Nouveau-Brunswick	14,1	13,5	15,2
Québec	14,0	11,9	15,8
Ontario	10,8	10,0	11,3
Manitoba	17,1	16,4	17,7
Saskatchewan	14,7	12,5	15,5
Alberta	8,7	8,1	9,8
Colombie-Britannique	9,5	8,7	9,8
Canada	13,0	12,2	13,5

En général, le taux de vacance est relativement stable. Il affiche une tendance à la hausse plus on s'éloigne du dernier remaniement puisque la base de sondage est moins à jour. Après chaque remaniement, le taux de vacance affiche une tendance à la baisse. Pour cet indicateur de qualité, certaines provinces se distinguent des autres en affichant des taux beaucoup plus bas ou plus élevés.

#### 8.2.4 Erreur de réponse

Cette erreur peut être attribuable à la conception du questionnaire, à la formulation des questions, à la compréhension du répondant, à la façon dont l'interview est menée ainsi qu'aux conditions générales dans lesquelles l'enquête est réalisée. Des erreurs de réponse peuvent se produire au moment où les renseignements

sont fournis, reçus ou entrés dans l'ordinateur. Toutefois, le mode de collecte informatisé permet de réduire certaines de ces erreurs, puisque certaines règles de vérification sont intégrées à l'instrument de collecte et que les conflits doivent être résolus au moment même de l'interview. Il se peut cependant que le répondant interprète mal la question, qu'il ne sache pas la réponse, qu'il ait oublié ou qu'il préfère déformer les faits pour des raisons qui lui sont personnelles. De plus, il arrive que les intervieweurs aient tendance à expliquer les réponses ou à les interpréter de manière différente. Les erreurs de réponse, comme les autres catégories d'erreurs, peuvent avoir une variance et un biais.

Les réponses par procuration (ou par personne interposée), que l'on obtient lorsqu'on recueille auprès d'un membre du ménage des renseignements se rapportant à un autre membre, peuvent aussi entraîner des erreurs de réponse.

Dans les enquêtes répétées, où l'échantillon est constitué d'un certain nombre de panels ou de groupes de renouvellement, l'espérance mathématique des estimations varie légèrement d'un groupe de renouvellement à un autre. Il se produit alors ce qu'on appelle un biais de renouvellement. En ce qui concerne l'EPA, ce biais atteint son plus haut niveau pour le sixième de l'échantillon qui en est à sa première interview. On peut obtenir l'indice de renouvellement en faisant le rapport entre une estimation calculée pour la partie de l'échantillon participant à l'enquête pour un certain nombre de fois (premier mois, deuxième, etc.) et l'estimation calculée pour l'échantillon entier.

Brisebois et Mantel (1996) ont calculé un indice de renouvellement modifié qui tient compte des différences des effets des erreurs dues à l'échantillonnage pour les six groupes de renouvellement. Leur étude a permis de relever plusieurs différences statistiquement significatives parmi les groupes de renouvellement, mais dont l'effet concret est mineur.

#### 8.2.5 Erreurs de traitement

Les erreurs de traitement peuvent se produire à diverses étapes de l'enquête, soit à la saisie, à la validation, à la vérification, au codage, à l'imputation, à la pondération ou à la totalisation des données.

L'utilisation d'un mode de collecte informatisé permet d'éviter des erreurs d'aiguillage dans le questionnaire, puisque maintenant l'application détermine la prochaine question à poser compte tenu des réponses préalablement entrées. De même, certaines règles de vérification sont intégrées au système de

collecte, ce qui permet de détecter et de corriger certaines divergences au moment de l'interview. L'utilisation d'un mode de collecte informatisé permet d'éviter des erreurs d'aiguillage dans le questionnaire, puisque c'est maintenant l'application qui détermine la prochaine question à poser compte tenu des réponses préalablement entrées. De même, certaines règles de vérification sont incorporées au système de collecte, ce qui permet de détecter et de corriger certaines divergences au moment de l'interview.

Durant le processus du traitement, on attribue un code de validation à chaque réponse saisie par l'intervieweur. Ce code indique si le répondant ne connaissait pas la réponse à la question, si la réponse saisie a été rejetée lors de l'application des règles de vérification, si le répondant a refusé de donner la réponse ou si la réponse saisie contient des données superflues qui doivent être effacées. Ce dernier code s'applique uniquement aux données provenant des ordinateurs du bureau central. Pour mieux contrôler les erreurs dues à l'étape de validation, la distribution de ces codes est examinée à l'occasion.

Le module de contrôle sur le terrain fournit également le taux de divergence au contrôle des formulaires. Une divergence est définie comme toute inscription effacée, modifiée ou ajoutée à une zone en blanc après soumission à certains contrôles pour en vérifier la validité. Le taux de divergence au contrôle, qui représente ainsi le pourcentage de divergences sur un questionnaire par rapport au nombre total d'entrées sur le questionnaire. Les taux sont calculés conformément aux divers procédés de vérification appliqués au bureau central pour l'EPA. Les taux de divergence au contrôle correspondants s'élèvent respectivement à 0,1 et 1 %.

Les codages automatiques et manuels des variables profession et secteur d'activité sont effectués au bureau central. Au premier mois d'interview ou en présence de tout changement pour ces deux variables, l'intervieweur recueille de l'information décrivant avec précision le genre d'entreprise, d'industrie ou de service où la personne travaille, et de l'information indiquant clairement et précisément le genre de travail ou la nature des fonctions. Le premier type d'information servira à déterminer le secteur d'activité, tandis que le second permettra d'identifier la profession. Une des premières étapes du traitement au bureau central consiste à coder automatiquement l'information descriptive recueillie pour les variables profession et secteur d'activité selon la classification normalisée pour ces variables, soit la

SCIAN et CTP. Les enregistrements ne pouvant être codés par le système automatique sont codés manuellement par une équipe de codeurs de l'EPA. Environ 14 000 enregistrements sont codés à la main chaque mois. Afin de contrôler la qualité du codage effectué manuellement, un plan statistique de contrôle de la qualité est appliqué mensuellement. Les trois mesures utilisées pour connaître l'efficacité de ce processus de contrôle sont le taux de vérification, qui est le pourcentage d'enregistrements vérifiés sur l'ensemble des enregistrements soumis au contrôle, la qualité moyenne à l'entrée (QME), qui est une estimation du pourcentage d'enregistrements contenant des erreurs avant d'effectuer le contrôle de la qualité, et la qualité moyenne à la sortie (QMS), qui est une estimation du pourcentage d'enregistrements contenant encore des erreurs de codage après le contrôle de la qualité. En 2005, la valeur moyenne de ces trois mesures était de 19,6 % pour le taux de vérification, 7,9 % pour la QME et de 4,8 % pour la QMS.

Le taux d'imputation est également un indicateur de qualité relatif au traitement des données. L'EPA produit à l'occasion des taux d'imputation. Ces taux sont ventilés par méthode d'imputation, par questionnaire et par question. Cet indicateur permet de contrôler la qualité d'imputation et de réagir en conséquence.

Pour éviter les erreurs susceptibles de se produire à l'étape de l'estimation et de la totalisation, on procède à un examen détaillé du résultat de ces activités, à une analyse de différents diagnostics produits automatiquement par le système et à une comparaison avec d'autres sources de données.

## **8.2.6 Erreurs lors de l'activité de collecte**

L'application de collecte produit des fichiers contenant une foule de renseignements sur ce qui se passe sur le terrain. À partir de ces fichiers, il est possible de produire une vaste gamme d'indicateurs de qualité des activités des intervieweurs sur le terrain. Sur une base régulière, l'EPA analyse les appels et les visites effectués par les intervieweurs, ainsi que la durée des interviews. Par exemple, on peut obtenir le temps moyen par interview personnelle ou téléphonique, le nombre de tentatives pour joindre un répondant, les meilleurs moments (heures et jours) pour réaliser une interview, etc. On peut également vérifier si les procédures de collecte sont suivies à la lettre par les intervieweurs.

L'application permet donc d'en connaître davantage sur le travail exécuté sur le terrain, de réagir en présence

de cas douteux (par exemple des interviews effectuées en moins d'une minute) et, par conséquent, de réduire certaines erreurs qui peuvent se glisser. Ces indicateurs peuvent également être utilisés pour améliorer le programme de formation donné aux intervieweurs et renforcer certains comportements liés à la planification des tâches ou à l'horaire de travail. Ces données permettent également d'en savoir plus sur les règles de vérification appliquées au moment de l'interview. Il est possible de savoir combien de règles de vérification ont été outrepassées après confirmation avec le répondant, combien de fois une règle de vérification a été appliquée et le nombre de fois qu'une observation a échoué aux règles.

Toutes ces mesures permettront de mieux comprendre ce qui se passe sur le terrain, et aideront à l'interprétation des résultats.

### 8.3 Rapports sur la qualité des données de l'EPA

Un certain nombre de rapports sont mis à la disposition des utilisateurs des données de l'EPA au bureau central à Ottawa ou dans les régions. De plus, ces documents sont consultés de façon régulière par les membres du Comité sur la qualité des données de l'EPA pour s'assurer de la qualité de l'enquête. Ils contiennent toute une gamme d'indicateurs de qualité à des niveaux géographiques différents et pour des périodes de temps plus ou moins longues.

*Rapport mensuel sur les opérations de l'enquête de l'EPA.* Tous les mois, la sous-section de la qualité des données de l'EPA produit un rapport sur la qualité des données de l'enquête du mois courant. Son objectif principal est de permettre de contrôler la qualité des opérations sur le terrain. C'est pourquoi la plupart des indicateurs de qualité y sont présentés par bureau régional. Certaines séries sont également présentées pour une période couvrant 26 mois pour mieux saisir les changements saisonniers et les changements mensuels en comparaison avec une année précédente. Le rapport contient les indicateurs de qualité suivants : taux de non-réponse (par bureau régional, composante, nombre de mois dans l'enquête, type de secteur), taux de vacance (par bureau régional, type de secteur), taux de glissement (par province et groupe d'âge-sexe), taille de l'échantillon, nombre de problèmes techniques et nombre de dossiers

*Tableaux de variance.* Les tableaux de variance, produits chaque mois, contiennent les estimations, les c.v., les variances et les effets du plan des principales caractéristiques de l'EPA à différents niveaux

géographiques. À ces indicateurs s'ajoutent la taille moyenne des ménages et les taux de glissement à des niveaux plus détaillés que ceux présentés dans le Rapport sur les opérations de l'EPA.

*Rapport de qualité.* Le Rapport de qualité de l'EPA est produit chaque année. L'objectif de ce rapport est de présenter un examen approfondi des mesures de la qualité associées à l'EPA. De plus, on analyse les mesures de la qualité sur une période de trente mois afin de déceler des tendances ou les effets de certains changements apportés aux activités ou au plan de sondage. Dans ce document, on reprend plusieurs indicateurs de la qualité présentés dans le Rapport sur les opérations, mais cette fois on s'intéresse à l'échelle provinciale plutôt que régionale. Contrairement au Rapport sur les opérations, le Rapport de qualité renferme des textes qui commentent les différents tableaux et graphiques. À ces indicateurs réguliers, s'ajoute à l'occasion un chapitre spécial sur un sujet d'intérêt particulier.

*Rapports spéciaux.* En plus des rapports produits régulièrement pour assurer et contrôler la qualité des données de l'EPA, des rapports spéciaux sont rédigés à l'occasion. Par exemple, une récente étude s'intéresse à l'effet possible d'une réduction de la période de collecte de dix jours à neuf jours sur les taux de réponses et les estimations.

### 8.4 Programmes d'assurance de la qualité

Au fil des ans, l'EPA s'est pourvu de plusieurs programmes pour assurer la qualité des données qu'elle diffuse. Sept de ces programmes sont présentés dans les lignes qui suivent.

*Recrutement.* Avant d'embaucher des candidats à des postes d'intervieweurs, on évalue leurs aptitudes et leur capacité à bien remplir les documents d'enquête. Avant même que la formation commence, on leur envoie un exemplaire de plusieurs documents. Ces documents décrivent le travail des intervieweurs à Statistique Canada (en soulignant quelles sont les responsabilités, les techniques et les compétences requises) et l'organisation de Statistique Canada.

*Formation.* La période initiale de formation des intervieweurs est de deux mois. Elle commence par un cours de trois jours en classe, pendant lequel on montre aux nouveaux intervieweurs comment utiliser l'équipement informatique, comment remplir les formulaires d'enquête et les documents administratifs. En outre, ces derniers y font des exercices pratiques, simulent des interviews et apprennent les techniques d'interview.

Ensuite les intervieweurs reçoivent deux jours de formation en milieu de travail pendant leur première semaine d'enquête et un jour ou deux pendant la deuxième semaine, au besoin. Pendant ce temps, un intervieweur principal les accompagne, les observe, leur indique comment mener les interviews et leur donne l'exemple en effectuant lui-même des interviews. Tous les intervieweurs participent également à des séances spéciales de formation et de recyclage en groupe, au moins une fois par an.

Le travail des intervieweurs est évalué dans le cadre d'autres programmes qui seront décrits plus loin. Selon le rendement de chacun, on détermine s'il y a lieu d'ajouter des cours d'autoformation ou des exercices de révision pour éclaircir certains points ou remédier à des faiblesses.

*Observation.* Le programme d'observation vise à réduire au maximum les erreurs que les intervieweurs pourraient commettre en donnant à l'intervieweur principal l'occasion d'observer ceux qui relèvent de lui, d'évaluer leur rendement et de déceler les problèmes. Chaque intervieweur est observé au moins une fois tous les vingt-quatre mois. C'est au bureau régional qu'on décide qui sera observé et quand il le sera, de manière à ce qu'on ne puisse pas deviner l'ordre dans lequel le programme d'observation se déroule. En dehors de ce programme, il est possible pour l'intervieweur principal d'observer un de ses intervieweurs s'il soupçonne un problème en particulier. L'intervieweur principal accompagne l'intervieweur pendant toute une journée et voit comment se passent les interviews sur place et les interviews téléphoniques. La deuxième journée, l'intervieweur principal vérifie les listes de grappe. Il envoie ensuite les résultats de l'observation au BR et rédige des rapports périodiques à l'intention du bureau central. L'intervieweur principal transmet à l'intervieweur le résultat de sa performance aussitôt que possible après l'observation.

*Rétroaction sur le rendement.* Tous les mois, le rendement des intervieweurs fait l'objet de rapports mensuels. Ces rapports ont trait aux coûts, aux taux de rejet à la vérification et aux taux de réponse. Les intervieweurs principaux sont régulièrement en contact avec leurs intervieweurs et ils portent à leur attention les résultats des divers indicateurs de rendement.

*Vérification des logements codés vacants.* Le programme de vérification des logements codés vacants a pour but de contrôler le travail effectué par les intervieweurs sur le terrain. Pour chacun des intervieweurs, au moins une fois tous les vingt-quatre mois, un

échantillon de logements codés vacants est sélectionné. L'intervieweur principal retourne à ces logements pour vérifier si les logements étaient effectivement vacants et les logements sont recodés (vacants ou autre). À la suite des résultats de ce programme, l'intervieweur recevra de la formation additionnelle si on le juge nécessaire. Cette information sert également à mesurer combien de ménages codés hors du champ de l'enquête ont contribué positivement au glissement, bien qu'il soit très difficile d'extrapoler pour l'échantillon complet puisque le choix des logements vérifiés est laissé au choix des bureaux régionaux.

*Programme de validation.* Le programme de validation a été conçu afin de contrôler la performance des intervieweurs et de fournir aux intervieweurs une rétroaction rectificative prenant la forme de formation additionnelle selon les faiblesses identifiées. Les intervieweurs sont soumis à la validation de façon aléatoire de sorte que, durant une année, chacun d'entre eux est choisi deux fois. Environ 2 % des ménages font partie de ce programme chaque mois (à l'exception des mois d'avril et de décembre où ce programme n'a pas lieu). La semaine suivant la semaine d'enquête, les intervieweurs principaux communiquent de nouveau avec les personnes qui ont fourni l'information durant la semaine d'enquête pour l'échantillon soumis à ce programme. Ils leur posent des questions de politesse pour obtenir l'information suivante : confirmation de l'adresse, souvenir d'avoir participé à l'enquête, moment de l'interview, attitude de l'intervieweur, etc. L'intervieweur principal profite également de l'occasion pour remercier le répondant pour sa participation et son temps.

*Contrôle du rendement des UPE.* Le rendement des UPE est contrôlé sur une base mensuelle afin de détecter tout écart entre le nombre de logements recensés sur le terrain et le nombre de logements utilisés dans l'élaboration du plan de sondage. Le plan de sondage utilise un nombre qui est dérivé des dénombremens effectués à partir des données du recensement précédent. Par conséquent, toute divergence importante, c'est-à-dire tout écart de 50 % (positif ou négatif) entre le recensement et la dérivation, fait l'objet d'une enquête. Tout d'abord, toute grappe ayant un rendement inattendu est portée à l'attention de l'unité responsable du contrôle de l'échantillon à Ottawa, qui vérifie les frontières de la grappe et le nombre de logements attendu. Si l'écart ne peut être expliqué au bureau central, la grappe est acheminée au bureau régional concerné pour être analysée en détail. Toutes les causes expliquant les écarts sont répertoriées pour références futures.

Ce contrôle joue un rôle important puisque si la taille de l'échantillon nécessite des changements, il est essentiel de connaître quelles régions sont sous-échantillonnées ou suréchantillonnées. De plus, les écarts enregistrés peuvent révéler des problèmes pour l'enquête qui pourraient entacher la qualité des données de l'EPA.

### 8.5 Comités de l'EPA

L'EPA a besoin de plusieurs groupes de coordination pour veiller au bon déroulement de l'enquête. Certains comités jouent un rôle permanent à l'EPA, tandis que d'autres sont actifs uniquement pendant le remaniement. Durant le dernier remaniement, le Comité de direction du remaniement était le principal comité de haut niveau. Il avait comme mandat de contrôler l'ensemble du remaniement, c'est-à-dire de faire en sorte que les activités liées au remaniement de l'échantillon se déroulent sans heurts.

Dans les lignes qui suivent, nous décrivons trois comités permanents. Leur mandat est de s'occuper de manière régulière des opérations permanentes et de l'évaluation de l'enquête.

*Comité des opérations.* Ce comité a pour mandat de passer en revue les activités survenues pendant les mois d'enquête et les circonstances entourant la tenue de l'enquête, de voir à ce que les opérations se déroulent bien, d'examiner les changements proposés et d'en recommander l'adoption; le tout afin que l'enquête continue d'atteindre ses objectifs. Ce comité, qui se réunit toutes les deux semaines, est présidé par un membre de la Division de la statistique du travail.

*Comité de direction des estimations de population.* Ce comité a pour mandat de contrôler les estimations de population postcensitaires requises par l'EPA. Ce comité évalue également les sources de données utilisées et les méthodes appliquées pour obtenir les

estimations à différents niveaux géographiques, et met en oeuvre plusieurs projets de recherche sur le sujet. Ce comité est présidé par un membre de la Division de l'analyse des enquêtes sur le travail et les ménages.

*Comité sur la qualité des données.* Le comité actuel, qui a officiellement vu le jour au printemps de 1972, avait à l'époque comme fonction de diffuser la qualité des données de l'EPA et de ses suppléments. Depuis ce temps, le mandat de ce comité s'est quelque peu élargi. Son mandat consiste maintenant à examiner et à évaluer la qualité des données de l'EPA sur une base mensuelle, à proposer et à réviser des projets de recherche et de développement visant la mise au point de méthodes pouvant influencer sur la qualité des données, et enfin à surveiller la recherche et le développement dans ce domaine. Ce comité est présidé par un membre de la Division des méthodes d'enquêtes auprès des ménages.

Pour assurer la meilleure qualité de données possible, le Comité sur la qualité des données examine périodiquement les différents indicateurs de qualité décrits plus haut. Le comité se réunit chaque mois pour étudier et évaluer la qualité des données mensuelles et pour faire des suggestions et des recommandations sur tout aspect susceptible d'amélioration. Les membres du comité disposent d'un certain nombre de documents pour mener leur tâche à bien. En suivant étroitement l'évolution des indicateurs, le comité peut intervenir immédiatement auprès des responsables des activités de l'EPA concernées afin de contrôler la qualité des données mensuelles. Le comité discute également de faits nouveaux susceptibles d'influencer la qualité des données venant d'être recueillies ou devant être recueillies dans l'avenir, en particulier les changements relatifs aux méthodes de collecte ou au questionnaire, des problèmes inhabituels sur le terrain, de la mise à l'essai continue des procédés et des méthodes, etc.

## Chapitre 9 Base de sondage de l'EPA pour d'autres enquêtes

### Introduction

La base de sondage et les échantillons de l'Enquête sur la population active servent à la collecte d'information dans le cadre de nombreuses enquêtes auprès des ménages. En général, ces enquêtes peuvent être décrites comme des enquêtes spéciales, des enquêtes supplémentaires ou des enquêtes utilisant des ménages qui ne font plus partie de l'EPA. Dans le cadre des enquêtes spéciales, la base de sondage de l'EPA sert à la sélection d'un échantillon de ménages distinct, généralement dans les UPE qui sont actives dans l'EPA. Dans le contexte des enquêtes supplémentaires, les ménages qui viennent tout juste d'être sélectionnés pour l'EPA sont interviewés. Les enquêtes utilisant des ménages qui ne font plus partie de l'EPA sont semblables aux enquêtes supplémentaires, mais les ménages sont sélectionnés à partir des ménages qui ne font plus partie de l'EPA. Des ménages actifs de l'EPA sont également utilisés au cours de certaines de ces enquêtes, mais les interviews sont menées à une date ultérieure. Les enquêtes spéciales et supplémentaires représentent une partie importante du programme des enquêtes-ménages de Statistique Canada et sont souvent parrainées par d'autres ministères.

Bien que les enquêtes spéciales soient menées auprès de ménages distincts, l'intervieweur peut souvent se servir des ressources de l'EPA puisque ces enquêtes sont généralement réalisées auprès des mêmes UPE. Les données recueillies dans le cadre de l'EPA peuvent également servir aux enquêtes supplémentaires et aux enquêtes utilisant des ménages qui ne font plus partie de l'EPA aux fins de sélection des répondants. Précisons que les enquêtes supplémentaires peuvent être de deux types : l'enquête supplémentaire dépendante, qui sonde les ménages de l'EPA alors que ceux-ci sont toujours interviewés pour l'EPA, et l'enquête supplémentaire indépendante, qui sépare la population active qui sera interviewée à un autre moment ou qui alloue davantage de temps que l'EPA pour la collecte des données. Quand un ménage ne participe plus à l'EPA, il demeure toutefois admissible aux enquêtes utilisant des ménages qui ne font plus partie de l'EPA pour une période pouvant aller jusqu'à deux ans. L'utilisation de la base de sondage et des échantillons de l'EPA pour d'autres enquêtes permet de réaliser des économies de coûts considérables.

Dans le contexte des enquêtes spéciales, un ensemble de départs aléatoires est réservé aux fins de sélection des

logements et à l'usage exclusif de celles-ci. Selon la répartition souhaitée, une ou plusieurs départs aléatoires peuvent être réservés de cette façon pour chaque strate de l'EPA. Dans certains cas, il pourrait y avoir des départs aléatoires dans des UPE qui ne seront pas actives dans l'EPA pour plusieurs années qui seront réservées pour les enquêtes spéciales. Les échantillons pour les enquêtes spéciales sont sélectionnés à l'aide de ces départs aléatoires. Cette stratégie réduit le fardeau du répondant, car le même logement n'est pas sélectionné pour l'EPA et les enquêtes spéciales.

La principale préoccupation dans le contexte des enquêtes supplémentaires et des enquêtes utilisant des ménages qui ne font plus partie de l'EPA est le fardeau du répondant. Les sujets ou les questions qui risquent d'être inacceptables pour les répondants ou qui risquent d'avoir une incidence, dans une certaine mesure, sur les réponses obtenues à l'EPA le mois suivant sont évités. Les suppléments doivent être conformes aux critères adoptés par Statistique Canada en matière de fiabilité des données et de confidentialité des réponses. Selon le sujet et le nombre d'enquêtes actives dans un mois donné, certains suppléments sont bien accueillis; ils augmentent la durée d'interview, mais d'un autre côté, ils permettent de varier l'expérience des ménages qui font partie de l'échantillon de l'EPA pour une durée de six mois.

Chacun des six groupes de renouvellement de l'EPA peut servir à produire des estimations. En général, ces enquêtes utilisent d'un à cinq groupes de renouvellement pour établir leur échantillon, selon le niveau de fiabilité nécessaire. Habituellement, on évite le groupe de départ de l'EPA (les naissances), c'est-à-dire celui formé de ménages sondés pour la première fois dans le cadre de l'EPA, car l'interview initiale de l'EPA prend plus de temps que les interviews ultérieures.

