



HAL
open science

Sondages aréolaires pour l'estimation de données démographiques en milieu urbain- Essai de définition d'une méthode de collecte intégrant l'information satellitaire, application aux villes des pays en développement

Olivier Barbary

► **To cite this version:**

Olivier Barbary. Sondages aréolaires pour l'estimation de données démographiques en milieu urbain- Essai de définition d'une méthode de collecte intégrant l'information satellitaire, application aux villes des pays en développement. Méthodes et statistiques. EHESS - Paris, 1988. Français. NNT: . tel-02479137

HAL Id: tel-02479137

<https://hal.science/tel-02479137v1>

Submitted on 19 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES

*SONDAGES AREOLAIRES POUR L'ESTIMATION DE DONNEES
DEMOGRAPHIQUES EN MILIEU URBAIN*

*ESSAI DE DEFINITION D'UNE METHODE DE COLLECTE
INTEGRANT L'INFORMATION SATELLITAIRE, APPLICATION
AUX VILLES DES PAYS EN DEVELOPPEMENT*

Thèse de doctorat

Soutenue le

par

Olivier BARBARY

Direction de thèse :
Marc BARBUT, EHESS
Jean Paul GILG, EHESS

Membres du jury :

REMERCIEMENTS

Cette thèse a fait l'objet d'un double encadrement : au Centre d'Analyse et de Mathématiques Sociales de l'EHESS et à l'ORSTOM, au sein de l'unité de recherche "Facteurs et mesures de la croissance urbaine" du département "Urbanisation et socio-systèmes urbains". Je voudrais donc remercier M. M. BARBUT, directeur du CAMS pour m'avoir honoré de sa confiance en acceptant de diriger cette recherche et M. J.P. GILG qui m'a également aidé de ses conseils. A l'ORSTOM, mes remerciements vont à M. E. LEBRIS, pour m'avoir accueilli dans le département et à MM. Ph. ANTOINE, J.P. DUCHEMIN qui ont favorisé la poursuite de mes travaux.

Je tiens à adresser ma profonde reconnaissance à l'ensemble des membres de l'équipe de recherche constituée à l'ORSTOM autour du projet "Télédétection et démographie" ; d'abord à F. DUREAU qui m'a communiqué son enthousiasme durant les trois ans qu'on dure cette recherche, mais aussi à tous les autres : B. LORTIC et A. MICHEL du département S.D.U., F. PELLETIER et M. SOURIS de l'unité d'Infographie du L.I.A., A. AING et I. RANNOU de l'unité de télédétection du L.I.A., qui ont chacun contribué à enrichir ce travail.

A Marseille, A. CHATELAIN, J.D. GRONOF et S. STEPAN m'ont accueilli chaleureusement au laboratoire de l'EHESS de la Vieille Charité lors de mes missions en 1985 et 1986 et m'ont fait profité de leur connaissance du milieu urbain marseillais, qu'ils en soient ici remerciés. De même à Quito, ma gratitude s'adresse à Ph. CAZAMAJOR, D. COURET, M.A. DE CASTRO, H. GODARD et J. VEGA, chercheurs du projet d'Atlas Informatisé de Quito. Leur expérience des réalités quiténiennes a grandement contribué au succès de l'application de la méthode à cette ville.

Cette recherche n'aurait pu être menée à bien sans la collaboration des superviseurs et enquêteurs qui ont participé aux trois enquêtes que l'équipe a réalisées à Quito. Je voudrais leur exprimer à tous ma reconnaissance, et plus particulièrement à MM. H. ALBAN, J. ANDRADE et J. ROJAS : leur efficacité et leur conviction sont responsables de la qualité des données sur lesquelles j'ai travaillé.

Enfin, que tous ceux qui ont contribué d'une manière ou d'une autre à la réalisation de ce travail, par leur conseils leurs encouragements ou les informations qu'ils m'ont communiquées, soient assurés de ma sincère reconnaissance. Je ne peux les citer tous ici, qu'il ne m'en tiennent pas rigueur et sachent que sans eux cette thèse n'aurait pas été écrite. Cependant je tiens à remercier tout spécialement Gina, Margarita et Katia VALENCIA, qui m'ont communiqué, durant ces deux ans passés à Quito, leur joie de vivre, leur goût et leur science de la "Salsa".

* * *

SOMMAIRE

	Pages
AVERTISSEMENT.....	5
INTRODUCTION.....	6
PREMIERE PARTIE : PROBLEMATIQUE, CONTEXTE THEORIQUE, CHOIX DES METHODES ET ETAPES DU TRAVAIL	20
1 : Le problème général : un sondage aréolaire stratifié, à deux degrés.....	21
2 : Sondages aréolaires et télédétection, état de la recherche.....	29
3 : Les questions et les outils statistiques dans les différentes phases de la recherche.....	54
DEUXIEME PARTIE : MISE AU POINT DE LA METHODE A MARSEILLE, STRATIFICATIONS ET APPROCHE DU PREMIER DEGRE DU SONDAGE.....	65
Introduction : objectifs généraux et principe de la méthode d'évaluation des plans de sondage.....	66
4 : Base de données, objectifs et plan d'expérience à Marseille.....	69
5 : Analyse typologique pour la construction d'une stratification morphologique de l'agglomération marseillaise.....	89
6 : Mise au point de la méthode d'évaluation des plans de sondage, première expérience, formulaire et programme informatique.....	131
7 : Evaluation de plans de sondage stratifiés pour l'estimation de caractères démographiques (résultats des test menés grâce au programme VARECH sur la base de données constituée à Marseille).....	167
8 : Apport du tirage aréolaire systématique.....	201
Conclusion.....	209

	Pages
TROISIEME PARTIE : DEFINITION ET MISE EN ŒUVRE D'UN PLAN DE SONDAGE AREOLAIRE SUR IMAGE SPOT POUR UNE ENQUETE DEMOGRAPHIQUE A QUITO.....	214
Introduction	215
9 : Objectifs généraux, plan d'expérience, réalisation des enquêtes préliminaires.....	216
10 : Enquête préliminaire "bâti / population", exploitation pour la mise au point du premier degré du sondage : échantillonnage des pâtés de maisons	232
11 : Enquête préliminaire "bâti / population", exploitation pour la mise au point du second degré du sondage : échantillonnage des ménages dans les pâtés de maisons.....	255
12 : Le plan de sondage et sa mise en œuvre pour l'enquête Démographie/ migrations.....	279
QUATRIEME PARTIE : EVALUATION DES PERFORMANCES ET DES CONDITIONS D'APPLICATION DE LA METHODE	318
13 : Précision de la méthode : les premiers résultats de l'enquête réalisée à Quito.....	320
14 : Domaines et conditions d'application	342
15 : Coûts et délais de la mise en œuvre.....	354
CONCLUSION : SYNTHESE DES PRINCIPAUX RÉSULTATS, PERSPECTIVES POUR LES RECHERCHES FUTURES.....	374
ANNEXES	383
BIBLIOGRAPHIE.....	622
Liste des tableaux, figures et photographies.....	629
PLAN DETAILLE.....	634

AVERTISSEMENT

Comme on le verra dans l'introduction, ce travail s'inscrit dans un projet de recherche pluridisciplinaire commencé à l'ORSTOM en 1985. Intégré dès cette date à l'équipe constituée autour de ce projet ¹, j'ai mené toutes mes recherches dans ce cadre. Étant donnée l'étroite imbrication des domaines sur lesquelles travaillait chacun des membres de l'équipe, cette thèse ne pouvait se concevoir sans l'intégration de certains textes présentant les travaux et les résultats des autres chercheurs participant au projet. De ces travaux et résultats dépendait l'avancement de nos propres recherches, nous y ferons constamment référence dans la suite. On trouvera donc dans cette thèse, des paragraphes construits à partir des éléments que nous ont fournis les autres participants au projet ou même directement repris de leurs textes. Il nous a en effet semblé inutile, lorsque la synthèse de certains travaux été faite par leur auteur même de réécrire moins bien que lui le contenu de son texte. Dans tous les cas, emprunts d'idées, d'éléments de résultats ou de textes déjà écrits, nous nous sommes efforcé de donner, dans le texte ou en note, la référence précise permettant de se reporter à l'écrit original de l'auteur.

¹ : L'équipe ORSTOM comprend, F. DUREAU (géographe-démographe), B. LORTIC (télédétection), A. MICHEL (urbaniste spécialisé en télédétection) et moi-même. Les travaux liés à l'infographie ont été réalisés par M. SOURIS (informaticien) et F. PELLETIER (infographie), les travaux photographiques par A. AING et I. RANNOU (photographes). Sur la partie du projet concernant Marseille, nous avons travaillé en association avec A. CHATELAIN (géographe) de l'EHESS Marseille.

INTRODUCTION

1. Petite chronique d'une croissance urbaine

Si nous écrivions, pour une certaine presse qui répond plus souvent des lieux communs qu'une information contrôlée sur place, un "reportage" sur une grande métropole du tiers monde, nous pourrions commencer ainsi :

"Ce qui frappe à l'arrivée à K., pendant qu'on se demande si le taxi parviendra jusqu'à l'hôtel sans se désintégrer, c'est l'image contrastée des fastes du centre d'affaires et des quartiers résidentiels, cotoyant la misère des faubourgs où s'entassent les exclus de la croissance économique. Comme beaucoup de capitales d'Amérique Latine, K. a connu, et connaît encore, un afflux de migrants incontrôlable et elle n'offre à la majorité d'entre eux que le chômage et les conditions de vie les plus précaires. K., comme Mexico et Rio de Janeiro, et au même titre que Calcutta ou Le Caire, est un déficit à la gestion urbaine etc..."

On peut bien sur reconnaître aux médias où foisonnent ces phrases à l'emporte pièce, le mérite d'attirer l'attention sur d'autres caractéristiques que celles que vantent les dépliants des agences de voyage. Cependant, si l'on s'éloigne du papier glacé de ces hebdomadaires pour vivre dans l'une de ces villes, on comprend très vite qu'il faut apporter à ces clichés certaines nuances. Comparée à ces mégalo-poles "à sensations", Quito, capitale de l'Equateur, avec un peu plus d'un million cent mille habitants, est une ville de taille moyenne qui apparemment ne connaît pas les mêmes problèmes. Elle offre effectivement aux hommes d'affaires pressés un centre économique et commercial moderne, aux touristes curieux un site grandiose et un quartier colonial pittoresque, et à ses habitants aisés des quartiers résidentiels agréables et de nombreux services. Mais, effet d'une croissance relativement lente, les quartiers populaires centraux sont assez anciens, bien structurés et n'ont rien de commun avec des bidonvilles.

Quito paradis urbain donc, ou du moins ville à visage humain? Qu'en est-il si l'on s'écarte des itinéraires touristiques et des lieux de vie des résidents étrangers ou nationaux fortunés?

Le centre historique de Quito, l'un des ensembles d'architecture coloniale les plus riches d'Amérique Latine, a été déclaré par l'UNESCO "patrimoine culturel de l'humanité". C'est, avec les îles Galapagos, la principale attraction touristique du pays et le point de passage obligé de toute visite à Quito. Lors du tremblement de terre qu'a connu la ville en mars 1987, le vieux centre a souffert d'importants dommages et on s'est ému. Cependant les accidents naturels ne sont pas les seules menaces qui pèsent sur ce quartier, la population y vit en fait dans des conditions difficiles. Si l'on franchit les portes, derrière les façades plus ou moins entretenues des vieilles demeures coloniales, les patios intérieurs sont souvent totalement délabrés et pourtant, ils abritent les plus fortes densités de toute l'agglomération. Quito est une ville très étendue, comptant peu de constructions élevées, où la densité est faible en

moyenne (environ 65 hab/ha¹) mais extrêmement variable. Dans les extensions périphériques récentes la densité ne dépasse pas 25 hab/ha, tandis que dans certains îlots du vieux centre elle peut atteindre et même dépasser 500 hab/ha. Dans ce quartier comptant beaucoup d'immeubles innocupés (édifices administratifs, commerciaux, religieux etc.) et où la plupart des immeubles d'habitation ne dépassent pas trois niveaux, on imagine l'entassement et les conditions de logement qu'affronte la population. Les propriétaires, qui pour la plupart ne résident plus dans le quartier, se soucient généralement peu d'effectuer les travaux d'entretien que réclament les appartements, avant tout destinés au rapport. Puisque ceux-ci trouvent preneurs en l'état, on fait dans bien des cas "commerce de sommeil".

Si le vieux centre est très peuplé, ce n'est pas parce que Quito manque de terrains constructibles. Vu depuis les premiers sommets du volcan Pichincha qui domine la ville, son territoire apparaît comme une mosaïque où alternent, sauf dans la partie centrale, espaces construits et espaces vides. On estime à près de la moitié de la surface considérée comme urbaine les disponibilités en terrains à l'intérieur de cette même surface². Pourtant la croissance spatiale de la ville se poursuit en dehors des limites d'extension des réseaux de voirie, d'électricité, d'eau courante etc., parfois dans des zones que l'on imagine mal pouvoir être raccordées un jour à ces réseaux. Le recensement de 1982 chiffrait à 132 000 habitants la population des secteurs dits "marginaux" de Quito et on estime que 35% de la population totale de la ville n'est pas reliée aux réseaux d'égouts, d'électricité, et d'eau courante³. Ces phénomènes d'extensions, appelées ici "invasiones", ont lieu dans bien d'autres grandes villes d'Amérique du Sud où ils sont parfois complètement incontrôlés.

A Quito il est relativement récent et semble ne pas faire l'objet d'une répression très virulente de la part des autorités. La presque totalité des coopératives de logements constituées lors de ces invasions sont maintenant "légalisées" et distribuent à leur adhérents des titres de propriété officiels. En revanche ces extensions, du fait qu'elles regroupent une population importante et que l'on considère comme facilement "manipulable", sont souvent le terrain de manœuvres politico-électorales rien moins qu'innocentes. Durant tout l'exercice du gouvernement antérieur, de 1984 à 1988, un conflit de terres a opposé les coopératives "Pisuli", jugée subversive, et "Jaime Roldos", pro-gouvernementale et très favorisée en terrains, viabilisation et équipements. Avant tout politique, et soigneusement entretenu, le conflit faisait régulièrement des victimes dans la population : au total dix huit morts en quatre ans, tous membres de la coopérative "Pisuli" à notre connaissance.

La dernière invasion en date s'est produite à la fin du mois de juin 1988, l'ancien gouvernement étant encore en place. Elle concernait 180 hectares, propriété du ministère de la santé, situés à environ 3500 mètres d'altitude sur les flancs du volcan Pichincha, dans ce qu'on appelle le bois protecteur de Quito. Sur ces terrains totalement isolés et sans aucune possibilité de viabilisation, la coopérative "Jaime Roldos", prévoyait d'installer 14 000 familles, regroupées dans une nouvelle coopérative qui se serait appelée "Leon Febres Cordero" (nom du Président de la République d'alors ...). Les lots, chacun de 100 m², auraient été vendus 40 000 sucres c'est à dire plus de trois fois le salaire minimum mensuel d'alors. Mais les 35 familles qui avaient commencé l'abattage des arbres, au dire de la presse sous l'œil bienveillant

¹ : Ce chiffre, et les indications de densités suivantes, sont ceux obtenus en novembre 1987 grâce à l'enquête par sondage effectuée à cette date dans le cadre du projet ORSTOM "télé-détection et démographie", enquête ayant portée sur 3169 ménages et dont le plan de sondage est l'aboutissement des recherches de l'équipe. Ce plan est exposé au chapitre 12.

² : F. CARRION, directeur de la planification de la municipalité de Quito, déclarait par exemple au journal "EL HOY" le 6 juin 1988 : " (...) A todo esto, hay que mencionar que casi el 50 % del suelo está vacante, o sea, que es manejado con fines especulativos." (...A tout cela il faut ajouter que presque 50% du sol est vaquant, c'est à dire qu'il est utilisé à des fins spéculatives).

³ : Chiffre cité par F.CARRION dans "EL HOY" du 6/6/88.

des policiers, n'avaient en fait payé qu'un droit d'inscription de 1500 sucres. L'aspect spéculatif de l'affaire n'a pas été, curieusement, au centre du compte rendu qu'en fit la presse ; c'est la menace écologique que représentaient ces déboisements pour l'environnement de la ville et les risques de coulées de boue, qui étaient mis au premier plan. Après que deux solutions de relogement des familles concernées aient été envisagées puis abandonnées, l'affaire, à notre connaissance, n'a plus été abordée dans les journaux, mais il semble que les déboisements se soient arrêtés¹.

Ainsi, que se soit dans le vieux centre surpeuplé ou dans les extensions périphériques récentes, les conditions de vie des habitants sont parfois difficiles. Dans les deux cas également, il s'agit d'une population aux faibles ressources contenant une proportion importante de migrants récents. La capitale attire en effet maintenant beaucoup de migrants, comme Guayaquil, première ville du pays par sa population et jusqu' à présent le cœur économique de l'Équateur, l'a fait depuis longtemps déjà. C'est que durant les années 70, Quito a connu une croissance économique soutenue du fait de son rôle moteur dans l'exploitation du pétrole amazonien découvert à la fin des années 60. Aujourd'hui donc, Quito se transforme et doit affronter des problèmes de gestion de l'espace complexes. Il s'agit peut être, pour revenir à notre terminologie à "sensations" du début, d'un défi à la gestion urbaine, mais le rapide tour d'horizon que nous venons de faire montre bien que l'ensemble des acteurs sociaux, chacun avec ses pouvoirs et ses moyens, agissent et proposent ou imposent des solutions. Défi oui, mais certainement pas mécanisme totalement incontrôlé, ou cause perdue.

2. La recherche et la gestion urbaine ont besoin de données actualisées, localisées et diversifiées

Si ces quelques exemples concrets occupent la première place de cet exposé, ce n'est pas pour noircir à plaisir le tableau d'une ville à laquelle nous nous sommes attachés parce que depuis deux ans nous l'avons parcourue et avons eu le temps de la connaître. Ce n'est pas non plus pour dénoncer la corruption de certains fonctionnaires et moins encore l'inefficacité de la gestion municipale ; les premiers ne sont pas une spécialité équatorienne et quant à la seconde, elle fait ce qu'elle peut avec ses faibles moyens et réglementer ou planifier n'a jamais suffi à contrôler le jeu des spéculations et ses conséquences de ségrégation spatiale. Nous voulions simplement, à travers l'évocation de certains problèmes que connaît Quito, mettre en évidence une triple nécessité d'information.

Tout d'abord, les chercheurs comme les gestionnaires et plus généralement tous les "usagers" de la ville, ont besoin d'une *information actualisée*. L'exemple de Quito, qui pourtant connaît un taux de croissance très raisonnable, 4,7 % annuel durant la dernière période inter-censitaire (1974-1982)², est à cet égard instructif. On ne peut pas observer ou planifier la croissance urbaine à partir des seules données que fournit le recensement, dans le meilleur des cas tous les dix ans. Ainsi d'après les recensements, en 1950 la ville comptait 210 000 habitants, 355 000 en 1962, 600 000 en 1974, 866 000 en 1982 et, d'après l'enquête que nous avons réalisée, 1 120 000 en 1987. Ensuite il faut également une *information localisée*. Disposer régulièrement, par exemple tous les ans, d'un chiffre global de population ne suffit pas non plus au besoin de l'administration ou de la recherche. Comme le montre l'exemple cité, savoir

¹ : Les informations citées à propos de cet événement sont extraites du quotidien "EL COMERCIO", qui durant la période du 20 juin 88 au 2 juillet 88, en fit un compte rendu suivi. L'autre quotidien quiténien "EL HOY" donnait des informations concordantes mais moins complètes.

² : Ce taux a peu varié durant la période 1950-1982 : d'après les recensements il est de 4,47 % entre 1950 et 1962, 4,47 % entre 1962 et 1974, 4,70 % entre 1970 et 1982. En revanche il semble croître légèrement depuis 1982 puisque d'après l'enquête faite en Novembre 1987, il s'établirait à 5,27 sur la période 1982-1987.

où, dans la ville, s'installent les nouveaux arrivants et où iront-ils au fur et à mesure de leur intégration peut être aussi important, pour le chercheur comme pour le gestionnaire, que de savoir combien ils sont chaque année.

Enfin, pour comprendre et gérer la ville, il faut une *information diversifiée*, presque à la demande. Pour cerner les motivations, les stratégies et les besoins des habitants, un questionnaire unique, aussi long et complet soit-il, ne suffit pas. C'est la diversification des enquêtes et de leurs thèmes qui peut seule répondre à la demande.

Ces trois caractéristiques indispensables des données nécessaires à l'étude et à l'administration des villes : actualité, localisation et diversité, déterminent le thème de notre recherche. La question centrale à partir de laquelle elle s'est développée est la suivante : en ville, quelle méthode de collecte de données démographiques serait à l'heure actuelle capable de satisfaire ces trois impératifs? Avant de répondre à cette question nous allons examiner rapidement en quoi les méthodes traditionnelles de collecte se révèlent souvent inadaptées à ces exigences.

3. Les méthodes traditionnelles de collecte ne suffisent pas à la demande d'information sur les populations urbaines

3.1 Le recensement

Le recensement de population est jusqu'à présent l'instrument de base de la planification, qu'elle soit nationale, régionale ou urbaine. Ses caractéristiques générales : périodicité et exhaustivité en font la source privilégiée d'informations relatives à la population. Ces deux caractéristiques semblent à première vue garantir les qualités d'actualité et de localisation que nous jugeons indispensables s'agissant des données urbaines. Qu'en est-il exactement? Pour répondre à cette question nous nous référerons principalement à l'ouvrage que D. TABUTIN a consacré à la collecte des données démographique¹.

- Périodicité des recensement et actualité des données

Les Nations Unies recommandent "d'effectuer les recensements à intervalles réguliers afin de disposer d'informations comparables et régulièrement espacées". Cependant :

"Les pays qui effectuent des recensements à périodicité régulière sont relativement peu nombreux : citons l'Angleterre (tous les dix ans depuis 1921), les Etats Unis (tous les dix ans depuis 1790), le Japon (presque tous les cinq ans depuis 1898) parmi les pays dits développés, l'Inde et le Pakistan (tous les dix ans depuis 1881 et 1901) et quelques pays d'Amérique Latine ou Centrale comme le Brésil et le Mexique depuis le début de ce siècle parmi les pays en développement."²

Comme on le voit, la régularité est une qualité rare des recensement, mais surtout seul le Japon a systématiquement un intervalle intercensitaire inférieur à dix ans : *partout ailleurs il est généralement variable entre huit et douze ans*³. A la longueur de la période intercensitaire s'ajoute le délai nécessaire à la saisie, vérification, correction et tabulation des données collectées avant la publication des résultats.

" Le déroulement complet d'un recensement, de ses phases préparatoires aux publications, dure en général 4 ou 5 ans. Sans même parler des coûts financiers, l'opération est donc d'envergure."⁴

¹ : D. TABUTIN : La collecte des données en démographie. ORDINA EDITIONS, LIEGE (1984).

² : D. TABUTIN, 1984. Op. cit., p. 85.

³ : Voir D. TABUTIN, 1984. Op. cit., p. 22 & 23.

⁴ : D. TABUTIN, 1984. Op. cit., p. 96.

En fait, comme le dit Ph. ANTOINE. "Les résultats, sous forme de tableaux seront en général disponibles au mieux trois ans après l'enquête de terrain"¹. *Ainsi il est clair que l'actualité des chiffres du recensement est toute relative et généralement très insuffisante pour permettre le suivi de la croissance urbaine dans le tiers monde.*

- Exhaustivité

" Un autre grand intérêt du recensement est la possibilité des études régionales ou locales : il faut attendre un recensement pour par exemple avoir une idée précise de la croissance urbaine, de l'état général du parc de logements dans les villes (...). Le recensement de la population est ainsi la principale base statistique permettant de faire le point en maints domaines et donc d'élaborer une politique."²

Cet intérêt des recensements, lié à leur exhaustivité, n'est pas contestable. Cependant, comme nous l'avons vu au paragraphe précédent, "l'état des lieux" détaillé qu'il permet de dresser rendra compte, au moment même où il sera connu, d'une situation vieille d'au moins trois ans. De plus la lourdeur des opérations que nécessite l'exploitation d'un recensement au niveau le plus fin, empêche souvent son utilisation pour des études régionales ou locales :

"L'objectif premier de ces recensements est inscrit dans la logique de leur financement : il répond au besoins de connaissances démographiques des organismes nationaux. Pour les administrations nationales, il devrait permettre d'élaborer des objectifs précis de planification et d'aménagement du territoire(...). *Or le recensement est rarement conçu comme un outil potentiel d'aide au développement. Aussi les informations locales et régionales ne sont elles traitées qu'accessoirement.* D'autant que leur simple distribution à tous les services concernés n'est souvent pas assurée."³

Lorsque malgré tout cette exploitation locale a lieu, il s'opère généralement, par rapport au questionnaire relatif à chaque ménage, *une importante réduction de l'information disponible en fin de compte au niveau du district de recensement.*

- L'information recueillie

La définition des thèmes couverts par le recensement est également une décision prise à l'échelon national.

"A l'intérieur même du pays, l'urgence des problèmes peut être vue différemment selon les ministères, et souvent, après de longues discussions, l'opération nationale qu'est le recensement sera un compromis entre diverses tendances. Dans pratiquement tous les recensements il y a cependant quelques points communs dans les objectifs : effectif de la population totale, sexe, âge, état matrimonial, le logement, la scolarisation et l'activité économique sont en général abordées (...)."⁴

Mais, dès que l'on s'écarte de ces rubriques de base, le recueil de l'information est loin d'être systématique : durant la période 1965-1971, par exemple le lieu de résidence habituel et le lieu de naissance n'étaient recueillis que dans 75% des recensements

¹ : Ph. ANTOINE, Recensement : une efficacité relative. Article paru dans "Actuel développement", n° 59, 1984.

² : D.TABURIN, 1984. Op. cit., p. 87.

³ : Ph. ANTOINE, 1984, article déjà cité.

⁴ : D. TABUTIN, 1984. Op. cit., p. 89.

effectués, la religion dans 45% des cas et le nombre d'enfants nés vivants dans seulement 30%¹.

En définitive, que conclure de ce rapide examen des principales caractéristiques des recensements? Ils restent certes indispensables à l'établissement des statistiques nationales et dans ce domaine, D. TABUTIN constate qu'ils s'améliorent, en particulier dans les pays en voie de développement. Ils permettent également, lorsque leur exploitation est poussée jusque là, d'obtenir des "instantanés" de l'état de la population concernant des unités géographiques fines mais généralement avec un nombre de variables réduit. *Cependant, surtout pour ces données locales, les délais d'exploitation et la réduction d'information sont tels que cette source ne peut répondre seule aux besoins de la gestion et de la recherche urbaine.*

3.2 Les enquêtes par sondages

Face au coût, à l'inactualité et au caractère nécessairement trop général des données fournies par les recensements, les enquêtes par sondages se multiplient dans tous les domaines où les phénomènes à observer évoluent vite et où on a besoin de multiples informations : démographie, économie, agriculture, santé etc. Traditionnellement, du moins en ce qui concerne les enquêtes visant les populations humaines, les méthodes de sondages employées, reposent sur l'existence de listes de ménages.

Ces méthodes peuvent-elles fournir une information actualisée, localisée et diversifiée? *C'est précisément l'intérêt du sondage de permettre pour un coût raisonnable, de recueillir en un temps très court une information sur un grand nombre de variables. Actualité et diversité de l'information sont donc assurées.* Mais deux problèmes demeurent :

- Il est très souvent difficile de disposer d'une liste de ménages à jour comme base de sondage.
- On fait généralement au sondage le reproche de ne pas pouvoir produire une information localisée.

- Le problème de la base de sondage

La plupart du temps, dans les enquêtes démographiques d'envergure, le sondage est effectué à partir de la liste de ménages ou d'individus que fournit le recensement. D'une part dans les villes des pays en développement de telles listes sont en fait assez rares. D'autre part il est bien évident que, même lorsqu'elle est disponible, cette liste n'est jamais complète.

Si l'on considère la date de référence du recensement, on estime, en France ou aux Etats Unis, entre 2 et 3% le taux d'omission global, mais ce taux peut atteindre 8 ou 10 % concernant des catégories de population minoritaires plus "difficiles à atteindre" (c'est généralement le cas des catégories sociales défavorisées)². De plus nous avons vu que dès sa parution, le recensement date au minimum de trois ans. Or certaines grandes villes des pays en développement croissent depuis plus de dix ans au rythme de 10% par an. Avec une telle croissance, la population a augmenté d'un tiers un trois ans, de deux tiers en cinq ans et double tous les sept ans. *Le coût de l'actualisation d'une liste de ménages datant d'au moins trois ans sera dans ces conditions très élevé.*

Il peut parfois être assumé par des organisme nationaux pour des enquêtes jugées prioritaires, comme en Inde, depuis 1950, avec le N. S. S. (National Sample Survey devenu en 1968 le S.R.S., Sample Registration Scheme), mais c'est beaucoup plus rarement le cas pour des administrations régionales ou locales. *Que dire alors du coût d'établissement d'une liste lorsque, ce qui n'est pas si rare, aucune source n'existe!*

¹ : D. TABUTIN, 1984. Op. cit., p. 93.

² : Voir D. TABUTIN, Op. cit., p. 108

- Le problème de l'information localisée

Il n'y a en fait *aucun obstacle théorique* à produire à l'aide d'un sondage des données dans des domaines géographiques restreints. On utilise pour cela des techniques d'estimation dans des domaines d'études sur lesquelles nous reviendrons dans les chapitres suivants. Le problème pratique est celui de la taille de l'échantillon à sélectionner lorsque qu'on fixe à l'avance la précision des estimations. GOURIEROUX le formule ainsi :

"Fondée sur la loi des grands nombres, la méthode des sondages suppose l'observation d'échantillons suffisamment importants. La taille de l'échantillon nécessaire pour obtenir une précision donnée de l'estimation est à peu près indépendante de l'importance numérique de la population étudiée. Ainsi, si la population est divisée en dix catégories pour lesquelles on veut effectuer des estimations distinctes, il faudra approximativement dix fois plus d'observations que pour obtenir la même précision sur l'estimation correspondant à la population toute entière."¹

On peut, pour se rapprocher des termes concrets dans lesquels se posent généralement les problèmes de taille d'échantillon, renverser la question : étant donné un budget alloué à l'enquête, c'est à dire une taille fixée de l'échantillon, quelle précision auront les estimations restreintes aux domaines géographiques d'intérêt? Si, comme le dit GOURIEROUX, cette précision ne dépend que peu de la taille des domaines considérés (encore qu'il ne faille pas descendre en dessous de certains seuils, mais cette contrainte n'est généralement pas gênante), *elle dépend en revanche beaucoup des différents choix effectués lors de la définition de la base de sondage, de la stratégie d'échantillonnage et des techniques d'estimations qui vont être employées*

4. Utiliser en ville les images satellites à haute résolution comme base de sondage, un projet de recherche à l'ORSTOM

Face au constat, que nous venons d'effectuer, des limites des méthodes classiques d'observation de la population dans les villes des pays en développement, que proposer pour tenter d'améliorer la collecte démographique en milieu urbain?

Le projet et ses hypothèses fondatrices

En 1985 a débuté à l'ORSTOM, au département Urbanisation et Socio-systèmes urbains, un projet de recherche pluridisciplinaire² dont l'objectif est de *mettre au point un système d'observation permanent des populations urbaines utilisant l'information apportée par les satellites à haute résolution* (SPOT et Thematic Mapper). F. DUREAU, qui est à l'origine de ce projet, lorsqu' elle en définit l'idée directrice, propose une réponse à la question que nous posions précédemment :

"Viser une amélioration des recensements exhaustifs offre peu d'intérêt puisque ceux-ci demeureront toujours trop longs à mettre en œuvre et à exploiter ; il apparaît certain que seul un système basé sur la technique des sondages peut remplir les conditions de souplesse et de rapidité nécessaires dans les villes à croissance démographique rapide.

¹ : C. GOURIEROUX : Théorie des sondages, P. 12. ECONOMICA, PARIS (1981).

² : L'équipe ORSTOM travaillant à ce projet est celle citée dans l'avertissement :

- Département D : F. DUREAU, A. MICHEL, O. BARBARY et A. CHATELAIN (chercheur associé, EHESS Marseille).

- Atelier de Télédétection : B. LORTIC, I. RANNOU

- Infographie : M. SOURIS, F. PELLETIER, C. ANTIGEOL

Dans ce cadre, une meilleure connaissance de l'espace intra-urbain peut être un élément fondamental d'amélioration des techniques d'enquête par sondage. D'une part, une bonne connaissance cartographique permet de disposer d'une base de sondage complète et à jour, composée d'ilôts clairement identifiés ; d'autre part, une bonne connaissance de la morphologie urbaine permet de stratifier la ville en augmentant sensiblement le gain de précision qu'introduit toute stratification, et d'obtenir des résultats selon un découpage spatial significatif de la ville enquêtée." ¹

L'idée centrale du projet repose donc sur deux hypothèses :

"1) il existe des relations étroites entre les caractéristiques morphologiques du milieu urbain et les caractéristiques démographiques et socio-economiques des habitants ;

2) on peut identifier sur les images satellite de type SPOT ou Thematic Mapper les caractéristiques morphologiques pertinentes pour l'observation démographique en ville." ²

L'application de ces deux hypothèses à la définition d'une méthode de collecte de données en milieu urbain intégrant la télédétection spatiale constitue un champ nouveau de recherche. Il couvre principalement les domaines de la télédétection urbaine et des sondages spatiaux en ville. Nous verrons plus loin, grâce à une étude bibliographique, que les travaux réalisés sur ces sujets sont en nombre très limité. En télédétection urbaine, ils concernent principalement le traitement multi-dates de l'occupation du sol, l'évaluation de la quantité de végétation et l'observation des franges urbaines. Quant au sondage spatiaux, leur application à la collecte démographique en milieu urbain est très rare ; pour ce type de sondage l'habitude est, comme nous l'avons vu, de tirer des échantillons de ménages ou d'individus dans des listes. Le sondage spatial est en revanche souvent pratiqué en statistique agricole, forestière ou minière. Les travaux réalisés dans ces domaines constituent évidemment des acquis méthodologiques précieux mais nous verrons, ce qui n'est pas étonnant, que les techniques qui sont alors mises en œuvre ne peuvent pas s'appliquer telles quelles au milieu urbain.