Dans certains cas, seule une partie des ménages d'un groupe de renouvellement est requise. Les logements sont retirés au hasard pour que le nombre de ménages soit réduit au nombre requis, comme c'est le cas dans le programme de stabilisation de l'EPA. On peut également procéder à une sélection parmi les ménages en prélevant un échantillon aléatoire ou en éliminant les individus selon certaines caractéristiques.

Dans un logement sélectionné, l'enquête peut s'adresser à tous les répondants admissibles selon les critères de l'EPA ou à des personnes en particulier. Il s'agit alors, soit de prélever un échantillon aléatoire

d'individus à l'intérieur des ménages choisis, soit de faire une présélection en consultant les documents de l'EPA ou en posant certaines questions afin de retenir les répondants ayant certaines caractéristiques du point de vue démographique ou du point de vue de l'activité.

La liste suivante présente certaines des enquêtes ayant utilisé la base de sondage ou des échantillons de l'EPA en 2005.

Enquête	Période de collecte des données - 2005
Enquêtes supplémentaires	
Enquête sur les voyages internationaux – Aéroports (EVI)	Avril à mai
Projet un avenir à découvrir (PUAD)	Avril à juin
Enquête canadienne sur l'utilisation d'Internet (ECUI)	Novembre
Enquête sur les voyages des résidents du Canada (EVRC)	Janvier à décembre (mensuelle)
Enquête sur les communautés	Février à juin
Enquête sur le service téléphonique résidentiel	Décembre
Enquête auprès des enfants du Nord	Février à juillet
Enquêtes après renouvellement de l'EPA	
Enquête sur la dynamique du travail et du revenu (EDTR)	Janvier à mars
Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)	Septembre à juin
Enquête longitudinale auprès des immigrants du Canada (ELIC)	Janvier à juin
Enquêtes spéciales	
Enquête sur la santé dans les collectivités canadiennes (ESCC)	Janvier à novembre (mensuelle)
Enquête sur les dépenses des ménages (EDM)	Janvier à mars
Enquête sur la sécurité financière (ESF)	Mai à juin
Enquête nationale sur la santé de la population (ENSP)	Février, juin, août, novembre

### 9.1 Exemples d'enquêtes spéciales et d'enquêtes supplémentaires

L'*Enquête sur la santé dans les collectivités canadiennes (ESCC)* est une enquête continue qui permet de recueillir tous les mois des données relatives à une variété de sujets sur la santé. L'échantillon de la base de sondage de l'EPA est complété à l'aide d'une base de sondage téléphonique et, dans quelques régions sociosanitaires, grâce à un système de composition aléatoire.

L'*Enquête sur les dépenses des ménages (EDM)* est une enquête-ménage annuelle qui a été introduite, dans sa configuration actuelle, dans le cadre du Projet

d'amélioration des statistiques économiques provinciales (PASEP) de Statistique Canada. L'Enquête sur les dépenses des ménages (EDM) continuera de remplir sa fonction traditionnelle de source d'information pour le calcul de l'indice des prix à la consommation. L'EDM est une enquête spéciale, ce qui signifie que les ménages de cette enquête sont choisis dans les UPE échantillonnées par l'EPA, mais les ménages faisant partie de l'EDM ne sont pas interviewés dans le cadre de l'EPA.

L'*Enquête sur la sécurité financière (ESF)* est réalisée occasionnellement en vue de recueillir de l'information sur la valeur nette (richesse) des familles canadiennes, soit la valeur de leurs actifs moins leurs dettes. Des adresses de régions géographiques où les revenus sont élevés sont également ajoutées à l'échantillon de la base de sondage de l'EPA.

L'*Enquête sur la dynamique du travail et du revenu (EDTR)* a été créée en 1993 pour étudier les processus qui influencent la vie économique des Canadiens. On utilise cette enquête pour observer les mouvements d'entrée et de sortie relativement à la situation de faible revenu, les transitions du marché du travail et la relation entre la dynamique familiale et le bien-être économique. Les panels sont choisis aux trois ans et chaque panel participe à l'enquête pendant six ans. À l'origine, chaque panel est formé de ménages ayant été récemment interviewés dans le cadre de l'EPA (renouvellement). À l'instar d'autres enquêtes longitudinales, l'EDTR suit les répondants choisis dans le temps.

L'*Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)*, qui a été lancée en 1994, porte sur un échantillon d'enfants dont le développement est suivi pendant un certain nombre d'années de l'enfance à l'âge adulte. Il s'agit d'une enquête complexe qui, à l'origine, était menée auprès de ménages de l'échantillon de l'EPA pour former un échantillon d'enfants. Dans le contexte de l'ELNEJ, en plus de la cohorte initiale d'enfants âgés de 0 à 11 ans échantillonnés lors du premier cycle, un nouvel échantillon d'enfants de 0 à 1 an est également sélectionné à chaque cycle ultérieur. Puisque seulement 3 % des ménages de l'EPA donnent naissance à des enfants dans une année donnée, l'ELNEJ utilise jusqu'à 20 groupes de renouvellement dans les petites provinces.

## Bibliographie

- Alavi, A. et Beaumont, J.-F. (2004). Nonresponse Adjustment Plans for the Labour Force Survey. Article présenté au Comité consultatif sur les méthodes statistique, mai 2004.
- Alexander, C.H., Ernst, L.R. et Haas, M.E. (1982). A System for Replacing Primary Sampling Units When the Units Have Been Exhausted. *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 211-216.
- Bocci, C., et Beaumont, J.-F. (2004). Longitudinal Hot-deck Imputation for Household Nonresponse in the LFS. Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- Bocci, C., et Beaumont, J.-F. (2005). A Refinement of the Regression Composite Estimator in the Labour Force Survey. Rapport interne, Statistique Canada.
- Brisebois, F. et Mantel, H. (1996). Month-in-sample Effects for the Canadian Labour Force Survey. *Recueil de la Section des méthodes d'enquête*, Société statistique du Canada.
- Brodeur, M., Montigny, G. et Bérard, H. (1995). Challenges in Developing the National Longitudinal Survey of Children. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Chen, E.J., et Liu, T.P. (2002). Choices of Alpha Value in Regression Composite Estimation for the Canadian Labour Force Survey: Impacts and Evaluation. Document de travail de méthodologie, HSMD-2002-05E, Statistique Canada.
- Chen, E.J., Gambino, J., Laniel, N. et Lindeyer, J. (1994). Design and Estimation Issues for Income in the Redesign of the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Chen, E.J., Lindeyer, J. et Laflamme, G. (2004). Les enjeux associés aux régions à coûts de collecte élevés considérés lors du remaniement de l'Enquête sur la population active canadienne. *Recueil de la Section des méthodes d'enquête*, Société statistique du Canada.
- Cillis, P. (2004a). LFS Mapper Functional Specification. Document interne, Division de la géographie, Statistique Canada.
- Cillis, P. (2004b). LFS Mapper User Guide. Document interne, Division de la géographie, Statistique Canada.
- Cochran, W.G. (1977). *Sampling Techniques*, 3<sup>ème</sup> édition, New York : John Wiley and Sons, Inc.
- Deville, J.-C., et Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Drew, J.D., Bélanger, Y. et Foy, P. (1985). La stratification dans l'enquête sur la population active du Canada. *Techniques d'enquête*, 11, 109-124.
- Drew, J.D., Choudhry, G.H. et Gray, G.B. (1978). Some Methods for Updating Sample Survey Frames and Their Effects on Estimation. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 62-71.
- Dufour, J., Simard, M., Allard, B. et Ray, G. (1996). Redesign of the Labour Force Survey Sample: Impact on Data Quality. Statistique Canada, document interne.
- EPA (2005). Enquête sur la population active, rapport annuel sur la qualité, janvier-décembre 2004. Statistique Canada, document interne.
- EPA-100 IPAO (2006). Manuel de l'intervieweur IPAO. Statistique Canada, document interne, 75030-1800.02
- EPA Listage initial pour les grappes du registre des adresses (2004). Guide d'autoformation pour les intervieweurs aux IPAO. Statistique Canada, document interne, 75030-6458.2.
- EPA Manuel de l'intervieweur des grappes du groupe RA 1 (2004). Guide d'autoformation pour les intervieweurs aux IPAO. Statistique Canada, document interne, 75030-6489.2
- Friedman, H.P., et Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Fuller, W.A., et Rao, J.N.K. (2001). Un estimateur composite de régression qui s'applique à l'Enquête sur la population active du Canada. *Techniques d'enquête*, 27, 49-56.
- Gambino, J., Kennedy, B. et Singh, M.P. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada : Évaluation et application. *Techniques d'enquête*, 27, 69-79.
- Gouzi, N., Turmelle, C., Thompson, G. et Rodrigue, J.-F. (2004). Processus de traitement des données du Registre des adresses. Document interne, Statistique Canada.

- Gray, G. (1973). Rotation of PSUs. Document interne, Statistique Canada.
- Hartley, H.O. et Rao, J.N.K. (1962). Sampling With Unequal Probabilities and Without Replacement. *Annals of Mathematical Statistics*, 33, 350-374.
- Kennedy, B. (1998). Weighting and Estimation Methodology of the Canadian Labour Force Survey. Document de travail de méthodologie, HSMD-98-002E, Statistique Canada.
- Kennedy B., Drew J. D., et Lorenz P. (1994). The Impact of Nonresponse Adjustment on Rotation Group Bias in the Canadian Labour Force Survey. Présenté au 5<sup>ième</sup> Atelier international sur la non-réponse dans les enquêtes auprès des ménages. Ottawa, Canada.
- Keyfitz, N. (1951). Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities. *Journal of the American Statistical Association*, 46, 105-109.
- Laflamme, G. (2003). Sélection et rotation des UPE : quelques précisions concernant la RAM, Statistique Canada, document interne.
- Laflamme, G., et Dochitoui, C. (2005). An Evaluation Study for the Sequencing of the LFS Primary Sampling Units. Statistique Canada, document interne.
- Lebrasseur, D., et Dion, S-M. (2005). L'approche du premier contact par téléphone dans l'Enquête sur la population active. *Recueil du Symposium 2005 de Statistique Canada: Défis méthodologiques pour les besoins futurs d'information*.
- Lemaître, G., et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. *Techniques d'enquête*, 13, 211-220.
- Lindeyer, J. (2004). LFS Mapper User Guide: Inset Changes, Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- Little, R. J. A. (1986). Survey Nonresponse Adjustment for Estimate of Means. *Revue internationale de la statistique*, 54, 139-157.
- Lorenz, P. (1996). Head Office Hot Deck Imputation System Specifications, Version 3. Document interne, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- Mantel, H., Laniel, N., Duval, M.-C. et Marion, J. (1994). Cost Modelling of Alternative Sample Designs for Rural Areas in the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Mian, I.U.H. et Laniel, J. (1994). Sample Allocation for the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Pandey, S., Alavi, A., et Beaumont, J.-F. (2003). Comparison of Integrated and Non-integrated Estimation Methods for GREG and Composite Estimators. Document interne, Statistique Canada.
- Rao, J.N.K. (2005). Évaluation de l'interaction entre la théorie et la pratique des enquêtes par sondage. *Techniques d'enquête*, 31, 127-151.
- Rao, J.N.K., Hartley, H.O. et Cochran, W.G. (1962). A Simple Procedure for Unequal Probability Sampling Without Replacement. *Journal of the Royal Statistical Society*, B, 24, 482-491.
- Revue chronologique de la population active. Statistique Canada. n° 71-F0004X au catalogue.
- Särndal, C.-E., Swensson, B., et Wretman, J. (1992). *Model Assisted Survey Sampling*. New-York: Springer Verlag.
- Satin, A., et Shastri, W. (1992). L'échantillonnage un guide non mathématique. Deuxième édition, Statistique Canada, numéro de catalogue 12-602-XPE.
- Sheridan, M., Drew, D. et Allard, B. (1996). Le taux de réponse et l'Enquête sur la population active canadienne: Fruit de hasard ou bonne planification? *Recueil de Statistique Canada Symposium 96 sur Erreurs non dues à l'échantillonnage*, 75-83.
- Simard, M. et Dufour, J. (1995). Impact de l'implantation des interviews assistées par ordinateur comme nouvelle méthode de collecte à l'Enquête sur la population active. Statistique Canada, document interne.
- Singh, A.C., Kennedy, B., et Wu, S. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada avec plan de sondage à renouvellement de panel *Techniques d'enquête*, 27, 35-48.
- Singh, A.C., Kennedy, B., Wu, S. et Brisebois, F. (1997). Composite Estimation for the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Singh, M.P., Drew, J.D., Gambino, J.G. et Mayda, F. (1990). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada. n° 71-526 au catalogue.

- Singh, M.P., Gambino, J. et Mantel, H. (1994). Les petites régions: problèmes et solutions (avec discussion). *Techniques d'enquête*, 20, 3-23.
- Sunter, D., Kinack, M., Akyeampong, E. et Charette, D. (1995). Redesigning the Canadian Labour Force Survey Questionnaire: Development and Testing. Statistique Canada, document interne.
- Swain, L., Drew, J.D., Lafrance, B. et Lance, K. (1992). La création d'un registre des adresses résidentielles pour améliorer la couverture du recensement du Canada de 1991. *Techniques d'enquête*, 18, 139-155.
- Tambay, J.-L. et Catlin, G. (1995). Plan d'échantillonnage de l'Enquête nationale sur la santé de la population. *Rapport sur la santé*, 7, 29-38.
- Thompson, G., et Rodrigue, J.-F. (2005) Evaluation of the AR Use in the LFS Frame. Statistique Canada, document interne.
- Thompson, G., et Turmelle, C. (2004). Classification of address register coverage rates - a field study. *Proceedings of the Research Methods Section*, American Statistical Association.
- Wolter K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.

## Annexe A.1 Glossaire

**De l'information plus détaillée sur un bon nombre de ces termes se trouve sur les sites suivants :**

*2001 Census Dictionary* <http://www12.statcan.ca/english/census06/reference/dictionary/index.cfm>

*Geography illustrated glossary* [http://geodepot.statcan.ca/Diss2006/Reference/COGG/Index\\_e.jsp](http://geodepot.statcan.ca/Diss2006/Reference/COGG/Index_e.jsp)

*Guide to the Labour Force Survey* <http://www.statcan.ca/english/freepub/71-543-GIE/71-543-GIE2007001.htm>

### **Aire de diffusion (AD)**

Petite région composée d'un ou de plusieurs îlots de diffusion avoisinants et regroupant de 400 à 700 habitants. L'ensemble du Canada est divisé en aires de diffusion. Elle est définie par le recensement de la même façon que les UPE de l'EPA sont définis par le SARTe. Elle est créée en tant qu'unité de diffusion des données du recensement.

### **Base aréolaire**

Base d'unité fondée sur une étendue aréolaire comme les îlots, les zones de diffusion du recensement ou la géographie similaire.

### **Chômage**

Les personnes en chômage sont celles qui, durant la semaine de référence :

- étaient mise à pied temporairement, mais qui s'attendaient à être rappelées et qui étaient prêtes à travailler, ou
- qui étaient sans emploi, mais qui cherchaient activement du travail dans les quatre dernières semaines et qui étaient prêtes à travailler, ou
- qui commenceront un nouvel emploi dans les quatre semaines suivant la semaine de référence et qui étaient prêtes à travailler.

### **Classification de la population active selon l'activité**

Chaque répondant de 15 ans ou plus est classé dans une catégorie de la population active (personne occupée, chômeur ou inactif), d'après les réponses qu'il a fournies à un certain nombre de questions durant l'interview.

### **Échantillonnage à deux étapes**

Dans l'échantillonnage à deux étapes, les unités sont sélectionnées en deux étapes. Les unités dans les deux étapes sont des entités distinctes (parexemple, les îlots dans la première étape et les logements dans la seconde). Au cours de la première étape de l'échantillonnage de l'EPA, les UPE sont sélectionnées à l'intérieur des strates. Dans la seconde étape, les logements sont sélectionnés à l'intérieur de l'UPE.

### **Échantillonnage systématique**

Méthode de sélection systématique d'un échantillon dans lequel le premier élément est sélectionné aléatoirement dans une population (origine choisie au hasard) et dont les éléments restants sont tirés à des intervalles échelonnés également (taux d'échantillonnage inverse). Avec un taux d'échantillonnage de 1 pour 10, et un point de départ au hasard à 7, la 7<sup>e</sup> unité est sélectionnée. Par la suite chaque 10<sup>e</sup> unité est choisie, dont la 17<sup>e</sup>, la 27<sup>e</sup>, la 37<sup>e</sup>, etc.

### **Échantillonnage systématique randomisé avec probabilité proportionnelle à la taille**

- Dans l'échantillonnage à probabilité proportionnelle à la taille (PPT) chaque unité d'échantillonnage a une mesure de taille (les logements dans le cas de l'EPA). La taille relative des unités détermine leur probabilité de sélection. Les plus grandes unités ont plus de chance d'être sélectionnées.
- L'échantillonnage systématique est une méthode de conception d'un échantillon dans lequel le premier élément est sélectionné aléatoirement dans une population (départ choisi au hasard) dont les éléments restants de l'échantillon sont tirés à des intervalles espacés également (taux d'échantillonnage inverse)
- La méthode utilisée par l'EPA est appelée randomisée puisque la liste des unités primaires d'échantillonnage est ordonnée aléatoirement avant l'échantillonnage systématique.

### **Effet du plan de sondage**

Le ratio de la variance réelle d'un estimateur selon le plan de sondage actuel et ce qu'il serait selon un plan de sondage aléatoire pour le même nombre d'éléments.

## **Emploi**

Les personnes occupées sont celles qui, au cours de la semaine de référence :

- ont fait un travail quelconque dans le cadre d'un emploi ou dans une entreprise, c'est-à-dire tout travail rémunéré accompli pour un employeur ou à son propre compte; cela comprend aussi le travail familial non rémunéré, c'est-à-dire un travail (non rémunéré) qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un cabinet de professionnel appartenant à un membre apparenté du même ménage et exploité par lui; ou
- avaient un emploi, mais n'étaient pas au travail à cause d'une maladie ou d'une incapacité, pour obligations personnelles ou familiales, pour des vacances, par suite d'un conflit de travail ou du fait de tout autre facteur (cela n'inclut pas les personnes mises à pied, celles qui étaient inactives entre deux emplois occasionnels et celles qui avaient un emploi devant commencer à une date ultérieure).

## **Groupe du Registre des adresses (RA)**

Extrait du RA utilisé par l'EPA pour l'identification de l'UPE et la séquence des numéros appliqués à chaque logement. L'intégralité et l'exactitude de l'extrait au niveau de l'UPE sont estimées afin d'assigner un des trois groupes de nombres à l'UPE.

- 1 : Excellente qualité, pas nécessaire de vérifier la liste des logements, sélectionner simplement à partir de la liste actuelle.
- 2 : Bonne qualité, mais envoyer la liste pour vérification avant la sélection des logements.
- 3 : Médiocre qualité ou pas de couverture du RA, donc liste retirée.

## **Îlot**

Zone limitée de tous les côtés par les routes et/ou les limites des zones géographiques normalisées sélectionnées. Les îlots sont utilisés pour générer des unités plus grandes telles que les zones de diffusion et les UPE de l'EPA.

## **Listage**

Processus par lequel les logements appartenant à une zone (UPE, AD) sont enregistrés sur papier ou sous forme électronique. Les cartes du secteur, ayant des limites claires, sont nécessaires pour déterminer où le listage est nécessaire. La plupart des listages sont séquencés selon un modèle particulier afin de permettre que tous les côtés de l'îlot soient examinés et pour pouvoir relocaliser une adresse particulière dans un autre mois ou année après la création de la liste de départ.

## **Logement**

Ensemble de pièces d'habitation qu'une personne ou un groupe de personnes habitent ou pourraient habiter. On appelle « vacant » les logements inoccupés. Dans le cas de l'EPA, un logement se compose de toute pièce d'habitation ayant une structure de séparation et une entrée particulière à l'extérieur de l'immeuble ou d'une entrée ou d'un escalier commun.

## **Logement collectif**

Un logement collectif désigne un logement de nature commerciale, institutionnelle ou communautaire où les gens résident, mais où le concept de logement unifamilial est difficile à appliquer. Il y a deux types fondamentaux de logements collectifs : institutionnel et non institutionnel.

## **Ménage**

Toute personne ou groupe de personnes vivant dans un logement. Un ménage peut comprendre les combinaisons suivantes : une personne vivant seule, une ou plusieurs familles, un groupe de personnes qui ne sont pas apparentées, mais qui partagent le même logement. Il faut noter que les résidents étrangers et les personnes qui ont leur résidence principale ailleurs ne sont pas visés par l'enquête.

## **Moyenne mobile de trois mois**

Moyennes des estimations à partir du plus récent trimestre, chaque mois.

## **Période de référence**

Période de temps utilisée dans les enquêtes dont les répondants doivent se souvenir et sur laquelle ils doivent répondre, par exemple : « combien d'heures avez-vous travaillé la semaine dernière? »

### **Population active**

Population civile de 15 ans et plus (à l'exclusion des pensionnaires d'établissements) qui, durant la semaine de référence de l'enquête, était occupée ou en chômage. Avant 1966, l'enquête portait sur les personnes de 14 ans et plus.

### **Population cible**

La population cible de l'EPA correspond à l'ensemble des personnes de 15 ans et plus qui résident dans les provinces du Canada, à l'exception des suivantes : les personnes qui vivent dans les réserves indiennes, les membres à temps plein des Forces et les pensionnaires d'établissements (par exemple, les personnes détenues dans les pénitenciers et les patients d'hôpitaux ou de maisons de repos qui ont résidé dans l'établissement pendant plus de six mois).

### **Région économique**

Une région économique (RE) est constituée d'un groupe de divisions de recensement (DR) entières (sauf pour un cas en Ontario). Ces régions sont créées comme une unité géographique normalisée et servent à l'analyse de l'activité économique régionale. Ces régions infraprovinciales sont reconnues par l'EPA et le recensement.

### **Région économique de l'assurance-emploi**

Ensemble de régions dans tout le pays définies par Ressources humaines et développement des compétences Canada (RHDC) en vue de la répartition équitable des prestations de l'assurance-emploi. L'EPA est responsable de produire des estimations récentes requises par RHDC afin d'établir des normes pour l'admissibilité au programme et pour la durée des prestations.

### **Région métropolitaine de recensement (RMR) ou Agglomération de recensement (AR)**

Une région métropolitaine de recensement (RMR) ou une agglomération de recensement (AR) est formée d'une ou de plusieurs municipalités adjacentes situées autour d'une grande région urbaine (appelée noyau urbain). Une RMR doit avoir une population d'au moins 100 000 habitants et le noyau urbain doit compter au moins 50 000 habitants. L'agglomération de recensement doit avoir un noyau urbain d'au moins 10 000 habitants. Pour être comprises dans une RMR ou une AR, les autres municipalités adjacentes doivent avoir un degré d'intégration élevé avec la région urbaine centrale, lequel est déterminé par le pourcentage de navetteurs établi d'après les données du recensement sur le lieu de travail.

### **Région rurale**

Les régions rurales comprennent tout le territoire situé à l'extérieur des régions urbaines. Ensemble, les régions urbaines et les régions rurales couvrent tout le territoire canadien.

### **Région urbaine**

Définition provenant du Dictionnaire du recensement : « Une région urbaine a une concentration démographique d'au moins 1 000 habitants et une densité de population d'au moins 400 habitants au kilomètre carré. » Les régions urbaines de l'EPA sont plus souvent délimitées par les strates et peuvent être assez différentes de celles établies par le recensement. Les régions urbaines trop petites pour la création de strates séparées sont considérées comme région rurale.

### **Registre des adresses (RA)**

Base de données contenant les adresses domiciliaires conservées majoritairement pour le recensement, mais utiles pour les enquêtes auprès des ménages. La base de données contient plus de 11 000 000 adresses, concentrées en majorité dans les grandes régions urbaines, mais comprenant la plupart des collectivités ayant des descriptions de logements de type urbain.

### **Renouvellement**

Remplacement périodique d'une unité par une autre. L'EPA effectue le :

- Renouvellement des logements (dans une UPE) après six mois dans l'enquête.
- Renouvellement des UPE après de deux à 50 ans dans l'enquête, avec une moyenne d'environ dix ans. Dans bien des cas, un remaniement de l'enquête aura lieu avant la date de renouvellement prévue pour l'UPE.

Les logements établis (ou les UPE qui les contiennent) qui se renouvellent le même mois sont appelés groupes de renouvellement. Chaque groupe se compose d'un sixième de l'échantillon. Ainsi, chaque mois, il y a un mélange de

logements qui ont leur première, deuxième, troisième, quatrième, cinquième et sixième interview. La plupart des strates ont six (parfois 12) UPE sélectionnées afin que chaque groupe soit aussi représentatif possible du total. Le groupe de renouvellement qui a la première interview un mois donné est appelé panel de naissances ou renouvellement des naissances. Les numéros des renouvellements sont assignés de manière à ce que le renouvellement 1 ait lieu en janvier et juillet, le renouvellement 2 en février et août, jusqu'au renouvellement 6 en juin et décembre. Le terme hors renouvellement indique l'introduction d'un logement ayant un numéro de renouvellement particulier pour un mois qui ne correspond pas au modèle normal pour le renouvellement des naissances.

### **Répartition**

Fait de répartir une taille d'échantillon déterminée en plusieurs zones provinciales ou infraprovinciales pour répondre à des contraintes liées aux coûts de la collecte et à la fiabilité des estimations.

### **Réseau routier**

Représentation numérique du réseau routier national du Canada. Il contient des renseignements comme les noms de rue, le type, l'orientation et les tranches d'adresses. L'application comprend, la cartographie, le géocodage, la recherche géographique, le tracé des secteurs et l'entretien de la base de données comme une source pour les noms de rue et les emplacements. Puisque les activités statistiques n'exigent pas une exactitude de position absolue, une relative exactitude de position a préséance dans le réseau routier. En conséquence, le réseau routier ne convient pas aux applications d'ingénierie, aux services de répartition d'urgence, aux applications juridiques et d'enquête.

### **Stratification**

La stratification regroupe les UPE créés par le programme SARTe en ensemble calibré. Les UPE d'un ensemble donné (strate) tendent à posséder des caractéristiques similaires. La stratification tient compte des limites géographiques comme les provinces, les REAE et les régions économiques. La strate est l'unité fondamentale de l'EPA et c'est dans cette unité que s'amorce la sélection de l'échantillon

### **Subdivision de recensement (SDR)**

Une municipalité ou une région jugée équivalente à des fins statistiques (p. ex., une réserve indienne ou un territoire non organisé). Le statut de municipalité est défini par les lois en vigueur dans chaque province et territoire au Canada.

### **Système automatisé de regroupement des territoires**

Un ensemble de programmes géographiques utilisés pour délimiter les UPE de l'EPA dans toutes les provinces. Programmes créés à l'origine pour délimiter les secteurs de dénombrement pour le recensement et ensuite modifiés par la Division de la géographie pour répondre aux besoins de l'EPA.

### **Taux d'activité**

Pourcentage de la population active totale par rapport à l'ensemble de la population de 15 ans et plus. Le taux d'activité pour un groupe donné (âge, sexe, etc.) correspond au pourcentage de la population active dans ce groupe par rapport à l'effectif total du groupe.

### **Taux de chômage**

Nombre de chômeurs exprimé en pourcentage de la population active (employé + sans emploi). Ce taux est une des principales statistiques produites par l'EPA.

### **Taux d'échantillonnage**

Rapport de la taille de l'échantillon à la taille de la population dans la base de sondage. Un échantillon de 1 sur 20 sélectionnerait 5 % des unités pour la collecte des données et aurait un taux d'échantillonnage à 0,05 ou un taux d'échantillonnage inverse de 20.

### **Taux d'emploi (rapport emploi-population)**

Nombre de personnes occupées exprimé en pourcentage de la population de 15 ans et plus. Le taux d'emploi d'un groupe particulier (selon l'âge, le sexe, l'état matrimonial, la province, etc.) correspond au nombre de personnes occupées dans ce groupe exprimé en pourcentage de l'effectif de ce groupe.

### **Taux de glissement**

Mesure de l'erreur entre une estimation de la taille de la population et la valeur correspondante fondée sur les projections du recensement. Il est égal à  $1 - (\text{ratio des estimations de la projection})$ .

**Taux de vacance**

La proportion de logements qui sont inoccupés. Les logements hors du champ de l'enquête comme les entreprises ou les logements démolis ne sont pas compris dans le dénominateur.

**Tenue de liste**

L'opération de listage de l'EPA s'effectue en deux étapes. La seconde étape est la tenue continue de la liste préexistante. La liste était générée à l'origine par le listage initial ou provenait directement du RA (groupe 1). L'étendue des changements est habituellement mineure excepté si une croissance significative se produit dans le secteur de l'UPE, ou si des erreurs significatives sont trouvées dans le travail de listage précédent. Les mises à jour sont envoyées directement à l'administration centrale sans l'intervention des interviewers principaux, sauf si le sous-échantillonnage est requis en raison d'une croissance significative.

**Unité primaire d'échantillonnage**

Les unités sélectionnées à la première étape de l'échantillonnage dans un plan de sondage à plusieurs phases sont appelées les unités primaires d'échantillonnage, ou UPE. L'EPA utilise le Système automatisé de regroupement des territoires (SARTE) pour créer ses UPE.

**Variance de l'échantillonnage**

Mesure de la variation d'une statistique, calculée comme la valeur moyenne de la différence au carré de cette statistique et de sa moyenne pour l'ensemble de tous les échantillons possibles.

## Annexe A.2 Abréviations

AD	Aire de diffusion (recensement)
AR	Agglomération du recensement
BR	Bureau régional
CU	Centre urbain
C.V.	Coefficient de variation
DR	Division du recensement
EPA	Enquête sur la population active
FSI	Fraction de sondage inverse
MSI	Ménages sélectionnés par l'intervieweur
PPT	Probabilité proportionnelle à la taille
RA	Registre des adresses
RE	Région économique
REAE	Région économique de l'assurance-emploi
RHC	Rao-Hartley-Cochran (méthodes des groupes aléatoires)
RHDSC	Ressources humaines et Développement social Canada
RMR	Région métropolitaine du recensement
SARTE	Système automatisé de regroupement des territoires
SD	Secteur de dénombrement (recensement)
SDR	Subdivision du recensement
SPPTA	Systématique avec PPT et ordre aléatoire
SR	Secteur de recensement
UPD	Unité de premier degré
UPE	Unité primaire d'échantillonnage
USE	Unité secondaire d'échantillonnage

## Annexe B Caractéristiques de la base de sondage et du plan de sondage

**Tableau B.1 Nombre de ménages compris dans la base de sondage et les tailles d'échantillon par province**

Province	Ménages compris dans la base de sondage	Ménages exclus de la base de sondage	nombre		
			Échantillon financé par Statistique Canada	Échantillon financé par RHDSC	Échantillon total
Terre Neuve et Labrador	188 136	907	1 986	18	2 004
Île-du-Prince-Édouard	50 654	144	1 421	0	1 421
Nouvelle-Écosse	357 712	2 310	2 609	353	2 962
Nouveau-Brunswick	281 526	2 295	2 604	235	2 840
Québec	2 970 336	9 466	5 457	4 618	10 075
Ontario	4 226 887	20 165	7 171	8 722	15 893
Manitoba	418 140	14 510	3 254	520	3 774
Saskatchewan	367 815	11 863	3 409	488	3 897
Alberta	1 103 438	10 553	4 030	317	4 347
Colombie-Britannique	1 510 709	25 552	4 111	1,329	5 440
Canada	11 475 353	97 765	36 052	16 600	52 653

**Tableau B.2 Nombre de ménages dans les strates de régions éloignées, les strates à haut taux de vacance et les strates à trois degrés selon la province**

Province	Strates régions éloignées		Strates haut taux de vacances		Strates à trois degrés	
	Strates	Ménages	Strates	Ménages	Strates	Ménages
	nombre					
Terre Neuve et Labrador	1	3 012	1	5 184	0	0
Île-du-Prince-Édouard	0	0	0	0	0	0
Nouvelle-Écosse	0	0	1	4 196	0	0
Nouveau-Brunswick	0	0	1	1 904	0	0
Québec	1	11 373	2	23 546	3	16 192
Ontario	2	21 595	3	43 991	2	48 575
Manitoba	1	5 310	1	5 119	0	0
Saskatchewan	1	3 988	1	7 077	0	0
Alberta	2	15 673	2	14 213	2	8 529
Colombie-Britannique	2	13 390	3	26 253	1	10 231
Canada	10	74 341	15	131 483	8	83 527

**Tableau B.3.1 Statistiques pour les strates de hauts revenus**

Région métropolitaine de recensement (RMR)	Strates à hauts revenus	Ménages dans les strates à hauts revenus	Prévalence des ménages à hauts revenus dans les strates à hauts revenus	Prévalence des ménages à hauts revenus dans la RMR	Ménages à hauts revenus dans la RMR qui sont assignés à une strate à hauts revenus
	nombre		pourcentage		
St. John's	1	4 462	21,4	5,5	26,9
Halifax	2	11 327	23,7	5,8	31,9
Moncton	1	5 116	14,9	4,2	38,5
Saint John	1	5 111	16,4	4,4	39,2
Saguenay	1	4 703	9,8	2,8	26,9
Québec	1	20 380	21,8	4,1	36,1
Sherbrooke	2	7 327	10,0	2,9	32,8
Trois-Rivières	1	4 055	16,4	2,8	40,3
Montréal	5	118 543	28,1	5,8	40,3
Gatineau	1	6 173	25,3	6,4	23,4
Ottawa	1	18 020	43,3	14,2	17,5
Kingston	1	3 234	22,6	6,2	20,2
Oshawa	1	6 243	33,3	10,5	18,6
Toronto	6	133 184	44,0	13,6	26,3
Hamilton	1	15 728	35,0	9,1	23,8
St. Catharines - Niagara	1	8 224	18,2	5,3	18,9
Kitchener	1	10 029	34,4	8,8	25,2
Brantford	1	4 966	17,1	6,1	31,2
Guelph	1	5 448	23,0	9,0	31,6
London	2	18 428	28,1	6,8	42,5
Windsor	1	8 237	31,5	10,2	21,6
Greater Sudbury	1	3 941	21,0	5,8	22,4
Thunder Bay	1	3 594	17,2	5,2	23,9
Winnipeg	3	19 942	26,6	5,1	38,3
Regina	2	7 830	24,2	5,8	42,5
Saskatoon	1	5 167	21,5	4,8	25,9
Calgary	2	30 814	43,0	12,1	30,5
Edmonton	2	33 571	28,4	7,7	34,6
Abbotsford	1	3 849	14,7	5,2	21,3
Vancouver	4	56 335	30,0	8,8	25,2
Victoria	1	7 113	22,6	5,8	20,5

**Tableau B.3.2 Statistique pour les strates d'immigrants**

RMR	Strates d'immigrants	Ménages dans les strates d'immigrants	Prévalence des ménages d'immigrants <sup>1</sup> dans les strates d'immigrants	Prévalence des ménages d'immigrants dans la RMR	Ménages d'immigrants dans la RMR qui sont assignés à une strate d'immigrants
	nombre		pourcentage		
Montréal	5	118 828	21,8	4,2	43,4
Ottawa	1	17 868	38,3	7,9	23,2
Toronto	6	132 908	37,9	10,9	28,3
Calgary	2	32 017	24,8	8,5	25,2
Vancouver	4	56 330	31,5	9,5	24,6

1. Un ménage d'immigrants est un ménage dans lequel au moins un membre est immigrant. Pour Montréal, Toronto et Vancouver, un membre est immigrant s'il a immigré dans les cinq dernières années. Pour Ottawa et Calgary, le membre est immigrant s'il est arrivé au pays dans les 10 dernières années.

**Tableau B.3.3 Statistiques pour les strates d'Autochtones**

Province	Strates d'Autochtones	Ménages dans les strates d'Autochtones	Prévalence des ménages d'Autochtones <sup>2</sup> dans une strate d'Autochtones	Prévalence des ménages d'Autochtones dans la province	Ménages d'Autochtones dans la province qui sont assignés à une strate d'Autochtones
	nombre			pourcentage	
Saskatchewan	6	28 741	38,7	13,1	19,8
Manitoba	7	27 018	34,6	11,8	21,1
Alberta	7	83 137	17,5	6,0	22,7
Colombie-Britannique	11	144 004	14,2	5,2	26,4

2. Un ménage autochtone est un ménage pour lequel au moins un membre est autochtone.

**Table B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales**

Province	UPE	Strates	Population	Logements	Ménages	Échantillon		
			nombre			%	nombre	%
<b>Terre-Neuve et Labrador</b>	<b>967</b>	<b>37</b>	<b>499 608</b>	<b>219 337</b>	<b>184 842</b>	<b>100,0</b>	<b>1 944</b>	<b>100,0</b>
REAE								
01	310	14	170 848	68 346	64 133	34,7	664	34,2
02	657	23	328 760	150 991	120 709	65,3	1 280	65,8
RÉ								
1010	462	20	242 562	102 117	90 001	48,7	932	47,9
1020	89	3	42 133	18 387	15 309	8,3	195	10,0
1030	183	7	100 326	44 592	36 982	20,0	380	19,5
1040	233	7	114 587	54 241	42 550	23,0	437	22,5
Centre urbain <sup>1</sup>								
St. John's (RMR)	314	...	175 918	69 118	64 831	35,1	671	34,5
Corner Brook (CU)	43	...	25 790	10 769	9 833	5,3	103	5,3
Autre urbain	69	...	35 819	14 131	13 142	7,1	138	7,1
Non urbain	541	...	265 081	125 319	97 036	52,5	1 032	53,1
Type de strate								
Régulier	909	35	472 835	198 394	175 196	94,8	1 870	96,2
Haut revenu	21	1	12 948	4 751	4 462	2,4	45	2,3
Taux de vacance élevé	37	1	13 825	16 192	5 184	2,8	29	1,5
<b>Île-du-Prince Édouard</b>	<b>286</b>	<b>22</b>	<b>134 876</b>	<b>55 845</b>	<b>50 654</b>	<b>100,0</b>	<b>1 378</b>	<b>100,0</b>
REAE								
03	286	22	134 876	55 845	50 654	100,0	1 378	100,0
RÉ								
1110	286	22	134 876	55 845	50 654	100,0	1 378	100,0
Centre urbain <sup>1</sup>								
Charlottetown (CU)	116	...	58 211	24 069	22 351	44,1	608	44,1
Summerside (CU)	32	...	16 200	6 593	6 356	12,5	173	12,6
Non urbain	138	...	60 465	25 183	21 947	43,3	597	43,3
Type de strate								
Régulier	286	22	134 876	55 845	50 654	100,0	1 378	100,0
<b>Nouvelle-Écosse</b>	<b>1 681</b>	<b>57</b>	<b>898 763</b>	<b>400 740</b>	<b>365 459</b>	<b>100,0</b>	<b>2 873</b>	<b>100,0</b>
REAE								
04	316	16	180 860	81 940	70 150	19,2	615	21,4
05	713	20	368 980	170 482	147 957	40,5	1 115	38,8
06	652	21	348 923	148 318	147 352	40,3	1 143	39,8
RÉ								
1210	241	10	141 585	63 366	55 199	15,1	486	16,9
1220	306	9	156 394	71 911	61 866	16,9	457	15,9
1230	226	6	120 048	52 514	47 166	12,9	369	12,8
1240	236	7	121 640	59 623	49 771	13,6	385	13,4
1250	672	25	359 096	153 326	151 457	41,4	1 176	40,9

1. Voir la note à la fin du tableau

Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
				nombre			%	nombre	%	
<b>Nouvelle-Écosse (suite)</b>										
Centre urbain <sup>1</sup>										
Halifax (RMR)	672	...	359 096	153 326	151 457	41,4	1,176	40,9		
Cape Breton (CU)	175	...	105 968	45 491	41 609	11,4	364	12,7		
Truro (CU)	79	...	43 455	19 270	17 434	4,8	123	4,3		
New Glasgow (CU)	71	...	36 341	15 826	14 333	3,9	102	3,6		
Autre urbain	82	...	38 999	18 063	16 831	4,6	130	4,5		
Non urbain	602	...	314 904	148 764	123 795	33,9	978	34,0		
Type de strate										
Régulier	1 601	54	856 537	378 069	349 936	95,8	2 760	96,1		
Haut revenu	58	2	32 115	11 910	11 327	3,1	92	3,2		
Taux de vacance élevé	22	1	10 111	10 761	4 196	1,1	21	0,7		
<b>Nouveau-Brunswick</b>	<b>1 414</b>	<b>54</b>	<b>721 400</b>	<b>310 412</b>	<b>282 671</b>	<b>100,0</b>	<b>2 754</b>	<b>100,0</b>		
REAE										
07	<b>666</b>	<b>25</b>	<b>351 009</b>	<b>147 304</b>	<b>139 156</b>	<b>49,2</b>	<b>1 236</b>	<b>44,9</b>		
08	251	12	119 028	51 510	462 250	16,4	647	23,5		
09	497	17	251 363	111 598	97 265	34,4	871	31,6		
RÉ										
1310	323	13	167 215	70 195	64 284	22,7	578	21,0		
1320	353	12	180 906	79 375	71 546	25,3	639	23,2		
1330	336	14	167 022	71 455	65 547	23,2	643	23,3		
1340	230	7	123 540	54 311	49 387	17,5	452	16,4		
1350	172	8	82 717	35 076	31 907	11,3	442	16,0		
Centre urbain <sup>1</sup>										
Moncton (CU)	225	...	117 681	48 920	46 801	16,6	420	15,3		
Saint John's (RMR)	243	...	122 678	51 775	48 262	17,1	433	15,7		
Fredericton (CU)	142	...	80 534	34 089	33 210	11,7	286	10,4		
Bathurst (CU)	46	...	23 815	10 443	9 550	3,4	86	3,1		
Miramichi (CU)	44	...	24 425	10 268	9 392	3,3	84	3,1		
Edmundston (CU)	43	...	22 038	9 879	9 014	3,2	125	4,5		
Autre urbain	45	...	20 922	9 123	8 529	3,0	88	3,2		
Non urbain	626	...	309 307	135 915	117 913	41,7	1 232	44,7		
Type de strate										
Régulier	1 357	51	686 589	297 914	270 540	95,7	2 645	96,0		
Cout élevé	11	1	5 038	2 040	1 904	0,7	17	0,6		
Haut revenu	46	2	29 773	10 458	10 227	3,6	92	3,3		
<b>Québec</b>	<b>14 598</b>	<b>208</b>	<b>7 204 393</b>	<b>3 219 007</b>	<b>3 033 553</b>	<b>100,0</b>	<b>9 773</b>	<b>5,9</b>		
REAE										
10	290	11	147 382	65 797	58 721	1,9	577	5,9		
11	1 415	15	681 517	310 287	303 795	10,0	672	6,9		
12	296	14	137 361	64 414	60 012	2,0	651	6,7		
13	289	11	145 548	60 578	55 426	1,8	638	6,5		
14	313	17	151 605	71 319	65 886	2,2	793	8,1		
15	960	14	466 604	203 012	187 623	6,2	747	7,6		
16	6 869	56	3 429 391	1 474 891	1 467 550	48,4	2 199	22,5		
17	1 996	19	949 056	478 823	392 611	12,9	924	9,5		
18	479	8	233 747	114 748	93 577	3,1	498	5,1		
19	888	12	448 950	200 062	178 335	5,9	727	7,4		
20	498	17	258 294	109 958	107 820	3,6	758	7,8		
21	305	14	154 938	65 118	62 197	2,1	589	6,0		

1. Voir la note à la fin du tableau

Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
			nombre				%	nombre	%	
<b>Québec (suite)</b>										
<b>RÉ</b>										
2410	178	7	94 994	42 050	37 643	1,2	409	4,2		
2415	430	7	200 630	94 444	81 830	2,7	403	4,1		
2420	1 350	18	637 677	299 210	286 618	9,4	637	6,5		
2425	772	16	383 376	161 768	150 565	5,0	817	8,4		
2430	600	19	285 613	136 776	120 369	4,0	1 006	10,3		
2433	442	7	217 931	94 488	87 619	2,9	249	2,5		
2435	2 474	33	1 277 949	527 886	507 854	16,7	1 219	12,5		
2440	3 929	26	1 814 117	844 746	841 285	27,7	1 153	11,8		
2445	633	5	343 419	135 811	139 591	4,6	271	2,8		
2450	725	5	386 824	167 892	146 756	4,8	286	2,9		
2455	884	8	460 952	218 799	181 720	6,0	369	3,8		
2460	634	17	315 181	147 139	131,340	4,3	862	8,8		
2465	286	7	142 329	63 462	57 788	1,9	329	3,4		
2470	545	14	252 338	120 600	110 115	3,6	753	7,7		
2475	536	15	275 845	114 451	108 207	3,6	754	7,7		
2480	145	3	89 460	40 178	35 773	1,2	206	2,1		
2490	35	1	25 758	9 307	8 480	0,3	50	0,5		
<b>Centre urbain<sup>1</sup></b>										
Chicoutimi (RMR)	305	...	154 938	65 118	62 197	2,1	589	6,0		
Québec (RMR)	1 415	...	681 517	310 287	303 795	10,0	672	6,9		
Sherbrooke (RMR)	317	...	153 811	72 218	66 731	2,2	803	8,2		
Trois-Rivières (RMR)	296	...	137 361	64 414	60 012	2,0	651	6,7		
Shawinigan (CU)	122	...	57 304	28 543	25 601	0,8	54	0,6		
Drummondville (CU)	140	...	68 451	30 266	28 948	1,0	70	0,7		
Granby (CU)	130	...	60 264	26 585	25 285	0,8	101	1,0		
Saint-Jean-sur-Richelieu (CU)	164	...	79 919	34 014	32 797	1,1	129	1,3		
Montréal (RMR)	6 869	...	3 429 391	1 474 891	1 467 550	48,4	2 199	22,5		
Gatineau (RMR)	498	...	258 294	109 958	107 820	3,6	758	7,8		
Rimouski (CU)	91	...	47 688	22 277	20 040	0,7	79	0,8		
Baie-Comeau (CU)	57	...	28 940	12 831	11 615	0,4	72	0,7		
Sept-Îles (CU)	43	...	24 721	11 313	10 204	0,3	63	0,6		
Rouyn-Noranda/Val-d'Or (CU)	130	...	67 657	30 835	28 165	0,9	158	1,6		
Autre urbain	721	...	396 528	177 781	166 701	5,5	761	7,8		
Non urbain	3 300	...	1 557 609	747 676	616 092	20,3	2 614	26,7		
<b>Type de strate</b>										
Régulier	13 070	189	6 394 430	2 847 917	2 703 149	89,1	9 082	92,9		
Haut revenu	779	11	458 544	167 589	163 747	5,4	441	4,5		
Taux de vacance élevé	131	2	51 669	63 903	23 354	0,8	51	0,5		
Immigrant	543	5	267 034	126 205	131 930	4,3	169	1,7		
Région éloignées	75	1	32 716	13 393	11 373	0,4	30	0,3		

1. Voir la note à la fin du tableau  
Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
			nombre				%	nombre	%	
<b>Ontario</b>	<b>19 905</b>	<b>332</b>	<b>1 143 966</b>	<b>4 537 466</b>	<b>4 403 105</b>	<b>100,0</b>	<b>15 416</b>	<b>100,0</b>		
REAE										
22	1 490	23	822 608	326 805	329 173	7,5	872	5,7		
23	779	10	401 080	180 391	157 812	3,6	841	5,5		
24	276	12	145 610	63 746	59 295	1,3	677	4,4		
25	1 932	23	1 028 112	502 899	395 046	9,0	1 259	8,2		
26	504	14	296 770	106 576	108 614	2,5	653	4,2		
27	7 631	74	4 747 842	1 693 381	1 785 752	40,6	2 640	17,1		
28	1 238	16	664 593	261 718	261 700	5,9	758	4,9		
29	744	17	377 009	159 032	152 795	3,5	847	5,5		
30	791	19	415 990	177 522	176 612	4,0	872	5,7		
31	437	15	254 636	98 157	93 642	2,1	724	4,7		
32	568	18	300 501	121 006	116 531	2,6	648	4,2		
33	717	16	415 997	159 339	155 664	3,5	702	4,6		
34	567	10	314 297	129 721	121 150	2,8	719	4,7		
35	908	22	501 012	192 682	186 326	4,2	1 056	6,9		
36	291	16	150 461	66 588	62 478	1,4	701	4,5		
37	250	13	120 353	53 405	49 001	1,1	608	3,9		
38	782	14	476 095	244 498	191 514	4,3	839	5,4		
REÉ										
3510	2 072	30	1 122 635	456 715	447 468	10,2	1 502	9,7		
3515	831	19	422 350	195 847	168 989	3,8	1 040	6,7		
3520	634	8	337 306	189 030	132 249	3,0	433	2,8		
3530	8 058	86	4 996 356	1 784 224	1 878 538	42,7	3 205	20,8		
3540	1 847	36	1 062 025	422 275	391 419	8,9	1 660	10,8		
3550	2 380	48	1 273 928	510 986	500 710	11,4	2 282	14,8		
3560	1 105	27	586 661	242 328	238 877	5,4	1 204	7,8		
3570	1 116	24	605 638	247 053	234 257	5,3	1 347	8,7		
3580	539	11	279 158	124 517	107 605	2,4	595	3,9		
3590	975	27	536 612	268 659	218 931	5,0	1 371	8,9		
3595	348	16	210 297	95 832	84 062	1,9	777	5,0		
Centre urbain <sup>1</sup>										
Corwall (CU)	101	...	57 581	24 634	23 747	0,5	125	0,8		
Ottawa (RMR)	1 471	...	810 462	322 544	325 020	7,4	849	5,5		
Kingston (RMR)	276	...	145 610	63 746	59 295	1,3	677	4,4		
Belleville (CU)	182	...	87 395	37 185	35 814	0,8	78	0,5		
Peterborough (CU)	198	...	100 738	43 893	39 985	0,9	208	1,3		
Kawartha Lakes (CU)	113	...	69 179	34 637	27 182	0,6	64	0,4		
Oshawa (RMR)	504	...	296 770	106 576	108 614	2,5	653	4,2		
Toronto (RMR)	7 631	...	4 747 842	1 693 381	1 785 752	40,6	2 640	17,1		
Hamilton (RMR)	1 238	...	664 593	261 718	261 700	5,9	758	4,9		
St-Catharines/Niagara (RMR)	744	...	377 009	159 032	152 795	3,5	847	5,5		
Kitchener (RMR)	717	...	415 997	159 339	155 664	3,5	702	4,6		
Brantford (CU)	166	...	86 417	34 881	33 849	0,8	315	2,0		
Norfolk (CU)	101	...	59 947	24 075	22 526	0,5	141	0,9		
Guelph (CU)	216	...	117 345	46 253	47 128	1,1	268	1,7		
London (RMR)	828	...	435 104	184 706	183 600	4,2	909	5,9		
Chatham-Kent (CU)	178	...	106 864	44 529	42 202	1,0	251	1,6		
Windsor (RMR)	582	...	308 735	124 402	119 397	2,7	662	4,3		
Sarnia (CU)	173	...	87 636	37 581	35 505	0,8	213	1,4		
Barrie (CU)	273	...	156 457	59 227	54 996	1,2	231	1,5		
North Bay (CU)	123	...	62 303	27 169	24 896	0,6	111	0,7		
Greater Sudbury (RMR)	300	...	155 219	68 690	64 351	1,5	723	4,7		
Sault Ste-Marie (CU)	160	...	78 049	34 539	31 919	0,7	140	0,9		
Thunder Bay (CU)	252	...	121 387	53 870	49 337	1,1	612	4,0		
Autre urbain	944	...	607 718	258 188	243 645	5,5	1 184	7,7		
Non urbain	2 434	...	1 276 609	632 671	474 186	10,8	2 055	13,3		