Enfin, le projet de recherche touche un troisième domaine d'investigation qui est celui de la gestion informatique des bases de données localisées. Il s'avère en effet indispensable, pour mener à bien les travaux en télédétection urbaine et en sondages spatiaux, de manipuler des données démographiques, morphologiques et satellitaires, qui ont chacune des modes d'implantation géographique très différents. Nous reviendrons plus tard sur cette question, mais disons tout de suite que la gestion de ces différents types de données est grandement facilitée par l'emploi d'un système informatique de gestion de base données adapté aux données localisées. C'est pourquoi dès leur début, nos recherches ont fait un usage intensif du système d'information géographique développé à l'ORSTOM ³ ; réciproquement, le système s'est enrichi de nouveaux modules permettant l'intégration et la gestion des données satellitaires.

¹ : F. DUREAU & O. BARBARY (1987), Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale, article à paraître dans la revue "Informatique et Sciences Humaines".

² : F. DUREAU (1987). Op. cit.

³ : Il s'agit du système SAVANE, développé par l'Unité d'Infographie. Ce système, organisé suivant le schéma des systèmes de gestion de base de données relationnelle, est doté de possibilités de manipulation graphique grâce à des opérations algébriques nouvelles.

Les axes de recherche et leurs objectifs

Ainsi, pour réaliser ce projet, trois axes de recherche ont donc été définis :

- 1 : L'interprétation des images satellite à haute résolution en milieu urbain.
- 2 : La formalisation mathématique des relations entre morphologie urbaine et population, et le développement des techniques de sondages spatiaux en démographie urbaine.
- 3 : L'intégration des données de télédétection dans les bases de données localisées.

C'est le second axe de recherche qui fut à l'origine de notre intégration à l'équipe en 1985 et qui fait aujourd'hui l'objet de cette thèse. Nous décrirons, dans le paragraphe suivant, les objectifs principaux que l'on peut fixer à notre thème de recherche mais, étant donné l'étroite inter-dépendance qui existe entre les trois axes, il nous semble utile d'examiner rapidement quels objectifs poursuivront les recherches concernant les deux autres thèmes.¹

"- L'interprétation des images satellite à haute résolution en milieu urbain : il s'agira de définir quelles sont les informations relatives à la morphologie urbaine contenues dans ces images et quels sont les traitements permettant d'extraire ces informations. L'utilisation d'images satellite SPOT et TM en milieu urbain nécessite une phase préalable d'études de fiabilité du document et de mise au point de méthodes d'analyse spécifiques ; outre les méthodes déjà opérationnelles basées sur la valeur radiométrique du pixel, devront nécessairement être développées des techniques basées sur des analyses de texture et de structure.

- L'intégration des données de télédétection dans des bases de données localisées a pour objectif une observation suivie et spatialisée des villes dans leurs différentes composantes, notamment morphologique et démographique. Ce troisième axe de recherche repose sur un double constat :

1) Les connaissances actuelles en télédétection urbaine étant très limitées, les confrontations spatiales avec les autres données de la base facilitent l'extraction de l'information fournie par les images et la validation des méthodes d'interprétation.

2) La multiplication des possibilités de mise en relation de données démographiques et morphologiques localisées par les systèmes d'information géographique facilitera sensiblement la mise au point d'un système d'observation démographique reposant sur la connaissance de la morphologie urbaine.

De plus, comme l'écrit justement A. BALLUT² à propos de la télédétection urbaine, le fait de *prendre en compte la localisation* peut constituer un atout important : "Il faudrait, dans les algorithmes de traitement, faire intervenir des notions relatives à la répartition géographique". De même, la localisation peut constituer un paramètre discriminant d'un point de vue démographique, au même titre que des variables décrivant la morphologie urbaine."

Ainsi l'utilisation et le développement du système d'information géographique mis au point à l'ORSTOM peut avoir des répercussions importantes sur les recherches en télédétection urbaine comme sur celles portant sur les sondages spatiaux.

¹ : Les éléments qui terminent ce paragraphes sont repris de F. DUREAU dans l'article déjà cité.

² : Voir BALLUT, 1980, p. 18.

Les étapes du projet

La première phase du projet s'est déroulée en France, sur la ville de Marseille. Commencée au début de 1985, les trois axes de recherches y étaient menés en parallèle. L'objectif principal de cette phase était d'accumuler, dans les meilleures conditions de données disponibles et d'environnement technique et scientifique, le maximum de résultats méthodologiques applicables à la définition de la méthode. En effet, étant donné le caractère tout à fait exploratoire du projet, il importait d'avoir des références fiables pour tester la validité des méthodes mises au point au fur et à mesure de l'avancement des travaux. C'est pourquoi nous avons choisi de travailler d'abord sur une ville française disposant d'une image Thematic Mapper, d'un recensement récent disponible au niveau du district de recensement et d'une couverture aérienne à grande échelle. Les travaux concernant cette phase de mise au point méthodologique se sont poursuivis jusqu'à la fin de l'année 1987.

Cependant, en septembre 1986, a débuté la seconde phase du projet : première application de la méthode à une ville d'un pays en développement. Depuis cette date, nous travaillons sur la ville de Quito (Equateur). Ce site est tout à fait favorable à la réalisation de cette phase : recensement de 1982 disponible au niveau de l'ilot, couvertures aériennes récentes, milieu urbain bien connu, bénéficiant d'une accumulation de données dans le cadre de l'Atlas Urbain de Quito, image SPOT et TM d'août et septembre 1986. Appliquant les méthodes développées sur Marseille, nous avons réalisé en 1987 à Quito une enquête démographique par sondage aréolaire. Le plan de sondage spatial retenu utilise la base de sondage constituée par les images satellite, stratifiée selon des critères morphologique eux-même issus de l'information satellitaire. Il s'agit donc d'une application en vraie grandeur du système d'observation démographique intégrant la télédétection.

Enfin, à propos de ces deux phases du projet, F. DUREAU fait la remarque suivante :

"S'il fut nécessaire, dans la phase de mise au point de la méthode, de travailler sur des sites disposant de nombreuses informations localisées intégrées dans une base de données, il importe de souligner que le champ d'application visé par ce projet est en réalité constitué par les villes ne disposant que de très peu de données. C'est précisément pour répondre à cet état de carence, le plus fréquent dans les villes des pays en développement, que nous cherchons à élaborer un système d'observation reposant largement sur la télédétection, devant permettre de produire des données démographiques rapidement et à moindre coût."¹

5. Les objectifs principaux de l'analyse statistique et de la recherche sur les sondages spatiaux

Nous avons défini au cours du paragraphe précédent l'objet de notre propre recherche : la formalisation mathématique des relations entre morphologie urbaine et population, et le développement des techniques de sondages spatiaux en démographie urbaine. Nous détaillerons tout au long de ce travail, de nombreux objectifs, spécifiques à chaque étape, et nous nous attacherons successivement à chacun d'eux pour tenter de proposer au bout du compte un plan de sondage applicable au contexte urbain ; l'intégration, dans la définition de ce plan, de l'information satellitaire permet, nous le verrons, d'atteindre un degré de précision des estimations intéressant. Avant de passer à cet exposé technique, nous voulons ici donner *les grandes lignes de la problématique* qui le sous-tendra, lesquelles déterminent le plan de cette thèse.

¹ : F. DUREAU (1987). Op. cit.

Nous aborderons ce sujet en deux temps, correspondant à nos deux objectifs principaux qui, bien qu'ils n'occupent pas par la suite le même volume dans l'exposé, n'en sont pas moins à nos yeux d'importance égale.

- Définition générale du plan de sondage

Le premier objectif est la définition des caractéristiques générales du plan de sondage. Cette définition suppose, avant tout, l'étude des conditions dans lesquelles va se dérouler l'enquête, c'est à dire la définition de l'objet de cette enquête et l'examen des contraintes et des possibilités liées au milieu auquel nous l'appliquerons et aux moyens dont nous disposerons. Ces trois facteurs : l'objet de l'enquête, les contraintes et les possibilités, pèsent toujours un grand poids dans la définition d'un plan de sondage, de plus la réflexion dont ils doivent être l'objet ne peut pas être le seul fait du statisticien. *Ce fut l'un des grands mérites d'une recherche pluridisciplinaire que de permettre une approche complète de ces problèmes.* Si notre propre travail est parvenu à prendre en compte les préoccupations des autres membres de l'équipe et à répondre, nous l'espérons, à leur demande, c'est à cette approche collective qu'il le doit.

La détermination de l'objet de l'enquête, c'est à dire la définition de la population étudiée et des informations que l'on souhaite obtenir concernant cette population, permettra de déterminer *l'unité statistique* sur laquelle portera le questionnaire. Puis la prise en compte des contraintes dues à l'emploi de l'image satellite comme base de sondage nous conduira à la définition des *unités d'échantillonnage qui seront sélectionnées au premier degré.* Enfin l'examen des possibilités d'extraction d'information morphologique de l'image satellite permettra de formuler le *corpus d'hypothèses nécessaire à la poursuite de la définition du plan de sondage.*

Ces hypothèses concernent :

- *les domaines géographiques* pour lesquels il semble possible de produire les estimations des caractéristiques démographiques de la population ;
- *les techniques statistiques* que l'on pourra utiliser pour tenter d'améliorer la précision de ces estimations en utilisant l'information sur la morphologie urbaine tirée de l'image.

Dans la première partie de la thèse¹, la définition générale du plan de sondage dont il vient d'être question sera suivi d'un bilan bibliographique au travers duquel nous tâcherons de dégager les acquis méthodologiques applicables à la problématique. Enfin, cette partie s'achèvera par l'exposé des problèmes techniques que nous devons résoudre et des choix méthodologiques faits pour parvenir à des solutions.

- Mise au point et évaluation des performances de la méthode

Le second objectif, qui nous occupera dans les deuxième, troisième et quatrième parties de la thèse, est celui de la mise au point de la méthode et de l'évaluation des performances du plan de sondage défini.

Dans la deuxième partie², à partir des données démographiques et morphologiques exhaustives disponibles sur la zone d'étude de Marseille, *nous mettrons au point une méthode d'évaluation* de la précision des différents plans de sondages dont nous considérons qu'ils présentent un intérêt pratique ou théorique pour la détermination finale de la méthode.

¹ : Chapitres 1 à 3

² : Chapitres 4 à 8.

Dans la troisième partie¹, nous appliquerons la méthode d'évaluation des plans de sondage mise au point à Marseille aux données recueillies lors des enquêtes préliminaires effectuées à Quito, afin de parvenir à *la définition complète du plan de sondage que nous appliquerons à l'enquête finale*.

Enfin la quatrième partie² sera consacrée à l'évaluation des performances de la méthode à partir de l'expérience menée à Quito : *à l'issue de l'enquête réalisée, l'estimation qui sera faite de la variance des estimateurs constituera une mesure de la précision que permet d'atteindre la méthode proposée ; nous donnerons également les éléments permettant d'évaluer les conditions matérielles, les coûts et les délais nécessaires à la mise en œuvre de la méthode de sondage à laquelle nous avons abouti.*

6. Pour une recherche appliquée : justification du principe empirique de la méthode

On s'apercevra très vite, à la lecture des pages qui suivent, que le travail effectué ici procède d'une méthode très empirique que nous allons présenter rapidement.

Notre recherche a pour objectif principal la vérification de l'hypothèse de départ du projet :

la prise en compte, dans la définition du plan de sondage, des différences morphologiques intra-urbaines, permet un gain appréciable de précision lors de l'estimation des caractéristiques démographiques des habitants.

Dans l'expérience faite à Marseille par exemple, cette vérification se fera grâce à un programme informatique qui, en appliquant le formulaire des principaux plans de sondage classiques, calcule la variance des estimateurs à partir des données exhaustives du recensement de population. La démarche expérimentale procède donc par évaluation chiffrée d'un ensemble d'alternatives possibles, cela afin d'effectuer en connaissance de cause les choix techniques conduisant à un sondage efficace. Ce sondage devra être adapté au milieu urbain et pratiquement réalisable sur document satellitaire.

Accumuler le maximum de conclusions pratiquement exploitables lors de l'application de la méthode à Quito, parfois sans chercher une formalisation mathématique du phénomène qui les sous-tend, tel est le but d'un travail qui se veut avant tout appliqué, quitte à rester exploratoire. C'est pourquoi on ne trouvera ici aucun développement théorique nouveau, ni même de références aux avancées les plus récentes de la théorie de l'échantillonnage spatial (par exemple l'application de la théorie des variables régionalisées de G. MATHERON³). Les lecteurs passionnés de théorie s'en trouveront probablement déçus, mais pour qu'au moins ils ne croient pas qu'il s'agit là d'un a priori rétrograde face à ces nouvelles techniques, il ne paraît pas inutile de tenter de justifier cette absence.

Une des voies importantes du développement récent des techniques de sondage, tant du point de vue théorique que dans des applications plus pratiques, est *l'étude mathématique des caractéristiques spatiales des variables que l'on cherche à estimer et leur modélisation*. Dans les sciences de la nature, en géostatistique ou dans le domaine des statistiques forestières, *cette étude se réfère le plus souvent à la théorie des processus stochastiques ponctuels ou fonctions aléatoires*. J. DELINCE en évoque ainsi l'objet : "plutôt qu'un test du caractère aléatoire de la population, ces techniques

1 : Chapitres 9 à 12.

2 : Chapitre 13 à 15.

3 : G. MATHERON : Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature. MASSON, PARIS (1965).

cherchent à décrire la structure (du processus) à différentes échelles et proposent des tests d'adéquation vis à vis de modèles quelquefois fort complexes"¹.

Outre son éventuelle complexité, l'utilisation de ces modèles nécessite généralement qu'on fasse des hypothèses sur la variation spatiale du phénomène étudié. Par exemple, GIUDICELLI, LANLY, OUAKAM et PIETRI, dans un article publié dans les Annales des Sciences Forestières², examinent le problème suivant : "L'estimation de l'erreur statistique dans les dispositifs de sondage systématique, largement utilisés en forêt, est un problème qui n'est pas soluble dans le strict cadre de la théorie classique des sondages. Conformément à une suggestion faite par COCHRAN et d'autres, les auteurs proposent une solution en considérant la forêt comme une réalisation d'un processus aléatoire stationnaire et isotrope". Plus loin, ils justifient ainsi les hypothèses faites :

"Dans le cas d'une forêt,

- l'hypothèse de stationnarité peut se justifier par une certaine stabilité, dans toute la forêt, des conditions d'apparition et de développement des arbres, s'il semble que cette hypothèse n'est pas réalisée - par exemple lorsqu'une certaine tendance apparaît due à la modification de facteurs écologiques (reliefs, sol, microclimat) - l'hypothèse de stationnarité peut être acceptée au niveau de strates judicieusement choisies ;

- L'hypothèse d'isotropie, vérifiée si la corrélation entre les valeurs de la caractéristique cherchée dans deux parcelles situées à une distance donnée est la même dans toutes les directions, n'est satisfaite dans une forêt que s'il n'apparaît pas une ou plusieurs directions privilégiées de la répartition de l'essence étudiée."

Malheureusement, dans le cadre qui nous occupe - l'étude des paramètres démographiques en milieu urbain - il apparaît peu raisonnable de penser que ces hypothèses (stationnarité et isotropie) sont vérifiées au niveau de l'agglomération entière. De même, si l'on cherche à définir des strates "judicieuses" à l'intérieur desquelles ces hypothèses seraient plus plausibles, on se heurte très vite au fait que l'hétérogénéité des caractéristiques démographiques des habitants reste élevée, même au sein de quartiers morphologiquement homogènes.

Plus généralement, les déterminants de la variation spatiale du peuplement humain des villes nous semble trop complexes et encore trop mal connus pour que l'on puisse faire, avec quelques chances de succès, des hypothèses aussi fortes que celles qui peuvent se justifier dans le cas des peuplements des milieux naturels ou dans celui de la géostatistique.

Pour en finir avec cette tentative de justification de la méthode, il est un dernier argument que l'on peut avancer. L'objet des statisticiens spécialistes de sondages dans le domaine des sciences de la nature est souvent de perfectionner en la raffinant la méthode d'estimation par échantillonnage spatial qui, dans ce milieu et pour le type d'information recherchée, a déjà fait ses preuves.

¹ : J. DELINCE : Estimations robustes de densité dans le plan, étude comparée de techniques d'échantillonnage, application en forêt. Université Catholique de Louvain, Faculté des Sciences Agronomiques (1983).

² : X.GUIDICELLI, J.P. LANLY, J.B. OUAKAM et M. PIETRI : Application de la théorie des processus aléatoires à l'estimation de la précision d'un inventaire forestier par échantillonnage systématique. Annales des Sciences Forestières, Institut National de la Recherche Agronomique (1972).

Notre objectif en revanche, dans un domaine où l'on en est qu'aux premières tentatives, est de mettre au point cette même méthode du sondage spatial en milieu urbain et d'évaluer la précision qu'elle permet d'atteindre, pour un coût donné, dans l'estimation de données démographiques.

Pour l'ensemble de ces raisons, on s'en tiendra donc dans l'ensemble de cette recherche aux méthodes déjà classiques du point de vue des théories mathématiques qui les sous-tendent mais qui, comme on le verra, conduisent néanmoins à des résultats intéressants.

PREMIERE PARTIE

**PROBLEMATIQUE, CONTEXTE THEORIQUE, CHOIX DES METHODES ET
ETAPES DU TRAVAIL**

Chapitre 1

LE PROBLEME GENERAL : UN SONDAGE AREOLAIRE STRATIFIE,

A DEUX DEGRES

Introduction : Quel sondage pour une nouvelle méthode de collecte démographique en ville?

Pour mettre au point une nouvelle méthode de collecte de données démographiques adaptée au milieu urbain du tiers monde, nous partons de l'idée que "seule une méthode basée sur les techniques de sondage peut remplir les conditions de souplesse et rapidité nécessaires"¹. Rappelons donc sur quel principe repose la méthode des sondages.

Comme chacun sait, il consiste, lorsque l'on veut recueillir une information sur une population considérée dans son ensemble, à n'observer qu'une fraction des individus de cette population appelée échantillon. De cette information partielle, il faut ensuite déduire des indications sur les caractéristiques de la population entière. Les techniques de sondages comportent donc toujours les deux phases suivantes.

- Le choix de l'échantillon que l'on devra observer : l'ensemble des règles déterminant ce choix est appelé "plan de sondage" et lorsque l'on définit ces règles en se référant à la théorie des probabilités, le sondage est dit probabiliste.

- Le choix de la formule mathématique qui combinera les informations obtenues sur l'échantillon pour fournir l'indication recherchée sur la population d'ensemble. Cette indication est appelée "estimateur" et le premier but de la théorie des sondages est de permettre que cet estimateur soit le plus proche possible de l'information recherchée.

Bien sûr la valeur fournie par l'estimateur n'est pas en général exactement égale à la vraie valeur recherchée. L'écart entre ces deux valeurs a deux composantes. La première, appelée "erreur d'échantillonnage", est celle qui résulte de l'observation partielle de la population. La seconde est introduite lors de la mise en œuvre de l'enquête du fait des non-réponses ou réponses fausses etc, et s'appelle "erreur de mesure". Dans la phase de définition du plan de sondage on s'attache à minimiser la première de ces sources d'erreur tandis qu'une bonne rédaction du questionnaire et un contrôle rigoureux du déroulement de l'enquête permettent de faire diminuer la seconde. Dans ce qui va suivre nous nous intéresserons principalement à l'erreur d'échantillonnage, domaine privilégié de l'intervention du statisticien.

Seuls les sondages probabilistes permettent d'estimer, à l'issue du processus de collecte, l'erreur d'échantillonnage et donc d'agir sur le plan de sondage pour la réduire. C'est pourquoi, dans cette recherche, on considérera exclusivement des sondages probabilistes. Mais, même dans ce domaine, l'éventail des possibilités est très

¹ : F. DUREAU & O. BARBARY, Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale, article à paraître dans "Informatique et Sciences Humaines".

vaste. Nous voilà donc amenés à la question centrale : Quel type de sondage pour une nouvelle méthode de collecte de données démographique en ville?

1.1 : Détermination de l'objet de la recherche : possibilités et contraintes

GOURIEROUX¹ définit, pour toute enquête par sondage, trois étapes principales dans le travail préliminaire à l'opération de collecte proprement dite :

- " 1) La détermination de l'objet de l'enquête,
- 2) L'étude des conditions générales dans lesquelles elle se déroule, permettant notamment d'examiner les contraintes auxquelles sera soumis le statisticien,
- 3) Le choix d'un système de travail pour collecter les données."

L'objet principal de notre travail est évidemment le troisième point cité, mais nous devons constamment prendre en compte les conclusions et les contraintes issues des deux premières étapes. La réflexion qui les sous-tend n'est pas, bien sûr, le seul fait du statisticien ; il s'agit là d'un des lieux privilégiés du travail pluridisciplinaire, auquel ont participé tous les membres de l'équipe et qui a également fait l'objet de consultations extérieures. Examinons dès maintenant les principaux points abordés lors de cette réflexion dont les conclusions déterminent pour une grande part le domaine de notre recherche.

1.1.1 Détermination de l'objet de l'enquête

Il s'agit ici de déterminer la population étudiée et l'ensemble des informations que l'on souhaite obtenir sur cette population. Une enquête démographique en milieu urbain vise bien sûr l'estimation de l'effectif total de population dans l'agglomération. Les limites géographiques de l'espace enquêté, fixées pour permettre l'actualisation de ce chiffre, doivent donc inclure les extensions les plus récentes de la ville. Mais les objectifs du type d'enquête qui nous occupe sont multiples : caractéristiques démographiques générales de la population, conditions de logement, d'emploi et de revenu des ménages, phénomènes de migration etc. Il apparaît clairement, au regard de cette liste, que "l'individu statistique" auquel on s'intéresse est le ménage. *L'univers objet de l'enquête est donc l'ensemble des ménages vivant à l'intérieur des limites actuelles de l'agglomération.*

1.1.2 Définition de la base de sondage

La première question à résoudre, lorsque l'on projette de réaliser une enquête par sondage, est celle de la constitution d'une liste d'unités d'échantillonnage ou, pour choisir une terminologie plus adaptée à la situation dans laquelle nous nous trouvons, d'une base de sondage. Le problème est généralement posé en ces termes² :

" L'utilisation des procédures mathématiques pour tirer l'échantillon des individus observés suppose l'existence d'une liste dans laquelle sont énumérés *tous les éléments de la population* . (...) Souvent on ne dispose pas de listes d'individus mais de listes de groupes d'individus. Celles ci peuvent cependant être utilisées pour tirer des échantillons d'individus de la manière suivante (tirage à plusieurs degrés). On tire un échantillon de groupes et pour chacun des groupes tirés, on constitue une liste des individus y appartenant. On tire alors

¹ : C. GOURIEROUX, Théorie des sondages, p 14 (Economica, 1981, PARIS).

² : C. GOURIEROUX, Théorie des sondages, pp 17 & 18 (Economica, 1981, PARIS).

dans ces listes des échantillons d'individus. (...) Les listes servant au tirage des échantillons sont souvent appelées bases de sondage."

Nous avons déjà évoqué en introduction les obstacles qui s'opposent dans bien des cas à l'établissement d'une liste de ménages à jour, c'est à dire sans omissions ni répétitions : obsolescence de la liste issue du recensement (lorsque celui ci existe), coût très élevé de l'actualisation des listes existantes. Ces obstacles, lorsque l'on s'intéresse à des villes à croissance rapide, comme le sont la plupart des villes du tiers monde, et que l'on veut produire des données à faible coût et avec une périodicité rapprochée, sont insurmontables. En l'absence de liste de ménages, nous nous trouvons donc dans la situation typique décrite par GOURIEROUX d'application des sondages à plusieurs degrés. Pour nous, la question de la base de sondage est donc la suivante.

Comment définir les groupes "d'individus statistiques", c'est à dire les groupes de ménages, qui constitueront les unités d'échantillonnage (ou unités primaires) et pour lesquelles l'image satellite servira de base de sondage? Compte tenu de la résolution au sol actuelle des image SPOT ou THEMATIC MAPPER, la seule unité spatiale repérable sans ambiguïtés est l'îlot physique ou pâté de maisons : dans la plupart des cas un polygone clairement délimité par la voirie qui l'entoure. Rapprochons cette définition de l'unité d'échantillonnage de la définition que donne J. DESABIE des sondages aréolaires¹ :

"L'unité de sondage au dernier degré² est une aire géométrique ou une aire déterminée par des limites naturelles : routes, court d'eau, etc. Toute unité statistique : ménage, personne humaine, point de vente, exploitation agricole, se trouve rattachée de façon univoque à une aire et à une seule. L'aire est une grappe particulière."

Le sondage que nous cherchons à définir sera donc un sondage à deux degrés, aréolaire au premier degré, les unités primaires étant les pâtés de maisons constituant l'agglomération. Au second degré, après avoir recensé les ménages dans chaque unité primaire sélectionnée, on enquêtera un échantillon de ménages dans chaque îlot.

1.1.3 Définition des domaines d'intérêts, stratification

Abordons maintenant les problèmes de la production, à partir d'un sondage, de *données localisées* et de la *réduction de l'erreur d'échantillonnage* tant au niveau global qu'au niveau des estimations concernant chacun des domaines spatiaux d'intérêt.

- Production de données dans des domaines d'intérêt

La gestion urbaine, comme nous l'avons vu, n'a pas seulement besoin de données concernant les caractéristiques moyennes de la population totale de la ville. Il lui faut aussi connaître les différences existant au sein de l'espace urbain pour être informée des spécificités démographiques, économiques, migratoires etc, de tel ou tel quartier. Or d'une part, les différences morphologiques existant dans le tissu urbain, constituent souvent, pour les aménageurs, un critère pertinent de division de l'espace. D'autre part, l'image satellite peut nous renseigner sur les caractéristiques morphologiques des différents types de quartiers composant l'agglomération. On peut,

¹ : J. DESABIE, Théorie et pratique des sondages, pp 117 & 118 (DONOD, 1966, PARIS).

² : Cette restriction n'est pas faite par d'autres auteurs lorsque qu'ils définissent le sondage aréolaire et nous adoptons pour notre part une définition plus large. Pour nous, le caractère aréolaire peut être le fait d'un ou plusieurs degrés d'un sondage complexe et non seulement du dernier comme le dit DESABIE. Nous verrons d'ailleurs au chapitre 2 que la définition de ce type de sondage est également variable sur d'autres points.

à partir de l'information satellitaire, obtenir un découpage de l'agglomération en zones morphologiquement homogènes et il peut être intéressant, du point de vue de la gestion urbaine, que l'enquête permette d'obtenir des informations sur chacune de ces zones. Quelle conséquence aura cette nouvelle contrainte sur le plan de sondage?

L.M. ASSELIN¹ aborde ainsi la question :

"Obtenir de l'information suffisamment précise pour certaines sous-populations est une préoccupation majeure présente dans toute enquête d'envergure. (...) On utilise alors la technique de l'estimation dans un domaine d'étude (...). Mais alors, le nombre d'unités statistiques dans chaque sous-population est complètement aléatoire et, dans certains cas il peut être insuffisant pour fournir des résultats valables. Il est donc préférable d'effectuer la sélection de l'échantillon d'une façon *indépendante* dans chaque sous-population, ce qui n'est possible que si la caractéristique qui définit les sous-populations est connue pour chaque unité d'échantillonnage."

- Réduction de l'erreur d'échantillonnage

Nous avons vu qu'une des hypothèses faites au départ du projet de recherche est qu'il existe, en milieu urbain, certaines relations statistiques entre les caractéristiques morphologiques des quartiers et les caractéristiques socio-économiques des habitants. Nous chercherons bien entendu à exploiter cette liaison pour tenter de réduire l'erreur d'échantillonnage en utilisant l'information qu'apporte l'image satellite sur la morphologie urbaine. Comment cela est-il possible? A nouveau, référons nous à ASSELIN² :

"Une autre préoccupation omniprésente dans le processus d'organisation d'une enquête par sondage, c'est de réduire l'erreur d'échantillonnage globale, par d'autres moyens que celui de l'augmentation de la taille de l'échantillon. (...) Cela est possible à condition de connaître, pour chaque unité d'échantillonnage, une caractéristique définissant des sous-populations qui présentent des particularités importantes relativement à l'enquête. (...) Si on laisse le hasard seul décider de la répartition de l'échantillon entre les sous-populations, il pourra y avoir sous-représentation ou sur-représentation incontrôlable de certaines sous-populations ou encore une répartition inefficace due à l'ignorance des particularités de chaque sous-population. La technique qui s'impose encore ici consiste à sélectionner l'échantillon d'une façon indépendante dans chaque sous-population, en ayant cette fois comme objectif premier, la réduction de l'erreur au niveau de l'ensemble de la population.

Lorsqu'une procédure de sélection est telle que la sélection est effectuée indépendamment dans différentes sous-populations, celles-ci sont appelées strates et le sondage est dit stratifié."

Ainsi, qu'il s'agisse de produire des résultats dans des sous-ensembles particuliers (domaines d'intérêt) de l'univers ou qu'il s'agisse de réduire l'erreur d'échantillonnage à effectif sondé constant, la solution à étudier en priorité est la stratification de la base de sondage d'après des critères de morphologie urbaine. C'est d'autant plus vrai que, comme nous le verrons plus en détail dans la suite, toute l'information nécessaire à cette opération peut être extraite de l'image satellite. Bien sur, la répartition optimale de l'échantillon entre les strates pourra être différente suivant que l'on a comme objectif la production d'estimations précises dans chaque domaine ou la réduction de l'erreur au niveau de l'ensemble de la population. Nous verrons dans la suite que là n'est pas le seul dilemme que nous devons affronter ; le sondage

¹ : L.M ASSELIN, Techniques de sondage avec application à l'Afrique, pp 173 (Gaëtan Morin 1984, QUEBEC).

² : L.M ASSELIN, Techniques de sondage avec application à l'Afrique, pp 173 & 174 (Gaëtan Morin 1984, QUEBEC).

idéal est souvent impraticable et les bons plans de sondage sont alors ceux qui réalisent les meilleurs compromis entre des objectifs contradictoires.

1.1.4 Techniques de sélection des unités aréolaires

Le panorama que nous dressons ici des "contraintes" qui vont peser sur la définition du plan de sondage ne serait pas complet si nous n'évoquions pas une dernière question : quelle doit être, dans chaque strate, la stratégie retenue pour sélectionner l'échantillon d'unités primaires, c'est à dire l'échantillon de pâté de maisons?

Le problème recouvre en fait deux questions distinctes :

- celui de la probabilité d'appartenance des unités aréolaires à l'échantillon ;
- celui de la technique de sélection proprement dite : aléatoire ou systématique.

Du point de vue pratique, la première question a des implications contraignantes. En effet, si l'on envisage de pratiquer une sélection équiprobable des unités primaires, il faut obligatoirement constituer une liste de ces unités, ce qui dans l'état actuel des possibilités d'interprétation de l'image satellite est impossible (nous reviendrons plus loin sur ce point). Plus généralement toute probabilité de sélection proportionnelle à une mesure de taille des unités autre que leur extension spatiale¹ - on pense par exemple au nombre de bâtiments contenus dans l'unité etc - se heurterait au même obstacle. De plus il faudrait alors disposer de cette mesure sur l'ensemble des unités primaires de la base ...bref, la gestion rigoureuse de telles probabilités d'appartenance à l'échantillon conduit, à l'heure actuelle, à des opérations impossibles à réaliser.

Examinons maintenant l'alternative qui existe au tirage sur liste. Il s'agit d'un tirage qu'on peut appeler "spatial" c'est à dire que l'on sélectionne, de manière aléatoire ou systématique, un échantillon de points de l'espace-base de sondage. Les unités primaires sélectionnées sont alors celles qui contiennent les points tirés et la probabilité de sélection d'une unité est proportionnelle à sa surface totale. Dans ce cas, les mesures de surface ne sont requises que pour l'ensemble de l'univers (ou plus exactement pour chacune des strates) et pour les unités primaires de l'échantillon. On les obtient, nous le verrons, soit directement sur l'image satellite, soit sur le terrain.

La solution que nous venons de décrire est toujours beaucoup plus simple à mettre en œuvre que ces concurrentes, d'autant plus si l'on choisit un tirage systématique à l'aide d'une grille. Allons plus loin, c'est pour l'instant la seule réellement praticable dans des villes, encore nombreuses, où n'existe aucune autre source d'information fiable que l'image satellite. C'est pourquoi c'est cette solution qui sera retenue lors de la première expérimentation de la méthode à Quito.

Quant au choix entre tirage aléatoire et tirage systématique, la réflexion qui doit l'orienter porte principalement sur le problème de la réduction de l'erreur d'échantillonnage et de l'estimation de cette erreur. On constatera dans le chapitre suivant que la plupart des spécialistes s'accordent à considérer qu'une sélection systématique bien conçue peut induire une réduction appréciable de l'erreur. Le problème de ce type de sondage réside dans la difficulté que l'on a pour construire un bon estimateur de la variance, c'est à dire une bonne mesure de l'erreur d'échantillonnage.