1. Voir la note à la fin du tableau

Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
				nombre				%	nombre	%
<b>Ontario (suite)</b>										
Type de strate										
Régulier	17 739	301	10 086 861	3 964 127	3 923 649	89,1	14 218	92,2		
Haut revenu	1 174	19	742 037	244 895	242 493	5,5	781	5,1		
Taux de vacance élevé	231	3	102 463	141 673	43 548	1,0	109	0,7		
Immigrant	617	7	445 725	154 294	171 820	3,9	257	1,7		
Régions éloignées	144	2	55 880	32 477	21 595	0,5	51	0,3		
<b>Manitoba</b>	<b>2 120</b>	<b>71</b>	<b>1 062 777</b>	<b>452 019</b>	<b>431 986</b>	<b>100,0</b>	<b>3 661</b>	<b>100,0</b>		
REAE										
30	1 318	41	671 216	280 231	283 789	65,7	1 917	52,4		
40	622	20	310 898	130 561	117 015	27,1	1 017	27,8		
41	180	10	80 663	41 227	31 182	7,2	727	19,9		
RÉ										
4610	158	9	83 427	31 904	29 044	6,7	359	9,8		
4620	98	3	51 703	19 248	18 201	4,2	189	5,2		
4630	218	6	100 658	46 576	40 770	9,4	314	8,6		
4640	84	4	44 093	16 490	15 236	3,5	156	4,3		
4650	1 233	38	621 765	262 002	267 337	61,9	1 800	49,2		
4660	148	4	79 463	37 334	29 031	6,7	356	9,7		
4670	94	3	41 685	20 794	17 672	4,1	163	4,5		
4680	87	4	39 983	17 671	14 695	3,4	324	8,9		
Centre urbain <sup>1</sup>										
Winnipeg (RMR)	1 318	...	671 216	280 231	283 789	65,7	1 917	52,4		
Brandon (CU)	82	...	41 037	17 966	17 185	4,0	134	3,7		
Autre urbain	87	...	41 737	18 050	16 985	3,9	309	8,4		
Non urbain	633	...	308 787	135 772	114 027	26,4	1 301	35,5		
Type de strate										
Régulier	1 806	60	902 291	902 291	371 650	86,0	3 157	86,2		
Autochtone	146	6	73 210	73 210	28 741	6,7	289	7,9		
Haut revenu	99	3	60 247	60 247	21 166	4,9	143	3,9		
Taux de vacance élevé	28	1	12 097	12 097	5 119	1,2	30	0,8		
Régions éloignées	41	1	14 932	14 932	5 310	1,2	42	1,1		
<b>Saskatchewan</b>	<b>2 006</b>	<b>72</b>	<b>928 348</b>	<b>415 768</b>	<b>369 964</b>	<b>100,0</b>	<b>3 780</b>	<b>100,0</b>		
REAE										
42	397	19	192,à 800	80 772	77 061	20,8	902	23,9		
43	429	17	225 772	94 639	91 053	24,6	799	21,1		
44	767	22	314 420	148 951	127 625	34,5	1 289	34,1		
45	413	14	195 356	91,406	74 225	20,1	790	20,9		
RÉ										
4710	580	25	266 594	116 521	106 809	28,9	1 183	31,3		
4720	239	8	104 048	48 261	41 789	11,3	472	12,5		
4730	571	26	282 498	119 785	113 074	30,6	1 011	26,7		
4740	211	5	83 672	41 261	35 488	9,6	339	9,0		
4750	377	7	179 105	84 786	68 816	18,6	744	19,7		
4760	28	1	12 431	5 154	3 988	1,1	31	0,8		
Centre urbain <sup>1</sup>										
Regina (RMR)	397	...	192 800	80 772	77 061	20,8	902	23,9		
Saskatoon (RMR)	429	...	225 772	94 639	91 053	24,6	799	21,1		
Moose Jaw (CU)	64	...	33 616	15 056	14 061	3,8	161	4,3		
Prince Albert (CU)	69	...	41 200	15 977	15 133	4,1	168	4,4		
Autre urbain	124	...	61 248	27 114	25 611	6,9	273	7,2		
Non urbain	923	...	373 712	182 210	147 045	39,7	1 477	39,1		

1. Voir la note à la fin du tableau

Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
			nombre				%	nombre	%	
<b>Saskatchewan (suite)</b>										
Type de strate										
Régulier	1 731	61	790 087	344 936	318 695	86,1	3 296	87,2		
Autochtone	139	6	68 922	29 379	27 018	7,3	283	7,5		
Haut revenu	67	3	40 105	13 377	13 405	3,6	139	3,7		
Taux de vacance élevé	41	1	16 803	22 922	6 858	1,9	31	0,8		
Régions éloignées	28	1	12 431	5 154	3 988	1,1	31	0,8		
<b>Alberta</b>	<b>5 317</b>	<b>85</b>	<b>2 948 116</b>	<b>1 165 969</b>	<b>1 155 109</b>	<b>100,0</b>	<b>5 416</b>	<b>100,0</b>		
<b>REAE</b>										
46	1 699	22	971 117	376 446	388 273	33,6	1 206	22,3		
47	1 711	22	938 129	372 728	378 920	32,8	1 392	25,7		
48	339	14	198 600	81 163	69 382	6,0	1 010	18,6		
49	1 568	27	840 270	335 632	318 534	27,6	1 808	33,4		
<b>RÉ</b>										
4810	447	8	233 502	91 700	88 312	7,6	303	5,6		
4820	371	9	179 432	72 365	66 546	5,8	439	8,1		
4830	1 824	23	1 040 645	402 812	414 375	35,9	1 291	23,8		
4840	116	4	71 928	31 069	27 314	2,4	268	4,9		
4850	281	5	151 286	60 696	61 543	5,3	333	6,1		
4860	1 776	21	974 496	388 697	392 588	34,0	1 476	27,3		
4870	363	10	208 643	83 851	73 839	6,4	766	14,1		
4880	139	5	88 184	34 779	30 592	2,6	540	10,0		
<b>Centre urbain<sup>1</sup></b>										
Medicine Hat (CU)	122	...	61 735	25 693	24 372	2,1	77	1,4		
Lethbridge (CU)	138	...	67 375	29 276	29 557	2,6	97	1,8		
Calgary (RMR)	1 699	...	971 117	376 446	388 273	33,6	1 206	22,3		
Red Deer (CU)	135	...	69 608	27 871	32 097	2,8	183	3,4		
Edmonton (RMR)	1 711	...	938 129	372 728	378 920	32,8	1 392	25,7		
Autre urbain	194	...	128 397	51 069	48 189	4,2	397	7,3		
Non urbain	1 318	...	711 755	282 886	253 701	22,0	2 064	38,1		
<b>Type de strate</b>										
Régulier	4 298	69	2 363 145	942 186	940 942	81,5	3 852	71,1		
Autochtone	422	7	214 115	87 886	83 934	7,3	1 173	21,7		
Cout élevé	64	1	27 485	10 935	9 661	0,8	63	1,2		
Haut revenu	295	4	200 078	66 279	67 105	5,8	180	3,3		
Taux de vacance élevé	26	1	11 089	11 718	4 552	0,4	20	0,4		
Immigrant	145	2	94 642	33 372	36 915	3,2	88	1,6		
Régions éloignée	67	1	37 562	13 593	12 000	1,0	40	0,7		
<b>Colombie-Britannique</b>	<b>7 311</b>	<b>124</b>	<b>3 838 992</b>	<b>1 615 758</b>	<b>1 558 403</b>	<b>100,0</b>	<b>6 377</b>	<b>100,0</b>		
<b>REAE</b>										
50	1 192	13	591 184	270 509	247 068	15,9	954	15,0		
51	242	11	147 895	52 706	53 668	3,4	573	9,0		
52	3 676	54	1 982 716	784 745	797 388	51,2	2 346	36,8		
53	644	18	307 340	139 992	136 267	8,7	914	14,3		
54	985	14	490 650	231 587	203 222	13,0	834	13,1		
55	572	14	319 207	136 219	120 790	7,8	756	11,9		
<b>RÉ</b>										
5910	1 384	32	671 630	310 359	289 090	18,6	1 542	24,2		
5920	4 181	68	2 269 754	904 810	906 594	58,2	3 154	49,5		
5930	886	9	447 080	199 614	186 651	12,0	697	10,9		
5940	306	3	144 104	70 895	60 417	3,9	260	4,1		
5950	292	6	157 742	67 890	60 320	3,9	327	5,1		
5960	79	2	51 954	21 849	19 490	1,3	149	2,3		
5970	75	1	38 726	16 448	14 306	0,9	89	1,4		
5980	108	3	58 002	23 893	21 535	1,4	166	2,6		

1. Voir la note à la fin du tableau

Nota : voir Annexe A.2 pour les abréviations.

**Tableau B.4 Caractéristiques du plan de sondage pour des régions infra-provinciales (suite)**

Province	UPE	Strates	Population		Logements		Ménages		Échantillon	
			nombre				%	nombre	%	
<b>Colombie-Britannique (suite)</b>										
Centre urbain <sup>1</sup>										
Kelowna (CU)	270	...	139 882	58 875	58 132	3,7	212	3,3		
Vernon (CU)	96	...	49 727	21 583	20 868	1,3	61	1,0		
Kamloops (CU)	160	...	84 721	36 040	33 770	2,2	147	2,3		
Chilliwack (CU)	120	...	67 359	27 360	26 049	1,7	109	1,7		
Abbotsford (RMR)	242	...	147 895	52 706	53 668	3,4	573	9,0		
Vancouver (RMR)	3 676	...	1 982 716	784 745	797 388	51,2	2 346	36,8		
Victoria (RMR)	644	...	307 340	139 992	136 267	8,7	914	14,3		
Nanaimo (CU)	179	...	85 008	37 349	35 395	2,3	147	2,3		
Prince George (CU)	159	...	84 929	34 400	32 085	2,1	190	3,0		
Dawson Creek (CU)	37	...	17 251	7 154	6 662	0,4	53	0,8		
Autre urbain	1 305	...	629 094	308 427	257 064	16,5	469	7,4		
Non urbain	423	...	243 070	107 127	101 055	6,5	1 156	18,1		
Type de strate										
Régulier	5 745	98	3 030 838	1 267 066	1 239 216	7,5	4 382	68,7		
Autochtone	733	11	346 444	153 511	147 029	9,4	1 505	23,6		
Cout élevé	78	2	33 910	17 398	13 741	0,9	54	0,8		
Haut revenu	330	6	197 680	69 909	68 487	4,4	228	3,6		
Niveau de vacance élevé	63	1	28 365	33 282	12 383	0,8	28	0,4		
Immigrants	274	4	165 880	58 754	64 157	4,1	142	2,2		
Régions éloignées	88	2	35 875	15 838	13 390	0,9	38	0,6		

1. Les limites des RMR suivent les Régions métropolitaines de recensement de 2006 (tel que définies en 2003). Autres centres urbains (CU) suivent les Agglomérations du Recensement de 2001, s'il y a lieu.

Nota : voir Annexe A.2 pour les abréviations.

## Liste des variables utilisées pour effectuer la stratification optimale

La liste des variables utilisées pour effectuer la stratification optimale est identique à celle utilisée lors du précédent remaniement.

Le choix des variables de stratification est adapté à chaque région faisant l'objet d'une stratification optimale. Pour chaque UPE de la région, on a obtenu les variables ci-dessous à partir des données du Recensement de 2001. Si une variable représentait moins de 2 % de la population totale, elle était écartée. Pour les catégories comme les services, si une sous-catégorie, telle les services financiers, comptait trop peu d'effectif, alors on a plutôt utilisé la variable globale. Une catégorie est considérée significative si elle représente plus de 2 % de la population.

Nombre de personnes occupées dans les secteurs suivants:

- agriculture
- foresterie ou pêcheries
- mines
- secteur manufacturier - biens de consommation
- secteur manufacturier - caoutchouc, plastique, cuir
- secteur manufacturier - textiles et vêtements
- secteur manufacturier - meubles, pâtes et papier, imprimerie, bois
- secteur manufacturier - métaux et minéraux
- secteur manufacturier - pétrochimie et chimie
- construction
- transports
- services - commerciaux
- services - financiers
- services - personnels / entreprises
- services - gouvernement

Nombre total de personnes occupées

Revenu total

Population de 15 ans et plus

Population de 15 à 24 ans

Population de 55 ans et plus

Nombre de ménages d'une personne

Nombre de ménages de deux personnes

Nombre de logements en propriété

Loyer brut total

Population ayant fait des études secondaires

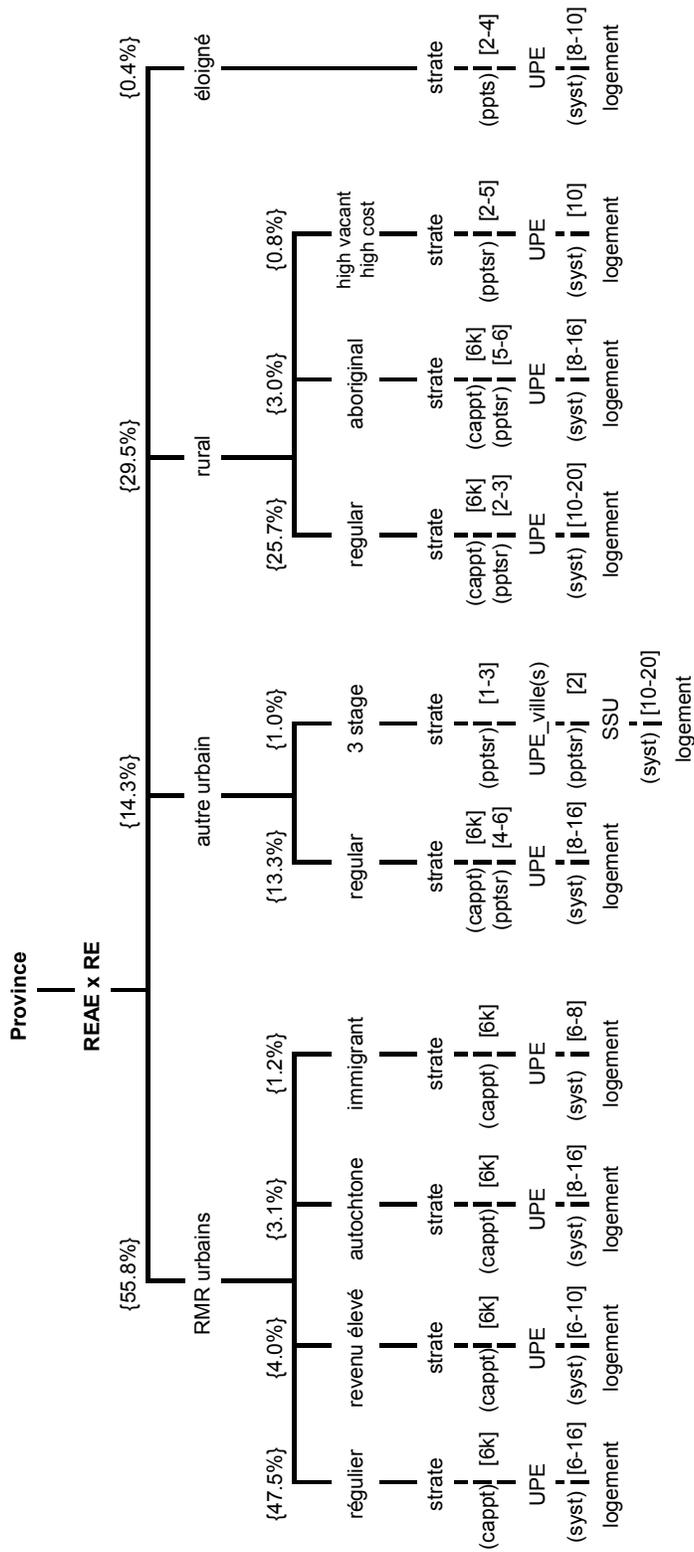
Nombre de personnes de langue maternelle anglaise

Nombre de personnes de langue maternelle française

Nombre de personnes de langue maternelle autre que l'anglais ou le français

# Annexe C

## Plan de sondage de l'Enquête sur la population active - depuis 2005



- = niveau de stratification
  - = degré d'échantillonnage
  - REAE région économique
  - RE région économique
  - RMR région métropolitaine de recensement
  - UPE unité primaire d'échantillonnage
  - USE unité secondaire d'échantillonnage
  - grappe ensemble de côtés d'îlots
- Quelques strates régulières et autochtones avaient trop peu de groupes pour justifier le capppt
- {%} = pourcentage de l'échantillon total (y compris l'échantillon complémentaire et le baisse de 3 %)
  - ( ) = méthode de sélection
  - [ ] = nombre d'unités sélectionnées (6k = multiple de 6)
  - ppt = probabilité proportionnelle à la taille
  - capppt = classement aléatoire ppt
  - pptsr = ppts randomisé
  - ppts = ppt systématique (liste ordonnée)
  - sys = ppt systématique

## Annexe D Exemples de cartes d'UPE (formule F01)

Avant de présenter certains exemples de cartes d'UPE, voici les détails de la légende de la F01, de haut en bas.

STATISTIQUE CANADA - STATISTICS CANADA SCHEMA DE LA GRAPPE / CLUSTER DIAGRAM	
NUMERO DE L'ECHANTILLON SAMPLE IDENTIFICATION <b>41005-01-999-9</b> GROUPE RA / GROUP AR <b>3</b>	
<small>MISE A JOUR DU RESEAU ROUTIER /            STREET NETWORK VINTAGE : 2004-02-19</small>	
COMPTE / COUNT : <b>999</b> ENQ / SVID : <b>10440</b> OCH / RST : <b>999</b> FS / ISR : <b>999</b> DATE : <b>YYYYMM</b> LFS.DATE.EPA : <b>0</b> ENDROIT / LOCATION :	Identification de l'UPE (strate : 41005; type de plan de sondage : 01; UPE 999 ; groupe de rotation : 9)  Groupe RA (3) Date du dernier réseau routier (2004-02-19) Nombre prévu de ménages dans l'UPE (999) Première enquête qui utilise l'UPE (10440) Début aléatoire et taux d'échantillonnage inverse à utiliser dans la sélection des ménages (999, 999) Date de la première utilisation (YYYYMM) Date de la première utilisation de l'EPA (habituellement vide) (0)
<b>WARWICK</b> SR / CT : <b>0000.00</b> AD / DA : <b>24390068</b> GRA REM / CLU REP : <b>1 de/of 2</b>	
COUT DE LISTAGE / LISTING COST HRS : KMS : AUTRES / OTHER : INTRVR# : DERN. LIGNE / LAST LINE :	
- - - - - GRAPPE / CLUSTER - - - - - SDR / CSD - - - - - ÎLOT / BLOCK - - - - - LIGNE ELECTRIQUE / POWER LINE - - - - - HYDROGRAPHIE / HYDROGRAPHY + + + + + CHEMIN DE FER / RAILROAD - - - - - RUE / STREET - - - - - ROUTE UTILITAIRE / UTILITY ROAD	Emplacement général fondé sur le Fichier de conversion des codes postaux (Warwick) Secteur de recensement selon la classification de 2001 pour repérer l'UPE (0000,00) Aire de diffusion prédominante selon la classification de 2001 (24390068)  Numéro de la carte et nombre de cartes pour cette UPE (1 de 2)  Formulaire des intervieweurs pour enregistrer les coûts de listage.
 1:35,800 0 0.6 1.2 1.8 Km EPA/LFS F01 DATE D'IMPRESSION / PRINT DATE : 15-09-2006	

Légende des traits géographiques :

Les îlots sont normalement délimités par un trait gris avec un point de départ (x encerclé) dans chacun.  
 L'aire à l'extérieur de la grappe est en vert, l'intérieur est en jaune.  
 Seules les routes utilitaires (chemins) sont distinguées séparément.  
 Les routes proposées ne sont pas facilement repérées et entretenues.

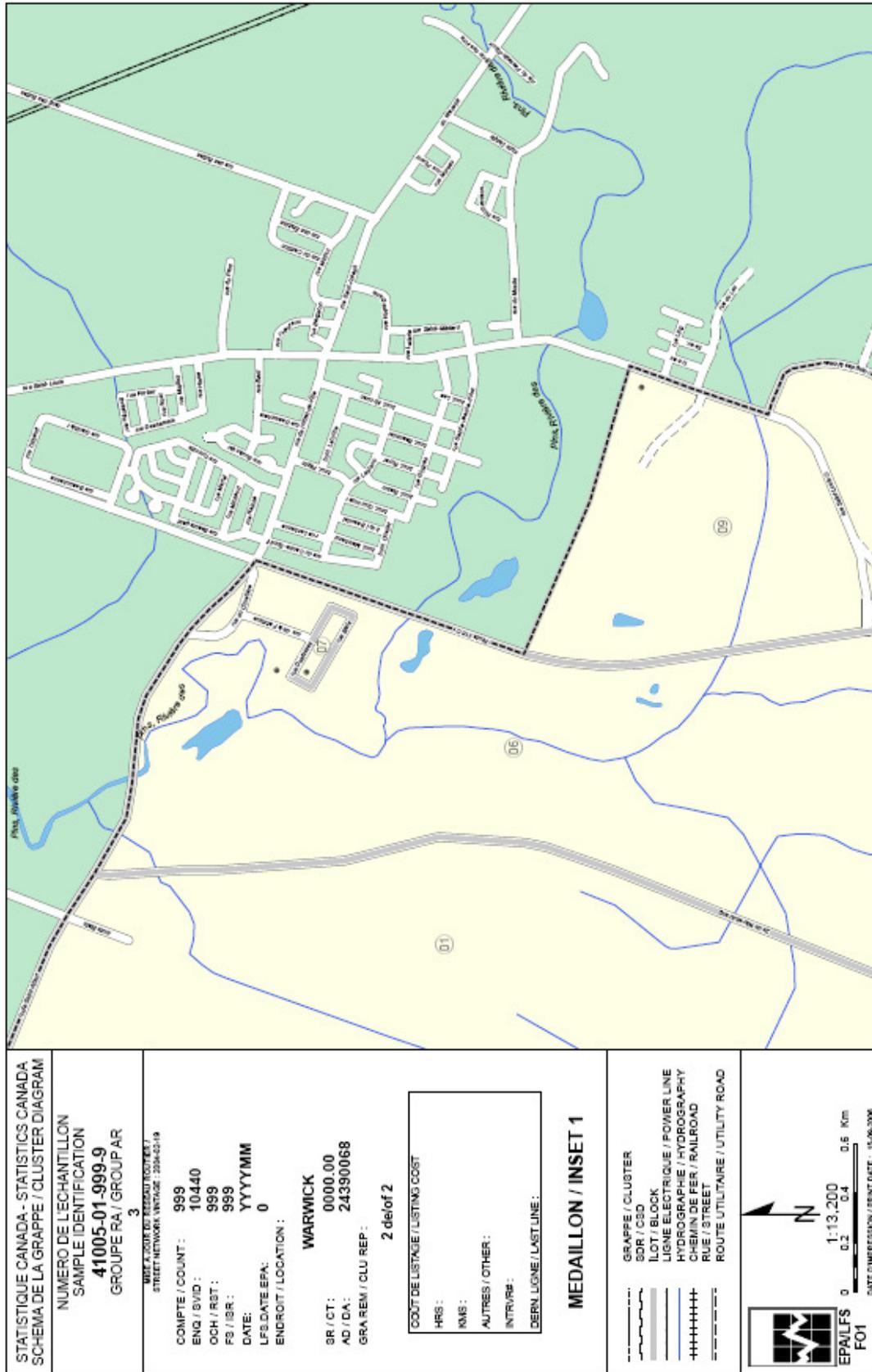
Le symbole nord pointe toujours vers le haut ou vers la droite pour faciliter la consultation de la carte lorsqu'elle se trouve dans le cartable de l'intervieweur.