L'ensemble des considérations exposées dans les paragraphes précédents ne prétend pas être un inventaire exhaustif des possibilités existantes pour mener un sondage à but démographique en milieu urbain. Nous avons simplement rendu compte d'une réflexion collective qui se devait de déboucher sur une solution praticable dans le

¹ : par mesure de l'extension spatiale des unités, on entend une mesure pouvant se ramener à une somme d'éléments de surface identifiables sur le terrain et sur l'image. Nous verrons que pour l'instant les possibilités se réduisent à la mesure de la surface totale.

cadre que se fixe le projet : une ville connaissant un fort accroissement de population et sur laquelle on ne possède pas d'autre base de sondage actualisée que l'image satellite. En suivant une logique principalement déterminée par les contraintes qui surgissent du contexte de la recherche, cette réflexion nous a permis de fixer les caractéristiques générales du sondage que nous mettrons en œuvre. Dans une terminologie un peu barbare on peut le décrire ainsi : un sondage systématique aréolaire stratifié, à plusieurs degrés, où la probabilité de sélection des unités primaires (les pâtés de maisons) sera proportionnelle à leurs surfaces. Mais il est certain que les choix techniques que suppose cette définition ont chacun des conséquences sur l'erreur d'échantillonnage. Ce sont ces conséquences qu'il va s'agir dans ce travail de mesurer.

1.2 Comment évaluer les performances de la technique envisagée?

Le sondage probabiliste présente sur ses concurrents un énorme avantage : si l'on peut dire, il "s'auto-évalue". L'estimation qui sera faite, à l'issue de l'enquête, de la variance des estimateurs constitue une mesure de l'erreur d'échantillonnage et l'expérience que nous avons réalisée à Quito permet donc une évaluation rigoureuse de la précision atteinte. Bien que nous considérons, à juste titre pensons-nous, que seule cette phase d'application "en vraie grandeur" permet de conclure sur la praticabilité et l'efficacité de la méthode, il faut faire à ce sujet plusieurs remarques.

Tout d'abord, considérant le peu d'acquis dans les domaines de la télédétection urbaine et des sondages aréolaires à but démographique, il eut été dangereux, voire suicidaire, de se lancer dans la réalisation de l'enquête finale sans avoir au préalable effectué des évaluations à partir de données existantes. C'est le but de la phase de mise au point méthodologique réalisée sur la ville de Marseille.

D'autre part le procédé systématique de sélection de l'échantillon retenu pour l'enquête finale de Quito, s'il permet d'améliorer la précision des estimations, pose en revanche des problèmes lorsqu'il s'agit d'évaluer cette précision. Diverse méthodes d'approximation existent ; nous reviendrons sur cette question mais disons tout de suite que celle que nous avons été en mesure d'appliquer aux données recueillies à Quito fournit probablement une indication pessimiste sur la variance des estimations.

Enfin, toujours en raison du peu d'expériences accumulées dans le domaine des enquêtes à objectifs multiples par sondage en milieu urbain¹, nous aurons beaucoup de difficultés, sur la seule base des résultats de l'expérience menée à Quito, à comparer du point de vue des performances la méthode que nous avons retenue à d'autres méthodes, différant par le plan de sondage ou utilisant d'autres bases de sondage.

L'évaluation des performances de la méthode à l'issue de son application à Quito nous semble donc indispensable mais, parce qu'elle ne pourra être que partielle, elle est également insuffisante. Nous nous efforcerons dans les chapitres qui vont suivre de la compléter et cela de deux façons.

Dans le chapitre 2, nous tenterons de confronter la solution que nous envisageons, d'une part aux idées exprimées sur la question par les spécialistes des sondages, d'autre part aux expériences déjà tentées dans les domaines de l'estimation de données urbaines utilisant une information sur la morphologie des quartiers et/ou des techniques de sondage.

Puis dans les deuxième et troisième parties, à partir de données existantes dans le cas de Marseille ou de résultats d'enquêtes préliminaires réalisée à Quito, nous essaierons de justifier l'ensemble des choix techniques effectués. Pour cela nous évaluerons les

¹ : Lorsqu'elles existent, ces expériences ne fournissent que rarement des indications rigoureuses sur la précision des résultats obtenus.

consequences de chacun de ces choix sur la précision des estimations et nous comparerons les performances de la méthode retenue avec celles permises par d'autres méthodes. Ils pourra s'agir de méthodes applicables dans les mêmes conditions de site, de données existantes et de thématique de l'enquête (ou dans des conditions voisines), mais aussi de méthodes présentant moins d'intérêt pratique dans le contexte qui est le nôtre mais que nous considérerons comme des références "théoriques" (par exemple le sondage équiprobable sur liste).

Conclusion : l'idéal et le praticable

Au terme de cet exposé de notre problématique générale, nous voudrions finir en situant notre démarche par rapport à un débat, qui n'est d'ailleurs pas spécifique au domaine des sondages, mais qui, bien qu'il soit fréquemment relancé, nous paraît souvent stérile. C'est celui de la place des mathématiques dans la recherche appliquée en sciences sociales. Ce n'est pas aux réflexes d'ostracisme disciplinaire qui alimentent parfois ce type de discussions que nous voulons nous référer ici ; les mathématiciens n'étant d'ailleurs pas toujours exempts de ce genre de comportements corporatistes. Considérons plutôt l'argument proprement dit : schématiquement, le statisticien aurait besoin de données "parfaites" pour formaliser correctement le problème posé et exercer son art. Les mathématiques seraient le domaine du théorique et de l'idéal, les sciences sociales celui du réel et du praticable. Pour ce faire un instant l'avocat du diable, illustrons l'argument grâce à un exemple empreinté précisément à la théorie des sondages.

Lorsque qu'on cherche à améliorer au maximum les estimations sur l'ensemble de l'univers sondé grâce à la stratification de la base de sondage, il se pose le problème de la répartition optimale de l'échantillon aux strates. NEYMAN est l'auteur d'une solution assez générale de ce problème dans laquelle¹ l'effectif sondé (n_h) dans la strate h doit être proportionnel à la quantité $N_h.S_h$ (où N_h est l'effectif de la strate, et S_h la variance du caractère à estimer dans cette strate). Comme le dit J.M. GROSBRAS², "la règle de NEYMAN suppose connue une information "riche" : les N_h et S_h . La connaissance des S_h souvent pose problème." En fait, pour appliquer stricto-sensu la théorie il faudrait, à la limite, connaître à l'avance ce que l'on est censé rechercher ! Faut-il conclure pour autant à son inutilité ?

Le cercle vicieux n'est, dans la plupart des cas, qu'apparent. GROSBRAS poursuit :

"Divers conseils peuvent être donnés :

- beaucoup d'enquêtes sont répétitives, on peut alors *utiliser les S_h estimé dans la dernière enquête connue*.
- souvent on a de bonnes raisons de croire que (...) les coefficients de variation intra- strates sont assez stables. (...) D'où une règle fréquemment employée : on répartit l'échantillon dans les strates proportionnellement à la somme de la variable étudiée, *ou d'une variable qui lui est corrélée*."

Ainsi, c'est bien la théorie, même lorsque sa stricte application est impossible compte tenu des données disponibles, qui oriente le choix du compromis à adopter. Plus généralement, la théorie, cet "idéal" du statisticien, doit servir, face aux situations concrètes proposées par la recherche en sciences sociales, à *concevoir et à évaluer le praticable*. C'est la démarche qui fut la nôtre tout au long de ce travail.

¹ : Il s'agit là de la solution simple dans le cas où il n'y a pas lieu de considérer des coût d'enquêtes variant selon les strates.

² : J.M. GROSBRAS, Méthodes statistiques des sondages, p 82 (Economica 1987, PARIS).

En conclusion, s'agissant d'observation quantitative d'une réalité sociale, il nous semble aussi illusoire de croire que le statisticien peut travailler seul, que de croire qu'un chercheur en sciences sociales peut faire l'économie des théories statistiques pour réaliser une observation fiable.

Chapitre 2

SONDAGES AREOLAIRES ET TELEDETECTION, ETAT DE LA RECHERCHE

Introduction

Dans la littérature consacrée aux sondages, les sondages aréolaires ne font pas l'objet de développements théoriques particuliers. C'est d'ailleurs tout à fait normal puisque leur caractéristique spécifique ne réside que dans le type d'unités statistiques que l'on considère : des aires géographiques. Dans l'absolu, le plan de sondage associé à un tirage d'aires peut être quelconque : à un ou plusieurs degrés, équiprobable ou à probabilité inégale, aléatoire ou systématique, sur liste ou "spatial" etc. Cependant, dans le chapitre précédent, nous avons défini d'une manière déjà beaucoup plus précise l'objet principal de notre recherche. Il s'agit d'un sondage aréolaire systématique stratifié à deux degrés, dont les unités primaires, de tailles inégales, sont sélectionnées avec probabilité proportionnelle à cette taille. Hormis "aréolaire", chacun des termes de cette définition fait l'objet de nombreux développements dans les ouvrages de référence.

Nous ne pouvons ici, étant donné l'ampleur du sujet, effectuer une recherche bibliographique exhaustive. Les ouvrages auxquels nous nous sommes le plus souvent référé tout au long de cette recherche sont au nombre de sept. Trois "classiques", COCHRAN (3^e édition 1977), YATES (4^e édition 1981) et DESABIE (1966), et quatre ouvrages plus récents, ceux de GOURIEROUX (1981), ASSELIN (1984), GROSBRAS (1987) et un ouvrage collectif¹ publié à l'occasion des Journées d'Etude en Statistique d'octobre 1986 qui avaient pour thème les sondages : DROESBEKE, FICHET, TASSI, éditeurs (1987). Dans la première partie de ce chapitre nous tentons, en faisant la synthèse des conclusions théoriques et pratiques auxquelles parviennent ces auteurs, de dégager les principes de l'évaluation de la méthode que nous proposons.

Nous nous intéresserons ensuite aux sondages aréolaires tels qu'ils sont pratiqués en statistique agricole, forestière et minière. C'est en effet dans ce domaine que l'expérience concernant ce type de sondage et les applications de la télédétection, est la plus ancienne et la plus riche. Mais, d'une part, l'unité spatiale enquêtée, souvent appelée segment, est alors généralement une aire de taille fixe, ce qui rend difficilement applicables au milieu urbain les résultats obtenus dans ces recherches. D'autre part, dans ce domaine, les utilisations qui sont faites des techniques de sondage et de l'image satellite pour l'estimation de surfaces se situent, comme nous le verrons, dans une problématique souvent différente de la nôtre. Dans cette partie, nous aborderons également rapidement les recherches récentes concernant l'échantillonnage par points dans le domaine des sondages forestiers et miniers.

¹: Cet ouvrage est constitué d'articles rédigés par divers auteurs. Dans ce qui suit nous nous référerons à ceux de J.C. DEVILLE, A.M. DUSSAIX et C. GOURIEROUX.

Enfin, dans une dernière partie, nous reprenons le bilan, dressé par F. DUREAU¹, des expériences déjà tentées pour estimer les caractéristiques des populations citadines en utilisant la morphologie urbaine.

2.1 Synthèse de sept ouvrages de théorie des sondages

Pour faire le tour de la problématique exposée au chapitre précédent, nous nous intéresserons successivement à huit thèmes, tels qu'ils sont traités dans les sept ouvrages retenus. Nous retiendrons en priorité les approches qui nous semblent adaptées à la situation dans laquelle nous tentons d'appliquer les diverses techniques. Le tableau 2.1 présente une synthèse des références possibles à ces huit thèmes dans les sept ouvrages.

2.1.1 Généralités et bases de sondages

Tous les ouvrages pris en compte débutent bien entendu par un (ou plusieurs) chapitres de généralités qui, après avoir exposé le principe de la méthode, s'attachent en général à en décrire les étapes, puis introduisent la notion de sondage aléatoire et son corollaire, l'erreur d'échantillonnage. Au titre des généralités, notons l'hétérogénéité des notations mathématiques adoptées par les différents auteurs : tous ne consacrent d'ailleurs pas un paragraphe à leur exposé systématique, ce qui ne facilite pas la tâche du lecteur.

C'est en fait dans les paragraphes d'introduction aux chapitres centraux qu'il faut chercher l'exposé des questions pratiques ou théoriques intéressantes. Nous en retiendrons trois, présentes de diverses manières dans tous les ouvrages, qui nous concernent plus particulièrement :

- Le problème de la constitution de la base de sondage

La première définition donnée d'une base de sondage est celle d'une liste "énumérant les individus de la population *sans omission* (ceci est indispensable puisque chacun des individus doit avoir une chance non nulle d'appartenir à l'échantillon) *ni répétition* (cela est préférable si l'on désire assurer facilement l'égalité des chances de sortie pour tous les individus" (DESABIE). Définition toute théorique, qui est en général suivie de l'énumération des défauts que peut présenter une base de sondage. YATES dresse la liste suivante, non limitative à notre sens :

- "(1) Inaccurate
- (2) Incomplete
- (3) Subject to duplication
- (4) Inadequate
- (5) Out of date"

D'une telle liste il ressort qu'il sera généralement très difficile d'obtenir une base de sondage adéquate sous forme de liste d'individus. YATES propose alors une série d'alternatives pour constituer des bases ("frames") d'échantillonnage des populations humaines :

¹ : Ce bilan est inclu dans l'article de F. DUREAU et O.BARBARY à paraître dans "Informatique et Sciences Humaines" : Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale, pp 1 à 6.

- (a) Lists of individuals in the population, or part of it, provided for administrative purpose.
- (b) Aggregates of census returns resulting from a complete census.
- (c) Lists of households or dwellings in given areas.
- (d) Town plans.
- (e) Map of rural areas.
- (f) Lists of towns, villages, and administrative areas, often with supplementary information of various types."

Dans YATES, l'examen de chacune des possibilités, qui suit la liste reproduite ci dessus, conduit dans bien des cas à un sondage à plusieurs degrés ("cluster sampling" ou "multi stage sampling"). Les autres auteurs introduisent également ce type de sondage en partant des difficultés de constitution d'une liste d'individus, et c'est aussi dans cette optique que nous avons choisi un sondage aréolaire.

- Les procédés de sélection de l'échantillon

L'alternative principale est toujours la même : sondage aléatoire ("random sampling") ou sondage systématique ("systematic sampling"). Cependant, alors que les auteurs d'expression française, présentent toujours le sondage systématique à partir du tirage dans une liste et s'en tiennent généralement là, les auteurs anglo-saxons font la distinction entre "systematic samples from list" et "systematic samples from areas" (YATES chap 3, COCHRAN chap 8). Si du point de vue strictement théorique les deux méthodes peuvent ne pas être distinguées, dans la pratique elle conduisent généralement à des plans de sondages très différents. En particulier la sélection par points (qu'elle soit aléatoire ou systématique) conduit à des probabilités d'inclusion proportionnelles à la surface des unités spatiales considérées. Nous reviendrons plus en détail sur cette question dans le paragraphe consacré aux sondages systématiques.

- Le problème de la réduction de l'erreur d'échantillonnage

A nouveau le problème est posé de façon similaire dans l'ensemble des ouvrages. DESABIE, par exemple identifie six méthodes principales d'amélioration de la précision qui n'entraînent pas d'augmentation du coût de l'enquête :

- " 1. choix judicieux des unités de sondage;
2. stratification et tirage systématique;
3. sondages à plusieurs degrés;
4. tirage des unités avec probabilités inégales;
5. utilisation dans les formules d'estimation d'informations obtenues par ailleurs : stratification a posteriori, estimation par le quotient, estimation par la régression;
6. renouvellement par partie de l'échantillon"

Hormis la dernière méthode, qui s'applique au problème de l'échantillonnage répété dans le temps, nous aborderons au cours de cette recherche les cinq autres techniques.

2.1.2 Le sondage équiprobable

Le sondage aléatoire équiprobable ("simple random sampling") est la méthode de base des sondages aléatoires. Dans tous les ouvrages, il sert de référence. GROSBRAS donne deux arguments pour justifier ce fait qui s'appliqueront tous les deux à notre démarche :

- "(...) Si dans la pratique on met en œuvre des plans plus élaborés, les concepts fondamentaux et les formules importantes découlent du sondage élémentaire. Par ailleurs il sert de première référence lorsqu'on cherche à évaluer les qualités d'une méthode quelconque. Il ne suppose, en effet, aucune connaissance a priori des variables à étudier, par conséquent toute autre

technique qui ne garantirait pas une précision théorique au moins aussi bonne à coût égal serait disqualifiée."

Comme nous le verrons, c'est sur le cas simple du sondage équiprobable, que nous mettrons au point, à Marseille, les outils qui serviront à l'évaluation des stratifications et des plans de sondages plus complexes. De plus, tout au long de cette évaluation, à Marseille comme à Quito, nous comparerons les performances des différentes méthodes applicables à celles qu'aurait, à effectif sondé égal, le sondage équiprobable. Cette démarche rejoint celle de KISH lorsqu'il définit le concept d'effet de sondage ("design effect") lié à un plan de sondage P comme le rapport de la variance de la statistique S que l'on cherche à estimer sous le plan P à la variance de S sous le plan aléatoire simple de même taille¹.

2.1.3 Les sondages aréolaires

Comme nous l'avons dit en introduction, les sondages aréolaires sont rarement traités à part dans les exposés généraux sur la théorie. Le plus souvent, ils apparaissent comme des cas particuliers de sondages à deux degrés où les unités primaires sont des aires géographiques. On trouvera donc la question abordée le plus souvent dans les chapitres consacrés aux sondages en grappes ou aux sondages à plusieurs degrés dans les ouvrages en français ("cluster sampling" et "multi stage sampling" en anglais). Dans notre échantillon, deux exceptions méritent pourtant d'être signalées, qui traitent des applications de ce type de sondages à la statistique agricole. On trouve dans YATES (§§ 4.23 à 4.30) des conseils pratiques pour constituer des bases de sondage aréolaire en milieu rural adaptées à différents objectifs : estimation de surfaces en culture, estimation de récoltes et de rendements, ressources forestière etc. Mais surtout, la seconde partie du livre d'ASSELIN est un compte rendu complet et détaillé de la conception et de la mise en œuvre d'une enquête agricole en milieu traditionnel africain qui repose sur un plan de sondage aréolaire. Le but principal est "d'évaluer la production vivrière annuelle des exploitations agricoles familiales" mais la structure de l'enquête (stratification, degrés de sondage, etc.) y est justifiée étape par étape ainsi que la taille et la répartition de l'échantillon ; l'exposé a donc un intérêt méthodologique qui dépasse largement le strict cadre de son domaine d'application.

2.1.4 La stratification

- Un double objectif

DESABIE introduit ainsi l'objectif de la stratification :

"La stratification a pour objet :

- a) Principal : d'améliorer la précision des estimations pour l'ensemble de la population.
- b) Secondaire : d'assurer une précision suffisante pour les estimations relatives à certains sous-ensembles de la population que l'on désire étudier isolément et auxquels on donne alors le nom de domaines d'études."

Ce sont effectivement ces deux objectifs que l'on retrouve chez tous les auteurs, parfois inversés dans leur ordre de priorité : COCHRAN favorise le fait de pouvoir produire des données dans les domaines d'études et de pouvoir gérer indépendamment des techniques de sondage différentes dans les sous-populations. ASSELIN, lui, met les deux objectifs au même rang et nous adoptons cette dernière optique lorsque nous stratifions l'image satellite de la ville sur des critères de morphologie urbaine.

¹ : Nous reprenons ici la définition du concept introduit par KISH, que donne J.C. DEVILLE dans l'ouvrage collectif de DROESBEKE, FICHET et TASSI.

- Une procédure d'inférence très simple

Signalons dès maintenant un résultat concernant les sondages stratifiés, fondamental pour la suite : en vertu d'une part de l'indépendance des tirages dans chaque strate et d'autre part de la linéarité des opérateurs "espérance" et "variance" lorsqu'ils s'appliquent à des sommes de variables aléatoires indépendantes, les estimateurs du total général et de la variance de son estimation s'obtiennent par sommation des estimateurs respectifs dans chaque strate¹.

- Le gain de précision sur l'estimation du total général

C'est toute la question de la pertinence des critères de stratification et de leur plus ou moins forte corrélation avec les caractères que l'on se propose d'estimer grâce à l'enquête. Toute une partie des évaluations qui vont suivre s'attache précisément à la mesure de ce gain dans le cas de la stratification d'une base de sondage urbaine selon des critères morphologiques. Citons cependant DESABIE, qui justifie ainsi des tentatives de stratification même "approximatives"

" On peut démontrer (...) que si les unités de l'univers sont réparties au hasard entre les strates, un échantillon prélevé par tirage exhaustif avec un taux de sondage uniforme dans l'univers ainsi stratifié aura une efficacité exactement égale à celle de l'échantillon élémentaire de même effectif. C'est dire qu'une stratification, même très imparfaite, est préférable à une absence totale de stratification."

Cette vue empirique du problème peut être complétée par l'approche fondée sur les équations d'analyse de la variance ou, si l'on préfère, la décomposition de la variance totale en variance inter-strate et variance intra-strate². C'est sur cette approche qu'est fondée la tentative de prédiction du gain apportée par la stratification morphologique synthétique réalisée à Marseille qui est présentée dans l'annexe A.2.2.

-Le problème de la répartition de l'échantillon entre les strates

Comme il a été dit au chapitre précédent, la théorie de la répartition optimale de l'échantillon entre les strates est l'œuvre de NEYMAN. On la trouvera exposée dans tous les ouvrages considérés ici mais nous avons vu qu'elle ne peut que rarement s'appliquer *strictement* dans les situations concrètes. C'est en particulier le cas lorsque l'enquête poursuit des objectifs d'estimations multiples et nous nous trouvons précisément dans cette situation. Se pose alors le problème du compromis à effectuer entre les allocations qui seraient optimales pour chacune des variables d'intérêt. En supposant disponible une information sur les variances intra-strates des différentes variables, YATES conclue de manière optimiste :

"When several quantities are under investigation, it will usually be found that their within-strata standard deviation are not in quite the same proportion. This, however, is not a serious problem in practice, since any sampling fractions which is somewhere near the optimal will give results which are nearly as accurate as those given by the optimal fraction. Consequently there is usually no great difficulty in choosing suitable sampling fractions which will reconcile the various conflicting requirements."

Une démarche semblable, à partir des données recueillies lors des enquêtes préliminaires faites à Quito, nous permettra en effet de définir une allocation aux strates "intéressante", du moins s'agissant de l'estimation de plusieurs totaux.

¹ : Voir par exemple ASSELIN (1984) pp 174-175.

² : Voir par exemple GROSBRAS (1987) § 4.1.3, pp 71-73.

2.1.5 Les sondages à plusieurs degrés

- Au départ, deux considérations pratiques

Deux arguments sont constamment avancés en faveur des sondages à plusieurs degrés :

1. la difficulté qu'il y a dans la plupart des situations d'application à constituer une liste d'individus statistiques ;
2. les problèmes de budget et de temps de transport liés à la dispersion géographique de l'échantillon dans le sondage à un seul degré.

Dans notre situation, c'est le premier argument qui est déterminant ; en revanche on conçoit que pour des enquêtes en milieu rural le deuxième ait également une grande importance. Le sondage à deux degrés solutionne le problème de la liste d'individus statistiques : il ne réclame cette liste qu'au sein des unités primaires sélectionnées ; la constitution de ces listes est une opération qui peut facilement se réaliser sur le terrain dans une première phase de l'enquête. Le prix à payer est une baisse de la précision par rapport au sondage de même taille dans une liste d'unités statistiques (COCHRAN p 233, YATES p 32, ASSELIN p 105, GOURIEROUX p 84, DESABIE pp 261-263, etc.).

- Un moyen terme entre le sondage à un seul degré, impraticable, et le sondage en grappe, imprécis

Si l'on veut pousser à l'extrême les deux avantages du sondage à deux degrés pour avoir un travail minimum de constitution de listes et une concentration géographique maximum de l'échantillon, on aboutit au sondage aréolaire en grappe. Pour qu'un tel sondage soit efficace, GOURIEROUX (p 114) dit :

"les grappes doivent être hétérogènes pour le caractère Y et les probabilités d'inclusion proportionnelles à la taille des grappes".

La deuxième condition apparaît immédiatement très difficile à réaliser lorsque la seule information disponible est l'image satellite. Nous reviendrons plus loin sur cette question. Quant à la première condition, s'agissant du milieu urbain, YATES (p 72) part d'une constatation qui nous semble de bon sens :

"A sample of dwellings can (...) be obtained by including all the dwellings in the selected blocks. In general, however, the variability between block and block is likely to be large even after careful stratification, since there is often considerable local segregation of different classes of the population. Consequently, two-stage sampling is in general advantageous, blocks being taken as the first-stage units and dwellings as the second-stage units."

Ainsi, même s'il existe une certaine hétérogénéité des ménages au sein d'un même îlot, cette variabilité à toute chance d'être faible comparée à celle existant entre les caractéristiques moyennes de la population de différents îlots. Nous serons effectivement amenés à cette conclusion lors de l'étude des résultats des enquêtes préliminaires effectuées à Quito. Le problème est alors de fixer les taux de sondage aux premier et second degré.

- Inférence, les variances associées aux différents degrés du sondage et leur estimation

Pour asseoir sur des bases solides le choix des taux de sondage, il faut connaître les parts de l'erreur d'échantillonnage globale dues respectivement à la variabilité inter-unités primaires (premier degré) et à la variabilité intra-unités primaires (second degré). Le problème est abordé de manière semblable par tous les auteurs, par exemple : DESABIE (p 264), ASSELIN (p 108 et suivantes), GROSBRAS (p 177 et suivantes). Les expressions des estimateurs et de leur variances peuvent être assez compliquées du fait des sommes multiples correspondant aux différents degrés du sondage. Il existe pourtant des règles simples pour construire ces quantités. ASSELIN les exprime très clairement et c'est à son livre que nous nous référerons le plus

souvent par la suite concernant ce sujet : ASSELIN, p 108 (proposition fondamentale pour la construction des estimateurs), p 109 (proposition fondamentale pour l'estimation des variances) et pages suivantes pour l'étude de cas particuliers de plans de sondages et l'estimation séparée de chacun des termes de la variance.

2.1.6 Les sondages à probabilités inégales

- Le problème

Dans les ouvrages qui nous intéressent, le problème des sondages à probabilités inégales est abordé de deux manières différentes.

Les auteurs de langue française introduisent généralement les probabilités inégales d'appartenance à l'échantillon comme une manière d'améliorer la précision du sondage en utilisant une information exogène dans l'étape de sélection. DESABIE, par exemple (p 226), l'exprime ainsi :

"Lorsque les unités sont de 'tailles' très inégales : établissements industriels, exploitations agricoles, etc., le tirage des unités avec des chances approximativement proportionnelles à leur taille permet de réaliser un important gain de précision."

ASSELIN (p 73) et GROSBRAS (p 43) ont la même optique.

L'école anglo-saxonne adopte, comme souvent, un point de vue plus pratique en présentant les probabilités d'appartenance inégales comme une conséquence des caractéristiques de la base de sondage et du mode de sélection retenu. Par exemple YATES (p 33) dit :

"If we have areas demarcated on a map, such as fields, and a point is located at random on the map, the probabilities of the point falling within the boundaries of the different fields are clearly proportional to the areas of the fields. Consequently areas can be selected at random with probabilities proportional to their size by the simple procedure of taking random points on the map."

COCHRAN (p 249) souligne simplement la fréquence des situations où l'échantillonnage a lieu dans un ensemble d'unités primaires ("cluster units") contenant des nombres inégaux d'unités secondaires ("subunits").

Pour notre part, nous envisagerons le problème sous ses deux angles. D'une part, nous voulons éviter d'avoir à constituer une liste d'ilots (unités primaires) et la technique de sélection pressentie nous place dans la situation décrite par YATES. D'autre part nous examinerons les différentes possibilités de définition d'une mesure de la taille de nos unités spatiales et nous nous poserons la question suivante : dans l'hypothèse où la gestion de probabilités proportionnelles à ces mesures est possible, laquelle d'entre elles fournit la meilleur précision des estimations?

- Inférence, les estimateurs d'Horvitz-Thompson

Dans le cas du sondage à probabilités inégales, la théorie distingue les plans avec remise (ou indépendants) des plans sans remise (ou exhaustifs). Les estimateurs (de totaux, moyennes, quotients) sont les mêmes dans les deux cas mais les expressions des variances diffèrent du fait que dans le cas du sondage exhaustif les probabilités d'appartenance se modifient au fur et à mesure du tirage. La théorie générale de ce mode de tirage a été faite en 1952 par HORVITZ et THOMPSON ; on la trouve exposée dans tous les ouvrages cités (à l'exception de YATES). En fait, les travaux d'HORVITZ et THOMPSON ne s'appliquent pas qu'au cas des probabilités inégales. Il s'agit d'une théorie générale de l'estimation sans biais dans le cas de l'échantillonnage exhaustif, quelques soient les probabilités d'appartenance (GOURIEROUX Chapitre 3). Si cette approche est satisfaisante théoriquement, elle s'avère sans grandes conséquences pratiques, du moins lorsque le taux est petit et la population-mère importante. Dans ce cadre, qui sera toujours le nôtre, on peut, comme nous le verrons, pratiquer le tirage

exhaustif, plus pratique et plus précis, tout en appliquant au problème de l'estimation de la variance, le formulaire plus simple du tirage indépendant ; la surestimation de la variance qui en résulte reste minime.

- Comparaison avec d'autres plans, choix de la mesure définissant la probabilité

Comme nous l'avons dit au début de ce paragraphe, il nous importe aussi d'évaluer comment l'utilisation d'un plan à probabilités inégales peut faire diminuer la variance des estimations.

Pour envisager plus concrètement la question, appelons Y le caractère à estimer et Z la mesure de taille des unités d'échantillonnage. En se basant sur les expressions des variances des estimateurs, il est facile de montrer, comme le fait ASSELIN (p 79), que le sondage à probabilité proportionnelle à Z est plus précis que le sondage équiprobable "si et seulement si il existe une corrélation positive entre les variables X et Y^2/X ".

Dans la situation qui nous intéresse, Y est par exemple la population totale P, et Z la mesure de taille des pâtés de maisons. Considérons tout d'abord la mesure de taille la plus naturelle et pour l'instant la seule praticable sur document satellitaire : la surface totale S du pâté de maisons. Pour bien faire ressortir toute l'importance que peut avoir le choix de la probabilité d'appartenance, nous allons faire un hypothèse tout à fait irréaliste qui est celle d'une densité de population $d=Y/S$ constante sur l'ensemble de la ville. Alors, comme Y vaut $S \times d$, la variable Y^2/X vaut $d^2 \times S$ et, comme d est constante, le coefficient de corrélation défini par ASSELIN (entre les variables S et $d^2.S$) vaut 1. Dans cette situation idéale, la variance de l'estimateur est nulle et il n'y a donc plus, dans le cas du sondage avec probabilité proportionnelle à la surface des pâtés de maisons, d'erreur d'échantillonnage.

Evidemment, les choses ne sont pas aussi simples, et même en supposant que l'on dispose d'une stratification de l'agglomération en quartiers à la morphologie très homogène, l'hypothèse d'une densité constante dans chaque strate ne tiendrait pas. Cela dit nous verrons qu'avec la stratification que nous appliquerons, l'homogénéité de la densité au sein de chaque strate sera suffisante pour que le sondage avec probabilité proportionnelle à la surface soit au moins aussi précis que le sondage équiprobable.

Ce qui précède montre que dans notre cas, définir une probabilité d'appartenance optimale pour l'estimation du total de population se ramène à trouver une mesure de la taille des pâtés de maisons la plus corrélée possible à l'effectif de population. Le problème est ensuite de savoir si on peut mettre en œuvre une stratégie de sélection des unités primaires qui respecte les probabilités d'appartenance définies par cette mesure de taille.

2.1.7 Utilisation d'informations exogènes dans l'estimateur

Pour poser le problème, nous résumerons ici l'exemple donné par GROSBRAS en introduction du chapitre qu'il consacre à l'estimation par le quotient¹ :

"La méthode d'estimation par le quotient peut aussi être appelée *redressement par variables quantitatives*

Pour voir comment elle fonctionne, imaginons le cas (fictif) d'un échantillon d'établissements industriels d'un secteur donné, réalisant un certain produit. (...)

On peut penser que le volume de la production est fortement lié à la consommation du courant industriel. La consommation moyenne, dans le mois,

des établissements de l'échantillon est $\bar{x} = 3\ 225$. Or on sait par une autre source que la consommation moyenne de tous les établissements du secteur considéré

est $\bar{X} = 3\ 350$. On constate donc, pour l'échantillon, une légère sous-représentation de la consommation de courant industriel. On en déduit qu'il

¹ : voir J.M. GROSBRAS, p 129 à 131.

existe vraisemblablement une distorsion du même ordre pour la variable production. On va donc redresser les résultats pour les caler sur la moyenne théorique 3 350."

Dans l'exemple cité, l'estimation du total de la production sera redressée, c'est à dire multipliée, par le facteur $3\ 350/3\ 225$, ce qui représente une correction de + 3,4 %. Comme le montre cet exemple, l'estimation par le quotient nécessite la connaissance d'une information supplémentaire concernant la totalité de l'univers qui ne peut bien entendu pas provenir de l'enquête. A propos de cette information, GROSBRAS fait la remarque suivante :

"Il est crucial que les données entrant dans \bar{x} soient bien dans le même champ que celles entrant dans \bar{X} (...). En d'autres termes, la source à laquelle on se réfère en prenant \bar{X} est bien compatible avec l'échantillon : il s'agit de la même population. (...) Par exemple, il faudra avant tout que la statistique \bar{X} des consommations moyennes de courant industriel concerne bien les établissements du même secteur que ceux de l'échantillon, mesuré pour la même période, etc."