L'échelle est utile pour mesurer la distance de conduite lorsque les caractéristiques géographiques manquent près de la limite.

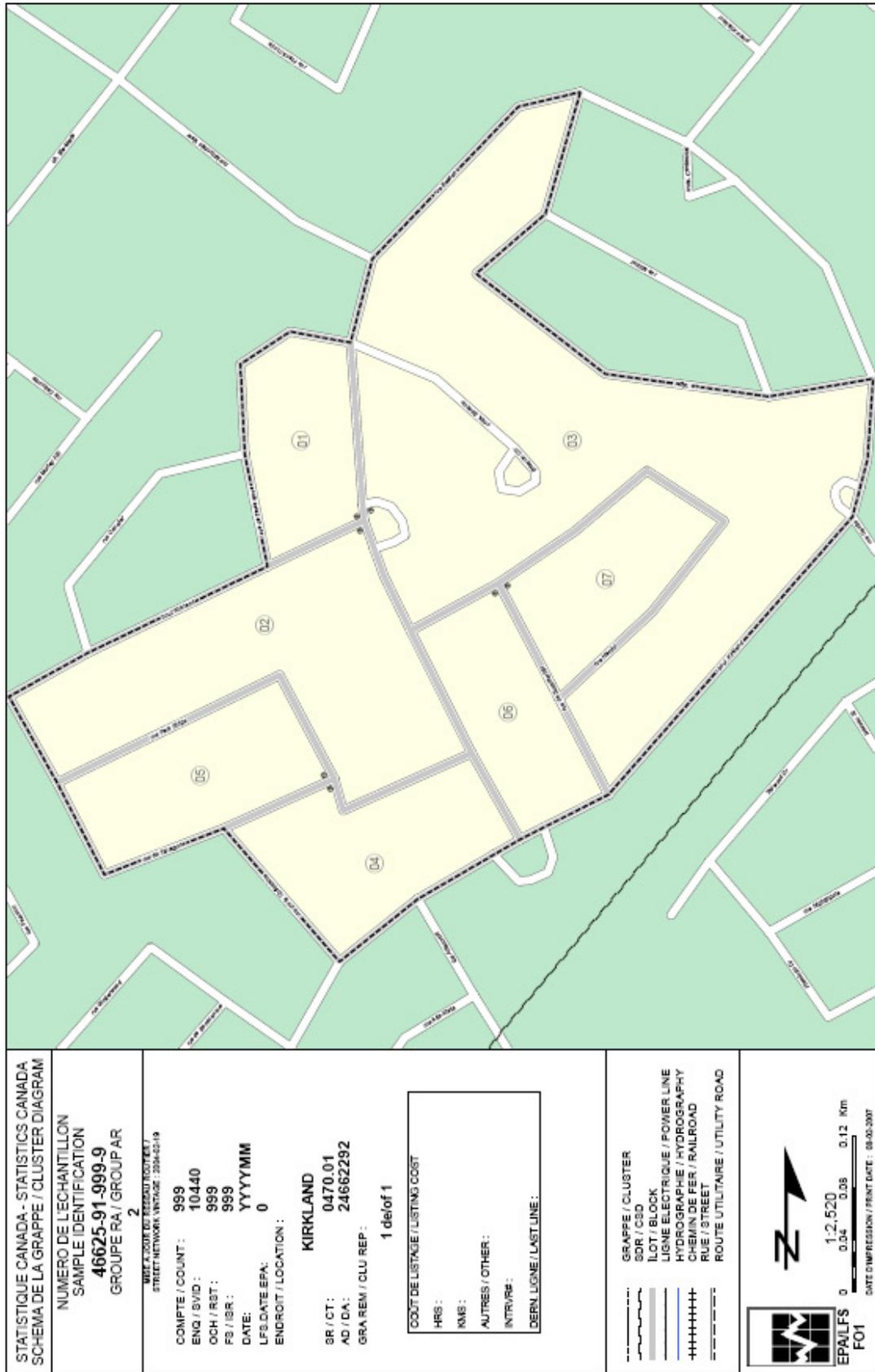
Date d'impression (15-09-2006)



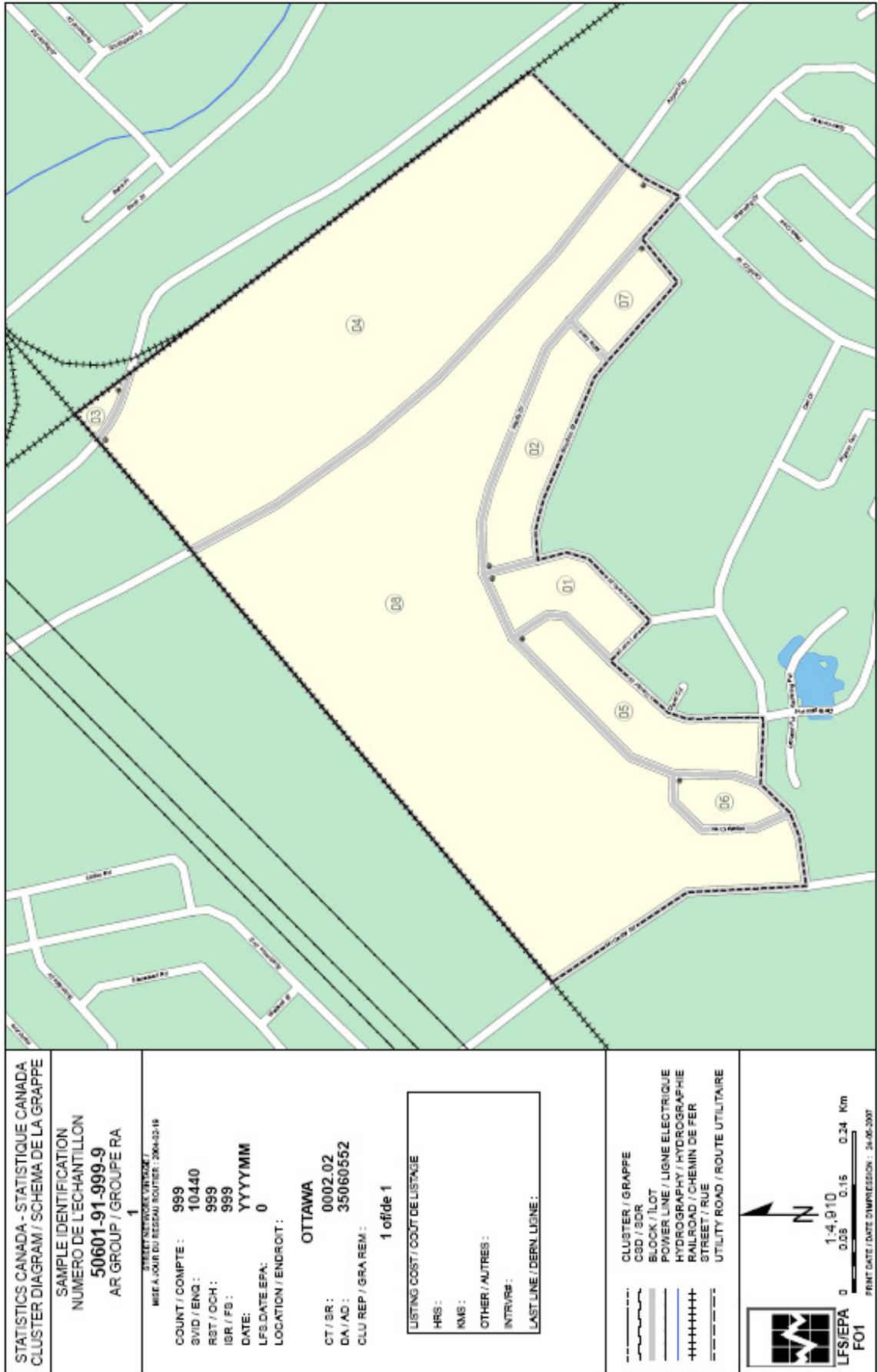
Exemple 1a : Le médaillon de l'exemple 1 montre que presque tout le petit village est hors des aires de listage. Notons que l'orientation est présentement au nord. Le médaillon 1 est la carte 2 de 2 pour cette UPE. Le reste de la légende est identique à la carte principale. Les routes utilitaires ne constituent normalement pas une limite d'îlots.



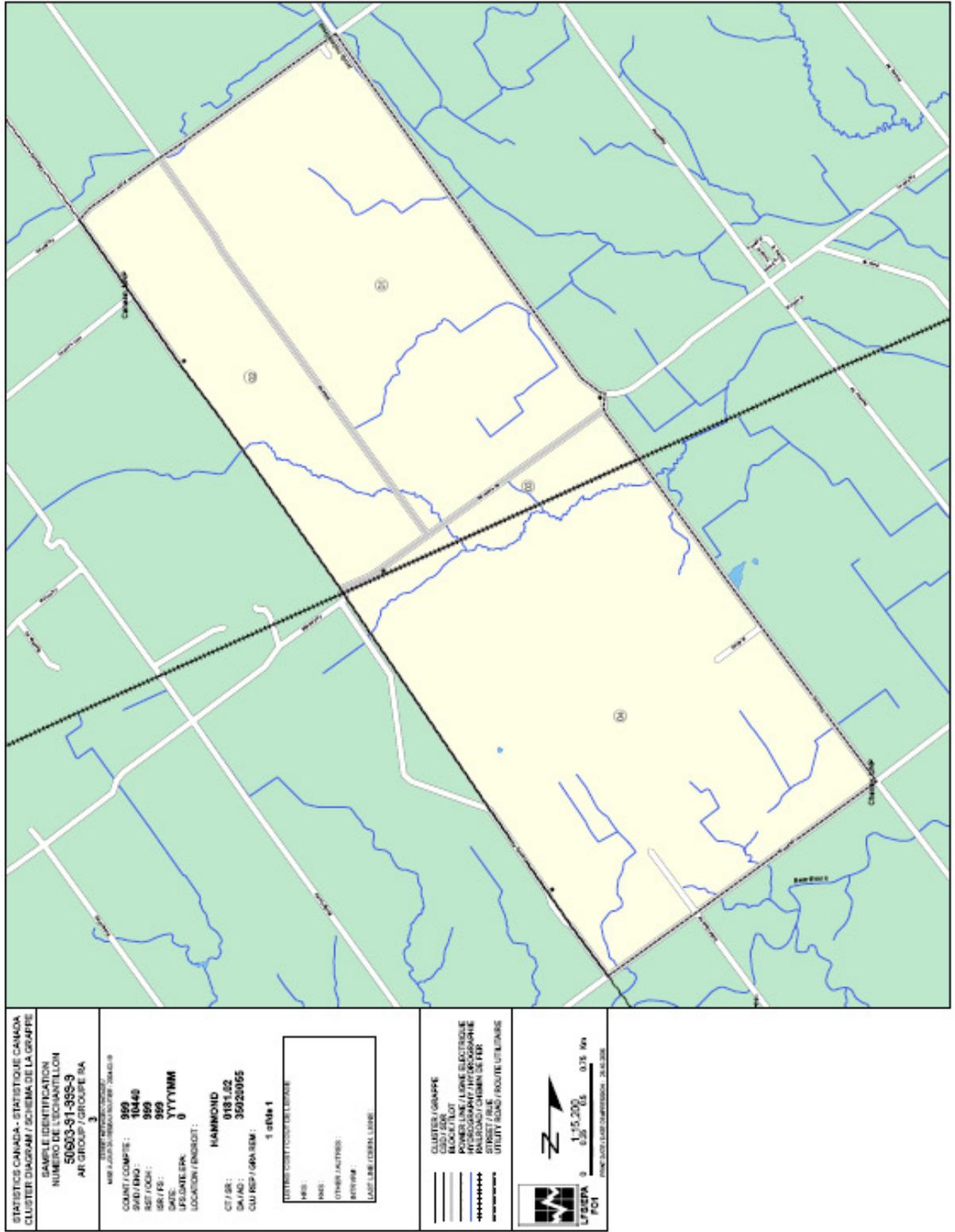
Exemple 2 : Taille de la carte 1, groupe 2 du RA, pas de médaillon, le nord est à droite. Notons que les points de départ sont regroupés pour réduire la distance à parcourir lors du listage. Certains flots ne constituent pas des flots de recensement (gros trait gris). Ces flots sont trop petits pour contenir des logements ou couvrent des zones non habitables comme l'espace délimité par l'échangeur ou pour indiquer de nouvelles routes depuis l'exercice de la formation d'îlots.



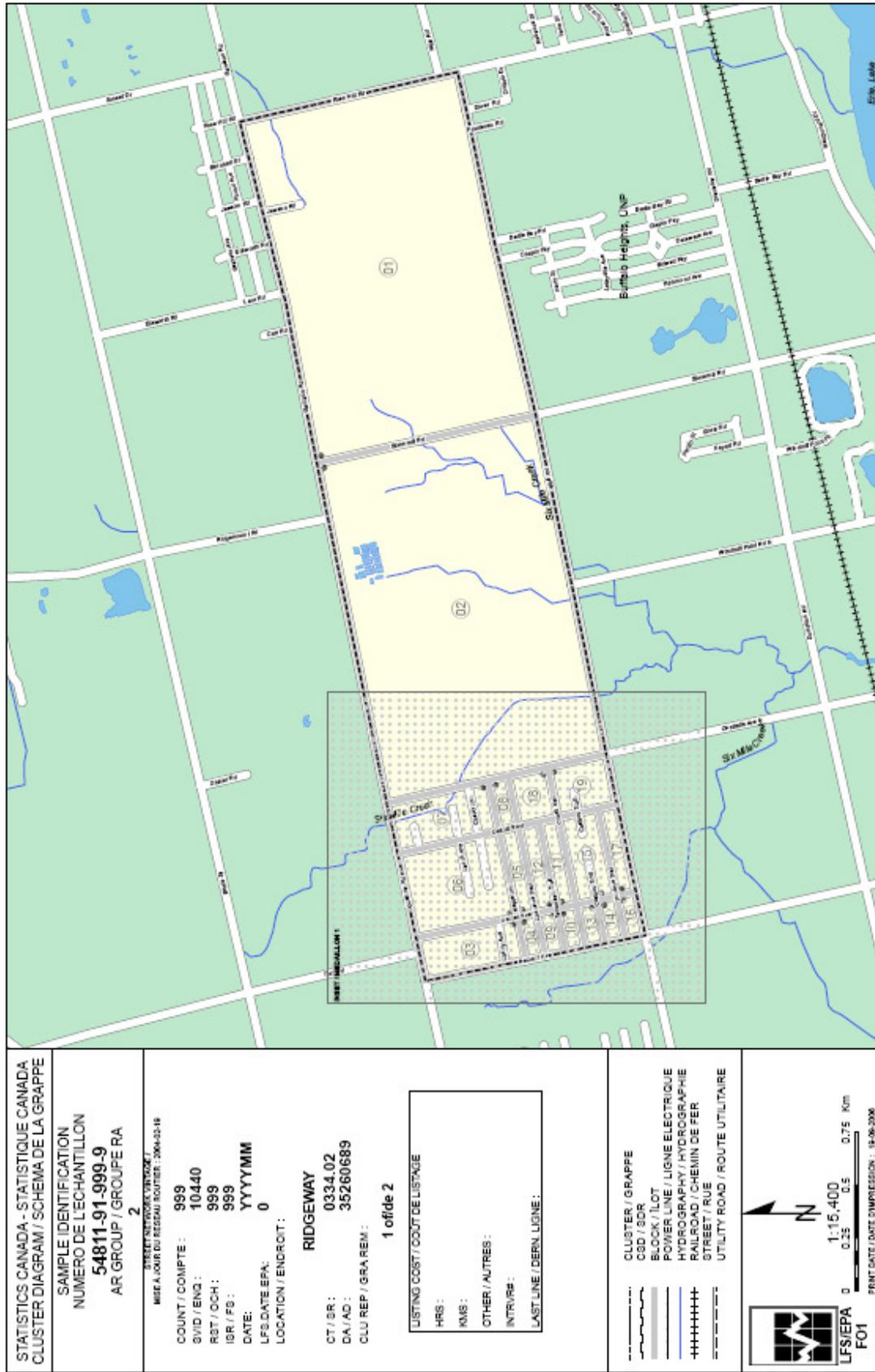
Exemple 3 : La carte de taille 1, groupe 1 du RA, pas de médaillon, le nord est en haut. Il n'est pas évident à partir de l'emplacement ou du réseau routier que cette UPE pourrait ou devrait être un groupe 1 du RA. Les chemins de fer peuvent être utilisés comme limites, mais générer de courts segments de route à inclure dans le listage.



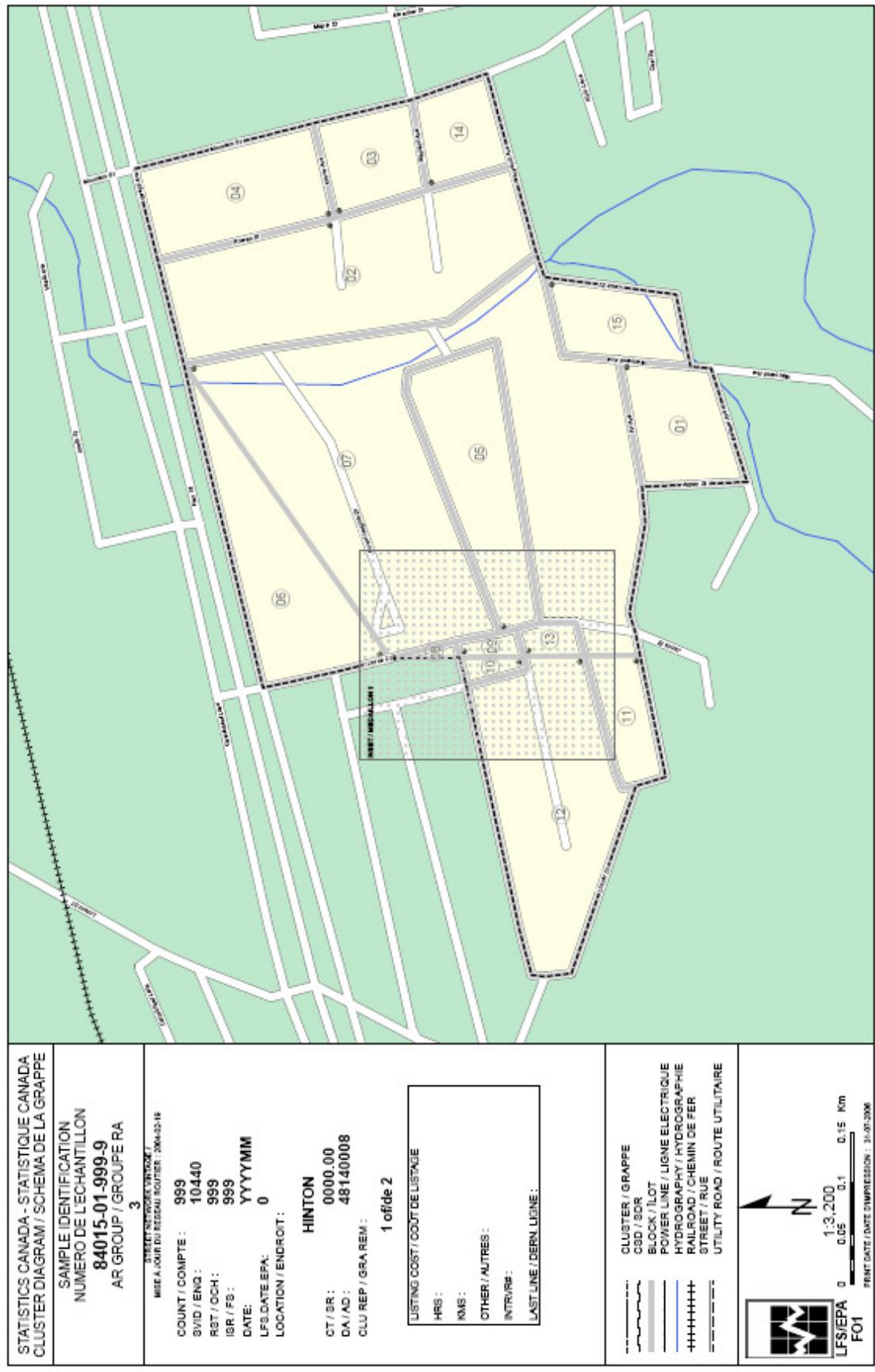
Exemple 4 : Taille de la carte 2 (17°x22"), groupe 3 du RA, pas de médaillon, le nord est à droite. Les cartes de plus grandes tailles sont plus difficiles à reproduire dans ces pages. Par conséquent, aucun exemple de carte de taille 3 (17°x34") n'est inclus dans ce document. La zone rurale n'est probablement pas comprise dans le Registre des adresses même si elle se trouve dans la RMR d'Ottawa. Les noms de localités non constituées permettent de repérer les zones à lister.



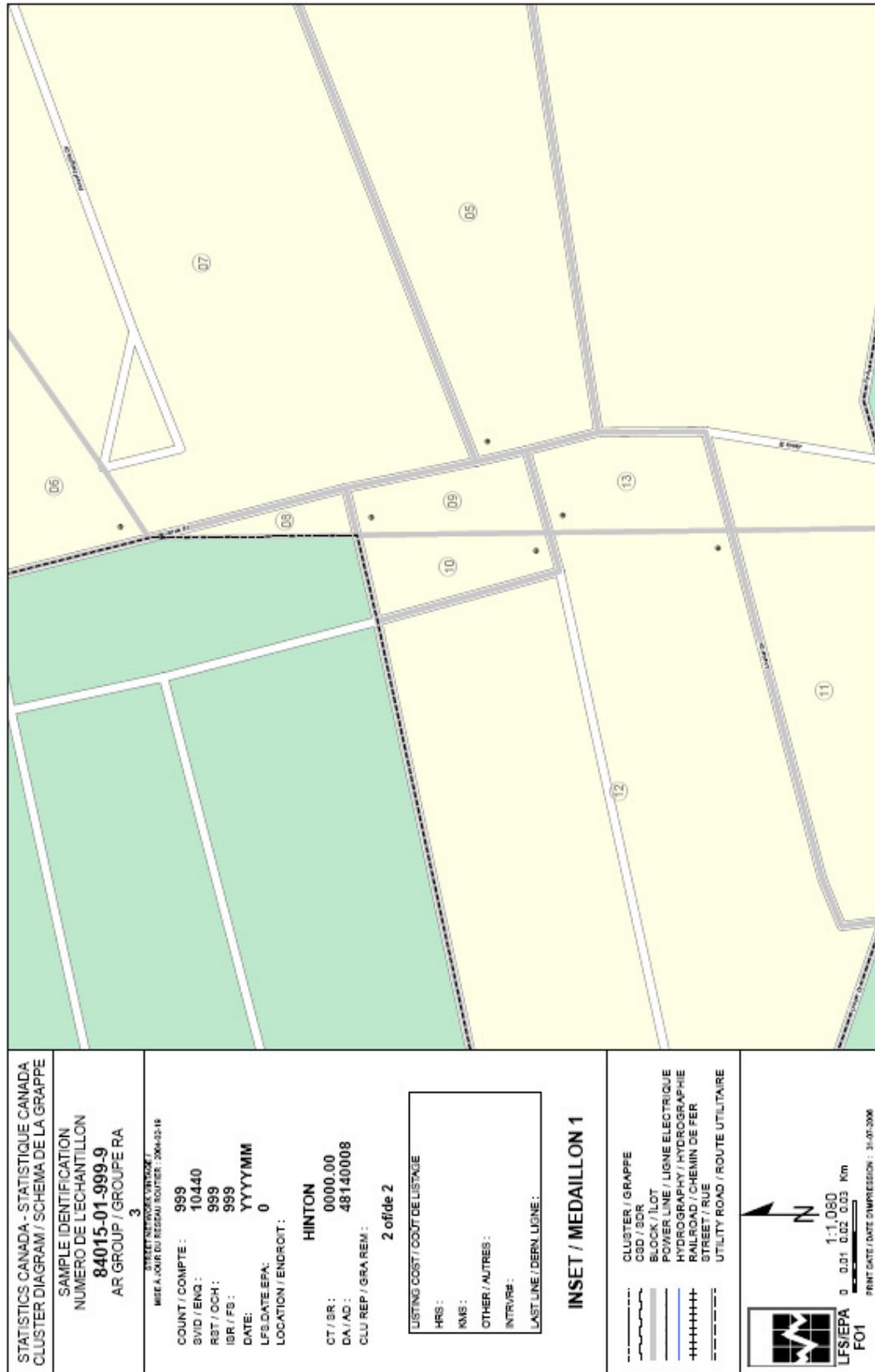
Exemple 5 : Carte de taille 1, groupe 2 du RA, un médaillon, le nord est en haut. Plusieurs petits îlots exigent un médaillon. Les tailles des cartes et les médaillons sont créés automatiquement, mais certaines corrections peuvent être nécessaires. Une carte de taille 2 n'ayant pas de médaillon pourrait suffire pour cette carte. Le médaillon de carte n'est pas inclus pour cet exemple.



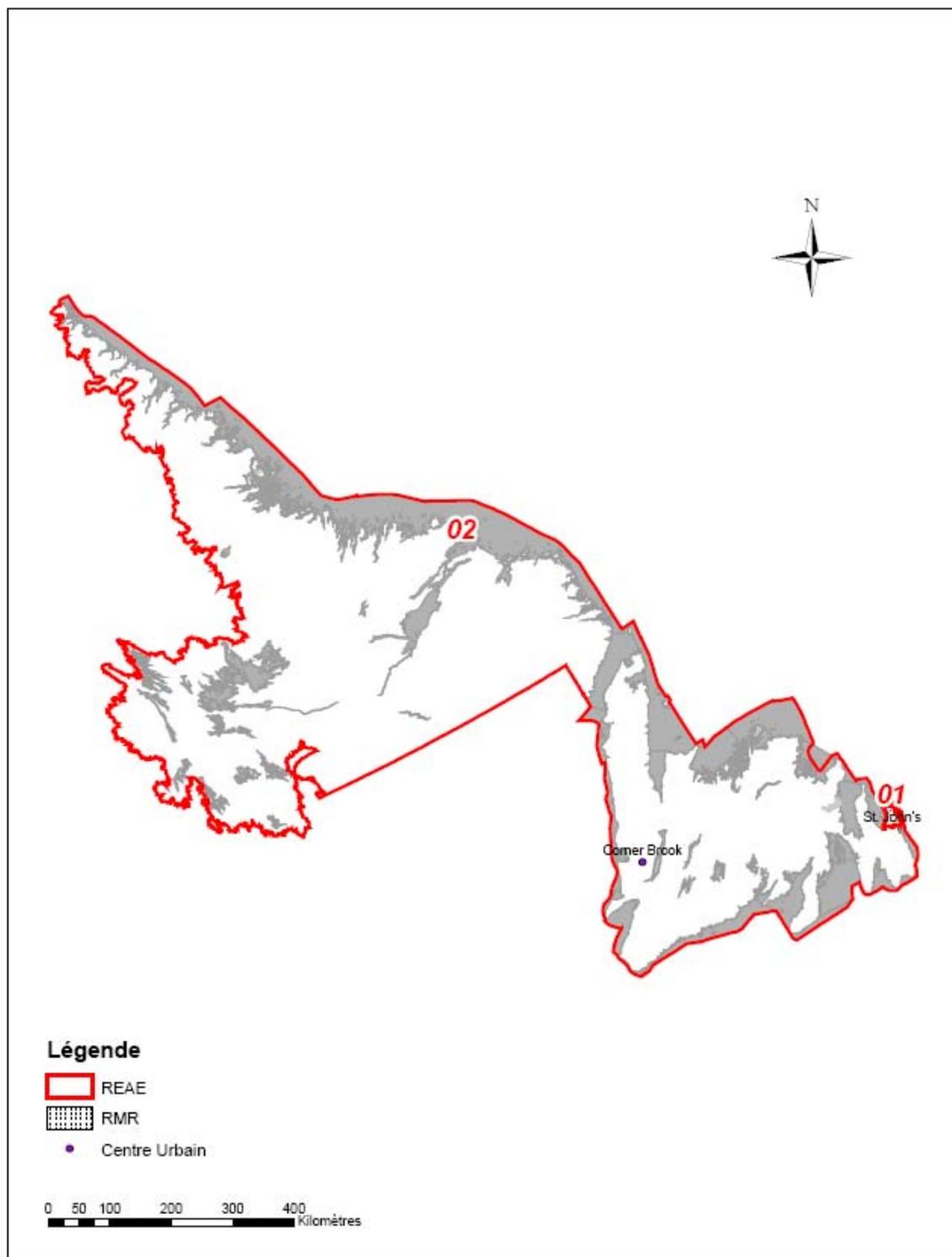
Exemple 6 : Taille de la carte 1, groupe 3 du RA, Un médaillon, le nord est en haut. Une ligne imaginaire engendre plusieurs îlots et exige un médaillon. D'autres routes ont été ajoutées trop tard pour être comprises dans la formation des îlots. La ligne imaginaire peut être une ancienne limite d'une SDR, la limite CÉF et même une ancienne limite de SD. De nouveaux îlots sont formés à l'occasion seulement. Pour les exemples plus complexes, les limites d'îlots, la numérotation et les points de départ sont retirés afin de faciliter la tâche du listage principalement pour le groupe 3 du RA.



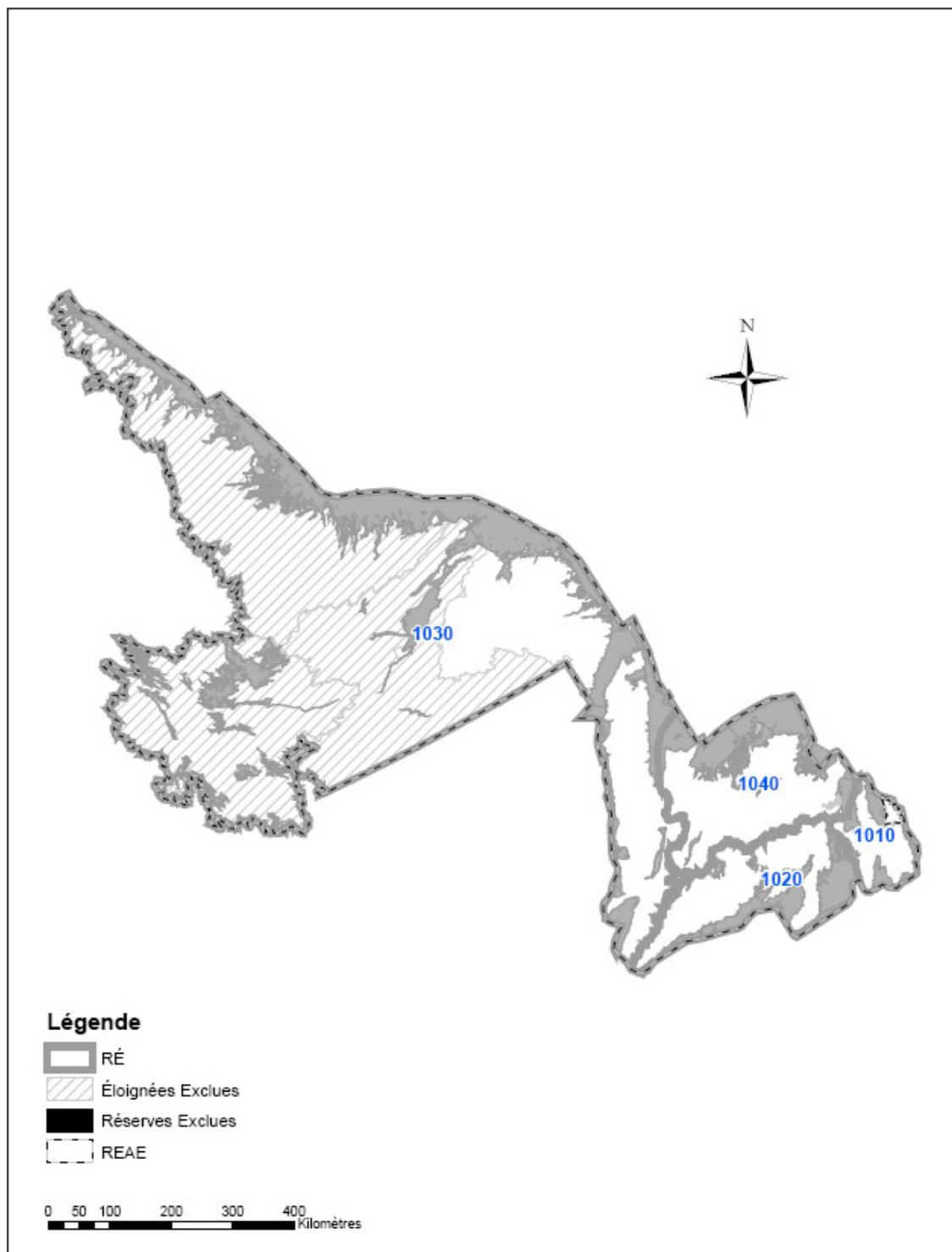
Exemple 6a : Médaille 1 pour l'exemple 6. Redéfinir les limites des îlots selon le réseau routier le plus récent exige de séquençer de nouveau les îlots et les logements du RA (s'il y a lieu) notamment par une fusion des îlots 09 et 10. L'îlot 08 disparaîtra en tant qu'entité séparée et redéfinira ainsi les limites de la grappe et le listage requis. Dans ce cas, les adresses de l'îlot 08 passeront probablement aux zones voisines à l'ouest. Tous les logements dans l'échantillon soulèvent la question du chevauchement dans le cas où l'UPE voisine est aussi sélectionnée.



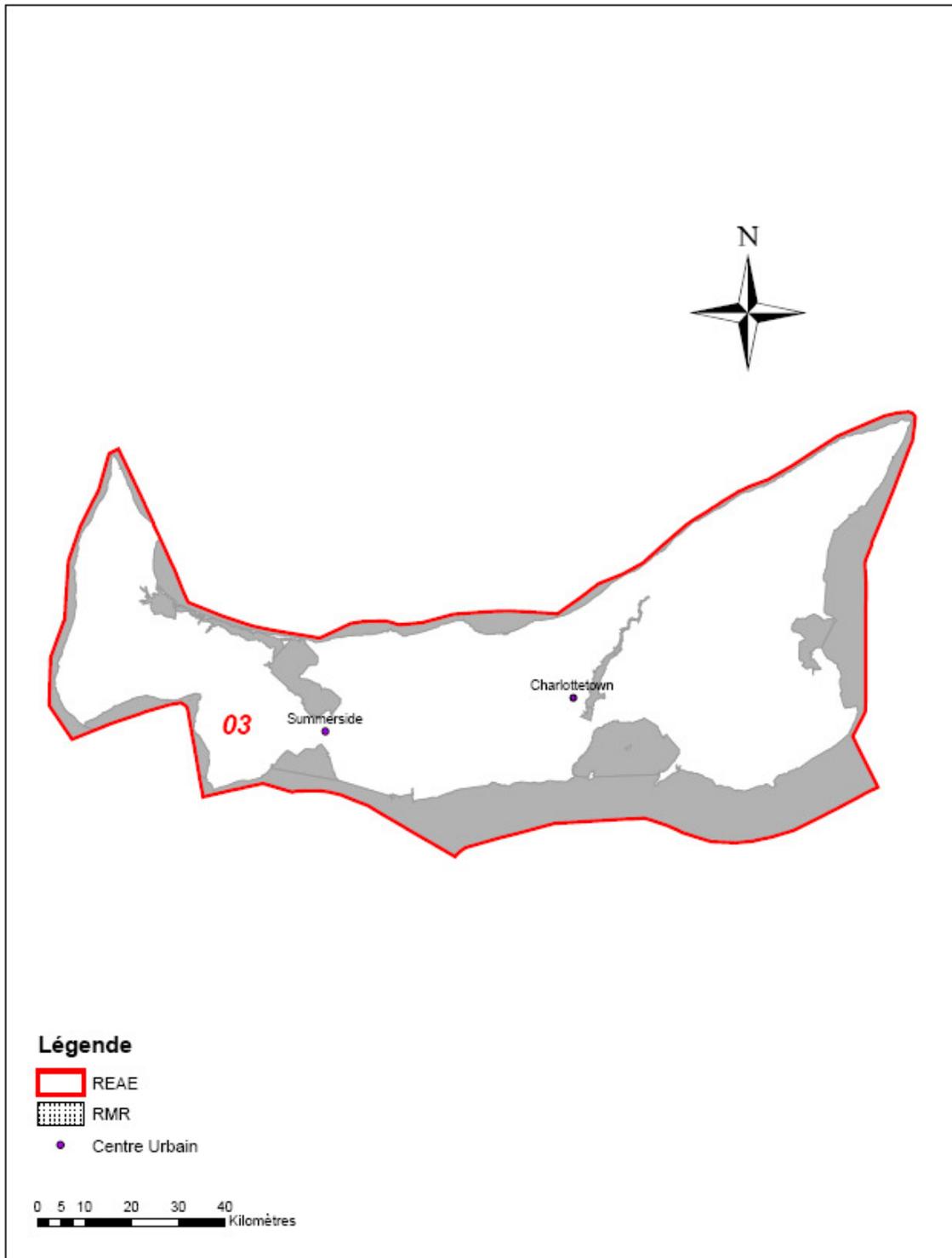
Annexe E Cartes provinciales  
Carte 1 Terre-Neuve et Labrador  
REAE et RMR



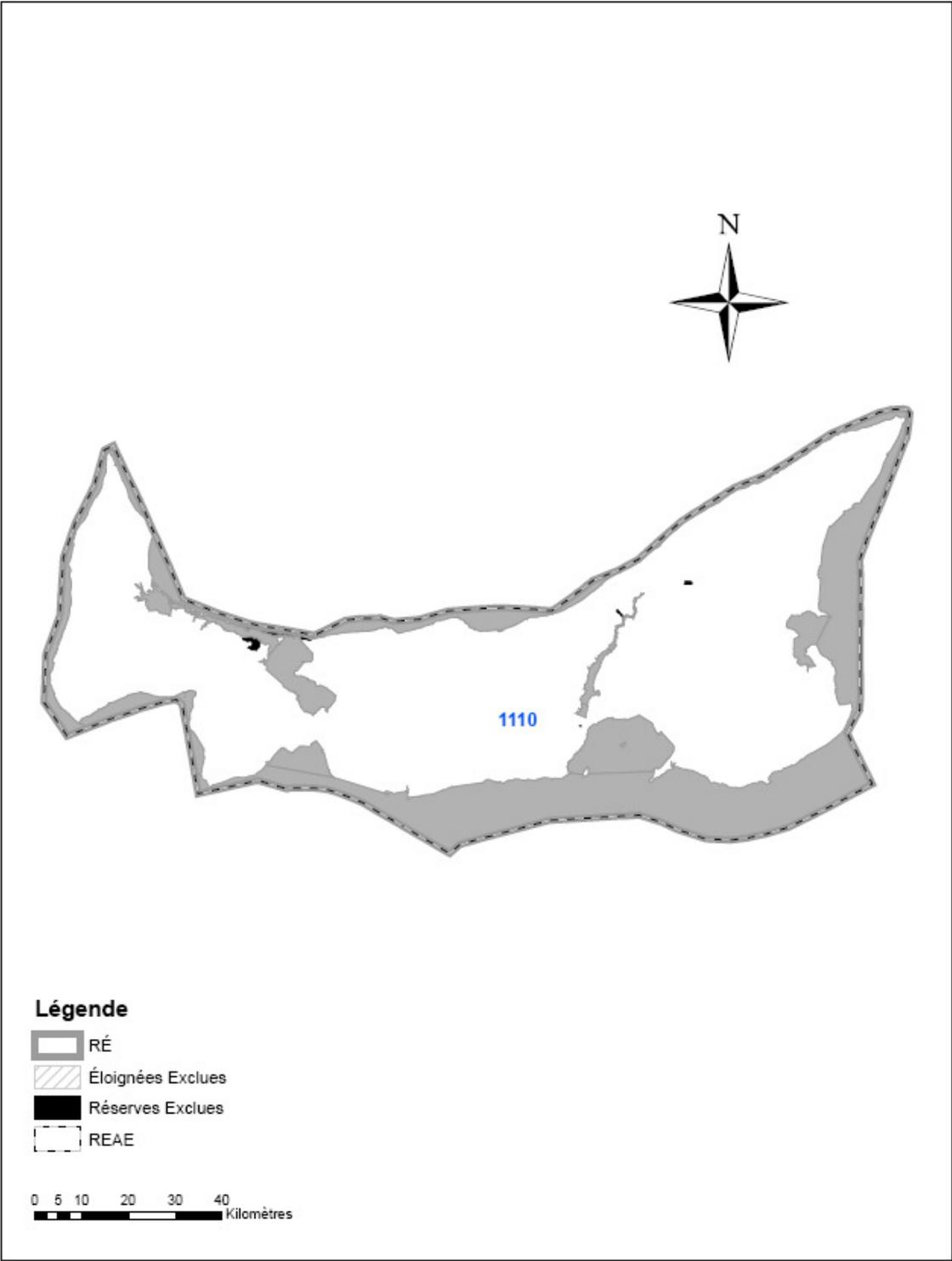
## Carte 2 Terre-Neuve et Labrador Régions économiques



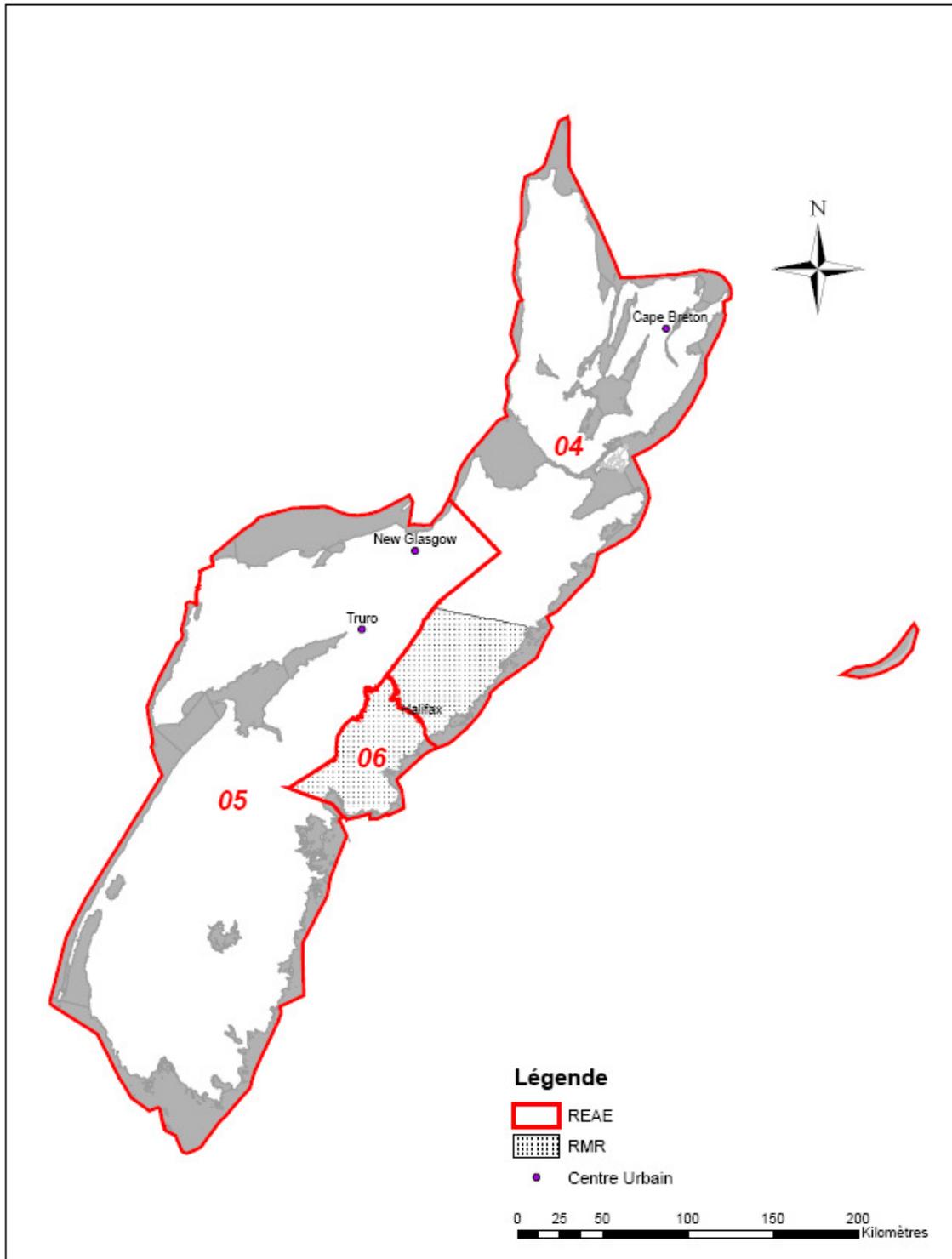
### Carte 3 Île-du-Prince-Édouard REAE et RMR



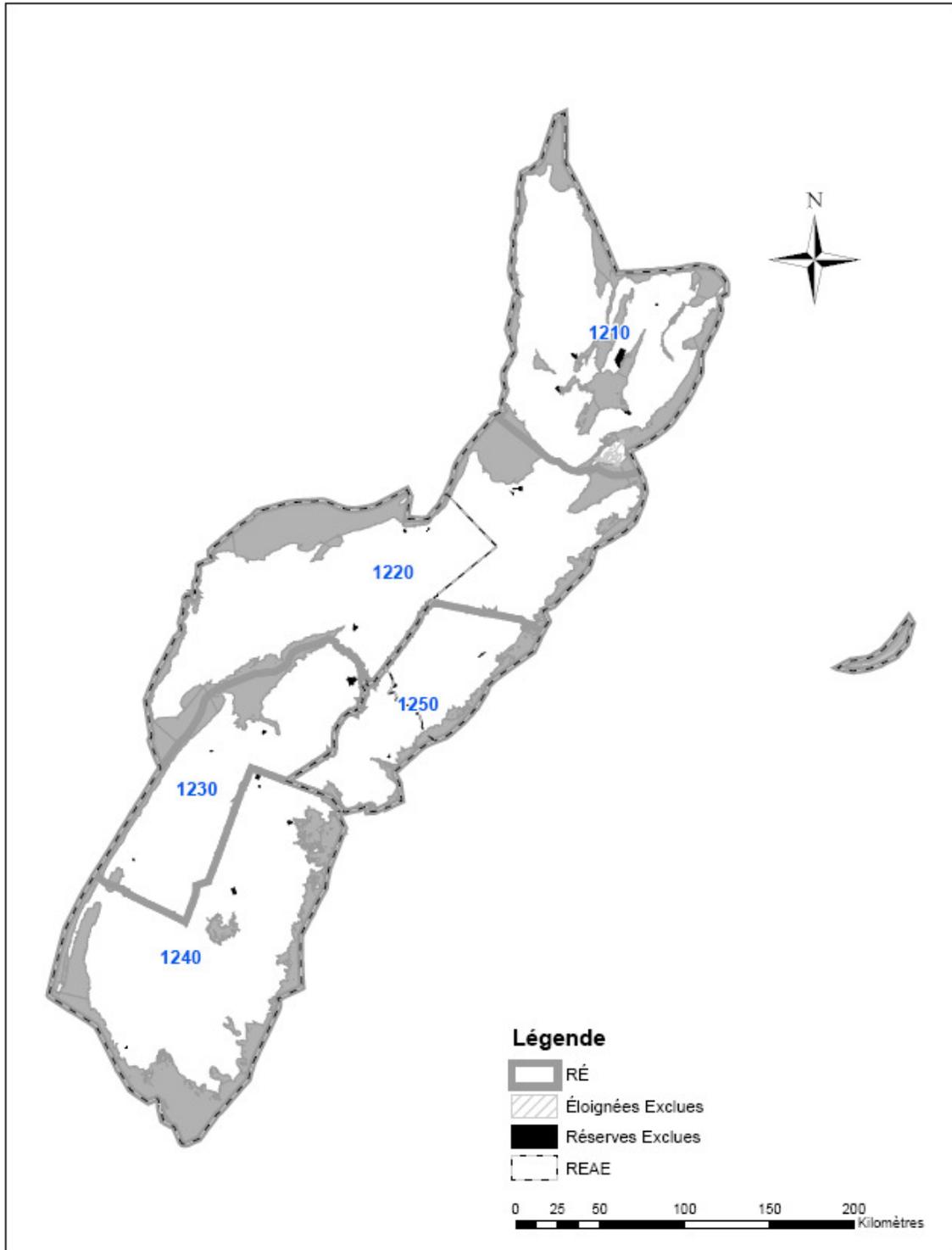
# Carte 4 Île-du-Prince-Édouard Régions économiques