Cette condition indispensable, constitue un obstacle difficilement surmontable à l'utilisation de ces techniques dans les conditions qui sont les nôtres ; les données exogènes disponibles dans les villes où nous voulons appliquer la méthode sont, rappelons le, rares, obsolètes ou partielles.

Cependant, le redressement à l'aide d'informations exogènes quantitatives n'est pas le seul intérêt de la théorie de l'estimation par le ratio. Citons à nouveau GROSBRAS :

"Remarquons enfin qu'une des difficultés techniques de l'estimateur par le quotient réside dans la présence de la variable aléatoire \bar{x} au dénominateur. Les méthodes que nous allons employer pourront également rendre service dans les cas où les phénomènes étudiés sont caractérisés par des quantités avec aléas au dénominateur.

Exemple :

Consommation moyenne d'un bien par ménage d'employé, le nombre de tels ménages dans l'échantillon étant aléatoire."

Comme on s'en rendra vite compte, la plupart des taux intéressant le démographe, le géographe ou l'urbaniste entre dans la catégorie décrite par GROSBRAS. En effet, étant donné le plan de sondage que nous appliquerons, le nombre de ménages, le nombre de migrants, le nombre d'actifs etc. observé sur l'échantillon, quelque soit d'ailleurs la définition que l'on retient de ces catégories, seront des quantités aléatoires.

Nous ferons donc un usage intensif des résultats provenant de la théorie de l'estimation par le quotient.

2.1.8 Les sondages systématiques

- La technique de sélection

Comme nous l'avons suggéré dans le paragraphe concernant les généralités et les bases de sondage, il nous paraît opportun, dans notre cadre d'application, de distinguer sondage systématique sur liste et sondage systématique spatial.

La première technique, dans sa version la plus simple (tirage systématique équiprobable), consiste à sélectionner, après avoir défini un pas entier P correspondant au taux voulu, un individu tous les P dans la liste (voir par exemple YATES p. 27). D'autres algorithmes de sélection dans une liste permettent de gérer des probabilités d'appartenance à l'échantillon proportionnelles à une variable de contrôle (voir par exemple GROSBRAS pp. 53-54). Ces techniques ont pour nous, étant donnée l'hypothèse faite d'absence d'une liste, peu d'intérêt pratique.

La technique du sondage systématique spatial ("systematic sampling in two dimensions") est présentée par COCHRAN (p. 227) dans les termes suivants :

"In sampling an area, the simplest extension of the one-dimensional systematic sample is the 'square grid' patern (...). The sample is completely determined by the choice of a pair of random numbers to fix the coordinates of the upper left unit".

Comme dans le cas du sondage aléatoire par points décrit par YATES, la probabilité de sélection d'une unité spatiale est ici proportionnelle à sa surface. C'est cette technique qui va nous intéresser tout au long de ce travail ; mais, comme nous allons le voir maintenant, le problème de son efficacité et de l'estimation de l'erreur d'échantillonnage qu'elle entraîne se pose dans les même termes que pour les sondages systématiques sur liste.

- Efficacité du sondage systématique

Dans la littérature, le problème de l'évaluation des performances des sondages systématiques est généralement posé dans les termes suivants. DESABIE dit par exemple :

" a) Le sondage systématique est rigoureusement assimilable à un sondage aléatoire réalisé par la méthode élémentaire, et par conséquent de même efficacité que ce sondage aléatoire lorsque le rangement des unités dans le fichier peut être considéré comme aléatoire.

b) Le sondage systématique est plus efficace que le sondage réalisé par la méthode élémentaire lorsque, comme ce sera généralement le cas, les unités de la population qui occupent des rangs voisins ont tendance à se ressembler. Ainsi, le fichier des feuilles de logement d'un recensement sera très souvent, à l'intérieur de chaque ville classé par quartier, par rue, un tirage systématique assurera donc une répartition géographique très satisfaisante de l'échantillon. Dans la mesure où beaucoup de caractères socio-économiques sont en relation avec le domicile, on conçoit que le recours au tirage systématique ait pour effet de diminuer la variance des estimations. Le gain de précision peut être très notable."

On noterait la même approche chez GROSBRAS (pp. 28, 29) et chez les autres auteurs retenus ici. DEVILLE et GROSBRAS¹ en résumant ainsi l'idée directrice (en notant Y le caractère à estimer) :

"(...) Le fond de ce raisonnement est le suivant : si Y_i possède une tendance (dans la liste), l'indice i lui même 'explique bien' Y et peut donc servir de variable de stratification. *Le tirage systématique permet de bénéficier d'une stratification implicite.*"

Mais comment estimer d'après les résultats de l'enquête dans quelle mesure la précision des estimations réalisées a bénéficié de cette stratification implicite ?

- Estimation de la variance dans le cas du sondage systématique

La difficulté d'estimer la variance des estimateurs associés à des plans de sondages systématiques vient du problème suivant : le sondage systématique est en fait un cas

¹ : Dans l'article sur les algorithmes de tirage rédigé par ces auteurs pour l'ouvrage collectif de DROESBEKE, FICHET et TASSI.

particulier de sondage en grappe mais dans lequel, comme le montre la figure ci dessous, on ne sélectionne qu'une seule grappe (le pas du sondage est noté ici k).

Grappes	Compositions des échantillons				
1	1	1+k	1+2k	...	1+ (n-1)k
⋮					
a	a	a+k	a+2k	-	a+ (n-1)k ← échantillon
⋮					
k	k	2k	3k	...	nk

La variance des estimateurs dépendant de la variance inter-grappe et le sondage n'en ayant sélectionné qu'une seule, on se trouve dans l'impossibilité d'estimer sa variance. D'après l'inventaire que nous avons effectué dans les ouvrages présentés ici, il existe à notre sens trois approches susceptibles d'apporter une solution au problème. Nous ne ferons ici que les citer, celle que nous utiliserons par la suite sera détaillée dans le chapitre 8 ; pour les autres, le lecteur intéressé peut se reporter aux ouvrages de référence en s'aidant pour cela du tableau 2.1.

1 : Approche par un sondage stratifié

Comme le suggère l'idée de stratification implicite exprimée par DEVILLE et GROBRAS, une "astuce" consiste à considérer l'échantillon issu d'un sondage systématique comme le produit d'un tirage stratifié où les strates regroupent les individus proches dans la liste. Dans le cas d'une sélection sur liste, on assimile alors le sondage systématique à un tirage stratifié de paires. Cette méthode est exposée dans la totalité des ouvrages considérés ici (voir tableau 2.1). Au chapitre 8, en nous référant à DESABIE, nous en ferons une adaptation simple au cas du tirage à l'aide d'une grille de points en assimilant ce dernier à un tirage stratifié sur un critère de proximité spatiale.

2 : Approche "modèles", $y_i = f(i)$

CHOCRAN (pp. 223 à 226) recense diverses tentatives (OSBORNE 1942, JOHNSON 1943, COCHRAN 1946, MATERN 1947 et YATES 1949) d'estimation de la variance de sondages systématiques fondées sur différentes hypothèses concernant le type de liaison existant entre la variable à estimer et l'indice des individus dans la liste.

3 : Approche par les méthode robustes d'estimation de la variance

Plus intéressantes à nos yeux que les précédentes sont les diverses méthodes d'estimation de la variance dites robustes c'est à dire permettant le calcul d'une approximation de la variance même lorsqu'on ignore les lois de probabilité des estimateurs. Citons les méthodes BRR (Balanced Repeated Replication), la méthode du JACKKNIFE et celle du BOOTSTRAP, la plus récente et la plus prometteuse. Nous n'avons malheureusement pas pu les appliquer, ne disposant pas des programmes correspondants.

2.1.9 Conclusion

Dans les sept ouvrages que nous avons retenus, les parts faites à la pratique et à la théorie sont très variables. Le livre de YATES, destiné avant tout aux praticiens, ne contient quasiment aucun développement mathématique, en revanche il fourmille d'exemples concrets qui couvrent l'ensemble des domaines d'application. Les livres de COCHRAN, DESABIE, ASSELIN et GROBRAS peuvent être considérés comme de bonnes

références techniques ; l'essentiel des démonstrations y figure et l'exposé est très clair. Les trois premiers fournissent également un grand nombre d'exemples d'applications, spécialement celui d'ASSELIN qui est, à notre connaissance, le seul ouvrage général où les procédures de sondage proposées sont évaluées de manière systématique à partir de données exhaustives. Cette démarche nous intéresse tout spécialement pour la phase marseillaise de l'expérience. Les deux derniers livres, GOURIEROUX et DROESBEKE, FICHET, TASSI ont une orientation beaucoup plus théorique, en particulier le deuxième qui fait une grosse part aux modèles et ne peut être abordé sans des connaissances très solides en statistique inférencielle. Ils permettent de faire le point sur les développements récents de la théorie mais ne contiennent que très peu d'exemples d'applications.

Tableau 2.1 : synthèse de sept ouvrages de référence consacrés aux sondages

Auteurs :	DESABIE	COCHRAN	YATES	GOURIEROUX	ASSELIN	GROSBRAS	DROESBEKE, FICHET TASSI & Coll
Ouvrages :	Théorie et pratique des sondages (1966)	Sampling techniques (1977)	Sampling methods for censuses & surveys (1981)	Théorie des sondages (1981)	Techniques de sondage avec applications à l'Afrique (1984)	Méthodes statistiques des sondages (1987)	Les sondages (1987)
Thèmes :							
I : Généralités, bases de sondage	Ch 1, 2 & 5	Ch 1	Ch 1, 2, & 3 (Pbls. gén.) Ch 4 (Base de sondage) Ch 5 (Exécution)	Ch I	Ch 1	Ch 1	§ 2.1, 2.2, 2.3 & 2.4
II : Sondages équiprobables	Ch 4	Ch 2 & 3	§ 3.2, § 6.1 à 6.4, § 7.1 à 7.5	§ II.D § III.A	Ch 2	Ch 2	§ 3.2
III : Sondages aréolaires		Pas d'exposé spécifique, voir :					
1 Probls. généraux.	pp 117-119	systematic	§ 3.14 & 4.9		p 18	§ 1.2.1	pp 157-159
2 Statistique. agric.		sampling, cluster sampling & § 8.13	§ 4.14 & 4.23 à 4.30		Toute la 2 ^e partie, étude de cas	Ch 9 ann. 3	
3 Milieu urbain			§ 4.13			p 159	
IV : Stratification	Ch 7	Ch 5 & 5A	Pas de Chap. spécifique		§ III.D	Ch 5	Ch 4§3.3
1 Objectifs, Problèmes généraux	pp 134-142	§ 5.1,5A.5 5A.6,5A.10 5A.13,5A.15	§ 3.3, 3.4	p 99	§ 5.1 & 5.4	§ 4.I	§ 3.3.1
2 Inférence	pp 142-144	§ 5.2 à 5.4, 5.10, 5.12 5A.12	§ 6.5 à 6.7 estimateurs § 7.6 & 7.6.1 variances § 8.3 à 8.5	pp 99, 100	§ 5.1.2 & 5.2	§ 4.II	§ 3.3.2
3 Gain de précision	pp 141, 161 à 166	§ 5.6, 5.7, 5.11, 5A.11			§ 5.2.1	§ 4.IV-3 & 4.IV-5	
4 Répartition de l'échantillon entre les strates, allocation optimale	pp 144-151	§ 5.5, 5.8, 5A.1, 5A.3, 5A.4	§ 3.5, 7.6, 8.5	pp 101-104	§ 5.3 & 6.4	§ 4.III & 4.IV	§ 3.3.4
5 Découpage optimal de l'univers en strates	pp 151-157	§ 5A.2, 5A.7	§ 8.15 5A.8	pp 104-107		§ 4.V & 4.VI	

Tableau 2.1 : synthèse de sept ouvrages de référence consacrés aux sondages(suite)

Auteurs :	DESABIE	COCHRAN	YATES	GOURIEROUX	ASSELIN	GROSBRAS	DROESBEKE, FICHET TASSI & Coll
Ouvrages :	Théorie et pratique des sondages (1966)	Sampling techniques censuses & surveys (1977)	Sampling methods for avec applications à l'Afrique (1981)	Théorie des sondages des sondages (1981)	Techniques de sondage (1984)	Méthodes statistiques (1987)	Les sondages (1987)
Thèmes :							
V : Sondages à plusieurs degrés	Ch 14 à 18	Ch 9, 9A, 10 11	Pas de chap spécifique	§ III.B, III.C, III.E, III.F	Ch 4	Ch 8	Ch 5
1 Objectifs, problèmes généraux	§ 14.1 à 14.4 § 15.1 Ch 17, § 18.1	§ 9.1 à 9.3, 9A.1, 9A.2, 10.1, 11.1	§ 3.8	§ III.B	§ 4.1.1 & 4.6	§ 8.I	§ 5.1 à 5.3, 5.6.1
2 Inférence, comparaison à d'autres techniques	§ 14.5, 14.6 § 15.2 à 15.9 CH 16,17,18	§ 9.4 à 9.7, 9A.3 à 9A.13 10.2 à 10.10 11.2 à 11.21	6.18, 6.19 estimateurs 7.17, 8.10 variances	§ III.C, III.F grappes § III.F plusieurs d'	§ 4.1.2, § 4.2 à 4.5	§ 8.II à 8.IV	§ 5.6.2 à 5.6.6
3 Estimation des différentes composantes de la variance	pp 287, 288, 300 à 304, 368, 376, 380 à 382	§ 10.4 à 10.7 § 11.3, 11.6	§ 8.10, 8.10.1, 8.11	pp 96 à 98 p 120	§ 4.2.4, 4.3.4		§ 5.6.3, 5.6.4
VI : Sondages à probabilité inégales	Ch 11 & 12	Ch 9A & 11	pas de chap spécifique	§ II.C, II.D, § III.A	Ch 3	Ch 3	pas de chap spécifique
1 Objectifs, problèmes généraux	p 226	§ 9A.1 à 9A.3, 11.1	§ 3.9, 3.10	§ II.C modèle stat. paramétriq.	p 73 § 3.1.1, 3.2.1	§ 3.I	§ 5.4
2 Inférence	pp 227 à 239	Ch 9A & 11	§ 6.16, 6.17 estimateurs § 7.15, 7.16 variances	§ II.D, III.A	§ 3.1.2 & 3.2.2	§ 3.II, 3.III, 8.III & 8.IV	§ 3.1 & 5.4
Avec remise	pp 227 à 229	§ 9A.3, 9A.4 § 11.9, 11.14	§ 6.16, 6.17, 7.15, 7.16	§ II.D	§ 3.1.2	§ 3.II, 8.III	§ 5.4 (p96)
Sans remise	pp 230 à 239	§ 9A.6 à 9A.13, 11.10		§ III.A	§ 3.2.2	§ 3.III, 8.IV	§ 3.1
3 Comparaison avec d'autres plans, définition des probabilités	Ch 12	§ 9A.4, 9A.5 11.4 & 11.14	§ 8.9 & 8.9.1	p 80	§ 3.1.3 & 3.3 Tableau pp 102 & 103	§ 3.II.5 & § 3.II.6 à 3.II.10	§ 5.4, pp 96 à 99

Tableau 2.1 : synthèse de sept ouvrages de référence consacrés aux sondages (fin)

Auteurs :	DESABIE	COCHRAN	YATES	GOURIEROUX	ASSELIN	GROSBRAS	DROESBEKE, FICHET TASSI & Coll
Ouvrages :	Théorie et pratique des sondages (1966)	Sampling techniques (1977)	Sampling methods for censuses & surveys (1981)	Théorie des sondages (1981)	Techniques de sondage avec applications à l'Afrique (1984)	Méthodes statistiques des sondages (1987)	Les sondages (1987)
Thèmes :							
VII : Utilisation d'information exogène dans l'estimateur	Ch 8, 9 & 10	Ch 6 & 7	pas de chap spécifique pbs. gen.: § 6.8, 8.6 & 8.6.1	Ch IV, approche bayésienne	§ 6.1	Chap 6	pas de chap spécifique
1 Estimateur par le ratio	Ch 9	Ch 6	§ 6.9 à 6.11 estimateurs § 7.8 à 7.10 variances	§ IV.C	§ 6.1.1 & 6.1.2	Ch 6	
2 Estimateur par la régression ou par la différence	Ch 10	Ch 7	§ 6.12 à 6.15	§ IV.B			§ 3.5 & 4.3.1
VIII : Sondages systématiques	pas de chap spécifique	Ch 8	pas de chap spécifique	§ III.E.3 pp 114-118	§ 2.2	§ 2.XIII	§ 10.4.1 à 10.4.3
1 Techniques et problèmes généraux	§ 5.2.1.2	§ 8.1, 8.2, 8.4, 8.14	§ 3.6, 3.14 & 6.20	pp 114 à 116	§ 2.2.1 & 2.2.2	§ 2.XIII	§ 10.4.1
2 Estimation de la variance							
Approche par stratification	pp 171-172 pp 323-325	§ 5A.12 § 8.2 (p 120)	§ 7.18	pp 117-118	§ 2.2.3 § 6.5.2	§ 2.XIII § 3.III.8	§ 10.4.2
Approche "modèle" $y_j = f(i)$		§ 8.5 à 8.11					§ 10.4.3
Approche méthodes robustes		§ 11.18 à 11.21				Annexe	Ch 7
IX : Exercices	non	oui	non	oui	oui	oui	non
X : Exemples de sondages réels	oui	oui	nombreux	non	nombreux	non	oui
XI : Calculs de variance d'estimation sur des pop. réelles	non	non	non	non	nombreux	non	non

2.2 Les sondages dans le domaine des statistiques agricoles, forestières et minières

Selon HOUSEMAN¹, aux Etats-Unis, les premières expériences d'application des sondages aréolaires probabilistes aux statistiques agricoles ont été faites dans l'état d'Iowa en 1938 et 1939. Si les américains furent pionniers en la matière, la méthode a connu, dès les années 40, une rapide extension dans divers pays (Angleterre, France, Inde, Norvège, Suède etc.). Pendant une longue période, les méthodes de sondage employées en statistiques agricoles et forestières ne se distingueront pas fondamentalement des méthodes appliquées à d'autres domaines. Cependant, dès le départ, leur objet central, l'estimation de surfaces réparties dans différentes catégories d'utilisation du sol, apparaît comme spécifique et en conséquence, la réflexion porte sur des concepts nouveaux : échantillonnage par points, segments, placettes etc.

Dans les années 60, la réflexion sur l'échantillonnage par points s'oriente vers l'étude de la corrélation spatiale existant entre les données à échantillonner (voir par exemple MATERN (1960)). Ces recherches débouchent dans les années 70 sur l'application de la théorie des processus aléatoires au problème de l'estimation dans les procédures de sondages spatiaux. La théorie des variables régionalisée, développée par MATHERON en 1971, et l'étude des correlogrammes spatiaux font, depuis lors, l'objet d'applications intéressantes dans les domaines des statistiques minières, forestières et agricoles.

Enfin, l'apparition de la télédétection spatiale bouleverse depuis une dizaine d'années le paysage des statistiques agricoles, forestières et d'occupation du sol en général.

La littérature concernant les différents thèmes que nous venons de citer n'est pas très abondante, néanmoins nous n'en ferons pas ici le recensement ; pour ne pas alourdir l'exposé, il nous semble suffisant d'aborder ces thèmes, qui ne sont pas directement liés à notre problématique, aux travers d'un échantillon d'ouvrages et d'articles qui posent les principaux problèmes.

2.2.1 Les sondages aréolaires en agriculture

L'ouvrage de Earl HOUSEMAN, auquel nous faisons référence au paragraphe précédent, est à notre connaissance la référence la plus complète concernant les sondages aréolaires en agriculture. Il permet de cerner la problématique de ce type d'application.

Comme nous le disions, l'objectif central est d'estimer l'extension de la culture i , à l'intérieur de l'aire d'étude S . L'ensemble de l'aire d'étude est alors divisée en N segments, en général de taille fixe s ($s = S/N$), définis par HOUSEMAN comme des portions de territoire aux limites bien définies sur le document servant de base de sondage (carte, photographies aériennes etc.). On prélève ensuite de manière aléatoire ou systématique, un échantillon E de m segments pour lequel l'enquête fournit les données x_{ij} qui sont les surfaces occupées par la culture i dans chaque segment j appartenant à E . L'estimateur de l'extension de la culture dans toute l'aire d'étude est obtenu par "expansion directe" :

$$\hat{X}_i = \frac{N}{m} \sum_{j=1}^m x_{ij}$$

Comme on le voit, la méthode s'inscrit dans le cadre simple du plan de sondage équiprobable sans remise. Le caractère spécifique de l'étude de HOUSEMAN a trait au problème de la définition du segment, c'est à dire de l'unité spatiale d'échantillonnage. Comment rattacher sans ambiguïtés aux différents segments les unités d'observations, en général des exploitations agricoles, lorsque l'enquête n'a pas pour objet unique

¹ : Earl E. HOUSEMAN : Area frame sampling in agriculture, Statistical Reporting Service - United States Department of Agriculture, Novembre 1975.

l'estimation des surfaces, mais vise également des caractéristiques attachées aux exploitations (cheptel, données socio-économiques concernant les exploitants, etc.)? En fonction du type de données recherchées, il propose trois définitions : le segment fermé, le segment ouvert et le segment pondéré, qui auront bien entendu chacune des conséquences sur la forme des estimateurs à employer.

Ce bref exposé suffit à montrer la différence, sinon fondamentale, du moins d'approche qui existe entre la démarche des statisticiens agricoles et la nôtre : pour nous, *l'unité aréolaire de taille variable et les probabilités de sélection inégales sont des données du problème*, en revanche le rattachement des unités d'observation à l'unité aréolaire d'échantillonnage ne pose aucun problème particulier lorsque l'on s'intéresse aux caractéristiques des personnes vivant dans les logements de l'ilot.

2.2.2 Les recherches récentes sur l'échantillonnage par points

Dans le domaine des statistiques d'utilisation du sol, mais aussi en statistique minière avec les foro-sondages, l'échantillonnage par points (ou plus justement le sondage spatial) occupe une place centrale dans les techniques d'estimation. En effet, dans ces disciplines, le problème est souvent d'estimer la répartition spatiale d'une ressource (culture, espèce animale, espèce d'arbre, minerais etc.) dans un espace à deux ou à trois dimensions (territoire géographique terrestre ou maritime, couche géologique).

J. DELINCE, dans l'introduction à son travail sur les techniques d'échantillonnage bi-dimensionnelles appliquées aux problèmes d'inventaires forestier, situe ainsi le cadre des recherches récentes dans ce domaine¹ :

"L'analyse des propriétés spatiales d'une population comporte deux aspects : l'étude de la *structure existant au sein de cette population* et la *quantification des paramètres du système*. (...) L'étude de la répartition spatiale s'est de plus en plus souvent référée à la théorie des processus stochastiques ponctuels. Dans ce contexte, on parle de l'analyse des propriétés de premier et second ordre. Les premières concernent *l'intensité du phénomène* ; la recherche d'estimateurs robustes y occupe une place prépondérante. Les propriétés d'ordre deux abordent la *structure du processus* (...) Plutôt qu'un test du caractère aléatoire de la population, ces techniques cherchent à décrire la structure à différentes échelles et proposent des tests d'*adéquation vis à vis de modèles* quelquefois fort complexes."

Le travail de DELINCE porte sur l'estimation de densités d'espèces dans des massifs forestiers, c'est à dire qu'il s'intéresse, dans la terminologie des processus aléatoires, aux propriétés du premier ordre. Il s'agit pour lui d'étudier comparativement la robustesse de différentes méthodes d'estimation face à la répartition spatiale.

Toujours dans le domaine de l'étude par sondage de la répartition d'espèces végétales, d'autres recherches sont menées, comme le signale DELINCE, basées sur des hypothèses d'adéquation de la structure du processus à des modèles théoriques. Citons par exemple les articles de GORDON et CAMARASA (1977)² et GORDON (1979)³

¹ : J. DELINCE : Estimation robuste de densité dans le plan, étude comparée de techniques d'échantillonnage, applications en forêt. Thèse de doctorat en Sciences Agronomiques, Université Catholique de Louvain (1983).

² : M. GORDON et J.M. CAMARASA : Un test hypergéométrique direct pour analyser l'hétérogénéité d'un échantillon "linéaire". Dans "Œcologica Plantarum", tome 12, n° 1 - 1977.

³ : M. GORDON : Analyse partielle des liaisons interspécifiques entre espèces prises trois à trois. Dans "Œcologica Plantarum", tome 14, n° 3 - 1979.

concernant l'échantillonnage linéaire, et l'article de GUIDICELLI, LANLY, OUAKAM et PIETRI¹ sur l'échantillonnage systématique dans le plan.

Le cadre théorique, décrit par DELINCE, des recherches récentes dans le domaine de l'échantillonnage forestier, s'applique également au domaine de l'échantillonnage minier qui a même eu en la matière un rôle pionnier. MATHERON (1965) établit une théorie des variables régionalisées basée sur les fonctions aléatoires (ou processus stochastiques) et propose son application aux sciences de la nature². Lui-même l'applique à l'estimation des concentrations de minerais dans les gisements miniers. Depuis, le domaine d'application de cette théorie ne cesse de s'étendre, et l'étude des corrélogrammes spatiaux liés au divers types de données que l'on souhaite échantillonner s'avère une direction de recherche très prometteuse. Citons, dans le domaine de l'échantillonnage par point pour des statistiques d'utilisation du sol l'article de SÆBØ³, ou dans celui de l'échantillonnage des données terrain pour l'interprétation d'image satellite l'article de DE KEERSMAECKER⁴.

Tout en reconnaissant le grand intérêt de ces recherches, nous avons choisi de ne pas effectuer une recherche parallèle dans le domaine de l'estimation des populations urbaines, ni même de tenter de lui appliquer des résultats obtenus dans d'autres types d'applications. Ce choix, que nous avons fait également pour des raisons de calendrier, a reçu des éléments de justification dans l'introduction générale, nous ne revenons donc pas ici sur ce sujet.

2.2.3 L'utilisation de la télédétection spatiale dans le domaine des statistiques d'utilisation du sol

Comme nous l'avons dit, la télédétection spatiale, parcequ'elle fournit aux statisticiens et aux thématiciens qui s'intéressent à l'utilisation du sol une nouvelle source de données, est venue modifier les techniques d'inventaires de ressources et d'estimations par sondages. Il nous paraît nécessaire de distinguer deux démarches, selon l'utilisation qui est faite des images satellites au sein de ces nouveaux dispositifs de production de données. Pour situer la question, nous prendrons à nouveau l'exemple de l'estimation de surfaces occupées par une catégorie données d'occupation du sol.

La première démarche est celle qu'adopte par exemple en France le SCEES (Service Central d'Etudes Economiques et Statistiques du Ministère de l'Agriculture). Dans un article publié en 1981, J. MEYER ROUX présente ainsi l'idée directrice de la méthode⁵ :

"À partir d'une connaissance a priori de la zone à étudier ou de certaines vérités terrain, on définit une méthode de classification des pixels. Si on s'intéresse à la catégorie C du territoire, tout pixel p_j est classé. Le produit cartographique final est une carte comportant les pixels classés en place avec une légende, C ou non C.

¹ : X. GUIDICELLI, J.P. LANLY, J.B. OUAKAM, M. PIETRI : Application de la théorie des processus aléatoires à l'estimation de la précision d'un inventaire forestier par échantillonnage systématique. Dans "Annales des Sciences Forestières", Tome 30, n° 2, 1972.

² : G. MATHERON : Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature. MASSON, PARIS - 1965.

³ : H. V. SÆBØ : Land use and environmental statistics obtained by point sampling. Dans "Meddelanden från Staten Skogsforskningsinstitut", 1- pp 317 à 340.

⁴ : M.L. DE KEERSMAECKER : Stratégie d'échantillonnage des données de terrain intégrées dans l'analyse des images satellitaires. Dans "L'Espace Géographique", n° 3 pp. 195-205, 1987 PARIS.

⁵ : J. MEYER ROUX : Cartographie et statistique en télédétection, SCESS 1981.

Le produit statistique est un listing comportant pour les sous zones intéressantes z_i du territoire T la surface x_i , avec $x_i = \sum_j p_{ij}$: somme des pixels de la zone z_i classés en catégorie C. La statistique au niveau de l'ensemble du territoire est obtenue également par sommation."

Dans ce système, c'est donc l'image classifiée du territoire à étudier qui fournit la statistique. Comme le souligne immédiatement MEYER ROUX, une telle démarche, appliquée telle quelle, a de gros inconvénients :

d'une part,

"La connaissance a priori du terrain est imparfaite et les vérités terrain recueillies sont très rarement représentatives de l'ensemble du territoire. Dans la plupart des nombreuses études réalisées sur le sujet les vérités terrain qui permettent d'obtenir des pixels d'entraînement sont surtout choisies pour leur facilité d'incorporation dans l'étude." ;

d'autre part,

"Il n'y a jamais eu de résultat statistique donné avec une variance associée puisque la statistique est issue d'un comptage exhaustif."

L'auteur propose donc une amélioration de la méthode basée sur un sondage de contrôle sur le terrain qui permet d'une part, de redresser l'estimation issue de l'image au moyen d'une régression effectuée sur l'échantillon entre "données image" et "données terrain", et d'autre part d'associer un intervalle de confiance à la statistique tirée de l'image (celui lié précisément à l'estimation de la droite de régression). Il n'en reste pas moins que la méthode réclame d'une part la couverture exhaustive du territoire étudié par l'image et d'autre part l'identification sur ces images du thème qui fait l'objet des estimations.

C'est la seconde condition qui, en l'état actuel des recherches sur l'interprétation des images satellites en milieu urbain, fait que nous ne pouvons pas appliquer ce type de méthode. Mais la première condition peut également se révéler un lourd handicap lorsqu'on cherche à produire des statistiques d'utilisation du sol relatives à de grands territoires et qu'on dispose d'un budget limité. C'est la raison pour laquelle H. ANDRIANASOLO¹ adopte une démarche différente :

"(...)Les types d'exploitation (d'images satellites) actuelles, aux résultats acquis, sont aussi fondés sur la nécessité de posséder une couverture totale de toute la région étudiée par les images, ce qui dans une optique d'opérationnalité dans un pays pauvre est un facteur de blocage. (...) Une contrainte s'imposant pour une réduction des coûts est alors celle de n'avoir recours qu'à un nombre minimum d'images pour l'obtention des résultats."

Partant de cette contrainte, ANDRIANASOLO utilise un nombre limité d'images couvrant des zones types du territoire malgache pour définir, d'après ces images, une stratification. Cette stratification est ensuite étendue aux zones non couvertes en s'aidant de la cartographie, existante, des unités physiques. L'étape finale du travail, l'enquête sur le terrain, a l'objectif suivant :

"Après stratification, le cadre de notre démarche exige qu'une estimation de l'extension moyenne de chaque culture à l'intérieur de segments, soit réalisée par strate. Extension moyenne qui servira ultérieurement pour les extrapolations, dans les surfaces non enquêtées, sur la base du nombre calculé de segments qu'elles possèdent."

Classiquement, puisqu'il s'agira d'estimer des surfaces en cultures, le sondage s'appliquera à des unités aréolaires de taille fixe (segments fermés) avec équiprobabilité de sélection dans l'échantillon. L'inférence se fera selon une démarche comparable à celle exposée dans le paragraphe précédent par MEYER ROUX

¹ : H. H. ANDRIANASOLO : Analyse statistique des données de télédétection, statistiques agricoles, application sur Madagascar. Thèse de doctorat en mathématiques appliquées et sciences sociales, EHESS, PARIS, 1987.

(estimateur par la régression), sinon que la statistique globale n'est pas obtenue par la classification d'images couvrant l'ensemble du territoire, mais par expansion directe à l'ensemble des segments que contient chaque strate. Une première application de la méthode proposée a pour objet l'estimation des superficies occupées par le riz de seconde saison dans la plaine de Marovoay.

De l'aperçu qui précède, il ressort que l'application de la télédétection à la production de statistiques d'utilisation du sol, repose avant tout sur le postulat suivant : il est possible, moyennant une quantité suffisante de vérités terrain, d'identifier sur l'image le thème auquel on s'intéresse et de parvenir à une classification suffisamment fiable pour produire, d'après elle, une statistique de surface occupée par le thème en question. Même si dans la seconde démarche que nous avons considérée, l'objectif de stratification de l'espace étudié semble se rapprocher d'un des objectifs de notre programme de recherche, on voit très vite que la similitude n'est qu'apparente. Pour ANDRIANASOLO, la stratification n'a pas pour but premier l'amélioration d'un sondage visant à produire par lui-même, la statistique recherchée ; il s'agit avant tout de définir les aires d'extension des statistiques partielles obtenues grâce aux images.

Quand au postulat qui fonde ces méthodes il est, depuis un certain temps déjà, justifié dans le domaine de l'occupation du sol en milieu rural, domaine où la taille des objets élémentaires qui composent le milieu et la bonne connaissance des caractéristiques de certains de ces objets, ont permis d'établir des signatures spectrales identifiables. Au tout début de l'interprétation des images satellites en milieu urbain, il serait évidemment déraisonnable de fonder une méthode de production de données localisées sur ce même postulat. En revanche, comme nous allons le voir maintenant, utiliser l'image satellite pour obtenir une information morphologique sur le milieu urbain et exploiter cette information pour améliorer un sondage à but démographique, est une idée qui peut bénéficier d'un certain nombre d'acquis méthodologiques.

2.3 Bilan des expériences dans le domaine de l'estimation de données urbaines à partir d'informations morphologiques¹

2.3.1 Utiliser la morphologie urbaine pour estimer les populations citadines : une méthode déjà ancienne pratiquée sur photographies aériennes

Depuis les années trente, des chercheurs étudiant le milieu urbain se sont intéressés aux relations entre les caractéristiques morphologiques des résidences et les caractéristiques démographiques et socio-économiques des habitants : les travaux des américains tels que KENZIE (1934) ou PARK (1937) ont reçu un écho certain en France, où se sont développées à partir des années cinquante des études sur le même thème (CHOMBART DE LAUWE, 1952).

Les relations entre la morphologie urbaine et les caractéristiques de la population citadine établies par ces travaux ont permis le développement de méthodes d'estimation de population à partir de l'information sur l'occupation du sol apportée par les photographies aériennes ; les caractéristiques de l'urbanisation dans les pays en développement renforcent l'intérêt de ces méthodes de production rapide de données socio-démographiques spatialisées.

¹ : Comme nous le disions en introduction, ce bilan a été rédigé par F. DUREAU pour l'article à paraître dans "Informatique et Sciences Humaines" : Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale, pp 1 à 6.

Comme le montre le Tableau 2.2¹, l'idée d'utiliser les éléments de connaissance sur les caractéristiques spatiales et morphologiques de la ville à des fins démographiques n'est pas nouvelle.

Tableau 2.2 - Utilisation de données de télédétection aérienne et spatiale pour des estimations démographiques

Auteur	Année	Ville (PAYS)	Document utilisé
GREEN	1956	Birmingham (USA)	Photo. aériennes N/B 1/7500
PORTER	1956	LIBERIA	Photo. aériennes N/B
HADFIELD	1963	Chicago (USA)	Photo. aériennes N/B 1/4800
BINSELL	1967	Chicago (USA)	Photo. aériennes Couleur 1/5240
HOLZ, HUFF et MAYFIELD	1969	40 villes du Tennessee (USA)	Photo. aériennes Haute Altitude
MURET (CRU)	1969	Auxerre (FRANCE)	Photo. aériennes N/B 1/2000
LINDGREEN	1970	Boston (USA)	Photo. aériennes IRC 1/20000
COLLINS et EL-BEIK	1971	Leeds (Angleterre)	Photo. aériennes N/B 1/10000
DUEKER et HORTON	1971	Washington (USA)	Photo. aériennes Haut. Alt. 1/50000
HSU	1971	Atlanta (USA)	Photo. aériennes N/B 1/5000
ANDERSON et ANDERSON	1973	23 villages du Kansas (USA)	Photo. aériennes 1/20000
VERNIERE	1973	Pikine (SENEGAL)	Photo. aériennes N/B 1/5000
KRAUS, SENER et RYERSON	1974	4 villes de Californie (USA)	Photo. aériennes IRC 1/60000
HORTON	1974	Washington (USA)	Photo. aériennes 1/50000
ALLAN et ALEMAYEHU	1975	Wolano (ETHIOPIE)	Photo. aériennes
D'ALLEUX	1975	Lagos (NIGERIA)	Photo. aériennes
HENDERSON et UTANO	1975	Albany (USA)	Photo. aériennes N/B 1/24000
OGROSKY	1975	18 villes, Puget Sound Region	Photo. aériennes IRC Haut. Alt. 1/135000
THOMPSON	1975	Washington (USA)	Photo. aériennes Haute Altitude
DAYAL et KHAIR ZADA	1976	AFGHANISTAN	Photo. aériennes
SUJARTO	1978	Bandung (INDONESIE)	Photo. aériennes
HENDERSON	1979	USA	Photo. aériennes 1/25000
LO	1979	Hong Kong	Photo. aériennes
LO et CHAN	1980	Sheung Shui-Fan (HONG KONG)	Photo. aériennes
CLAYTON et ESTES	1980	Goleta Valley, St. Barbara (USA)	Photo. aériennes IRC 1/63360
ADENIYI	1983	Lagos (NIGERIA)	Photo. aériennes 1/20000
IAURIF, ROC, SIDAU et SEE	1983	Niamey (NIGER)	Photo. aériennes 1/15000
NEVES DE OLIVEIRA et al.	1984	Sao José Dos Campos (BRESIL)	Photo. aériennes N/B 1/10000
OLORUNFEMI	1984	Ilorin (NIGERIA)	Photo. aériennes N/B 1/8000
POLLE	1984	Teheran (IRAN),	Photo. aériennes 1/10000 et 1/9000
WATKINS	1985	Colombo (SRI LANKA) Boulder, Colorado (USA)	Photo. aériennes N/B 1/20. et 60.000
IAURIF	1986	Bouaké (COTE D'IVOIRE)	Photo. aériennes
SADOL	1968	USA	Image radar
WELLAR	1969	San Antonio et Houston (USA)	Image du satellite GEMINI
REINING	1973	NIGER et HAUTE VOLTA	Image LANDSAT
MURAI	1974	Tokyo (JAPON)	Image LANDSAT
LO et WELCH	1977	Villes de 500000 à 2500000 H. (CHINE)	Image LANDSAT

(D'après ADENIYI P.O., 1983, Table 1, Page 546, et recherche bibliographique personnelle)

¹ : Le tableau 2.2, repris de l'article à paraître dans "Informatique et Sciences Humaines", rassemble les expériences faisant l'objet d'un article ou d'une simple citation dans la littérature scientifique française et anglo-saxonne.

Dès 1956, GREEN a eu recours à des photographies aériennes à basse altitude (échelle 1/7500) pour estimer la population de la ville de Birmingham (USA). Les photographies aériennes permettent de dénombrer les logements ; l'estimation de population est obtenue en multipliant ce nombre de logements par le nombre moyen de personnes par logement, fourni par le dernier recensement. Cette méthode détaillée, qui repose donc sur le *dénombrement des logements* a été reprise par de nombreux auteurs, et a fait l'objet de quelques applications dans les pays en développement, la première ayant eu lieu au Libéria (PORTER, 1956).

Deux expériences menées aux USA permettent de se rendre compte du degré de précision de la méthode : il s'agit des travaux de HSU (1971) et WATKINS (1985), dont le Tableau 2.3 résume la méthodologie et les conclusions. Pour les populations totales des villes, les niveaux de précision sont tout à fait corrects ; par contre, les estimations de WATKINS par quartier font l'objet d'erreurs relatives importantes, pouvant atteindre 16,57 %. La conclusion de WATKINS correspond, de fait, à celle qui a donné lieu au développement d'une autre méthode d'estimation de population à partir de photographies aériennes : l'amélioration de la précision passe par la prise en compte des différenciations internes à la ville.

L'idée de base de cette seconde famille de méthodes est *qu'il existe des relations entre les caractéristiques morphologiques du milieu urbain et les caractéristiques démographiques et socio-économiques des habitants*. Le principe de la méthode est d'exploiter l'information exhaustive fournie par les couvertures aériennes :

- *soit pour recueillir rapidement, par sondage, des données relatives aux populations urbaines*: les différents travaux menés en appliquant cette méthode ont mis en évidence que cette technique de sondage spatial permet un allègement de l'échantillon d'enquête et une spatialisation des résultats. Nombreux sont les bureaux d'étude qui ont appliqué la technique mise au point par VERNIERE¹ pour obtenir des estimations démographiques sur les villes dont ils avaient à réaliser un plan d'aménagement, ou des programmations d'équipement ; tout récemment, l'IAURIF a testé, avec succès, cette méthode pour l'estimation de caractéristiques socio-économiques des habitants des différents quartiers de Niamey².

- *soit pour actualiser l'effectif de population d'une ville ayant fait l'objet antérieurement d'un recensement de population* (Exemple : KRAUS, 1974. Voir Tableau 2.3) : les photographies aériennes servent à déterminer la superficie occupée par chaque type de morphologie urbaine. Connaissant, par un recensement antérieur, la densité de population par type de morphologie urbaine, on en déduit l'effectif total de population pour la ville, en faisant l'hypothèse de la constance dans le temps de ces densités démographiques par type morphologique. Dans le cas de Pikine, VERNIERE a pu vérifier la validité de cette hypothèse : l'application des coefficients de 1970 aux années antérieures (1961, 1963, 1967) a donné de bons résultats (cf Tableau 2.3).

Signalons enfin une troisième famille de méthodes, qui reposent sur la *relation entre superficie de la tache urbaine et population totale d'une ville* : le développement de ces méthodes est directement liée à l'engouement, depuis la fin des années cinquante, des géographes américains pour la modélisation. Les estimations de population basées sur la superficie des villes sont toutes fondées sur l'application des modèles mathématiques reliant superficie et population : pour leur part, LO et WELCH (Voir Tableau 2.3), travaillant sur des images LANDSAT, obtiennent de bons résultats sur les grandes villes chinoises en utilisant trois modèles très souvent appliqués, ceux de NORDBECK (Loi de croissance allométrique, 1965), TOBLER (1969) et HUXLEY (1932).

¹ : Voir résumé de la méthode dans le Tableau 2.3 et, pour l'exposé détaillé VERNIERE, 1978.

² : Voir IAURIF, ROC, SIDAU, SEE, 1983.

Tableau 2.3 : Quelques exemples d'estimations de population à partir de photographies aériennes et images satellite

Auteur	Ville	Niveau Géo.	Source d'inform.	Référence	Méthodologie	Résultats
FERU (1971)	Atlanta (USA)	Populations urbaine et rurale de la circonscription	Photographies aériennes N/B 1/5000 Rec. Nb moyen de pers. par log.	Recensements démo.	Comptage des logements sur les photo. aériennes Estimation de la population en appliquant une équation linéaire intégrant Nb de log. et Nb moyen de pers. par log.	Erreur relative sur la population totale : +/- 3%
WATKINS (1955)	Walden (USA)	3 quartiers de 2500 à 6200 H. en 1970 et 1950	Photo aériennes N/B 1970 : 1/20000 1950 : 1/5000 Rec. : Taux de logements vides et Nbre moyen de pers. par logement (Taux uniques pour les 3 quartiers, calculés pour l'habitat indiv. et le collectif)	Recensements démo. 1970 et 1950	Comptage des logements des 3 quartiers aux 2 dates sur les photo. aériennes Estimation de la population en appliquant une équation linéaire intégrant Nb de log. Taux de log. vides, Nb moyen de pers. par logement (pour l'habitat individuel et pour l'habitat collectif)	Erreur relative sur le nombre de logements : entre -4,91% et -0,37%, par quartier -0,70% et -3,26% pour le total des 3 q. Erreur relative sur la population : entre -6,64% et +10,57%, par quartier -1,96% et +8,54% pour le total des 3 q. <u>Conclusions</u> 1) Bon dénombrement des logements sur photographies aériennes Habitat individuel : 1/20000 à 1/40000 Habitat collectif : 1/5000 à 1/10000 2) Pour améliorer la précision calculer les taux de log. vides et le Nb moyen de pers. par logement, par quartier de la ville
VERNIERE (1973)	Pikine (SÉNÉGAL)	Zones morpho. homogènes de la banlieue de Dakar (Pop. Tot. 135.000 Hab.)	Photo aériennes N/B 1/5000 (1961, 1963, 1967, 1970) Enquête démo sur échantillon de 8 carrés de sondage (1970-71) : effectifs de population	Recensements OMS, ISEA, administratif Enquête stat. nationale	Zonage en espaces homogènes sur photo. aériennes de 1970 (morphologie urbaine) Calcul des Coef. de surface bâtie habitable (CSBH) pour 38 carrés de sondage, sur les photo. aériennes de 1970. Pour les 8 carrés où la pop. est connue, calcul du coef. Densité de pop./CSBH. En supposant que ce coef. est constant dans chaque zone homogène, estimation des densités et des effectifs par zone en 1970 et pour les années antérieures.	Erreur relative sur la population totale : entre -5,6% et +3%, selon les années <u>Bilan et recommandations</u> 3 intérêts de la méthode : 1) Spécialiser les résultats d'une enquête démo. globale 2) Actualiser des chiffres globaux de population entre 2 recensements 3) Moyen de connaissance autonome en l'absence de toute autre donnée, base pour une enquête par sondage
KRAUS et al (1974)	4 villes de Californie (USA)	Villes entières de 58.000 à 250.000 Hab.	Photo aériennes IRC 1/50.000 Rec. densité de population pour 3 types de quartier : Habitat individuel Habitat collectif Caravanes	Recensement démo.	Délimitation sur photo. des 3 types de quartier, pour chaque ville. Calcul des surfaces correspondantes. Calcul de la densité moyenne de chaque type de quartier (par ville) : moyenne des résultats obtenus sur une série de sondages aléatoires. Estimation de la population de chaque ville en appliquant l'équation : Pop. totale d'une ville = moyenne des densités par types de quartier pondérée par la surface de ceux-ci.	Erreur relative sur la population totale : entre -9,17% et +7%, selon les villes -4,51% pour le total des 4 villes <u>Exclusions d'amélioration</u> 1) Utiliser des photo. aériennes à une échelle > 20 000 2) Ne pas exclure les quartiers à vocation industrielle et commerciale 3) Affiner la classification selon le type d'occupation du sol
LO et WELCH (1977)	31 villes chinoises	Villes entières > 500.000 Hab.	Images LANDSAT : compositions colorées (C3 et C7) Cartes stéréos de 150 villes de Taiwan : calcul des équations linéaires et log. entre pop. et surf. des villes (1951-66)	Recensements démo. des 30 villes de Chine	Mesure de la superficie de chaque ville sur l'image LANDSAT au 1/500 000 Estimation de la population de chaque ville en appliquant les 2 séries d'équations mises au point sur les villes de Taiwan	Bonne concordance entre résultats calculés par les modèles et ceux des recensements démographiques. <u>Exclusions d'amélioration</u> 1) Utiliser une image au 1/250 000 2) N'appliquer le modèle qu'aux villes couvrant une surface > 10 Km ² 3) Utiliser un modèle linéaire pour les villes > 250 000 Hab. et logarithmique pour les villes plus petites

Extrait des articles de ISU (1971), VERNIERE (1973), KRAUS et al. (1974), LO et WELCH (1977), WATKINS (1955)

D'autres auteurs ont tenté, sans grand succès, de sophistication la méthode en introduisant la théorie des places centrales dans leurs équations : c'est le cas, en 1969, de HOLZ, HUFF et MAYFIELD (Tableau 2.2).

La plupart des estimations de population réalisées avec cette méthode, à partir de photographies aériennes, demeurent très imprécises, du fait de l'incertitude sur la délimitation de l'aire urbanisée ; il semble, au vu des meilleurs résultats obtenus par LO et WELCH, que l'information très réductrice de LANDSAT¹ convienne mieux à ces méthodes globales.

Ainsi, depuis trente ans environ, ont été accumulées de nombreuses expériences d'utilisation de la morphologie urbaine pour la production rapide de données démographiques. Les niveaux d'utilisation de la morphologie urbaine sont variés : depuis les méthodes les plus globales, ne retenant de la morphologie urbaine que la surface urbanisée, aux méthodes détaillées reposant sur des comptages de logement, en passant par les méthodes semi-globales, basées sur les typologies de quartiers. Ces méthodes, pour la plupart mises au point par des scientifiques travaillant sur des villes de pays développés, ont trouvé depuis la fin des années soixante un écho certain parmi les urbanistes opérant dans des villes des pays en développement, où elles ont aussi fait la preuve de leur efficacité.

2.3.2 Des photographies aériennes aux images satellite à haute résolution

2.3.2.1 Quelques rares expériences d'estimation de la population totale de villes à partir de la tache urbaine mesurée sur image satellite

Comme le montre le Tableau 2.2, rares sont les expériences d'utilisation des images satellite pour la production de données démographiques : à partir de notre recherche bibliographique, nous n'avons eu connaissance que de cinq applications, toutes basées sur la méthode globale d'estimation de la population totale d'une ville à partir de la superficie urbanisée mesurée sur image satellite, LANDSAT dans la plupart des cas.

Cet état de fait est directement lié au degré de résolution des images LANDSAT (80 m x 80 m), seules disponibles jusqu'en 1985 : cette résolution insuffisante a freiné l'utilisation de l'imagerie satellitaire en milieu urbain, caractérisé par une forte hétérogénéité et une faible dimension des éléments constitutifs. Avec l'imagerie LANDSAT, il n'était pas envisageable d'observer les différenciations morphologiques internes à la ville, et donc d'appliquer d'autres méthodes de production de données démographiques que les méthodes globales reliant population et surface de la ville.

2.3.2.2 Raréfaction des couvertures aériennes et progrès de la télédétection spatiale, système d'observation continue de la morphologie urbaine

Le recours aux photographies aériennes devient de plus en plus problématique. En effet, les couvertures aériennes des villes des pays en développement sont de plus en plus rares en raison de l'ampleur des moyens qu'elles réclament : *la répétitivité de l'observation aérienne devient une réalité de moins en moins effective dans les villes des pays en développement*. Par contre, cette répétitivité est offerte par les nouveaux satellites SPOT et THEMATIC MAPPER, qui ont une résolution suffisante (respectivement 10 / 20 mètres, et 30 mètres) pour observer assez finement le tissu urbain.

¹ : Le satellite LANDSAT enregistre sur des unités élémentaires appelées "pixels" d'environ 80 mètres de côté : les valeurs radiométriques enregistrées sur un pixel donné sont la résultante de toutes les composantes de ce pixel, un seul pixel pouvant être composé de portions de voirie, toiture, parking, espace vert...etc.

Les images satellite présentent des qualités indéniables, inhérentes à leur mode d'enregistrement et à la nature du document diffusé :

- répétitivité des observations (T.M. : 16 jours - SPOT : 1 à 4 jours)
- exhaustivité spatiale
- coût raisonnable : environ 15 fois moins que les photographies aériennes
- caractère numérique des données
- richesse de l'information enregistrée, avec possibilité de vision stéréoscopique (SPOT).

Il semble donc nécessaire de *tester dès à présent les possibilités de mise en oeuvre de méthodes d'observation démographique intégrant les données morphologiques observables sur image satellite*, afin de tenter d'apporter des éléments de solution aux carences de l'observation démographique classique ou par photographies aériennes. Il s'agit de tirer parti des qualités des images satellite en s'appuyant sur l'acquis méthodologique constitué par les travaux antérieurs de production de données démographiques à partir de photographies aériennes.

L'idée centrale reste la même : *utiliser l'information exhaustive apportée par les images satellite pour recueillir rapidement, par sondage, les données relatives aux populations urbaines*. De plus, outre l'utilisation de la télédétection comme moyen d'amélioration des techniques de sondage, on peut envisager d'utiliser la télédétection comme *moyen de connaissance autonome de la dynamique urbaine* : certains phénomènes démographiques ont une traduction physique immédiate (l'immigration, par exemple), et l'évolution des formes d'occupation de l'espace constitue une des dimensions importantes de la croissance urbaine.

Chapitre 3

LES QUESTIONS ET LES OUTILS STATISTIQUES DANS LES DIFFERENTES PHASES DE LA RECHERCHE

Introduction

Dans les deux premiers chapitres nous avons posé les deux questions principales autour desquelles s'organise notre travail :

1. Quelles techniques de sélection de l'échantillon et d'inférence appliquer à une enquête démographique par sondage en ville?
2. Comment utiliser l'information morphologique sur le milieu urbain qu'apportent les images satellite à haute résolution pour mettre en œuvre et améliorer un tel sondage?

En partant des contraintes que nous impose le domaine d'application fixé à la méthode et en faisant référence à la littérature sur le sujet, nous avons déjà cerné les principaux problèmes que pose la première question. Nous allons voir maintenant comment s'inscrit dans la problématique définie jusqu'à présent la seconde question. Nous récapitulerons ainsi les différents thèmes sur lesquels va porter la recherche et détaillerons les questions techniques¹ qui se posent à propos de chacun d'eux ; puis nous tenterons de faire l'inventaire des méthodes mathématiques qui leurs sont applicables. Enfin nous verrons comment ces thèmes seront abordés dans les principales étapes du travail et quels choix méthodologiques ont été faits dans chacune d'elles.

3.1 : Récapitulation des questions

3.1.1 : Questions liées à l'utilisation de l'image satellite comme base de sondage

Ces questions ont en fait été abordées dans le premier chapitre dans lequel nous avons considéré tout du long de l'exposé que nous ne disposons pas d'autre base de sondage que celle issue de l'image c'est à dire l'ensemble des unités aréolaires qui y sont repérables. Nous l'avons vu alors, cette contrainte conduit à un sondage aréolaire à deux degrés où l'unité primaire est l'ilot physique tel qu'on pourra l'identifier sur l'image et, de manière concordante, sur le terrain. Cet ilot regroupe un certain nombre d'unités statistiques (ménages), dont nous n'enquêterons, pour des raisons d'efficacité du sondage, qu'une fraction (sondage au second degré). Ainsi *le premier thème de recherche auquel nous conduit l'utilisation de l'image satellite comme base de sondage est celui des sondages à plusieurs degrés*

¹ : Dans le paragraphe 3.1, ces questions seront numérotées de façon continue afin de faciliter les références qui y seront faites dans la suite du chapitre.

Puisque la question du nombre de degrés du sondage et de la définition des unités à sélectionner à chacun de ces degrés est résolue par la prise en compte des contraintes liées au milieu étudié et à la nature de la base de sondage, les questions restant à résoudre sont les suivantes.

1. Quels taux adopter au premier et au second degrés du sondage?

2. Le taux au second degré doit-il être constant pour l'ensemble des unités primaires?

3. Quels estimateurs faut-il employer au différents degrés?

Le second thème de recherche, lié à l'emploi de l'image satellite comme base de sondage, est celui du sondage par points. En effet, nous avons déjà évoqué les difficultés, pour l'instant difficilement surmontables, que pose la constitution d'après l'image, d'une liste de toutes les unités aréolaires présentes dans la zone urbaine étudiée. Il n'est pas inutile, pour prendre bien conscience du problème, de détailler un peu les principaux obstacles à une telle opération.

⌘ D'une part, les techniques de traitement d'images satellite en milieu urbain ne permettent pas, dans l'état actuel de leur développement, d'extraire le réseau de voirie avec une fiabilité suffisante pour que l'on puisse réaliser cette opération automatiquement. L'établissement d'une liste d'ilots par interprétation visuelle de l'image, en supposant que l'œil fasse moins d'erreurs que l'ordinateur, représenterait un travail énorme et très fastidieux. D'autre part, qu'il s'agisse d'interprétation visuelle ou "automatique", dans certaines zones mal dotées en infrastructures, le repérage sans ambiguïtés des limites des ilots est extrêmement difficile. L'absence d'une voirie repérable sur l'image conduit nécessairement à définir, dans ces secteurs, des pâtés de maisons gigantesques. Outre qu'il en résulterait une hétérogénéité de la taille des unités primaires qui peut avoir de fâcheuses conséquences sur la précision des estimations, le recensement des ménages dans ces unités primaires, lorsqu'elles sont sélectionnées dans l'échantillon, demanderait alors un gros travail.

On voit bien, dans ce contexte, quel sont les avantages du sondage par points. Tout d'abord on ne constitue pas de liste d'unités primaires : les points de l'image sélectionnés désignent directement les ilots à enquêter. Ensuite, lorsque le repérage des limites d'un pâté de maisons pose problème sur l'image, on localise sur le terrain le point sélectionné (à quelques dizaines de mètres près) et on peut effectuer sur place, sans ambiguïtés, le repérage de l'ilot.

Le prix de ces avantages pratiques est, comme nous l'avons vu, une probabilité de sélection des unités proportionnelle à leur surface et, pour les unités dont les limites ne sont pas visibles sur l'image, de devoir effectuer la mesure de cette surface sur le terrain.

Enfin, concernant le sondage par points, deux questions principales se posent :

4. Quelle conséquence a la probabilité d'appartenance à l'échantillon liée à ce type de sondages sur la précision des estimations?

5. Peut-on adopter une stratégie de sélection des unités primaires qui définisse une probabilité proportionnelle à une mesure de l'extension spatiale des unités meilleure, au sens de la précision des estimations, que la surface totale de ces unités?

3.1.2 : Questions liées à l'utilisation de l'information morphologique extraite de l'image pour améliorer le sondage

Au chapitre précédent nous avons considéré trois manières d'améliorer la précision d'un sondage en prenant en compte une information exogène :

- par stratification de la base de sondage selon cette information
- par intégration de cette information dans l'étape d'inférence
- par gestion de probabilités d'appartenance à l'échantillon basées sur cette information.

Ces trois techniques peuvent en principe s'appliquer pour utiliser l'information morphologique extraite de l'image. Mais une remarque s'impose : dans cette optique d'utilisation d'une information exogène, nous devons distinguer entre ce qui serait théoriquement possible si l'on disposait de l'information morphologique "vraie", telle qu'elle apparaît au sol ou sur des photographies aériennes à grande échelle, et ce qui est réalisable à partir de l'information morphologique tirée de l'image. Examinons maintenant les questions que soulève l'emploi de chacune de ces techniques.

- La stratification

Comme nous venons de voir, il faut distinguer l'information morphologique au sol de celle extraite de l'image. Sur ce thème, la recherche devra donc répondre à trois questions :

6. Comment construire une stratification à partir de l'information morphologique?

7. Quelle est, en termes de précision des estimations, l'efficacité de ce type de stratification? Quels sont les critères morphologiques les plus efficaces?

8. En quoi l'image peut-elle rendre compte de ces critères? Quelle est l'efficacité des stratifications issues de l'image satellite?

Il est clair que les réponses à ces questions se conditionnent mutuellement, cependant nous verrons que les outils d'analyse statistique mis en œuvre pour les résoudre diffèrent de l'une à l'autre.

- Le redressement des estimations

A nouveau, on formule deux questions, liée chacune à un type d'information morphologique :

9. Existe-t-il, entre variables démographiques à estimer et variables morphologiques exogènes, des liaisons statistiques suffisamment fortes pour que l'on puisse envisager le redressement des estimations tirées de l'enquête à l'aide de variables morphologiques (estimation par le quotient ou par la régression)?

10. Si la réponse à cette dernière question est positive, peut-on obtenir à partir de l'image l'information morphologique avec un niveau de finesse suffisant?

- Les probabilités d'appartenance à l'échantillon

Comme on l'a vu au chapitre précédent, la précision d'un sondage à probabilités inégales dépend directement, dans la situation qui est la nôtre, du degré de corrélation existant entre la variable à estimer et la "mesure de la taille" des unités primaires qui

définit leur probabilité d'inclusion dans l'échantillon. Le problème, qui rejoint celui lié à l'emploi du sondage par points, se pose donc dans les mêmes termes que pour le redressement à l'aide de variables morphologiques :

11. Existe-t-il une mesure de la taille des pâtés de maisons, liée à des critères morphologiques (proportion de surface bâtie, élévation moyenne des bâtiments, etc), suffisamment corrélée aux variables démographiques pour qu'une probabilité de sélection des unités primaires proportionnelle à cette mesure soit avantageuse?

12. Peut-on, à partir de l'information extraite de l'image, et d'une stratégie de sélection des unités aréolaires, gérer une telle probabilité?

3.1.3 : Questions indépendantes de l'utilisation de l'image satellite

Du fait de l'orientation thématique de la recherche (démographie) et son contexte (des villes où l'on dispose de peu d'informations), l'enquête que nous mettrons en œuvre aura deux caractéristiques particulières :

1. elle sera réalisée à partir d'un sondage par points (caractéristique liée, comme nous l'avons vu, à l'emploi d'une base de sondage spatiale)
2. Le questionnaire sera à objectifs multiples, c'est à dire que l'enquête aura pour but d'estimer un grand nombre de caractères distincts de la population.

A propos de la première caractéristique, et cela quelque soit la base de sondage spatiale retenue (image satellite, photographies aériennes, plan cadastraux ou autres sources cartographiques), se pose le problème du mode de sélection des unités d'échantillonnage : doit-il être aléatoire ou systématique? De l'étude bibliographique faite au chapitre 2, il ressort que le tirage systématique peut être très avantageux lorsque les individus proches dans la base de sondage ont tendance à se ressembler du point de vue des variables à estimer, le sondage bénéficiant alors de ce que nous avons appelé une stratification "implicite". Dans le cas du sondage par points dans une base de sondage spatiale, la proximité entre individus de la base est une proximité géographique. En milieu urbain, l'hypothèse d'une ressemblance entre individus proches dans la base, c'est à dire d'une certaine similitude entre les caractéristiques démographiques d'îlots proches les uns des autres, apparaît tout à fait justifiée. Mais nous avons également constaté au chapitre 2 que l'estimation de la variance du sondage systématique pose problème. Concernant la technique de sélection à appliquer aux unités aréolaires, la question est donc la suivante :

13. Comment estimer la variance du sondage systématique et évaluer le gain de précision qu'il permet par rapport au sondage aléatoire?

La deuxième caractéristique de l'enquête, celle de viser des objectifs multiples, a pour conséquence que le sondage devra permettre d'estimer des variables différentes soit par leur nature mathématique (totaux, moyennes, quotients), soit par la catégorie de population à laquelle elles sont attachées (femmes, hommes, migrants, actif, population d'une tranche d'âge donnée, etc). Ces différentes variables peuvent présenter des "comportements statistiques" spécifiques s'agissant par exemple des corrélations avec les variables morphologiques, de leur variations spatiales etc.. Ces spécificités se répercuteront selon toute probabilité sur la précision des estimations produites. En fait, le problème apparaît comme transverse à toutes les questions posée jusqu'ici :

14. Quelque soit la technique envisagée pour améliorer la précision du sondage, comment cette technique profite-t-elle aux différents types de variables qu'il s'agit d'estimer?

La synthèse des recherches effectuées sur l'ensemble des questions évoquées ci dessus fournira les réponses aux deux questions finales :

15. Quel plan de sondage utilisant l'information satellitaire peut-on mettre en œuvre pour mener une enquête démographique en milieu urbain des pays en voie de développement?

16. Quelle est la précision des estimations réalisées à partir de cette méthode?

3.2 : Inventaire des méthodes mathématiques applicables

Dans l'introduction générale, nous avons identifié trois axes de recherches à développer parallèlement pour mener à bien le projet de recherche :

1 : L'interprétation des images satellite à haute résolution en milieu urbain.

2 : La formalisation mathématique des relations entre morphologie urbaine et population, et le développement des techniques de sondages spatiaux en démographie urbaine.

3 : L'intégration des données de télédétection dans les bases de données localisées.

Les deux premiers axes de travail sont le lieu privilégié de l'utilisation d'outils mathématiques et statistiques.

- Premier axe

Le but poursuivi est l'identification sur l'image satellite de l'information morphologique pertinente pour l'étude de la population. Les outils mathématiques utilisables dans ce but, qui s'appliquent aux données spectrales, relèvent, très schématiquement, des domaines de la reconnaissance des formes, de la morphologie mathématique et des techniques de classification automatique supervisée ou non. L'application de ces méthodes en milieu urbain constituent donc un sujet de recherche en soi. Une part importante est faite à ces techniques dans la thèse que A. MICHEL a rédigée à l'issue du travail qu'il a effectué au sein du projet concernant le premier axe de recherche.

Ces thèmes ne sont pas au centre de mon travail, axé, nous l'avons vu, sur le problème de l'optimisation des plans de sondage. Cependant, étant donné les liens étroits existant entre ces deux axes, nous serons très souvent amenés à faire état des résultats obtenus par B. LORTIC et A. MICHEL. En particulier, il est clair d'ores et déjà que les réponses aux questions 8, 10 et 12, telles qu'elles sont formulées au paragraphe précédent, sont entièrement conditionnées par le travail effectué dans le domaine de l'interprétation des images. Cette mise au point faite, examinons maintenant les méthodes applicables à l'axe de recherche qui nous est propre.

- Deuxième axe

Les quatorze questions que nous avons formulées au paragraphe précédent peuvent être rassemblées en trois groupes, d'ailleurs non disjoints, selon qu'elles se rapportent à l'un, ou plusieurs, des trois objectifs suivants.

- Identifier les variables morphologiques présentant un intérêt du point de vue de l'observation démographique, en confrontant les données de population, issues du recensement à Marseille ou des enquêtes préliminaires à Quito, avec les

données morphologiques provenant de l'étude des photographies aériennes ou d'enquêtes sur le terrain. Construire à partir de ces variables une stratification morphologique synthétique. (Questions 6 et 14)

- Proposer des modèles mathématiques explicatifs des relations entre démographie et morphologie dans les villes. (Questions 3, 4, 5, 7, 9, 11, 13 et 14)

- Définir les stratégies de sondage intégrant l'information morphologique adaptées à une enquête démographique en milieu urbain et évaluer leur précision. (Questions 1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13 et 14)

Expliquons-nous rapidement sur ce regroupement. Il a pour but unique de faire apparaître trois thèmes principaux relevant de *méthodes d'analyse statistique différentes*. Les vraies questions techniques à résoudre restent celles définies au paragraphe précédent, mais certaines d'entre elles peuvent faire l'objet de deux approches correspondant chacune à un type de méthode. Par exemple pour statuer sur l'efficacité d'une technique d'amélioration du sondage par intégration d'informations exogènes, on peut tenter de construire à partir de ces informations un modèle explicatif de la variable à estimer (deuxième thème). Mais pour aborder la même question on peut également, si les données s'y prêtent, calculer directement la variance du sondage correspondant à cette technique d'amélioration, et la comparer à celle du sondage ne l'intégrant pas (troisième thème).