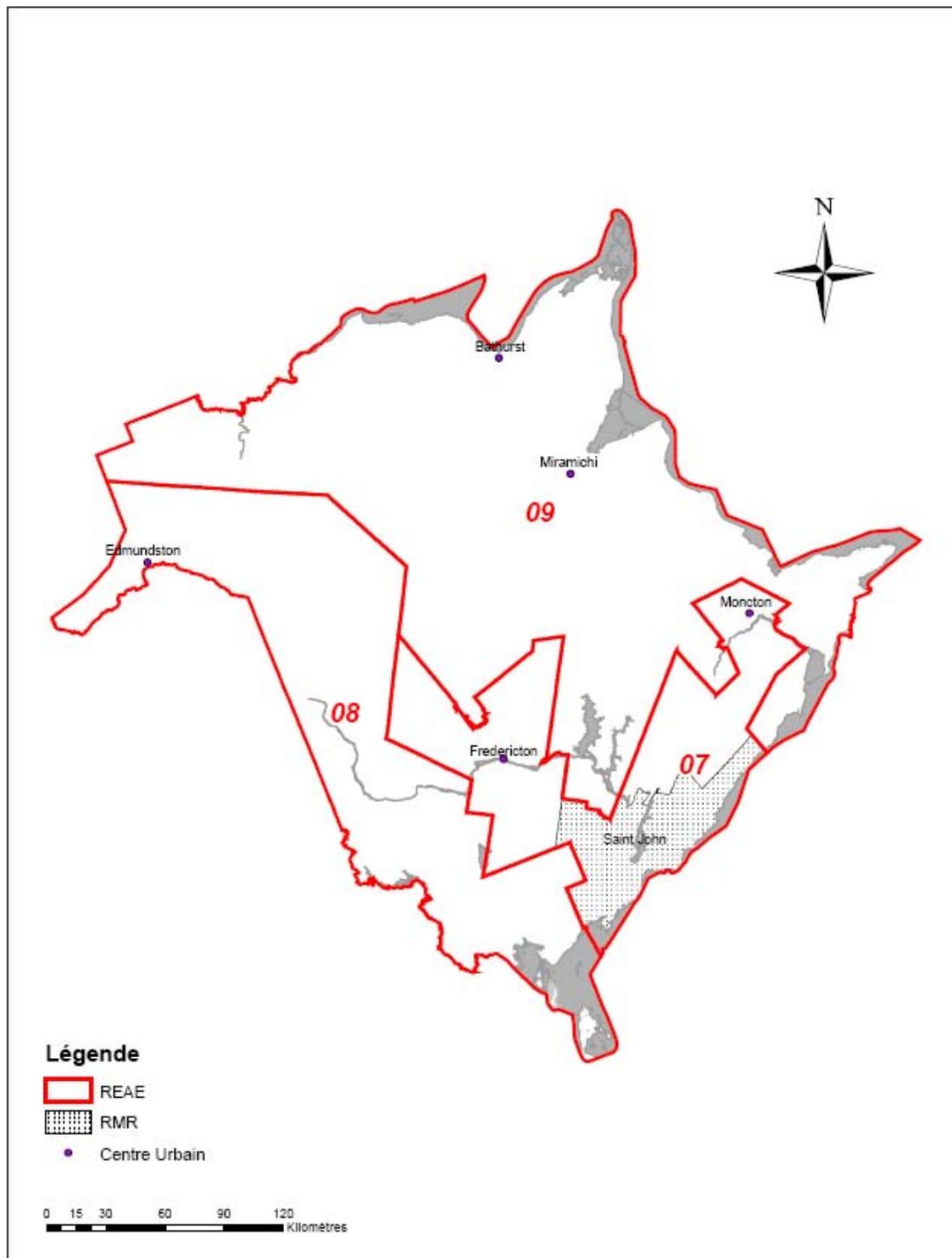
# Carte 5 Nouvelle-Écosse REAE et RMR



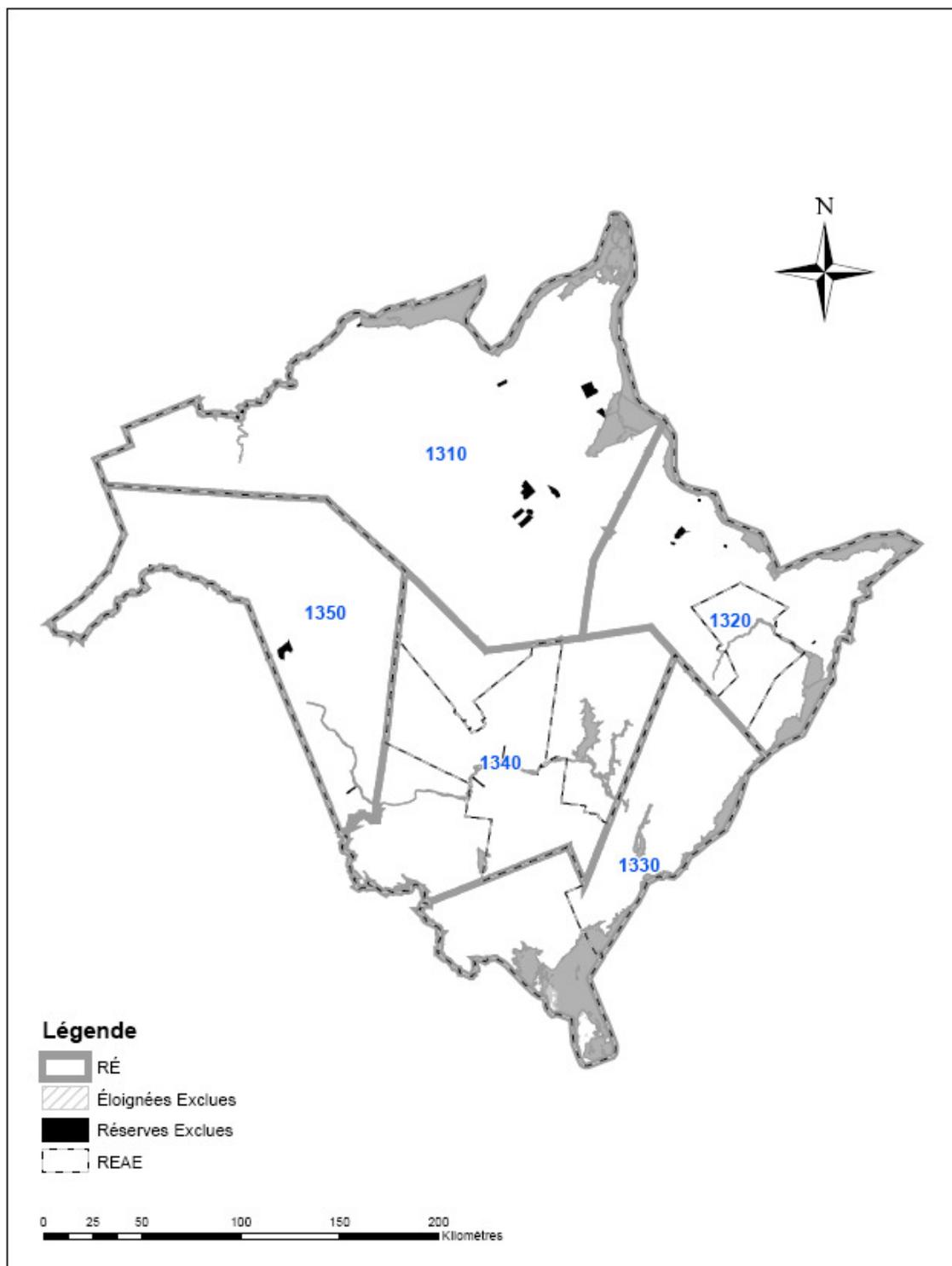
# Carte 6 Nouvelle-Écosse Régions économiques



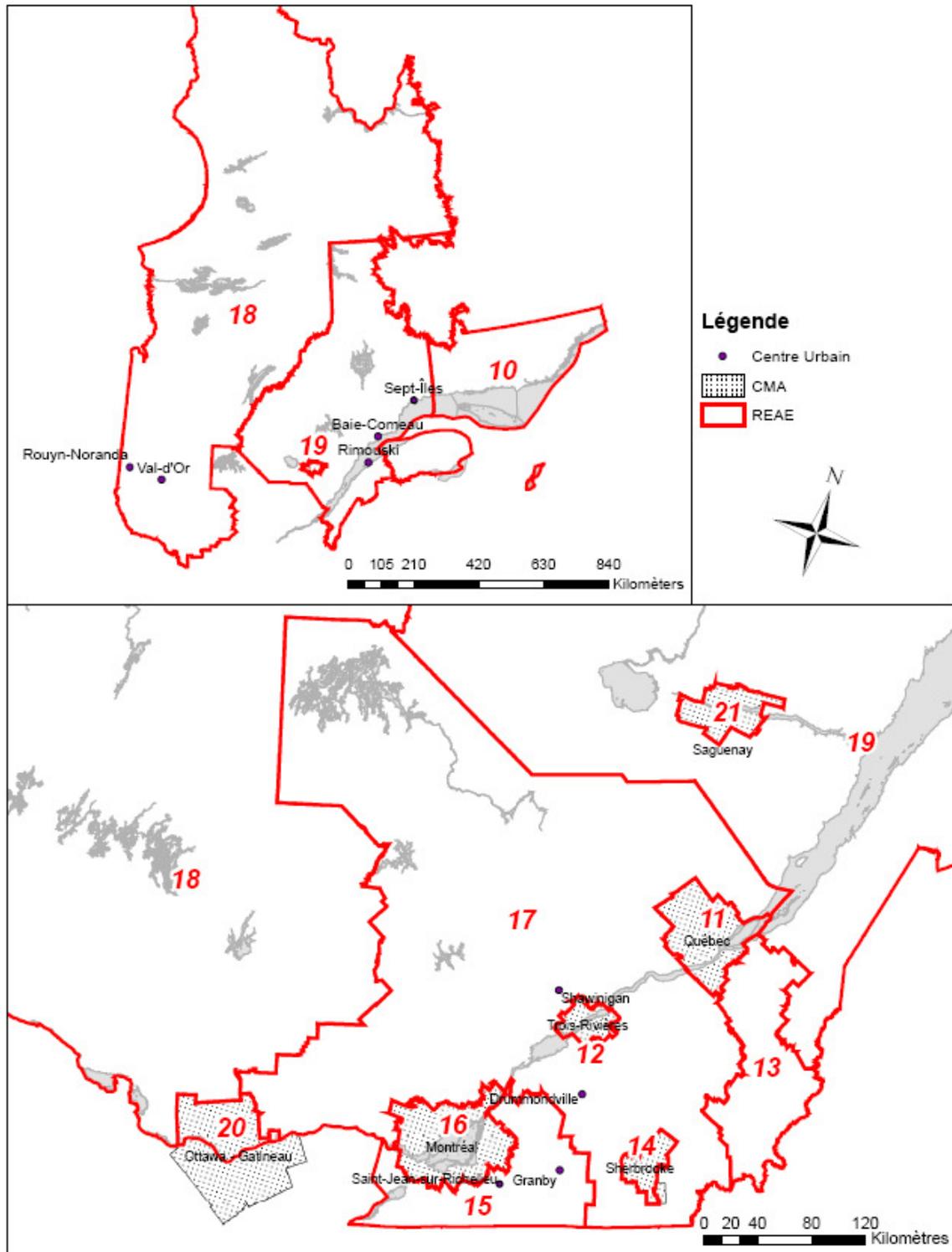
# Carte 7 Nouveau-Brunswick REAE et RMR



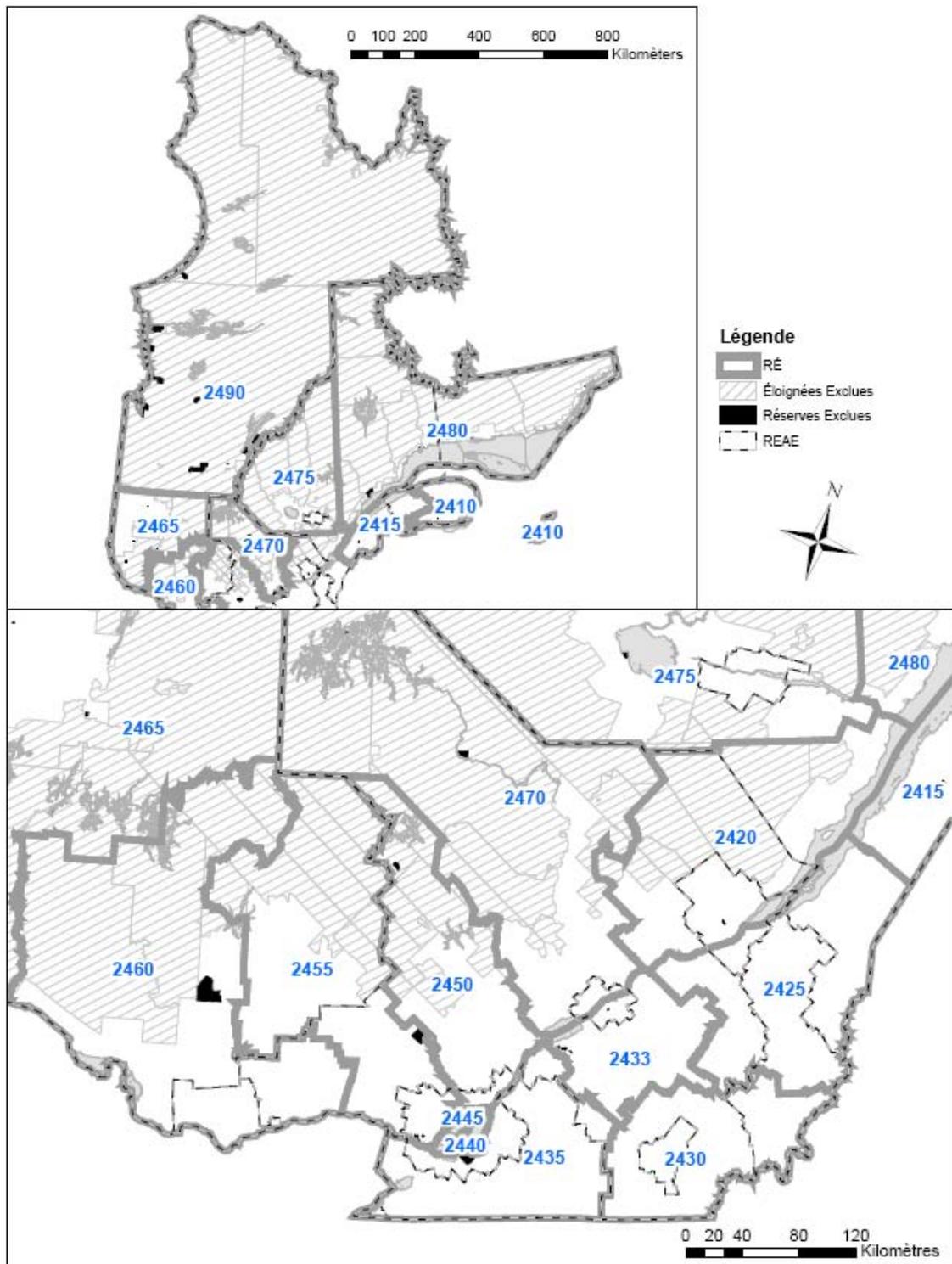
## Carte 8 Nouveau-Brunswick Régions économiques