Le champ des méthodes mathématiques potentiellement applicables à ces trois thèmes est très vaste. Elles relèveront en tous cas de trois domaines :

- 1) L'analyse des données, comprise ici dans l'acception la plus large du terme, non restreinte à l'une ou l'autre des écoles. Ce sont ces méthodes que nous appliquerons au premier thème.
- 2) Les techniques de régression, ajustement et modélisation mathématique qui s'appliquent aux second thème.
- 3) La théorie des sondages probabilistes qui permet d'aborder le troisième thème.

L'inventaire suivant, non exhaustif, permet cependant de citer les principales méthodes dans chacun des domaines. Volontairement nous ne l'avons pas limité aux méthodes que nous emploierons nous-même par la suite¹.

L'analyse des données

1. L'analyse exploratoire des grands tableaux de données au sens défini par J.P. BENZECRI. Analyse factorielle des correspondances appliquée aux tableaux de contingence ou aux tableaux mis sous forme disjonctive complète (correspondances multiples), classification ascendante hiérarchique et l'ensemble des méthodes d'aide à l'interprétation regroupées par l'ADDAD.
2. Les méthodes de classification non hiérarchique comme les nuées dynamiques (E. DIDAY) ou les méthodes anglo-saxonnes de "cluster analysis".
3. L'analyse de similitude (C. FLAMENT) et plus généralement les méthodes combinatoires d'analyse de données : classification, analyse booléenne, méthodes de sériation.

¹ : Pour la même raison que s'agissant des questions nous numéroterons les méthodes de façon continue.

4. L'analyse bayésienne et les méthodes inférentielles d'analyse de données.

Les techniques de régression et de modélisation mathématique

5. Méthodes de régressions linéaires ou polynomiales

6. Analyse de la variance

7. Discrimination logique

8. Analyse discriminante

La théorie des sondages

9. Sondages sans biais, aréolaires, stratifiés, à plusieurs degrés, à probabilités inégales, systématiques etc.

10. Estimation bayésienne

11. Méthode d'estimation des variables régionalisées et krigeage, plus généralement, approche stochastique des problèmes d'échantillonnage.

Dans les méthodes citées, nous ferons évidemment des choix tout au long de cette recherche. Ces choix tiennent à plusieurs raisons :

- des considérations théoriques ou pratiques sur le bien fondé de l'application d'une méthode à la problématique et au contexte qui sont les nôtres - dans l'introduction générale, nous nous sommes fondé sur des arguments de ce type pour écarter l'approche par les processus stochastiques (méthode 11) ;
- la nature et le volume des données à traiter, plus ou moins adaptés à telle ou telle méthode ;
- enfin et surtout, pour de strictes raisons de compétence, nous n'avons retenu que les méthodes que nous savions pratiquer efficacement.

Ce sont de ces choix et des différentes étapes du travail dont il va être maintenant question.

3.3 : Les étapes de la recherche et les choix méthodologiques effectués

Comme nous l'avons vu, pour le projet de recherche Télédétection et démographie, l'ORSTOM a choisi deux sites de travail qui correspondent en fait à deux phases du projet : la phase de mise au point de la méthode a lieu à Marseille, la phase de première application à une ville de pays en développement à Quito. Les objectifs, les plans d'expérience et les données associés à chacune de ces phases seront exposés en détail au début des parties II et III de la thèse¹. Nous ne ferons ici que résumer les objectifs principaux de ces deux phases et nous examinerons quels choix méthodologiques ont été effectués pour aborder les problèmes liés à chacune d'elle. Le tableau 3.1 présente une synthèse de ces choix : il permet de savoir dans quels chapitres et paragraphes des parties II et III sont abordées les questions définies au paragraphe 3.1 et quelles méthodes d'analyse leur ont été appliquées.

¹ : voir pour la phase marseillaise l'introduction à la deuxième partie et le chapitre 4, et pour la phase réalisée à Quito l'introduction à la troisième partie et le chapitre 9.

3.3.1 . Marseille, juin 85 - juillet 86 : phase de mise au point de la méthode

Pour mettre au point et valider une nouvelle méthode de collecte de données démographiques, la phase de mise au point à Marseille aura pour objectif principal d'apporter des réponses à deux questions :

- *quelle est l'information morphologique reliée à la population et comment l'intégrer dans la définition du plan de sondage pour optimiser sa précision?*
- *cette information morphologique est elle contenue dans l'image satellite ? si c'est le cas, quelle chaîne de traitements permet de la mettre en évidence?*

Seule la première question nous intéressera ici, la deuxième ayant trait à l'interprétation de l'image satellite.

A Marseille, nous disposons d'une part des données démographiques exhaustives du recensement et d'autre part de la couverture par photographies aériennes complète de la zone d'étude. En revanche, l'analyse de l'information satellitaire n'en n'est qu'à ses prémices. La question doit donc être abordée, dans un premier temps, en considérant l'information morphologique issue de la photo-interprétation et de l'observation du terrain. Ce sont les résultats obtenus dans cette étape qui permettront d'orienter la recherche en télédétection vers des critères morphologiques pertinents. A nouveau, selon le type de méthode applicable, distinguons deux phases :

- Première phase : *Définition d'une typologie et d'une stratification de l'espace urbain sur des critères morphologiques (question 6 du § 3.1)*

Le travail correspondant comporte les étapes suivantes :

1 : Recenser les caractères morphologiques observables, soit sur les photographies aériennes soit au sol, et proposer pour chacun une mesure ou un codage numérique.

2 : Etudier statistiquement les variables morphologiques obtenues et sélectionner un groupe de variables qu'on suppose liées au thème démographique et dont l'expression numérique offre une répartition multimodale ou plus généralement discriminante.

3 : Trouver une méthode d'analyse typologique qui fournisse sur ce type de données une stratification de la base de sondage, ici les îlots INSEE, en ensembles morphologiquement homogènes.

Comme on le voit les méthodes appliquées ici sont celle de l'analyse descriptive des données, mono ou multi-dimensionnelles. Ce sont les techniques d'analyse factorielle des correspondances multiples et de classification ascendante hiérarchique qui sont apparues les mieux adaptées aux données analysées, par conséquent elles ont été retenues pour la phase d'analyse typologique.

- Deuxième objectif : *Mesurer le gain de précision qu'apporte une stratification morphologique dans l'estimation par sondage de données démographiques et plus généralement, tester l'ensemble des méthodes d'amélioration possibles au premier degré du sondage (questions 4, 5, 7, 9, 10, 11, 12, 13 et 14 du § 3.1)*

Dans l'expérience à Marseille, les tests des différents plans de sondage n'impliquent pas, comme nous l'avons vu, la réalisation d'enquêtes démographiques. La base de sondage sera toujours l'ensemble des îlots INSEE et les résultats sont obtenus en effectuant sur la base entière le calcul de la variance des estimateurs. Cela permet de mesurer réellement la précision des différents plans mais, en contrepartie, on n'a pas la liberté du choix de l'unité spatiale à enquêter: on est tenu au découpage en districts de recensement fait par l'INSEE. De surcroît, on ne peut étudier par cette méthode et à

partir de ces données, que le premier degré du sondage. On ne dispose en effet à Marseille d'aucune donnée concernant individuellement les ménages.

Dans cette phase, la méthode est, on le voit, très simple. Le travail principal consiste à trouver l'expression mathématique, et à programmer le calcul, des variances associées à chacun des plans de sondage étudiés. L'outil est donc ici un programme informatique qui met en œuvre la théorie classique des sondages : sondages stratifiés, sondages à probabilités inégales, sondages systématiques (approche par le sondage stratifié) estimateurs par le quotient etc.

C'est ce même outil que nous appliquerons, à la fin de l'expérience marseillaise, à l'évaluation des premières stratifications issues de l'information satellitaire.

3.3.2 : Quito, juillet 86 - septembre 88 : première application de la méthode

La phase de mise au point de la méthode à Marseille s'attache exclusivement à la définition du premier degré du sondage. Dans la phase de première application de la méthode à une situation réelle de ville d'un pays en développement, il faut donc tout d'abord mener des recherches complémentaires. Pour disposer de tous les éléments nécessaire à la définition de la méthode, nous fixons à ces recherches deux objectifs principaux :

- 1. la vérification et éventuellement l'adaptation des résultats concernant la stratification et le premier degré du sondage obtenus au cours de l'expérience faite à Marseille.*
- 2. La définition de la base de sondage et du plan de sondage complet -premier et second degrés - que nous devons appliquer lors de l'enquête finale.*

Pour mener à bien ces deux tâches nous réaliserons tout d'abord, avec les autres membres de l'équipe, deux enquêtes préliminaires. La première aura pour but de recueillir sur un échantillon d'îlots représentatif des différents types de quartiers existant dans l'agglomération, l'information morphologique nécessaire à l'initialisation des classifications de l'image satellite qui fourniront la stratification. La seconde, réalisée sur un sous-échantillon du premier échantillon, collectera l'information démographique relative aux ménages résidant dans les îlots, les données obtenues devant permettre de statuer sur les taux de sondage qui devront être pratiqués aux premier et second degrés du sondage de l'enquête finale. L'ensemble des données collectées, c'est à dire le croisement des informations démographiques et morphologiques, serviront également à vérifier que les résultats obtenus à Marseille concernant les relations bâti-population et plus généralement l'amélioration du sondage au premier degré par la prise en compte de l'information morphologique, restent valides. Lorsque ce ne sera pas, ou pas tout à fait le cas, les résultats des enquêtes devront permettre "d'ajuster ces résultats" aux spécificités du site d'application.

Les méthodes d'analyse statistiques appliquées aux données recueillies lors de ces deux enquêtes préliminaires pour satisfaire les deux objectifs de cette phase du travail seront également de deux types.

1. Les techniques de regression et modélisation mathématique, parmi lesquelles nous avons choisi de n'utiliser que les plus simples : regression linéaire simple ou multiple et analyse de la variance à un ou plusieurs critères.
2. La méthode du calcul des variances associées aux divers plans de sondage considérés que permet de mettre en œuvre les outils informatiques mis au point lors de l'expérience marseillaise. Ces outils ont été étendus aux plans de sondage à deux degrés

pour permettre de les appliquer au problème de la détermination des taux aux premier et second degrés. A Quito nous les appliquons bien entendu aux seules données, partielles, dont nous disposons grâce aux enquêtes préliminaires.

Mise à part l'évaluation des gains de précision permis par les stratifications (question 6 du § 3.1), qui se fera toujours en employant la seconde méthode, tous les autres problèmes (questions 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13 et 14 du § 3.1) ont été abordés simultanément avec les deux méthodes, afin d'obtenir confirmation des résultats de l'une par les résultats de l'autre.

DEUXIEME PARTIE

**MISE AU POINT DE LA METHODE A MARSEILLE, STRATIFICATIONS ET
APPROCHE DU PREMIER DEGRE DU SONDAGE**

Introduction

OBJECTIFS GENERAUX ET PRINCIPE DE LA METHODE D'EVALUATION DES

PLANS DE SONDAGE

Comme il a déjà été dit, l'idée directrice de cette recherche est d'utiliser l'image satellite tout d'abord comme base pour la sélection d'un échantillon d'unités aérolaires à enquêter, puis comme source d'information pour obtenir une stratification de la base de sondage sur des critères de morphologie urbaine. Dans ce cadre, notre propre travail consiste à évaluer la précision des estimations que permettent les plans de sondages bâtis sur cette idée et à construire, sur la base de ces évaluations, un sondage, le plus précis possible, qu'on puisse facilement mettre en œuvre. Dans le chapitre 3, on a vu que la recherche se déroule en deux étapes principales. La première, essentiellement méthodologique, s'est déroulée en France à Marseille, ville sur laquelle nous disposons d'informations exhaustives sur la démographie et la morphologie de l'agglomération qui ont permis de mesurer les variances réelles des estimations obtenues avec différents sondages.

Cette deuxième partie est consacrée au compte rendu du travail effectué durant cette première étape et à l'exposé des résultats obtenus. Dans cette introduction, après avoir dressé la liste des objectifs fixés à cette expérience, nous revenons rapidement sur la méthode empirique qui a été appliquée tout au long du travail, puis nous indiquons dans quel(s) chapitre(s) les principaux points sont abordés.

- Les objectifs

Une des hypothèses fondatrices du projet, et qu'il va s'agir ici de vérifier, est que l'emploi d'une stratification morphologique améliore beaucoup la précision de l'estimation par sondage des caractères démographiques. De plus, comme nous le verrons par la suite, la technique du tirage aérolaire systématique sur document satellitaire induit un plan de sondage particulier dont il va falloir évaluer les performances en terme de précision des estimations produites. Pour mener à bien l'expérience marseillaise, il nous faut donc accomplir les étapes suivantes :

1 : Constituer et intégrer dans le système d'information géographique SAVANE, la base de données morphologiques, démographiques et satellitaires nécessaire à l'expérience.

2 : Effectuer l'analyse typologique des données morphologiques pour construire une stratification de l'espace urbain Marseillais.

3 : Mettre au point le formulaire et l'outil informatique nécessaire à l'évaluation des plans de sondage stratifiés.

4 : Mesurer, à partir des données du recensement général de la population à Marseille l'efficacité d'une stratification morphologique de la ville pour estimer par sondage les caractéristiques démographiques des habitants.

5 : Définir, après évaluation des différentes possibilités, les caractéristiques que doit avoir un plan de sondage utilisant comme base l'image satellite.

6 : Identifier les principales spécificités du plan de sondage défini et analyser les conséquences qu'elles ont sur la précision des estimations et la mise en œuvre pratique de l'enquête.

- La méthode

Comme nous l'avons déjà signalé, nous avons adopté, tout au long de cette recherche, une méthode empirique. En effet, l'un des buts poursuivis, est de vérifier, par expérimentation à partir des données exhaustives fournies par le recensement, l'hypothèse de départ du projet : la prise en compte, dans la stratification de la base de sondage, des différences morphologiques intra-urbaines, permet un gain appréciable de précision dans l'estimation des caractéristiques démographiques des habitants. Cette vérification, ainsi que la plupart des autres résultats du travail, sont obtenus grâce au programme VARECH (voir chapitre 6) qui, en appliquant le formulaire des principaux plans de sondage classiques, *permet le calcul exact sur les données du recensement de population des variances des différents estimateurs*. Ainsi, la démarche expérimentale suivie ici adopte comme moyen principal l'évaluation chiffrée d'un ensemble de possibilités. Elle doit permettre d'effectuer en connaissance de cause les choix techniques conduisant à un sondage efficace en milieu urbain et pratiquement réalisable sur document satellitaire. Comme nous l'avons dit dans l'introduction générale : un des buts de ce travail est d'accumuler le maximum de constatations pratiquement exploitables lors de l'application à Quito, sans toujours en chercher l'explication et sa formalisation mathématique.

- Plan de l'exposé

La première étape¹ du travail à Marseille est donc la constitution, sur cette ville choisie comme premier site d'expérimentation, d'une base de données morphologiques et démographiques localisées. Cette base comprend des informations issues des photographies aériennes, l'image satellite Thematic Mapper et une sélection d'indicateurs démographiques fournis par le recensement général de la population effectué par l'INSEE à Marseille en 1982. L'ensemble de ces informations est géré informatiquement par le système d'information géographique SAVANNE développé par l'Unité d'Infographie de l'ORSTOM. Le chapitre 4 rend compte de ce travail, effectué par l'ensemble de l'équipe, et précise les plans d'expérience mis au point pour satisfaire les différents objectifs évoqués ci-dessus.

Le travail d'analyse typologique des données morphologiques issues des photographies aériennes fait l'objet du chapitre 5. Il conduit comme on le verra à une stratification morphologique synthétique de l'agglomération marseillaise qui servira ensuite à deux titres :

1. Permettre le tirage d'un échantillon de districts INSEE représentatif des divers types de morphologie urbaine présents à Marseille, sur lequel s'effectue l'étude précise des relations entre la morphologie observée au sol (ou sur photographies aériennes) et

¹ : Etape réalisée de juin 85 à avril 86.

l'information spectrale satellitaire. Cette étude doit conduire les spécialistes en télédétection de l'équipe à produire, directement d'après l'image satellite, des stratifications utilisables pour réaliser des enquêtes par sondage.

2. Permettre, à partir des données du RGP 82 à Marseille, d'évaluer l'efficacité d'une stratification morphologique de la ville pour estimer par sondage les caractéristiques démographiques des habitants. Ces mesures seront répétées pour divers plans de sondage, différant soit par la probabilité de sélection des unités statistiques, soit par l'estimateur employé.

Le chapitre 6 décrit l'étape de mise au point de l'outil informatique nécessaire pour aborder ce second objectif et la suite des évaluations.

Pour cerner le problème en introduisant le principe du test et un formulaire simple, on considère dans la première partie du chapitre des plans de sondage équiprobables et des estimateurs sans biais. Ensuite le formulaire est étendu aux plans à probabilités inégales et aux estimateurs par le ratio. Il ne faut en effet pas perdre de vue que, dans la situation concrète d'application de la méthode, on ne disposera pas, a priori, d'une liste des unités composant la base de sondage. Le tirage s'effectuera à partir d'une grille de points plaquée sur une image satellite; cette image ne comportera que les limites externes de l'espace urbain et celles des strates morphologiques. Avec cette technique, et cela quelque soit l'unité spatiale considérée, le tirage s'effectue avec des probabilités proportionnelles à la surface de ces unités. Le programme que l'on décrit à la fin du chapitre permet donc notamment de comparer ce plan de sondage, imposé par la technique d'échantillonnage et que l'on devra appliquer à Quito, à son concurrent classique, le plan équiprobable sans remise.

Les trois premières étapes achevées on peut, à partir de la base de données constituée et de l'outil informatique mis au point, aborder les trois dernières étapes qui ont trait à l'évaluation proprement dite des possibilités existantes pour définir une technique de sondage adaptée et efficace. C'est l'objet des chapitres 7 et 8 qui traitent donc de l'application du programme VARECH aux données démographiques issus des chiffres du recensement général de la population de 1982 à Marseille ; dans cette phase, on exploite également, pour stratifier la base de sondage, les données morphologiques et satellitaires intégrées dans la base.

Le chapitre 7, dans ses deux premières parties, présente les résultats obtenus pour plusieurs indicateurs démographiques à l'aide de divers plans de sondages et évalue l'intérêt d'une stratification morphologique synthétique de la base de sondage. Dans sa troisième partie on s'intéresse aux stratifications fondées sur le seul critère morphologique de la densité du bâti, stratifications construites d'après l'information tirée de la photo-interprétation ou d'après l'information satellitaire, et on mesure le gain de précision qu'elles apportent.

Le chapitre 8 examine l'apport spécifique du caractère systématique du sondage c'est à dire la bonne répartition spatiale des unités sondées ; il tente par une méthode d'approximation, d'évaluer le gain de précision attaché à ce type de sondage.

Enfin dans la conclusion, en faisant la synthèse des principaux résultats, on identifie les acquis qui a permis ce travail pour la définition de la technique de sondage à mettre en oeuvre à Quito et on évoque les grandes lignes du travail restant à effectuer.

Chapitre 4

BASE DE DONNEES, OBJECTIFS ET PLAN D'EXPERIENCE A MARSEILLE

Introduction

Comme on l'a souligné dans le chapitre 3, la phase marseillaise de l'expérience correspond avant tout à une étape de développement méthodologique pour les trois principaux axes de recherche que comprend le projet :

- l'intégration de données satellitaires dans le système d'information géographique mis au point par l'unité d'infographie de l'ORSTOM
- l'interprétation des images satellites à haute résolution en milieu urbain
- la mise au point et l'évaluation de sondages spatiaux pour les enquêtes démographiques en milieu urbain.

La zone d'étude, les données utilisées et l'outil informatique qui permet la gestion de ces données, et une partie de leur traitement (le SIG SAVANNE), sont souvent communs aux trois axes de recherche. En revanche, les buts poursuivis sont spécifiques à chaque axe mais nous avons vu qu'ils sont également très dépendants les uns des autres. Le travail effectué sur le troisième axe, dont nous exposons les objectifs généraux en introduction à cette deuxième partie et dont il sera question dans les chapitres suivants, s'inscrit donc dans un schéma d'expérience global, conçu pour permettre une avancée simultanée dans les trois directions. C'est pourquoi il nous semble utile de présenter la zone d'étude, les données utilisées et leur intégration dans la base gérée par SAVANNE, puis de décrire rapidement le plan général de l'expérience pour y situer ensuite nos propres objectifs et les détailler : recherche d'une typologie des districts de recensement fondée sur l'information morphologique et mise au point (et évaluation) de plans de sondage utilisant cette information comme base pour la stratification¹.

4.1 : La zone d'étude, les données intégrées et le système de gestion de ces données

4.1.1 : La zone d'étude²

La zone d'étude (voir localisation sur la Figure 4.1) a été déterminée de manière à *couvrir les différents types de quartiers de Marseille*, c'est-à-dire les quartiers industriels et de grands ensembles ouvriers du nord, les quartiers résidentiels du sud,

¹ : Ce chapitre est, en grande partie, la recompilation des annexes 2,5,7 et 8 du rapport final rédigés pour l'ATP "Télétection" du CNRS par Françoise DUREAU et Marc SOURIS. L'ensemble du rapport, et donc toutes les annexes citées dans ce chapitre, peuvent être consultés au CNRS ou au département SDU de l'ORSTOM.

² : Paragraphe repris de l'annexe 5 au rapport ATP/CNRS, p. 59 du tome 2.

Figure 4.1 - Localisation de la zone d'étude sur Marseille



le vieux centre, et les zones pavillonnaires de l'est de la commune. Cette zone rassemble 3534 districts de recensement.

4.1.2 Les données intégrées¹

Dans cette première phase méthodologique du programme de recherches, les travaux réalisés reposent sur *trois types d'information, quasi-concomitantes* :

- données *démographiques* issues du recensement de mars 1982
- caractéristiques *morphologiques* observées sur photographies aériennes à grande échelle et au sol (août 1982)
- *image satellite* Thematic Mapper (février 1983) .

Notons qu'à Marseille, il n'est pas réalisé de collecte de données démographiques d'après un sondage aréolaire bâti sur image satellite, comme ce sera le cas à Quito. Le recensement de 1982 est utilisé comme référence pour mesurer le degré de précision des résultats obtenus en mettant en oeuvre différentes stratégies de sondage.

Etant donné l'importance qu'ont ces données dans la mise au point de la méthode que nous proposons, il n'est pas inutile, en les présentant chacune de manière un peu plus détaillée, de faire l'analyse de l'information disponible en insistant sur les caractéristiques thématiques, statistiques et surtout spatiales des données, qui prennent une dimension particulière dans le cas du traitement informatisé de données localisées.

Données démographiques

Le recensement, réalisé en mars 1982 sur l'ensemble du territoire français, a donné lieu à différents types de résultats, en fonction des étapes du schéma d'exploitation des questionnaires : le sondage au 1/20 puis le sondage au 1/4 et enfin les résultats exhaustifs ; ces derniers sont issus de la fusion d'un extrait des variables du fichier du 1/4, et du fichier provenant de la saisie des 3/4 des bulletins collectés restants. La saisie des 3/4 n'est que partielle ; en fait, des deux pages du questionnaire que chacun de nous est censé avoir rempli pour le recensement, ne sont saisies sur les 3/4 des questionnaires, et donc exploitées pour l'exhaustif, que les informations mentionnées dans le tableau ci dessous.

Données individuelles saisies pour les résultats exhaustifs du recensement de 1982 (INSEE)

Sexe
Situation de famille
Date de naissance
Nationalité (français, étrangers)
Département de résidence au 1.01.1975
Indicateur de scolarisation des enfants
Type d'activité (actif ayant un emploi, chômeur, élève ou étudiant, autre inactif)
Statut de la profession exercée (salarié, non salarié)
Caractère agricole de la profession exercée (agricole, non agricole)
Département et commune du lieu de travail des actifs ayant un emploi

Il s'opère donc une importante réduction de l'information collectée lors de l'exploitation exhaustive mais, cette exploitation étant la seule qui permette de

¹ : Paragraphe repris des annexes 5 et 6 au rapport final ATP/CNRS, pp. 59, 131 à 137 du tome 2.

travailler à un niveau géographique suffisamment fin, on est dans l'obligation d'y recourir pour toute étude démographique précise en milieu urbain. C'est donc sur les résultats exhaustifs que nous travaillons sur Marseille où nous disposons de *la bande magnétique des résultats exhaustifs sur la commune*, qui comprend, outre l'effectif de population, les répartitions de cette population selon les variables citées dans le tableau, prises indépendamment ou en les croisant.

Examinons maintenant les *caractéristiques du découpage géographique correspondant aux résultats fournis par l'INSEE à partir de l'exploitation exhaustive du recensement*. Pour réaliser le recensement de population, l'INSEE comme tous les instituts statistiques nationaux divise l'ensemble du territoire français en districts de recensement. Plusieurs critères interviennent dans ce découpage :

- *Un critère administratif*, lié à la fonction première du recensement qui doit fournir des chiffres de population légale pour les différentes unités territoriales françaises : ainsi, les districts de recensement doivent respecter les limites des sections cadastrales, des cantons et des communes.

- *Un critère démographique* : chaque district doit pouvoir être recensé par un seul agent recenseur dans le laps de temps imparti pour le recensement. C'est pourquoi on considère en France qu'un district doit correspondre à environ 800 personnes.

- *Un critère physique* : les voies de circulation automobile ou ferroviaire, les cours d'eau et autres limites naturelles sont en principe privilégiées par l'INSEE pour découper le territoire en unités de recensement.

La prise en considération de ces différents paramètres aboutit à un découpage spatial de l'ensemble du territoire en unités de taille très inégale, cette dernière variant grossièrement comme l'inverse de la densité de population. L'introduction de contraintes administratives et démographiques dans le principe de découpage *fait perdre, dans un certain nombre de cas, la notion d'îlot urbain au sens "physique" du terme*, c'est-à-dire le pâté de maisons.

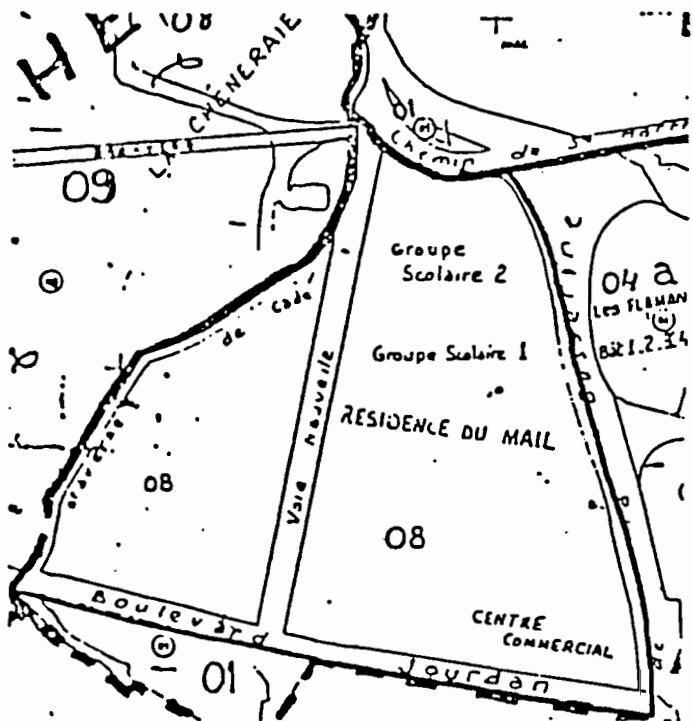
La figure 4.2 illustre quelques unes de ces divergences :

- un district de recensement peut englober plusieurs pâtés de maisons (exemple 1),
- un district de recensement peut être composé d'un ou plusieurs pâtés de maisons, auquel on a ajouté une portion d'un autre pâté de maisons (exemple 2),
- une limite peut diviser un seul et même bâtiment en deux districts de recensement (exemple 3),
- dans la majeure partie de Marseille, les ronds points au centre des places constituent chacun un district de recensement (exemple 4), tandis que d'autres portions du territoire urbain ne sont affectées à aucun district.

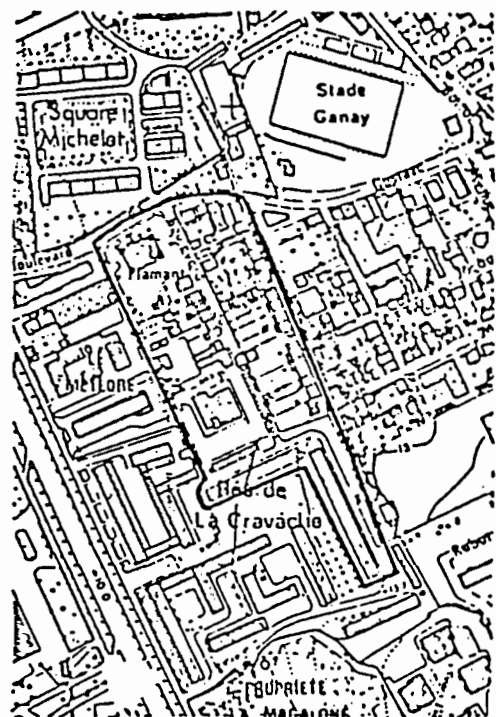
Il n'existe donc pas de relation bi-univoque entre le découpage physique du territoire déterminé par les bâtiments, rues, voies ferrées et cours d'eau, et le découpage de l'INSEE ; certains "héritages" de limites légales aboutissent à la situation la plus paradoxale, où une limite traverse un bâtiment d'habitation.

Il n'en demeure pas moins, après ces quelques remarques "critiques", que l'information censitaire est pour nous la référence indispensable à l'évaluation rigoureuse de la précision des procédures d'échantillonnage. Bien sûr les unités spatiales sur lesquelles est donnée l'information démographique (districts INSEE) ne coïncident pas toujours exactement avec celles que l'on identifiera sur l'image satellite comme unités primaires de sondage (îlots physiques). Mais, outre qu'il n'existe pas d'autre source à ce niveau de finesse géographique, on peut considérer que les différences existant entre les deux "bases" auront peu de conséquences sur les valeurs des estimations et de leurs variances.

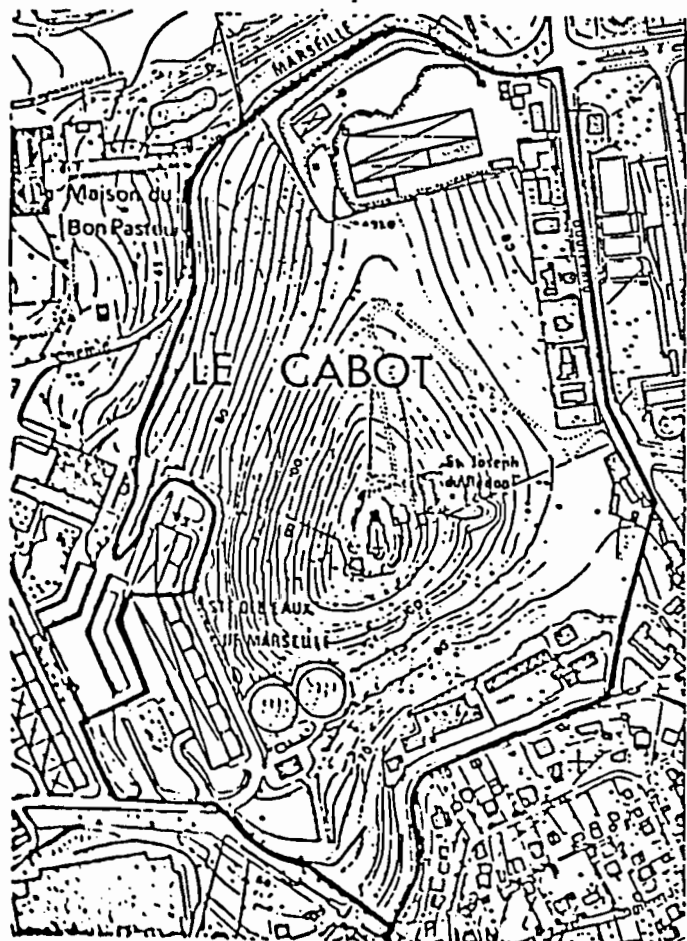
Figure 4.2 - Marseille . Découpage INSEE en districts de recensement



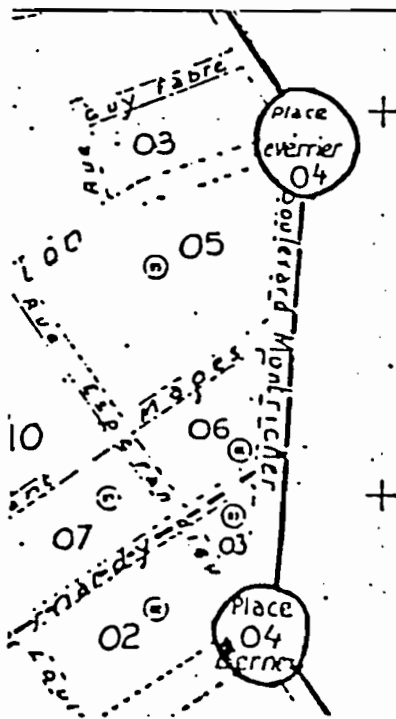
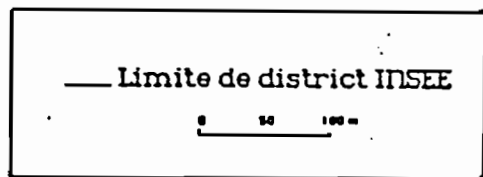
Exemple 1



Exemple 2



Exemple 3



Exemple 4

Données sur la morphologie urbaine

La connaissance de la morphologie urbaine sur l'ensemble de la zone d'étude repose sur *l'analyse visuelle d'une mosaïque de photographies aériennes au 1/23 000* (mission IGN d'août 1982), avec contrôle terrain .

L'analyse de la mosaïque photographique a été réalisée, classiquement, en deux étapes : zonage en unités homogènes, puis description des 542 zones ainsi dégagées selon des caractères morphologiques. Les traitements d'analyse des données sur le fichier descriptif des zones (voir chapitre suivant) et les vérifications sur le terrain ont permis de définir une *typologie morphologique en huit postes*.

A l'issue de l'analyse visuelle et du traitements d'analyse typologique, nous disposons donc d'une information homogène sur l'ensemble de la zone d'étude, décrivant la morphologie urbaine en 1982 , de façon synthétique (typologie en 8 postes), et analytique (onze descripteurs pour chacune des zones) ; *le découpage spatial dans lequel s'inscrit cette information lui est propre, et ne tient aucun compte du découpage établi par l'INSEE, ni de celui des îlots*. Il est néanmoins certain que le tracé des rues introduit une discontinuité sur la photographie aérienne, et a donc tendance à influencer sur le tracé des zones homogènes.

L'image Thematic Mapper

L'image dont nous disposons a été *enregistrée en février 1983 par le capteur Thematic Mapper*; la date d'enregistrement n'est pas la plus propice pour l'observation de la végétation, élément souvent très pertinent pour la discrimination des types d'habitat en milieu urbain. Mais, dans le but de minimiser le décalage temporel avec le recensement (mars 1982), nous l'avons préférée à une image de printemps enregistrée plus récemment, en juin 1984.

La principale caractéristique de l'information contenue dans les images satellite vient de leur mode d'enregistrement, qui se fait sur des unités élémentaires appelées "pixels", ayant une certaine surface (parallélogramme de 30 mètres de côté environ pour Thematic Mapper) : *les données satellitaires se présentent donc toujours sous une forme maillée*.

L'information dont nous disposons avec l'image Thematic Mapper correspond à l'enregistrement réalisé par le satellite dans six longueurs d'onde pour chacun des pixels de 30 mètres de côté, et dans l'infra-rouge thermique sur des pixels de 120 mètres. Chacune des unités élémentaires de l'image est donc décrite par *sept valeurs radiométriques*, pouvant chacune varier, théoriquement, de 0 à 255.

La complexité de l'interprétation d'image satellite en ville est liée à l'hétérogénéité et à la faible taille des objets constitutants : les valeurs enregistrées sur un pixel donné sont la résultante de toutes les composantes du pixel, un seul pixel pouvant être composé de portions de voirie, toiture, parking, espace vert etc. Etant donné le principe d'acquisition des données satellitaires, on ne peut comprendre la formation du signal enregistré par le satellite qu'en se référant à l'occupation du sol sur l'ensemble du pixel (et même des pixels voisins).

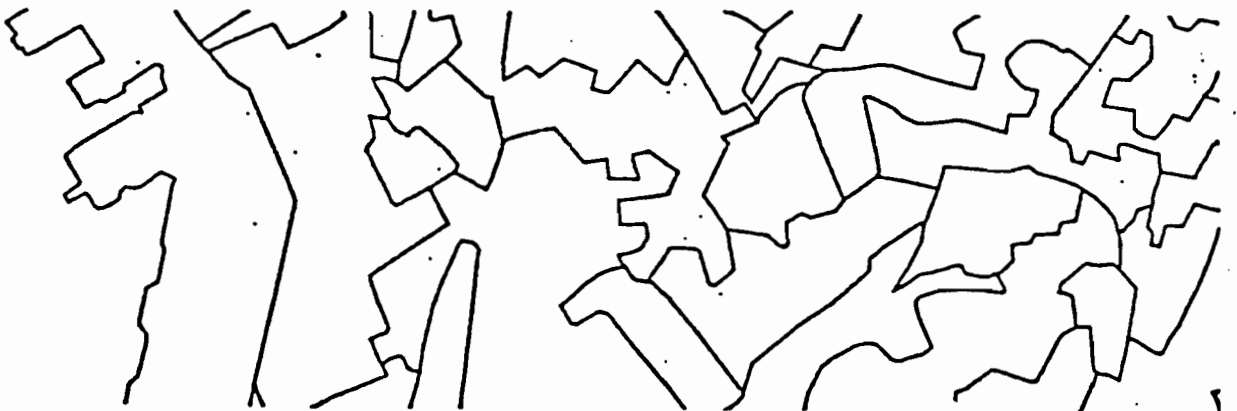
Comme le montre la Figure 4.3, *les trois types de données localisées à intégrer dans la base s'inscrivent dans des découpages géographiques différents*, ce qui justifie le recours au système TIGRE pour les mettre en relation et les gérer . En effet, un logiciel de cartographie automatique classique ne pourrait effectuer des traitements sans avoir ramené toutes les données à un découpage géographique unique, ce qui ferait perdre une bonne partie de son intérêt au croisement de données localisées tel que nous le concevons dans ce programme.

Figure 4.3 - L'implantation géographique des trois types de données à intégrer dans la base de données sur Marseille



Découpage INSEE en districts de recensement

0 250 500 m



Zonage morphologique

0 250 500 m

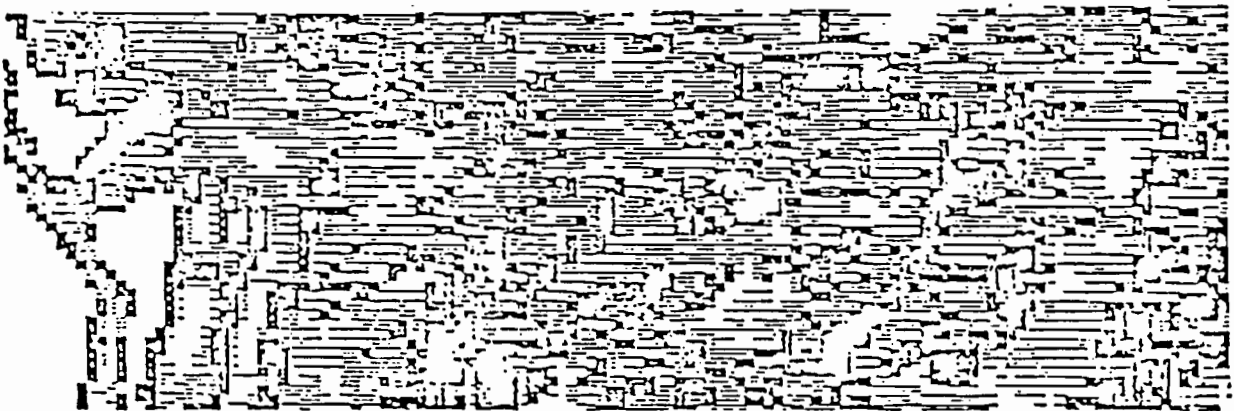


Image Thematic Mapper

0 250 500 m

Comme le met en évidence l'analyse des informations à intégrer dans la base de données sur Marseille, *ces différents découpages résultent bien sûr du mode d'implantation spatiale propre à un phénomène, mais aussi de la méthode de collecte et d'exploitation de l'information décrivant ce phénomène*. Ce deuxième facteur déforme nécessairement la réalité des faits que l'on cherche à observer : les données à intégrer dans la base correspondent à un certain niveau de médiatisation, donc à une certaine modélisation du réel, que nous maîtrisons plus ou moins selon la source d'information dont elles sont issues.

4.1.3 Le système de gestion de ces données et les étapes de la constitution de la base¹

Le système SAVANNE

L'ensemble de ces informations, connues sur toute la zone d'étude, est intégré dans une base de données gérée par le système d'information géographique SAVANNE. En effet, comme on l'a vu, les différents axes de recherche du programme impliquent la mise en relation de données localisées issues de sources d'information variées, et de nombreux traitements de données géographiques. C'est pourquoi nous avons choisi d'intégrer les trois types d'information dans une base de données (relative aux 3534 districts INSEE), malgré l'important travail que cette opération réclame : ainsi, nous nous trouvons dans une situation d'expérimentation quasi-idéale, en pouvant *croiser toutes les données localisées, chacune étant conservée dans son propre mode d'implantation spatiale*.

Ce n'est pas ici le lieu d'une description détaillée du système SAVANNE mais il nous semble important d'en faire une présentation rapide, du moins au niveau fonctionnel, puisque nous aurons souvent recours, dans la suite de ce travail, aux possibilités qu'il offre².

Le système SAVANNE, développé par l'unité d'infographie de l'ORSTOM, est un système d'information géographique organisé suivant le schéma des systèmes de gestion de base de données relationnelles, et enrichi d'opérations algébriques nouvelles pour la manipulation relationnelle de la localisation. Il réunit ainsi les possibilités des système de gestion de base de données et celles des système d'information géographique. Pour introduire brièvement trois concepts fondamentaux du vocabulaire des bases de données auxquels nous nous référerons par la suite, disons simplement que le modèle relationnel consiste à représenter les entités et associations du monde réel par des *relations*, ensembles de valeurs d'*attributs* décrivant les objets de ces entités ou associations, et à manipuler ces relations pour répondre à différentes *interrogations*. L'architecture du système est celle d'un calculateur de base de données simplifié : il possède un dictionnaire des données, indiquant les relations et attributs ainsi que leurs caractéristiques, un dictionnaire des accès permettant de gérer les niveaux externes et *un langage de commande qui permet à l'utilisateur d'interroger la base et de manipuler les données*.

Du point de vue fonctionnel, les modules de base du système sont les suivants:

- Un module de saisie graphique par digitaliseur
- Un module de description et de saisie de l'information descriptive
- Un module d'intégration des données dans la base
- Un module d'interrogation de la base de données

¹ : Paragraphe issu de la compilation des annexes 2 et 8 du rapport final ATP/CNRS, pp. 21, 26 à 28 et 147 à 149 du tome 2.

²: Ces éléments sont repris de l'annexe 2 au rapport ATP final (août 87) rédigé par Marc SOURIS : Le système Tigre, un système d'information géographique organisé selon le modèle relationnel. Pour plus d'information on peut se reporter à cette annexe, ainsi qu'à l'annexe 1 du même rapport : Base de données et information géographique, l'extension du modèle relationnel.

- Un module de cartographie automatique
- Des modules de visualisation d'images sur différents périphériques (écran graphique, traceur)
- Un module de création de nouveaux attributs
- Un module d'intégration de données satellitaires
- Un module de calculs géométriques sur écran (distances, périmètres, surfaces...)

La constitution de la base de données sur Marseille s'est effectuée en trois étapes que nous allons maintenant évoquer.

Digitalisation des fonds de cartes

La numérisation des cartes des districts de recensement et du découpage morphologique a été réalisée par l'Unité d'Infographie en une période de six mois environ (y compris contrôles et corrections).

La figure 4.4 présente le zonage morphologique numérisé au 1/25 000, tel qu'il est restitué sur table traçante à une échelle du 1/75 000 : sur cet exemple, ne figurent que les arcs (contours de zones) sans identification des zones par leur numéros.

La figure 4.5 correspond à un extrait du découpage INSEE (saisie réalisée au 1/2 000 ou 1/5 000 selon les quartiers de Marseille), restitué au 1/5 000, avec le code de district INSEE et le numéro de zone attribué lors de la saisie.

Intégration de l'information graphique

Après la phase de numérisation, les coordonnées table des périmètres de zones sont transformées en coordonnées géographiques (longitude, colatitude par rapport à un point de référence propre à la base), à partir des coordonnées géographiques connues pour les points de calage saisis en même temps que le fond cartographique. Ces coordonnées constituent l'information graphique de la relation "Population" de la base "Marseille".

Concernant le zonage morphologique reporté sur carte IGN au 1/25 000, il a été digitalisé en une seule feuille. L'intégration de cette information graphique de la relation "morphologie" et sa comparaison avec l'information décrivant le découpage INSEE a tout de suite révélé une incohérence sensible entre les deux découpages.

Etant donné l'échelle de digitalisation (1/5 000 et 1/2 000 pour le découpage INSEE, 1/25 000 pour la morphologie), on peut affirmer sans aucun doute que l'information graphique décrivant le découpage INSEE est bien plus fiable que celle relative à la morphologie : nous avons donc procédé à une correction des points de calage du zonage morphologique pour rendre cohérents les deux fichiers graphiques, INSEE et morphologie. C'est donc selon ces points de calage corrigés qu'a été réalisée l'intégration dans la base de l'information graphique décrivant le découpage morphologique.

Intégration de l'information thématique

L'information thématique relative aux fonds de carte numérisés constitue les "attributs", pour reprendre le vocabulaire des bases de données relationnelles : dans le cas de Marseille, on peut distinguer deux relations (*morphologie, et population*), avec deux séries d'attributs, l'une correspondant au zonage morphologique, l'autre aux districts du recensement.

Figure 4.4 - Marseille. Zonage morphologique d'après l'analyse visuelle d'une mosaïque photographique au 1/23 000 de 1982 (unité d'infographie, ORSTOM)

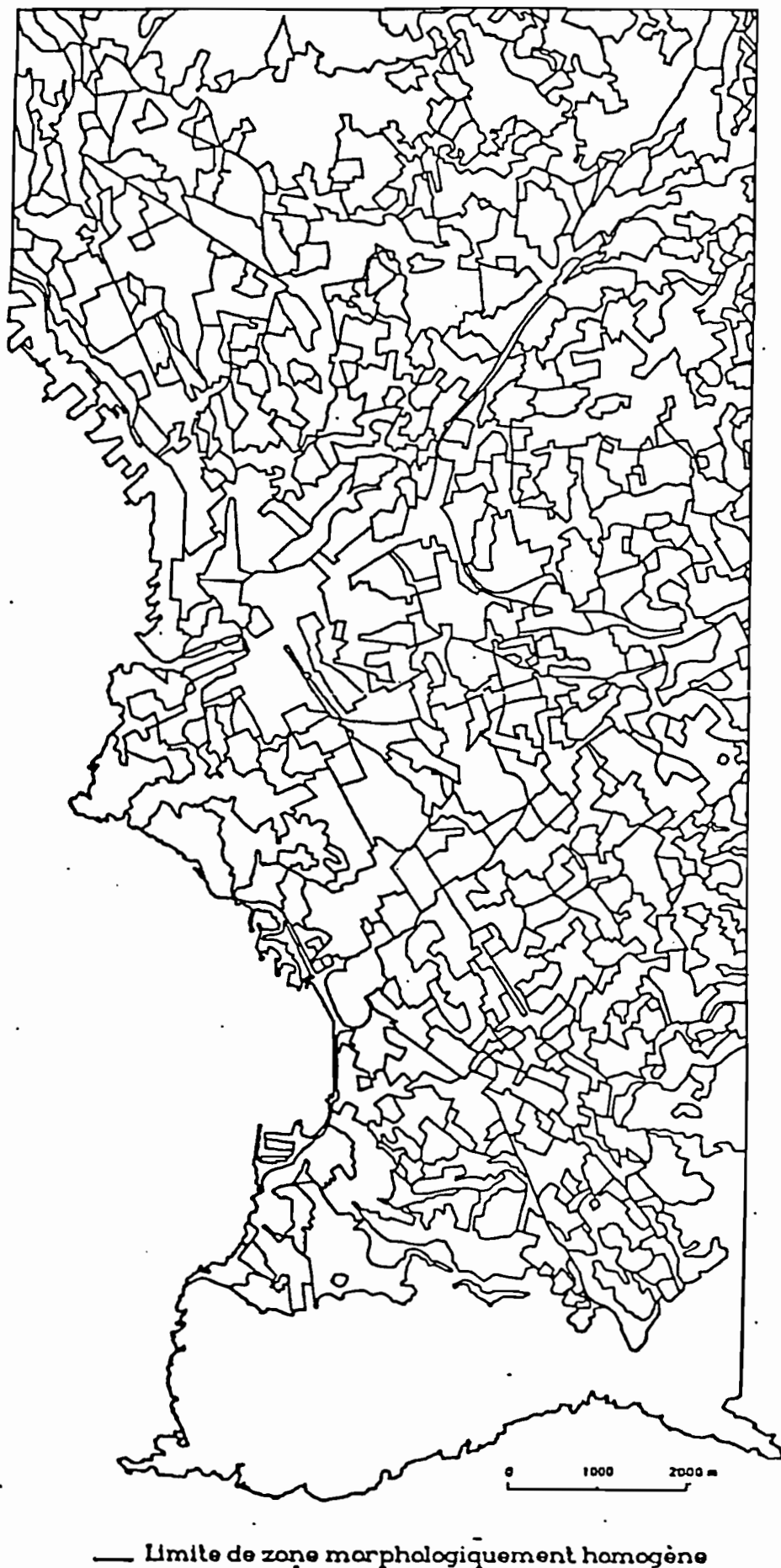
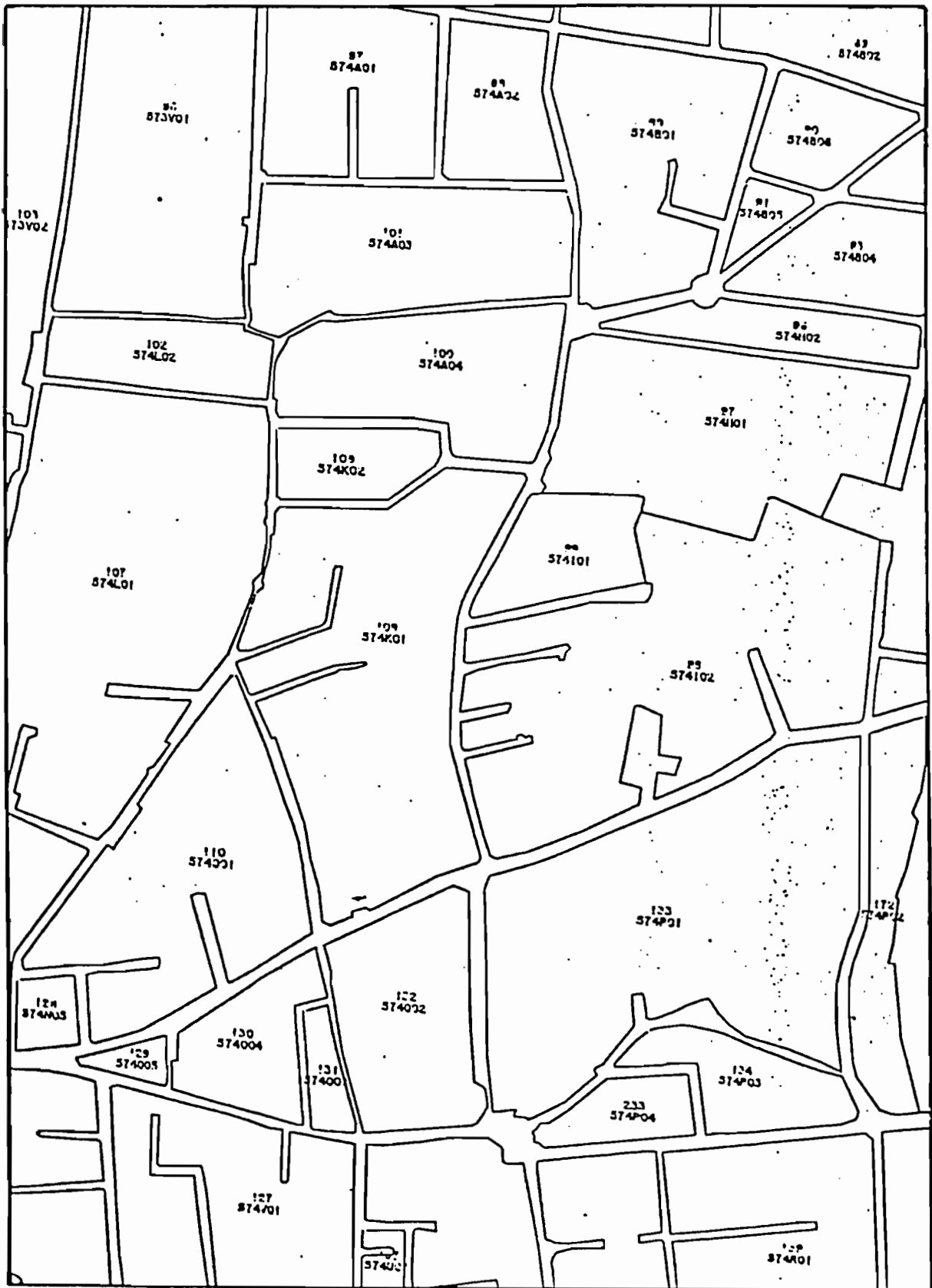


Figure 4.5 - Marseille. Découpage INSEE en districts de recensement, restitution au 1/5 000 (unité d'infographie, ORSTOM)



La liste des attributs a été définie par notre application dans le cadre de ce programme de recherche :

- *morphologie* (voir tableau 4.6) : vingt cinq variables décrivant analytiquement la morphologie, et un indicateur synthétique du type morphologique.

On a intégré dans la base les variables brutes, telles qu'elles ont été produites par la photo-analyse, les variables dérivées (exemple : rapport longueur / largeur) ainsi que les variables seuillées et recodées. Nous avons donc une certaine redondance d'information, une même variable pouvant apparaître trois fois dans la base, à des stades différents de recodage.

- *population* (voir tableau 4.7) : pour chaque district INSEE, on s'intéresse à neuf variables décrivant les nombres de logements et les effectifs de population selon différentes définitions liées à des critères de résidence, ainsi que huit variables décrivant quelques aspects importants des structures de population (voir en annexe A.2.2 les définitions INSEE des différentes catégories de population)

Tableau 4.6 - Relation morphologie. Liste des attributs

- 1 . Numéro de zone morphologique
- 2 . Densité
- 3 . Surface minimum de l'élément bâti
- 4 . Surface maximum de l'élément bâti
- 5 . Forme de l'élément bâti majoritaire
- 6 . Longueur de l'élément bâti majoritaire
- 7 . Largeur de l'élément bâti majoritaire
- 8 . Valeur dominante du bâti
- 9 . Valeur sous-dominante du bâti
- 10 . Valeur dominante du non bâti non voirie
- 11 . Valeur sous-dominante du non bâti non voirie
- 12 . Hauteur du bâti
- 13 . Surface de l'élément bâti majoritaire
- 14 . Rapport longueur/largeur de l'élément bâti majoritaire
- 15 à 24 : Attributs 2 à 4, et 8 à 14 seuillées et recodés
- 25 . Type morphologique (8 types + non bâti)

Tableau 4.7 - Relation population. Liste des attributs

1. Population municipale
2. Population sans double compte
3. Nombre de ménages ordinaires
4. Nombre de ménages collectifs
5. Population comptée à part
6. Nombre de résidences principales
7. Nombre de résidences secondaires
8. Nombre de logements vacants
9. Nombre d'immeubles
10. Pourcentages d'étrangers (Hommes + femmes)
11. Pourcentage d'étrangers (Hommes > 19 ans)
12. Pourcentage de chômeurs (Hommes > 19 ans)
13. Taille moyenne des ménages
14. Pourcentage de ménages d'une personne
15. Pourcentage de ménages de 6 personnes et plus
16. Pourcentages de 0-4 ans révolus (Hommes + femmes)
17. Pourcentage de 60 ans et plus (Hommes + femmes)

La base de données sur Marseille contient donc les données morphologiques et démographiques nécessaires au programme de recherche ; l'intégration des données satellitaires (image TM recalée) vient compléter le dispositif. Cette intégration a nécessité des travaux de développement informatique sur les deux systèmes, : système SAVANNE de l'Unité d'Infographie, et système de l'Unité de traitement Numérique de l'Atelier de Télédétection .

4.2 : Le plan général de l'expérience

Après une phase de travail consacrée à la constitution de la base de données sur Marseille, les trois directions de recherche du programme sont menées en parallèle, leurs résultats se conditionnant mutuellement ; le principe de l'expérimentation menée sur Marseille est résumé dans la figure 4.8. Ce schéma général de l'expérience détermine l'organisation de la fin de ce chapitre ; il appelle donc à quelques commentaires. Premièrement si l'on effectue une lecture verticale du graphique, on distingue trois colonnes principales, séparées par des tiretés gras, correspondant chacune à un des axes principaux de recherche. Notre intervention propre, portant sur les sondage spatiaux, est résumée dans la colonne de droite où apparaissent trois groupes de tâches à réaliser sur lesquels nous reviendrons. Les deux colonnes restantes figurent les étapes du travail sur l'intégration des données dans le système d'information géographiques (colonne centrale) et sur la recherche de méthodes d'interprétation des images satellite (colonne de gauche). Une lecture horizontale permet d'apprécier les différents "aller et retour" qui s'effectuent entre les trois axes. En ce qui nous concerne, l'échange d'informations et de résultats a lieu surtout avec l'équipe chargée de l'axe "développement de méthode d'interprétation d'image satellite", tandis que nous nous contentons, dans cette étape marseillaise, d'utiliser les avancées réalisées dans le domaine de l'intégration des données au S.I.G. Pour préciser un peu ces deux points, nous allons identifier et détailler les tâches principales qui figurent sur le schéma dans les colonnes de gauche et de droite.

Axe : "développement des méthodes d'interprétation des images satellite à haute résolution en milieu urbain"

Dans la colonne gauche du schéma, la distinction est faite entre les tâches visant d'une part à l'intégration de l'information satellitaire dans la base de donnée, regroupées sous l'appellation "travail préliminaire", et d'autre part la mise au point/validation des méthodes d'extraction de l'information morphologique sur image satellite, qui constitue, du moins par rapport à l'objectif fondamental du projet, la tâche centrale. L'ensemble du travail préliminaire s'effectue sur la totalité de la zone d'étude délimitée sur la figure 4.1. En revanche, tous les travaux méthodologiques correspondant à la tâche centrale ne sont pas réalisés sur l'intégralité de cet espace. En effet, il est nécessaire d'en *extraire un échantillon pour la mise au point de méthodes* relatives à l'interprétation d'images satellite : si l'on ne procédait pas ainsi, on ne pourrait pas vérifier la validité des traitements mis au point, c'est-à-dire tester qualitativement et quantitativement leur efficacité. *Travailler sur un échantillon permet de contrôler en permanence les méthodes développées* en les testant sur un espace disposant d'une information de référence : les données de population et de morphologie urbaine dans le cas présent. Pour que cette méthode de travail soit vraiment efficace, il importe de disposer d'un échantillon représentatif des différents types d'occupation du sol présents dans l'ensemble de la zone : c'est une des raisons pour laquelle nous réaliserons une *typologie de la morphologie urbaine*, d'après l'information observée sur photographie aérienne et contrôlée au sol, dont il sera question ultérieurement et qui fournira également une première stratification de la zone d'étude. Précisons dès à présent que, pour ce travail en télédétection urbaine, l'identification des types morphologiques sur image satellite n'est pas considérée comme un objectif à atteindre.

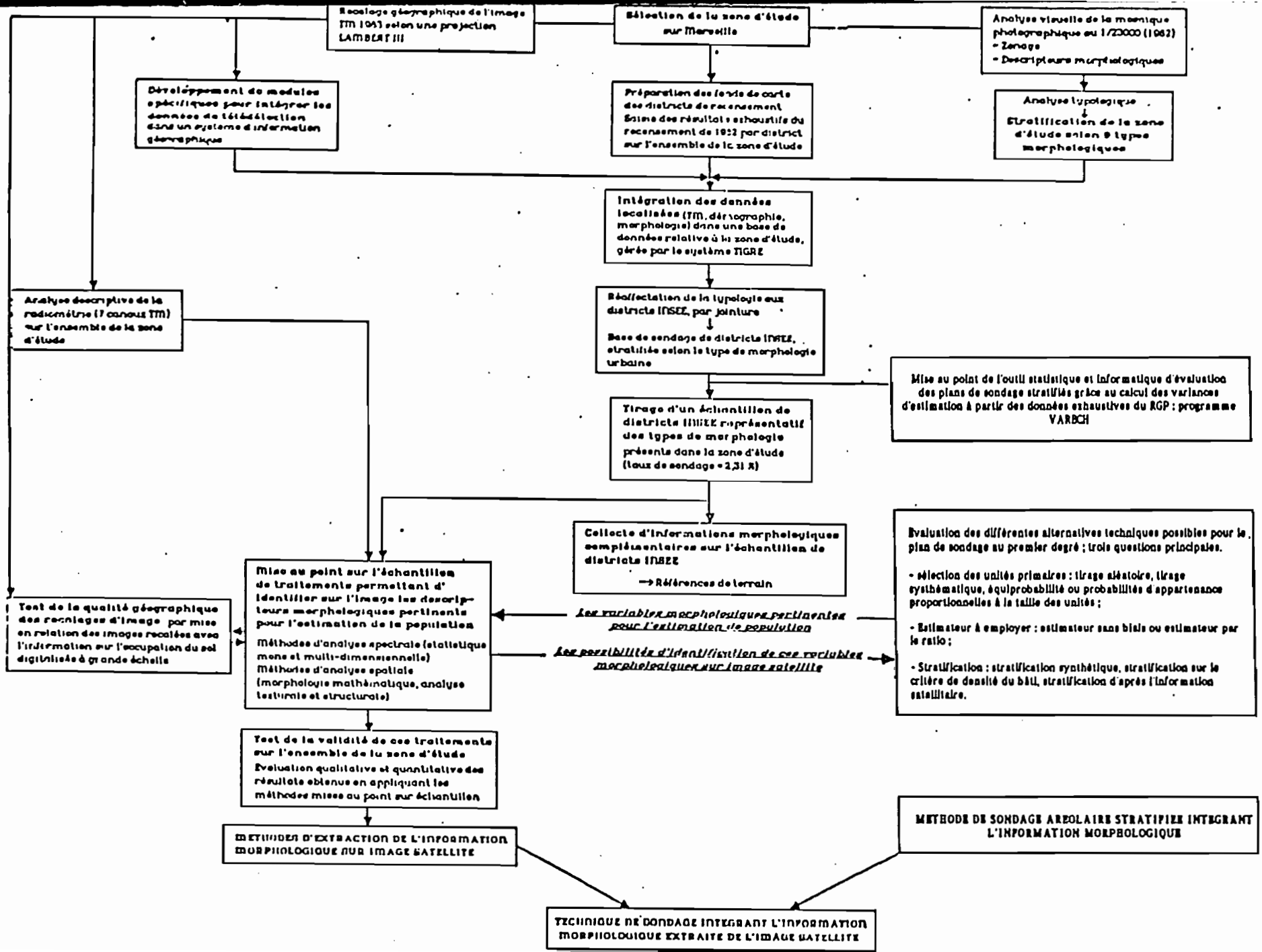


Figure 4.8 - Plan d'expérience sur Marseille

En effet, la possibilité d'identifier sur l'image satellite telle ou telle strate de l'agglomération marseillaise ne peut en aucun cas fournir la méthode de stratification de l'image d'un autre site : seule une *démarche de type analytique*, s'efforçant d'identifier des caractéristiques morphologiques simples de l'espace urbain peut permettre la *portabilité* de la méthode.

Axe : "mise au point et évaluation de sondages spatiaux pour les enquêtes démographiques en ville"

La première tâche identifiée dans cet axe, que nous appellerons également travail préliminaire, est donc l'établissement d'une typologie de la zone d'étude sur des critères morphologiques. Elle répond, comme on l'a vu, à une double nécessité :

- permettre la sélection d'un échantillon représentatif des différents types d'occupation du sol présents dans l'ensemble de la zone, échantillon sur lequel seront étudiées les possibilités d'extraction d'information morphologiques à partir de l'image satellite.

- fournir l'information morphologique qui servira à l'établissement des premières stratifications testée sur la base de sondage de Marseille.

Ainsi ce travail d'analyse typologique s'avère indispensable pour aborder les points centraux, tant pour les recherches en télédétection que pour celles portant sur les sondages.

En ce qui concerne les sondages spatiaux, pour mettre au point un système d'observation démographique intégrant l'information satellitaire, nous avons vu qu'il fallait répondre, à partir de l'expérience marseillaise, à un certain nombre de questions. Rappelons les deux principales :

- Quels sont les critères morphologiques pertinents pour établir la stratification spatiale d'une ville en vue d'une enquête démographique par sondage ?

- Quelle est la stratégie de sondage la plus efficace ? Comment doit-on stratifier ? Quel doit être le mode de tirage de l'échantillon (aléatoire ou systématique, selon quelle fréquence spatiale) et l'estimateur à utiliser ?

Ce sont ces questions qui définissent la tâche centrale de la colonne droite dans le schéma général de l'expérience à Marseille. Ici apparaît l'utilisation que nous ferons du S.I.G. : dans la recherche des éléments de réponse à ces questions, le système SAVANNE permet de mettre en relation et de traiter conjointement des données démographiques et morphologiques localisées. Concrètement, il effectuera la réaffectation des caractéristiques morphologiques tirées de la photo-interprétation et de l'analyse typologique aux districts de recensements INSEE. Cette information servira à tirer l'échantillon de districts INSEE représentatif des types morphologiques et à établir les premières stratifications de la base de sondage à Marseille.