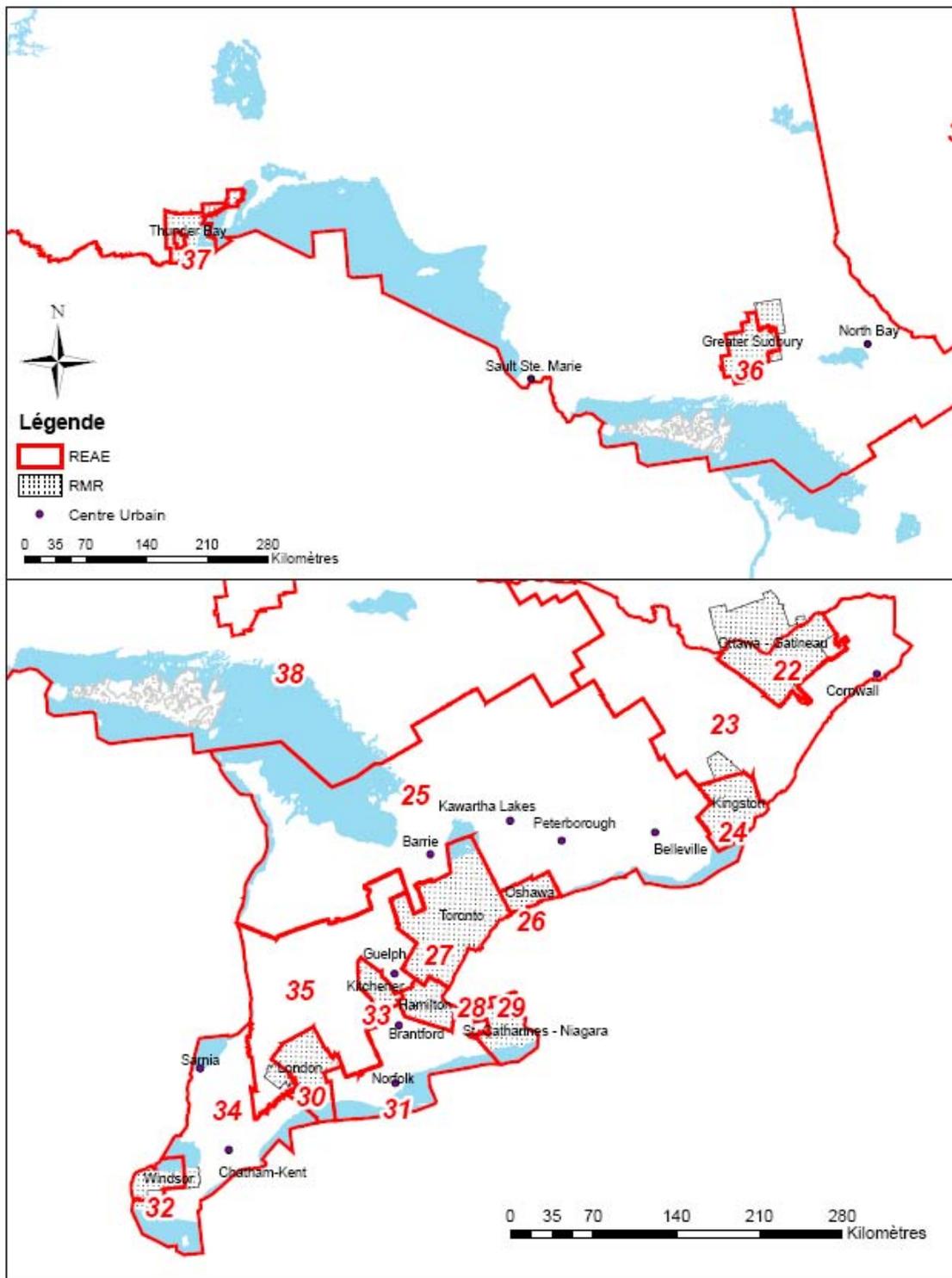
## Carte 9 Québec REAE et RMR



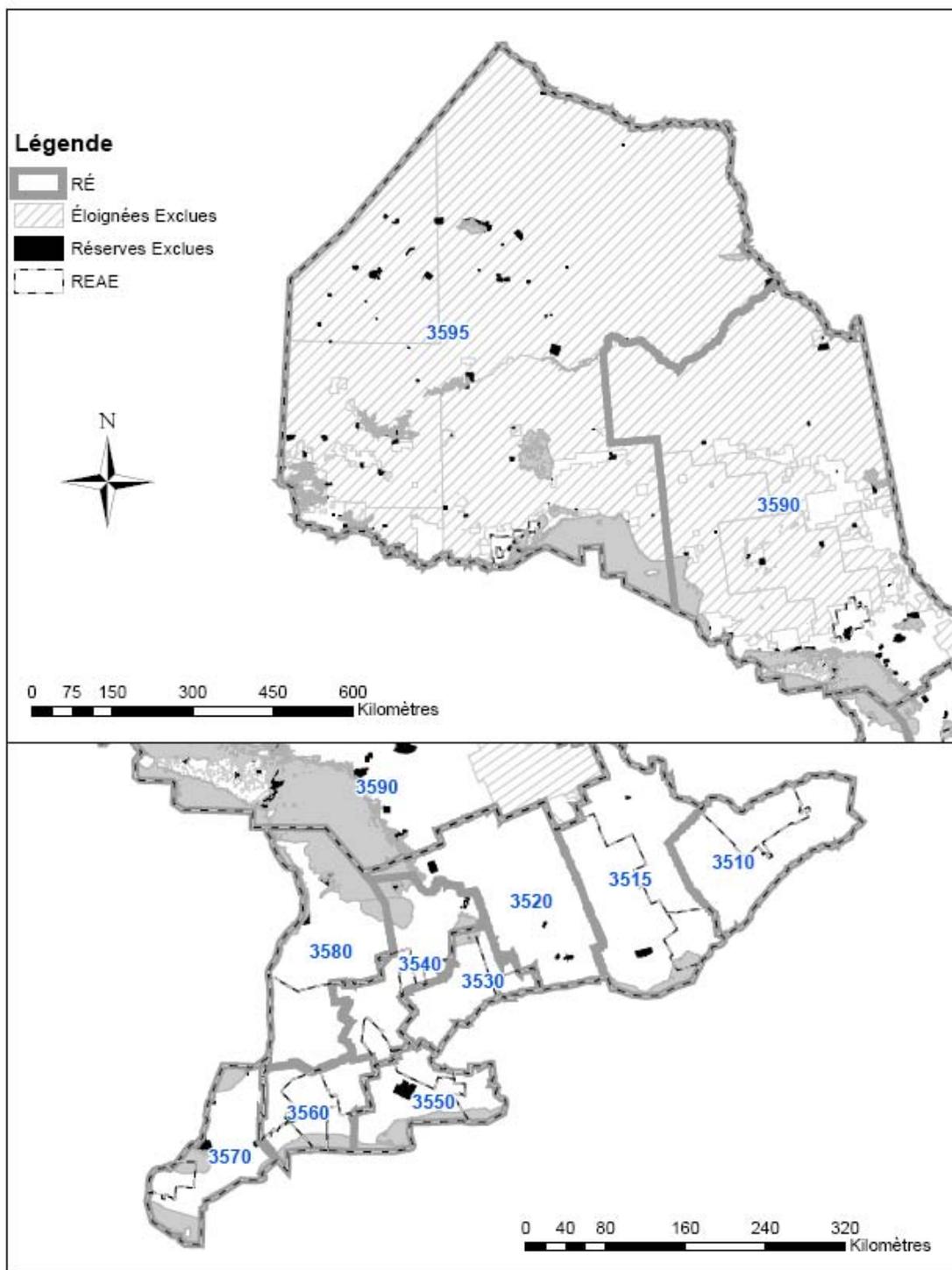
## Carte 10 Québec Régions économiques



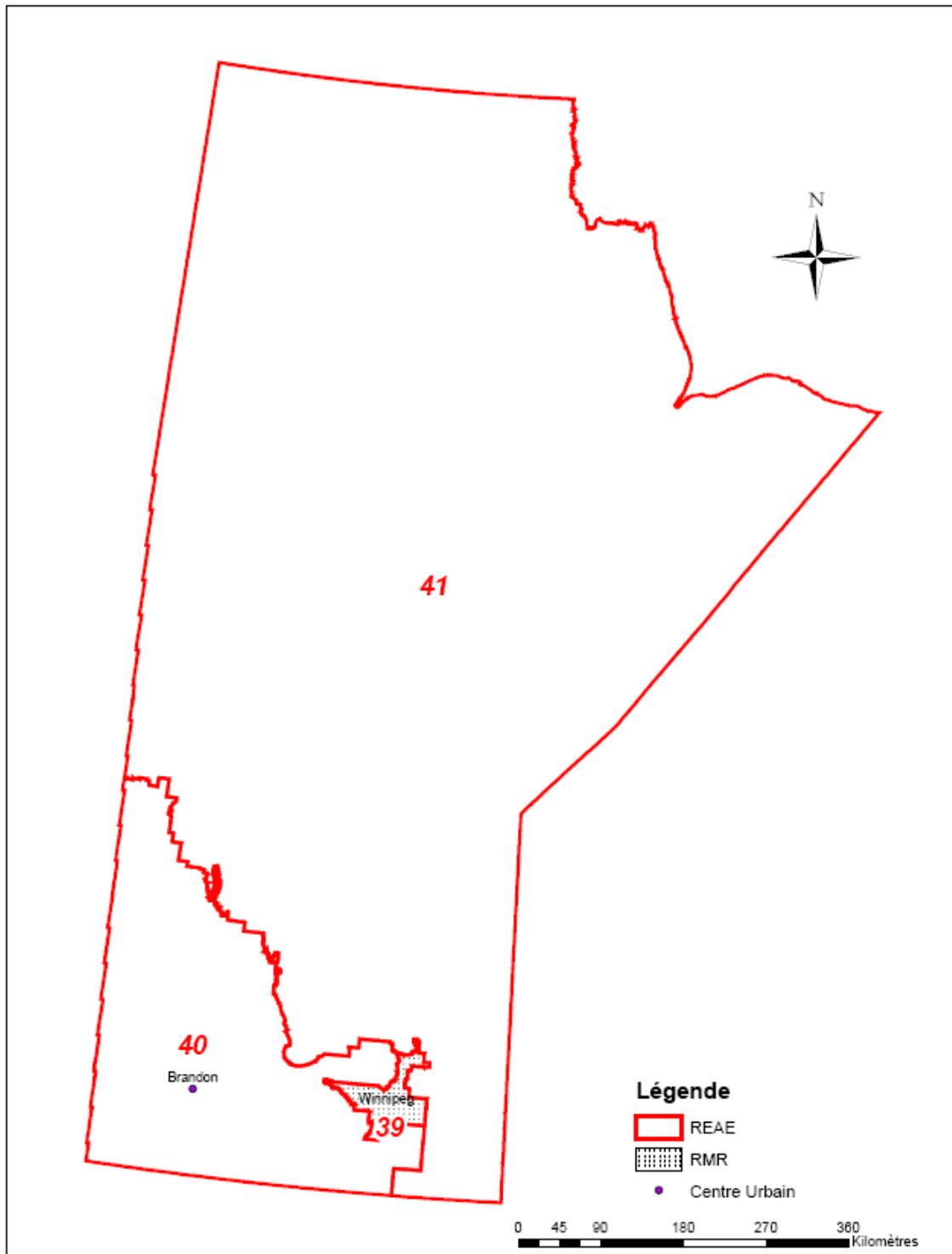
# Carte 11 Ontario REAE et RMR



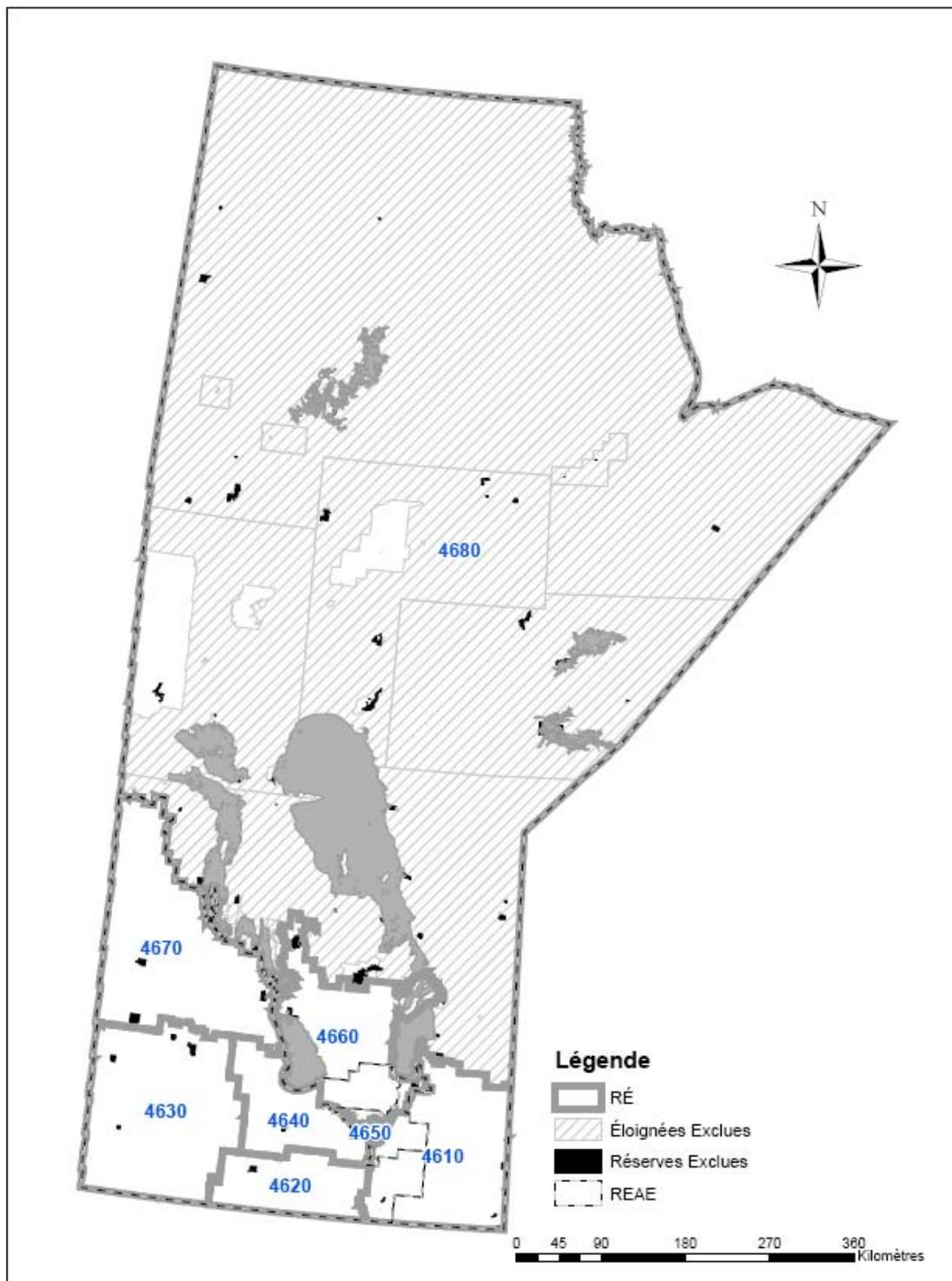
## Carte 12 Ontario Régions économiques



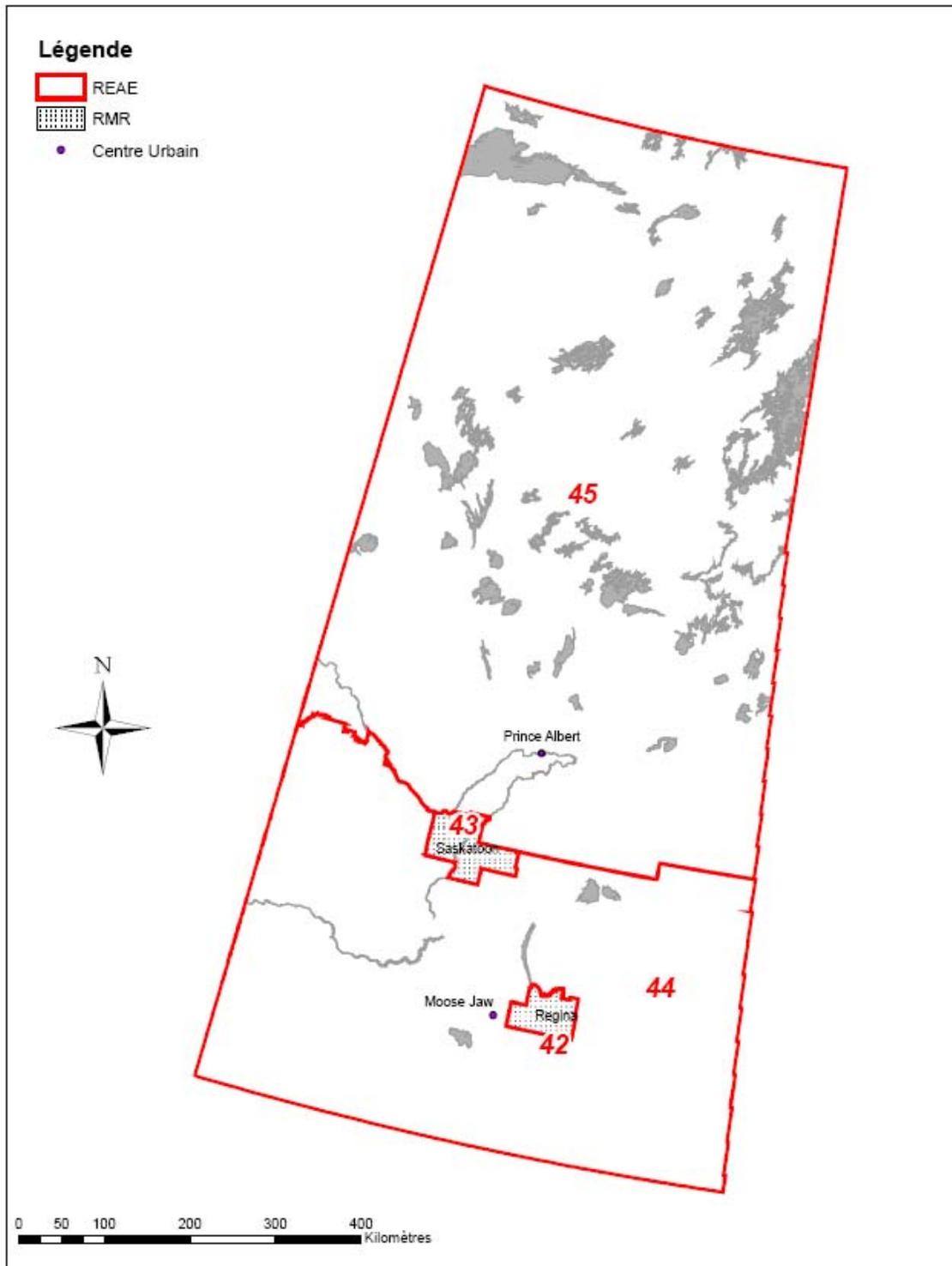
### Carte 13 Manitoba REAE et RMR



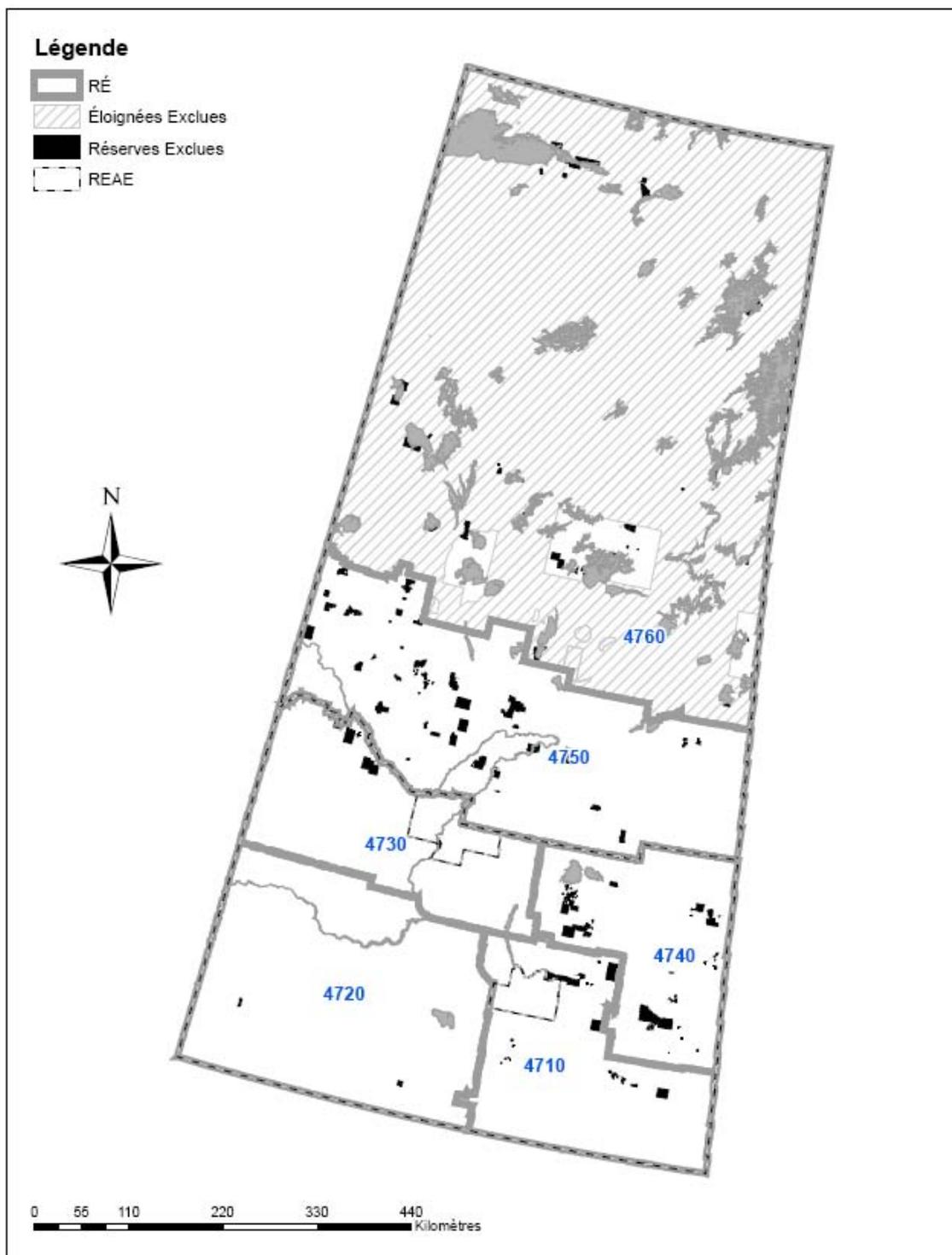
# Carte 14 Manitoba Régions économiques



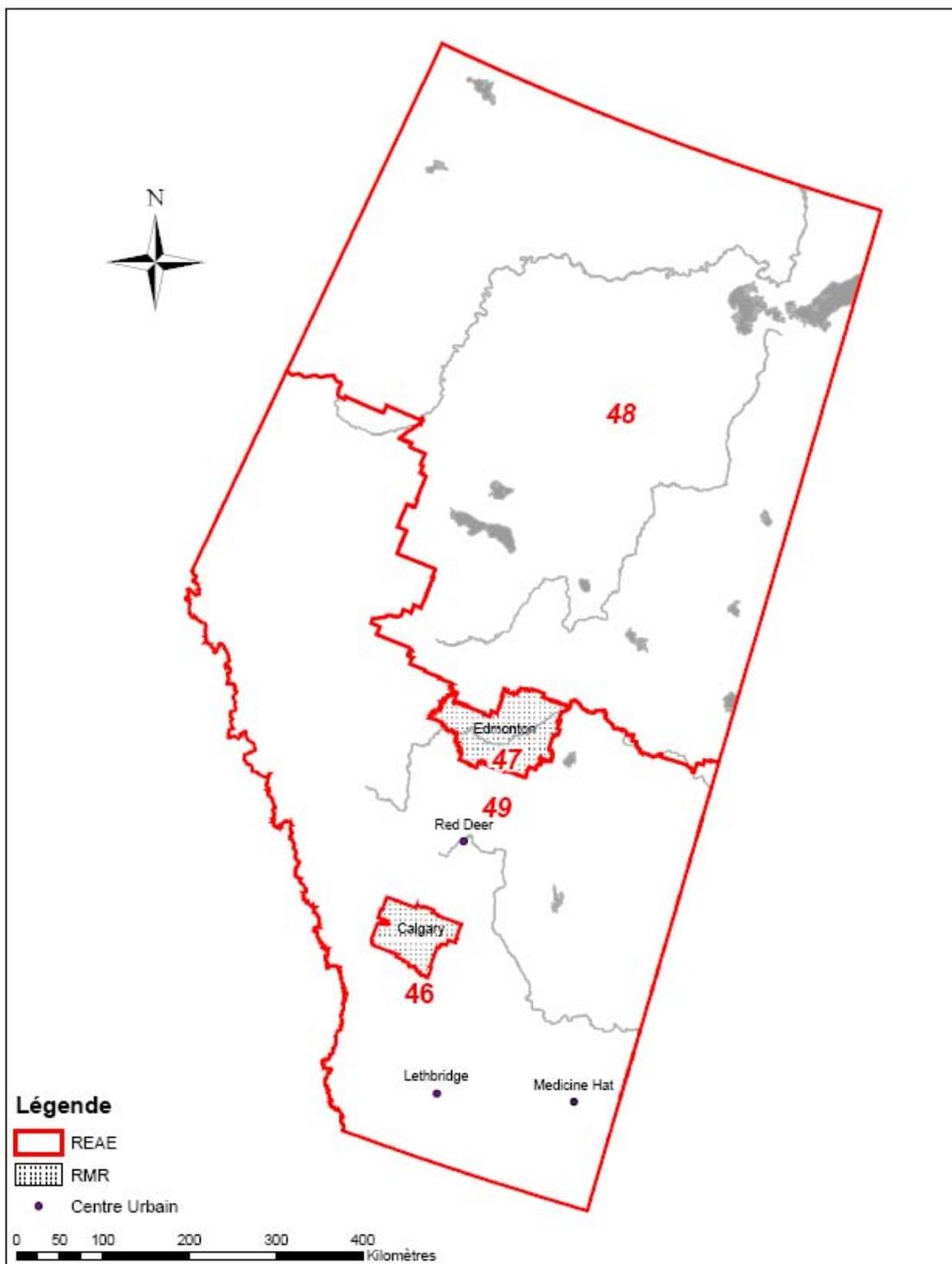
# Carte 15 Saskatchewan REAE et RMR



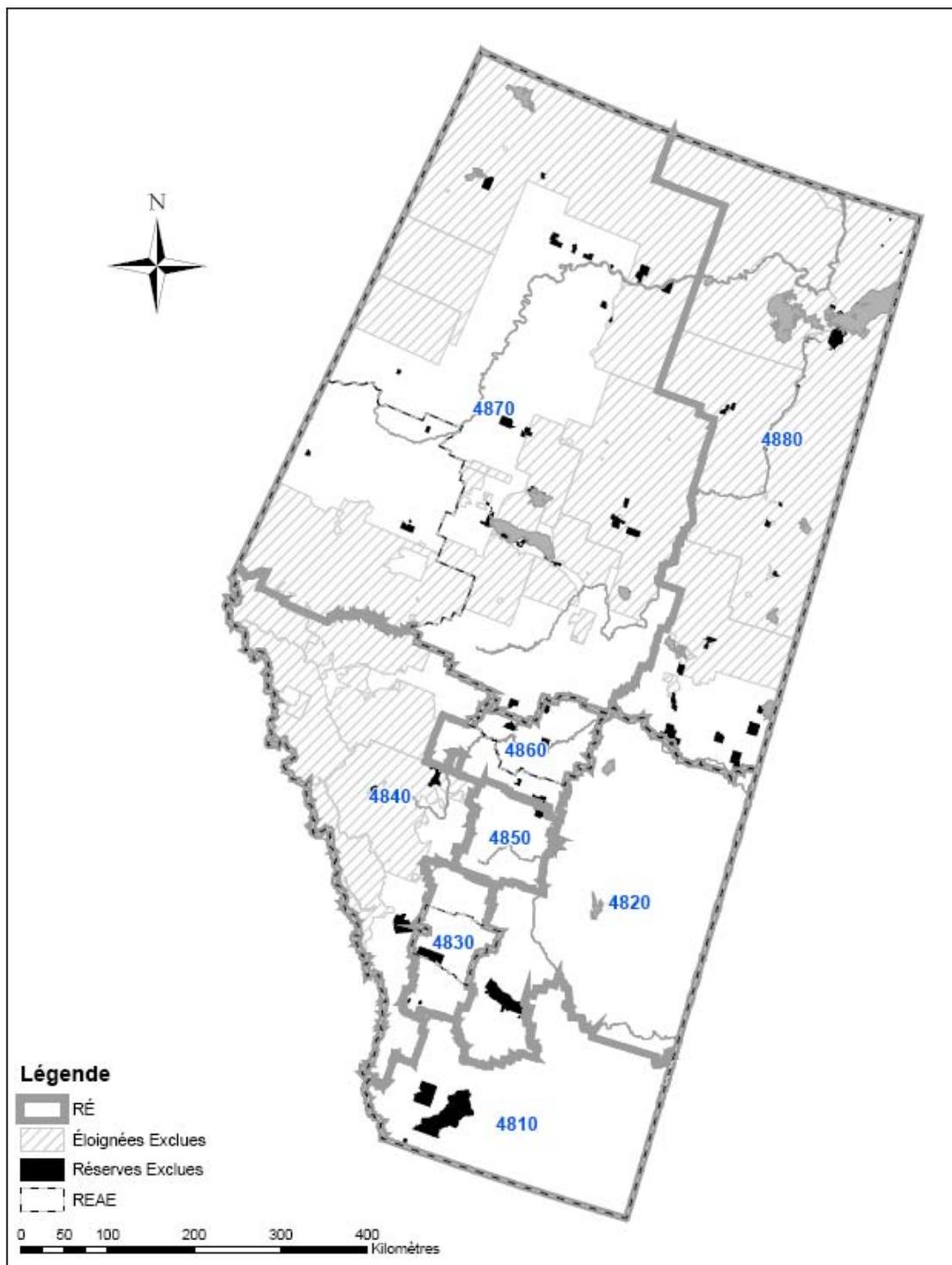
# Carte 16 Saskatchewan Régions économiques



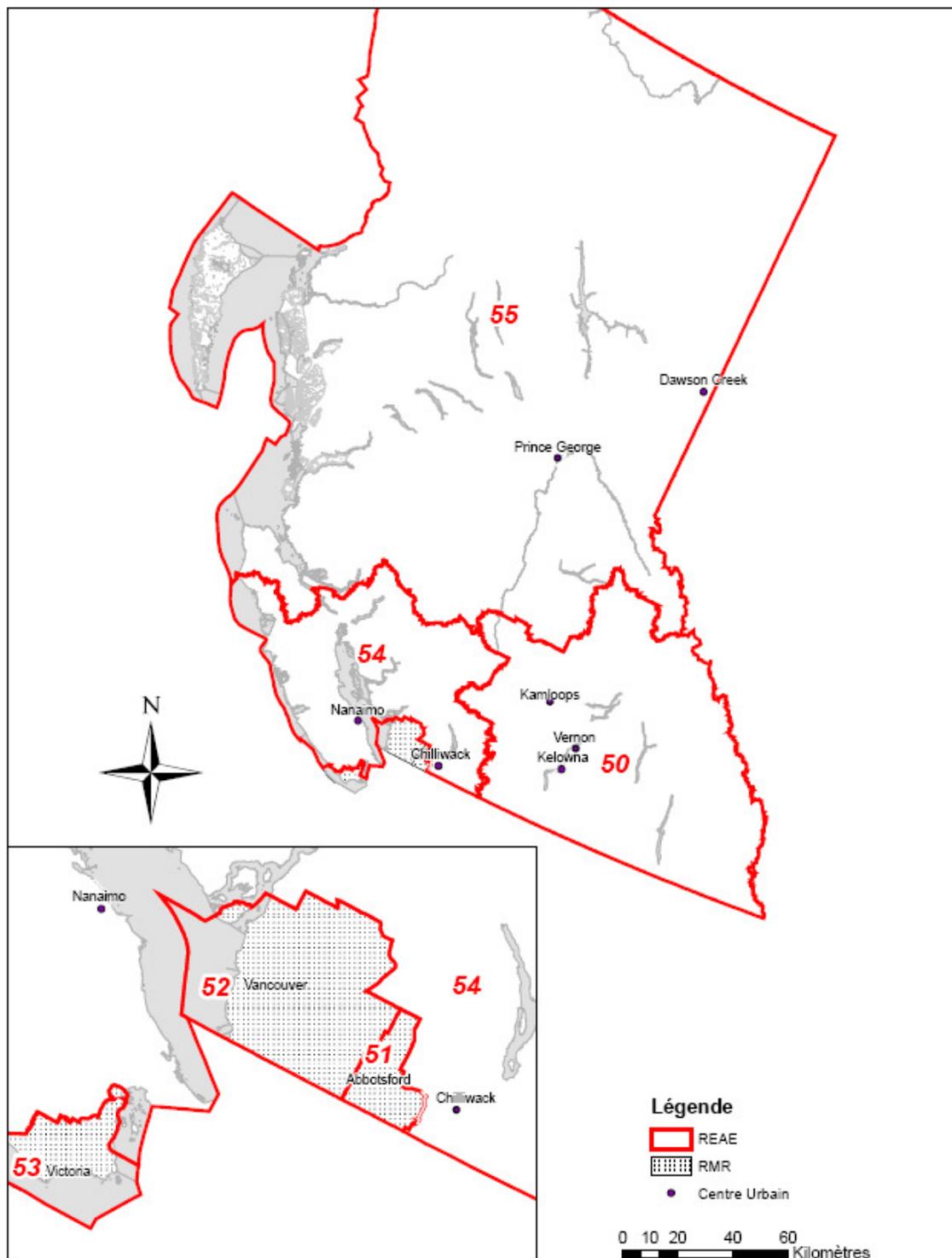
# Carte 17 Alberta REAE et RMR



# Carte 18 Alberta Régions économiques



# Carte 19 Colombie-Britannique REAE et RMR



## Carte 20 Colombie-Britannique Régions économiques