Le principe d'évaluation des différentes stratégies de sondage, que nous appliquerons durant toute l'expérience à Marseille, est toujours le même. Il s'agit de soumettre les données exhaustives du recensement de population à un programme informatique qui, en appliquant les formulaires associés à un plan de sondage donné, calcule la variance exacte des estimateurs ; il n'est donc pas ici question d'estimation de cette variance. Donnons quelques exemples d'application de ce principe, liste d'ailleurs incomplète comme on le verra par la suite. Il servira à comparer des plans de sondages équiprobables à des plans de sondages où la probabilité de sélection est proportionnelle à la taille des unités d'échantillonnage. Il servira aussi à comparer les estimateurs sans biais aux estimateurs par le ratio et enfin, il servira à évaluer l'apport d'une stratification morphologique de la base de sondage à la précision des estimations et à comparer entre elles différentes stratifications. A ce propos, signalons que, comme dans la démarche adoptée en télédétection urbaine, *les critères morphologiques doivent également être testés analytiquement comme critères de stratification*. En effet, si la stratification "synthétique", issue de l'analyse typologique

effectuée à partir de l'ensemble des caractères morphologiques décrivant chaque zone, fournit un ordre de grandeur pour le gain de précision permis par un sondage stratifié, cette stratification n'a de sens qu'à Marseille où au mieux sur une ville très comparable. Seuls les caractères morphologiques de base pris isolément (par exemple la densité du bâti ou la taille moyenne des bâtiments etc...) peuvent être identifiés sur n'importe quelle ville. Il faut donc également évaluer les stratifications construites à partir de chacun de ces descripteurs analytiques, du moins pour les principaux d'entre eux.

Les grandes lignes du déroulement de l'expérience à Marseille étant définies pour les trois axes de recherche, nous abordons maintenant de façon plus détaillée les étapes de notre propre travail à Marseille : l'analyse typologique des données morphologiques et la mise au point/évaluation des plans de sondages stratifiés.

4.3 : L'étape d'analyse typologique des données morphologiques

Rappel des objectifs

Les objectifs de l'analyse des photographies aériennes sont, rappelons-le, au nombre de trois :

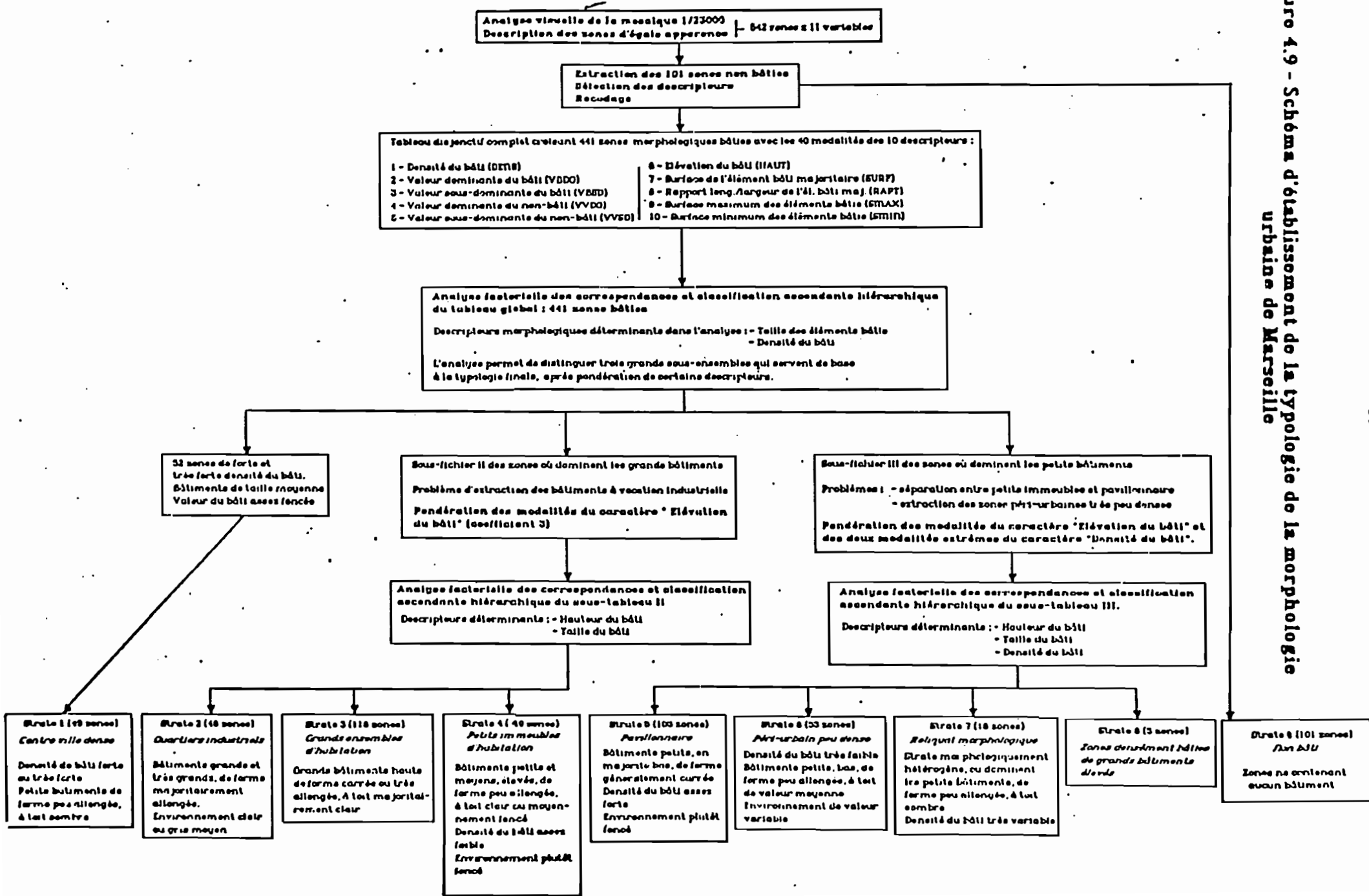
- *disposer d'une connaissance exhaustive de certains paramètres de la morphologie urbaine sur l'ensemble de la zone d'étude*, afin de pouvoir mettre au point les techniques d'interprétation d'image satellite et de tester leur validité : dans ce cas, l'information analytique sur la morphologie urbaine issue de la photo-analyse contrôlée par des vérifications au sol, sert de référence de terrain.
- *tirer un échantillon de travail représentatif des types d'occupation du sol présents dans notre périmètre d'étude de Marseille* : cette opération est nécessaire si l'on veut que la technique de mise au point des méthodes sur échantillon soit efficace. Il s'agit donc d'établir une typologie de la morphologie urbaine à partir des descripteurs observés sur photographies aériennes : cette typologie sert de base de stratification pour le tirage d'un échantillon de travail représentatif .
- *tester l'efficacité d'une stratification morphologique dans l'estimation de population par sondage* : le principe est de mesurer le gain de précision lié à la stratification morphologique pour différents plan de sondage.

Si le premier objectif repose sur les descripteurs analytiques directement issus de la photo-analyse, les deux autres supposent d'établir une typologie de la morphologie urbaine à partir de ces descripteurs.

Déroulement du travail

La figure 4.9 indique les différentes étapes du travail d'obtention et d'analyse typologique de l'information morphologique. Le premier travail consiste à extraire l'information morphologique du support choisi, les photographies aériennes, à la contrôler sur le terrain et à lui donner, par un codage approprié, une forme à partir de laquelle on puisse lui appliquer les méthodes classique de l'analyse des données.

Figure 4.9 - Schéma d'établissement de la typologie de la morphologie urbaine de Marseille



Les photos aériennes choisies ont été réalisées en août 1983 à Marseille dans le cadre de la couverture systématique du littoral français (mission IPLI - Lot 23 - échelle voisine de 1/23 000 - panchromatique - noir et blanc). Elles satisfont aux conditions fixées : le décalage temporel est réduit au minimum puisque les photographies aériennes ont été prises à une date intermédiaire entre le recensement et l'image Thematic Mapper, et l'échelle est tout à fait satisfaisante pour l'utilisation prévue ici. Une mosaïque photographique non contrôlée a été réalisée à l'ATOB ; c'est sur ce document qu'est faite la photo-analyse : zonage en unités d'égale apparence puis, description de chaque zone dans un tableau croisant les 542 zones obtenues avec 11 variables analytiques de la morphologie urbaine. Ces descripteurs morphologiques sont de diverse nature, au sens mathématique du terme : on trouve des variables métriques (mesure de longueur et largeures des bâtiments par exemple) aussi bien que des variables qualitatives (valeur de gris sur la photo ou niveau de densité du bâti). Afin d'harmoniser le tableau de description morphologique des zones pour pouvoir ensuite le recoder en un tableau disjonctif complet et soumettre ce derniers au traitements, on a procédé à une étude statistique des distributions de chaque descripteur et au seuillage des variables quantitatives. Notons qu'au cours de cette étude, un des onze descripteurs originaux a été abandonné parcequ'il n'introduisait manifestement pas de discrimination entre les zones.

Enfin, signalons que les informations extraites de la photo-analyse ont été contrôlées, par sondage, lors d'une enquête sur le terrain début novembre 1985.

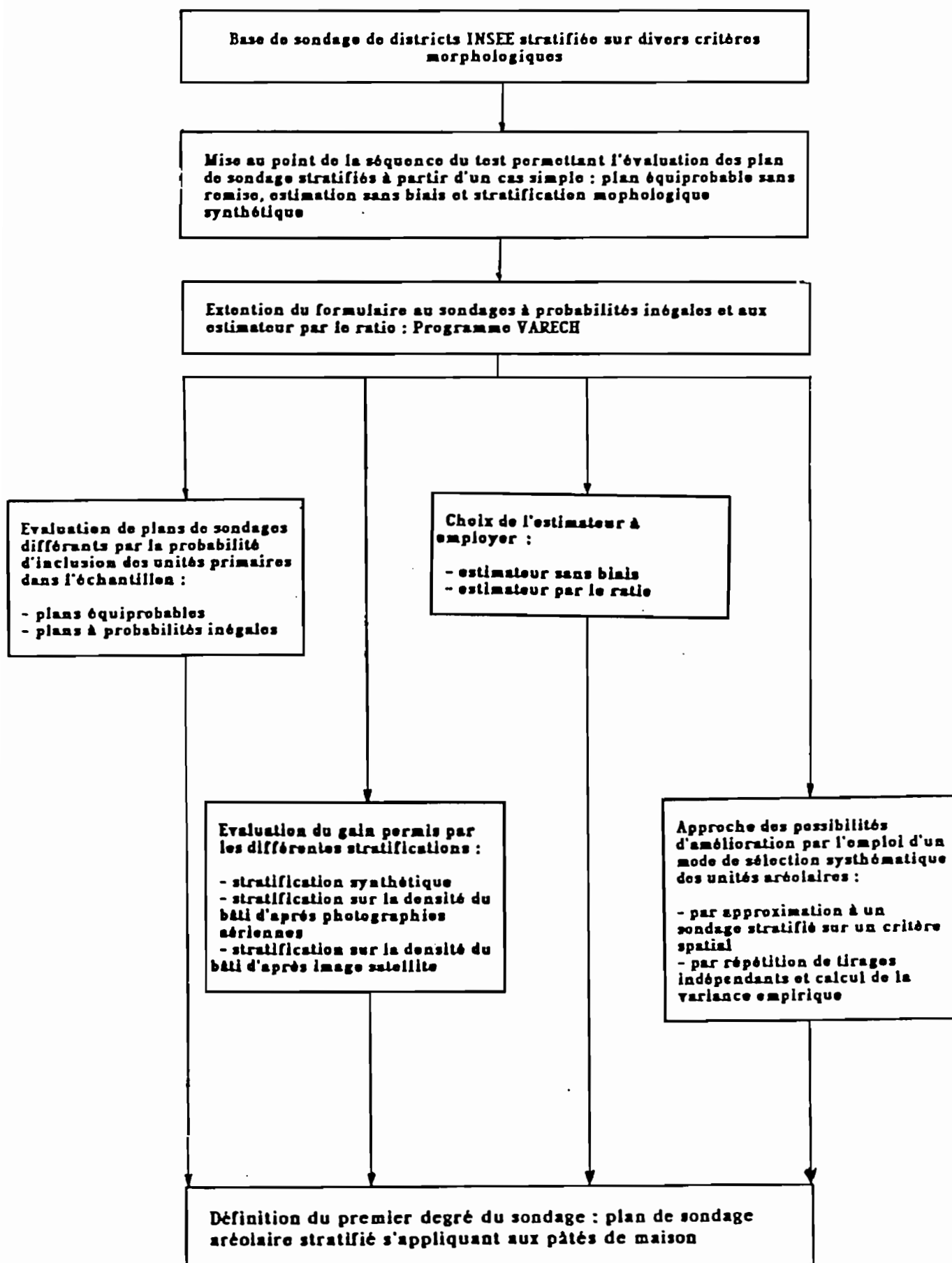
L'ensemble de ces opérations, préliminaires à l'analyse typologique proprement dite, sont décrites en détail dans l'annexe A.2.1 : "Obtention, analyse statistique et codage de l'information morphologique".

Une fois mise l'information morphologique sous une forme "exploitable", l'étape de classification des zones peut commencer. Soulignons tout d'abord que les espaces non bâtis (101 zones), identifiés lors de l'analyse visuelle de la mosaïque, sont exclus dès ce moment de la procédure d'analyse typologique. Le tableau disjonctif soumis aux traitements ne comprend par conséquent que 441 zones. Les 101 zones restantes constituent la strate non bâtie. Pour classer les 441 zones contenant des bâtiments et caractériser morphologiquement les classes, on utilise des outils désormais classiques en analyse des données : analyse factorielle des correspondance (AFC) et classification ascendante hiérarchique (CAH). Cependant, comme c'est souvent le cas lorsqu'on analyse des données dans un but thématique précis (ici établir une stratification pertinente du point de vue morphologique et démographique), il s'établit une sorte de dialectique entre le statisticien et le thématicien. Les résultats de la première analyse tyologique ont ainsi été confrontés aux objectifs des démographe et spécialistes en télédétection de l'équipe et, à partir des critiques formulées, on a ajusté le processus de classification pour obtenir une meilleure stratification. On aboutit finalement à une démarche de classification en deux temps schématisée dans la figure 4.9. L'exposé complet de ce travail d'analyse typologique fait l'objet du chapitre 5.

4.4 : L'étape de mise au point et d'évaluation de plans de sondages stratifiés

L'objectif général de cette deuxième phase de l'expérimentation menée à Marseille est, rappelons-le, de définir le plus complètement possible le plan de sondage au premier degré. C'est-à-dire qu'on va devoir procéder à l'évaluation et à la comparaison de différentes techniques de stratification, de sélection des unités primaires dans l'échantillon et d'estimation. Comme on l'a souligné dans l'introduction de cette troisième partie, nous appliquons pour résoudre ces problèmes une méthode empirique dont le principe est le suivant : pour chaque alternative technique envisagée, et, à chaque fois, pour une série de caractères démographiques, on calcule à partir des données exhaustives du recensement, les variances d'estimation sous le

Figure 4.10 - Schéma de l'étape de mise au point et d'évaluation de plans de sondages stratifiés à Marseille



plan de sondage correspondant. C'est ensuite la comparaison de ces variances qui permet d'effectuer l'ensemble des choix techniques.

Le travail d'évaluation comprendra donc deux phases :

1. la mise au point du formulaire et de l'outil informatique correspondant qui permettra de répéter les calculs pour l'ensemble des alternatives envisagées.
2. L'application de cet outil aux données du recensement et l'analyse des résultats pour parvenir aux conclusions.

La figure 4.10 indique les principales étapes correspondant à ces deux phases et leur enchaînement.

La phase de développement de l'outil commence par la mise au point de la séquence du test sur un cas simple : le sondage équiprobable sans remise avec estimation sans biais de totaux ou de moyennes et utilisation de la stratification morphologique synthétique résultat de l'analyse typologique des données issues de la photo-interprétation. Ce travail fait l'objet de la première partie du chapitre 6. On étend ensuite le formulaire au cas des plan de sondage à probabilités inégales et aux estimateurs par le ratio (deuxième partie du chapitre 6). Le résultat de cette phase est le programme VARECH présenté dans la troisième partie du chapitre et en annexe A.2.3.

Dans la phase d'application, on s'intéresse en premier lieu au problème du choix de la probabilité d'inclusion (équiprobabilité / probabilité proportionnelle à la taille des U.P.) et de l'estimateur à employer (première et deuxième parties du chapitre 7). On quantifie et compare ensuite les gains de précision permis par les différentes stratifications mises au point : stratification morphologique synthétique, stratifications sur le critère de la densité du bâti obtenue par photo-interprétation ou par classification de l'image satellite (troisième partie du chapitre 7). Enfin on aborde dans le chapitre 8 la question de l'influence sur la précision des estimations d'un mode de sélection systématique des unités primaires.

L'ensemble de l'expérience marseillaise trouve sa conclusion dans la définition du premier degré d'un sondage aréolaire stratifié qui nous semble adapté à l'utilisation de l'image satellite comme base de sondage spatiale, mais comme on le verra la mise au point complète et l'évaluation des performances de la méthode ne pourront être menées à bien que dans la suite de l'expérience qui se déroulera à Quito.

Chapitre 5

ANALYSE TYPOLOGIQUE POUR LA CONSTRUCTION D'UNE STRATIFICATION MORPHOLOGIQUE DE L'AGGLOMERATION MARSEILLAISE

Introduction

Afin d'établir une typologie des quartiers marseillais, nous avons exploité, grâce à des outils maintenant classiques en analyse des données (AFC et CAH), l'information morphologique tirée de la photo-interprétation concernant 441 zones de l'espace urbain de Marseille qui contiennent des bâtiments. A l'issue de l'analyse monodimensionnelle des descripteurs morphologiques qui a permis d'effectuer une première sélection et de coder les variables retenues en un ensemble de descripteurs qualitatifs, on dispose d'un tableau disjonctif complet croisant 441 zones bâties d'égale apparence avec les 40 modalités possibles des 10 descripteurs morphologiques (voir fig. 5.1).

Dans la première partie de ce chapitre, le tableau est soumis successivement à l'analyse factorielle des correspondances et à une classification ascendante hiérarchique maximisant le moment d'ordre 2 de la partition associé à la distance du χ^2 (programmes ANCOR et CAH2CO de l'ADDAD exécutés sous S.A.S. au CIRCE). Les résultats de cette première classification sont ensuite critiqués du point de vue de l'objectif thématique poursuivi, grâce au retour aux photographies aériennes et à une vérification sur le terrain. La deuxième partie aborde l'analyse typologique des grands sous-ensembles dégagés lors de la première classification ; les mêmes outils sont appliqués aux deux sous-fichiers après pondération des descripteurs selon leur importance thématique. On parvient ainsi à la stratification morphologique synthétique en huit classes décrite dans la troisième partie. Enfin, la quatrième et la cinquième partie tentent de dégager les conclusions pratiques et méthodologiques de ce travail et les perspectives pour la suite de l'expérience à Marseille.

Tableau 5.1 : extraits du tableau de données morphologiques, variables originales et descripteurs qualitatifs recodés

OBS	NUMERO	OS	SIXIM	SIXAX	FORM	LONG	LARG	VORO	VORSO	VYOO	VVSOD	HAUT	SURF	RAPT	DENSR	VBORR	VBSOR	VVORR	VVSOR	HAUTR	SIXINR	SIXAXR	SURFR	STRATE
154	262	2	50	300	2	300	80	3	2	5	2	4	240	375	1	2	1	4	1	2	3	4	3	3
155	268	4	25	3500	2	250	80	2	4	5	4	4	200	313	3	1	3	3	3	2	5	4	3	3
156	278	5	80	400	2	200	100	3	4	3	4	2	200	200	4	2	3	2	1	1	4	4	2	3
157	284	3	15	250	2	250	50	3	2	4	4	4	125	500	2	2	1	4	2	2	3	3	3	3
158	285	3	50	400	2	250	50	2	2	4	5	5	125	500	2	1	1	3	2	2	3	3	3	3
159	287	2	25	150	2	200	60	2	2	3	5	4	120	333	1	1	2	4	2	2	3	3	3	3
160	298	4	15	300	2	300	80	2	2	3	5	4	240	375	3	1	1	4	2	2	4	4	3	3
161	301	3	80	200	2	350	80	3	2	4	4	5	280	438	2	2	1	3	2	2	4	4	3	3
162	308	3	25	500	2	100	100	2	2	4	4	4	100	100	2	2	2	4	2	2	4	4	3	3
163	310	5	15	2500	2	200	100	3	4	3	5	4	200	200	4	2	3	2	2	1	5	4	3	3
164	313	3	40	400	4	200	80	3	2	3	2	4	160	250	2	2	1	2	2	2	3	4	3	3
165	315	4	80	300	2	300	50	2	2	3	4	4	150	600	3	1	2	3	1	2	4	4	3	3
166	325	4	80	1600	4	250	50	2	4	3	5	4	125	500	3	1	3	2	2	4	5	3	3	3
167	332	4	15	200	2	200	50	3	3	3	5	4	100	400	3	2	2	4	4	1	3	3	3	3
168	336	4	80	300	4	200	80	3	3	3	5	4	160	250	3	2	2	4	2	2	4	3	3	3
169	337	4	25	250	2	250	50	3	2	3	5	4	125	500	3	2	1	4	4	2	4	3	3	3
170	338	3	40	300	2	200	50	2	2	5	4	4	100	400	2	1	1	4	3	2	4	4	3	3
171	339	3	80	400	4	200	50	3	2	4	5	4	100	400	2	2	1	4	4	2	4	4	3	3
172	340	2	25	200	2	250	80	2	2	3	4	4	200	313	1	1	2	2	2	2	4	4	3	3
173	349	4	80	320	2	350	80	2	2	3	4	4	280	438	3	1	2	2	1	2	4	4	4	3
174	351	4	25	300	2	200	50	2	2	3	5	4	100	400	3	1	2	4	1	2	4	4	3	3
175	356	3	25	750	2	350	80	2	2	3	5	4	280	438	2	1	2	2	4	2	4	4	4	3
176	363	4	25	200	2	250	60	2	2	3	3	5	150	417	3	1	1	2	2	2	4	3	3	3
177	385	3	25	300	2	300	50	2	2	3	4	4	150	600	2	1	2	3	2	2	4	3	3	3
178	386	4	25	400	2	400	50	3	2	4	5	4	200	800	3	2	1	4	2	2	4	4	4	3
179	389	4	60	400	2	300	50	3	2	3	2	4	150	600	3	2	1	2	2	2	3	4	3	3
180	390	2	25	400	2	400	50	2	3	4	5	4	200	800	1	2	2	4	2	2	4	4	4	3
181	392	4	25	150	2	200	60	3	3	5	3	5	120	333	3	1	2	4	2	2	4	4	4	3
182	396	4	25	350	2	400	50	2	3	5	3	4	200	800	3	1	2	4	2	2	4	4	4	3
183	397	4	50	250	2	400	50	3	2	4	3	4	200	800	3	2	1	3	2	2	3	4	4	3
184	398	4	20	150	2	200	70	2	2	5	5	5	140	286	3	1	2	4	2	2	3	3	3	3
185	402	3	25	200	2	350	50	2	4	4	5	4	175	700	2	1	3	3	4	2	3	3	3	3
186	403	3	25	250	2	400	50	3	3	4	5	4	200	800	2	2	2	3	4	2	2	3	4	3
187	407	3	40	320	4	500	100	2	2	3	5	4	500	500	2	2	2	2	2	2	3	5	4	3
188	408	3	20	150	2	300	50	4	4	3	5	4	150	600	2	3	2	4	4	2	1	3	3	3
189	421	4	50	150	2	300	80	2	2	3	5	4	240	375	3	1	2	2	4	2	3	4	4	3
190	423	2	25	500	2	300	40	2	2	4	4	4	120	750	1	1	2	3	1	2	4	4	3	3
191	430	2	80	120	2	120	60	2	4	4	5	3	72	200	1	1	1	4	1	1	2	2	2	3
192	439	2	25	200	2	150	150	2	4	4	2	4	225	100	1	1	3	3	1	2	2	3	3	3
193	440	4	25	250	2	300	50	2	4	3	4	4	150	600	3	1	3	2	2	2	3	3	3	3
194	444	4	80	1600	2	200	50	2	3	3	2	4	100	400	3	1	2	2	2	2	4	5	3	3
195	446	3	50	600	2	500	50	2	3	2	2	4	250	1000	2	2	2	2	2	2	3	4	4	3
196	448	3	50	400	2	300	50	2	3	4	3	4	150	600	2	1	2	3	2	2	3	4	3	3
197	457	3	100	200	2	300	50	2	2	3	2	5	150	600	2	1	1	2	1	2	4	3	3	3
198	459	3	80	200	2	110	110	2	3	4	3	4	121	100	2	1	2	3	2	2	4	3	3	3
199	465	3	25	200	2	250	250	3	3	4	3	4	625	100	2	2	2	2	2	2	4	3	3	3
200	466	4	75	300	2	300	50	3	3	3	5	5	150	600	3	2	2	2	2	2	4	4	3	3
201	467	4	25	450	2	400	50	2	3	3	5	4	200	800	3	1	2	2	2	2	4	4	3	3
202	469	4	25	200	4	200	70	2	3	4	3	4	140	286	3	1	2	3	2	2	4	4	3	3
203	472	4	25	500	2	400	50	3	2	3	5	4	200	800	3	2	1	2	2	2	4	4	3	3
204	485	2	25	250	2	500	50	3	3	4	5	4	250	1000	1	2	2	3	2	2	4	4	3	3

5.1 : Analyse du tableau global

5.1.1 : Analyse factorielle des correspondances

Examen des valeurs propres

Comme toujours lorsqu'il s'agit du tableau disjonctif complet associé à une correspondance multiple, les valeurs propres exprimées en pourcentages de la trace de la matrice diagonalisée (égale à la somme des valeurs propres), reflètent mal la part d'inertie totale expliquée par chacun des axes. Dans l'ensemble des analyses factorielles qui vont suivre, on choisira donc, comme mesure de cette quantité, les valeurs propres issues de l'AFC. du tableau de Burt associé, également rapportées à leur somme. Ces valeurs propres appelées Δ_i ne sont en fait que les carrés des λ_i issus du tableau disjonctif. Le tableau ci-dessous fourni la correspondance entre les Δ_i et les λ_i pour les 10 premiers facteurs. La valeur propre λ_1 est exclue puisqu'elle vaut toujours 1.

rangs	λ_i	$\Delta_i = \lambda_i^2$ (%)	$T_i = \frac{\Delta_i}{\sum \Delta_j}$ (%)	cumul	histogramme
2	.3182	.1013	28.14	28.14	*****
3	.2151	.0463	12.86	41.00	*****
4	.1879	.0353	9.8	50.8	*****
5	.1622	.0263	7.3	58.1	*****
6	.1490	.0222	6.17	64.27	****
7	.1434	.0206	5.72	69.99	****
8	.1374	.0189	5.25	75.24	****
9	.1251	.0156	4.33	79.57	***
10	.1168	.0136	3.78	83.35	***
11	.1121	.0126	3.5	86.85	**

Les 7 premiers facteurs (correspondant aux valeurs propres 2 à 8) peuvent être considérés comme intéressants : ils ont une contribution à l'inertie du nuage supérieure à deux fois la contribution moyenne, mais on ne rendra compte ici que des 4 premiers qui permettent d'interpréter la partition en 8 classes issue de la classification.

Examen des facteurs et des plans factoriels

Premier facteur:

Dans l'ordre des contributions croissantes à l'inertie de l'axe, 11 modalités jouent un rôle significatif dans son positionnement. La contribution moyenne à l'inertie est de 2,32 % et on a retenu ici les contributions supérieures à 4%. L'ensemble des modalités citées expliquent 66.7 % de l'inertie de l'axe.

rangs	modl.	fact.	CTR%	cumul	poids
1	SUR1	1065	110	110	140
2	HAU2	-750	82	192	200
3	VBD1	-753	65	257	160
4	SMA1	1073	63	320	70
5	RAP4	-1434	58	378	50
6	RAP1	650	54	432	180
7	HAU1	550	52	484	240
8	SMA4	-784	48	532	100
9	SUR4	-1061	46	578	60
10	SUR3	-713	45	623	120
11	SMI4	-1199	44	667	40

Le premier facteur, qui explique environ 28% de l'inertie du nuage, apparaît comme classant grossièrement les zones selon la taille des éléments bâtis, leur forme et leur hauteur : l'ensemble des modalités caractérisant les petites tailles de bâti, les formes carrées ou peu allongées et les faibles hauteurs, sont groupées à droite de l'axe (coordonnées positives) et inversement, les grands bâtiments allongés et dont la hauteur dépasse en général deux niveaux se situent à gauche de l'axe (voir figures 5.2 et 5.3). Cette opposition explique 82% de l'inertie de l'axe si l'on prend en compte les modalités des cinq variables : SURF, SMIN, SMAX, RAPT, HAUT, 37,5% étant dû aux modalités de rangs 1 et 2 qui caractérisent le demi-plan droit et 44% aux modalités de rangs supérieurs.

Deuxième facteur :

Avec le même principe de sélection que pour le premier, les modalités contribuant au second facteur sont les suivantes :

rangs	modl.	fact.	CTR%	cumul	poids
1	VBD3	1775	206	206	60
2	DEN4	1618	146	352	50
3	RAP2	900	96	448	110
4	SMI1	913	80	528	100
5	RAP1	-583	63	590	180
6	DEN5	2418	62	653	10
7	VVS2	686	52	705	100
8	SUR2	676	44	749	90

Le second facteur (13% de l'inertie du nuage) classe les zones suivant la densité du bâti, la forme et la valeur dominante dans l'espace bâti. Le nuage s'étire du côté positif de l'axe où sont regroupées les zones denses. Le bâti est ici moyennement allongé et de valeur sombre : on est autour des modalités DEN4, DEN5, RAP2, VBD3, VVS1, VVS3, VVS4. A l'inverse les valeurs négatives du deuxième facteur sont le fait des zones peu denses ou à densité moyenne (modalités DEN1, DEN2, DEN3) dont les bâtiments sont soit carrés soit de forme très allongée (modalités RAP1 ou RAP4) et où la valeur du bâti est claire (VBD1 ou 2), (figures 5.2 et 5.3).

Cette opposition explique 73% de l'inertie de l'axe dont 58% est dû aux modalités de coordonnées positives : DEN4, DEN5, RAP2, VBD3, SMI1, VVS2.

Figure 5.2 : Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2)

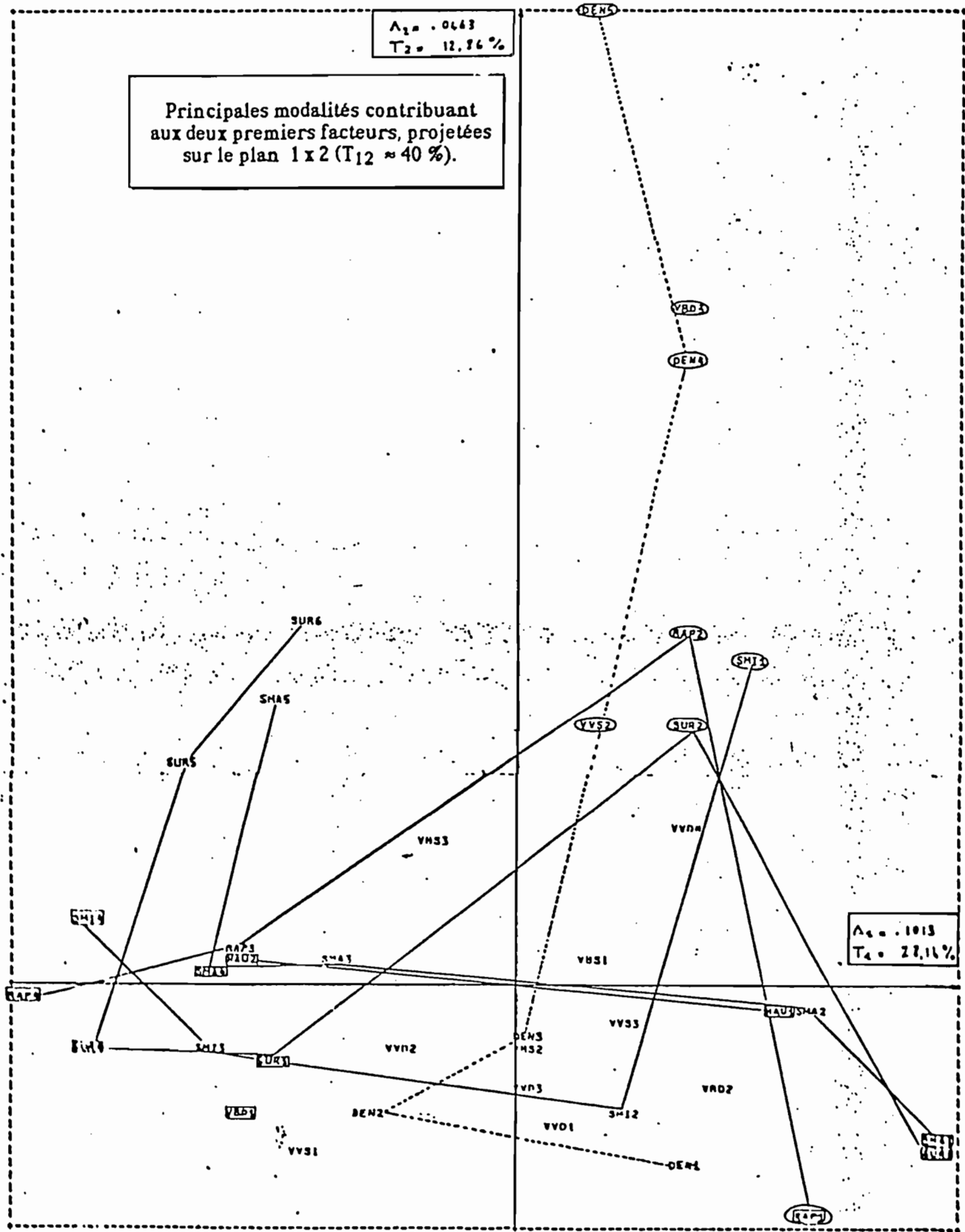
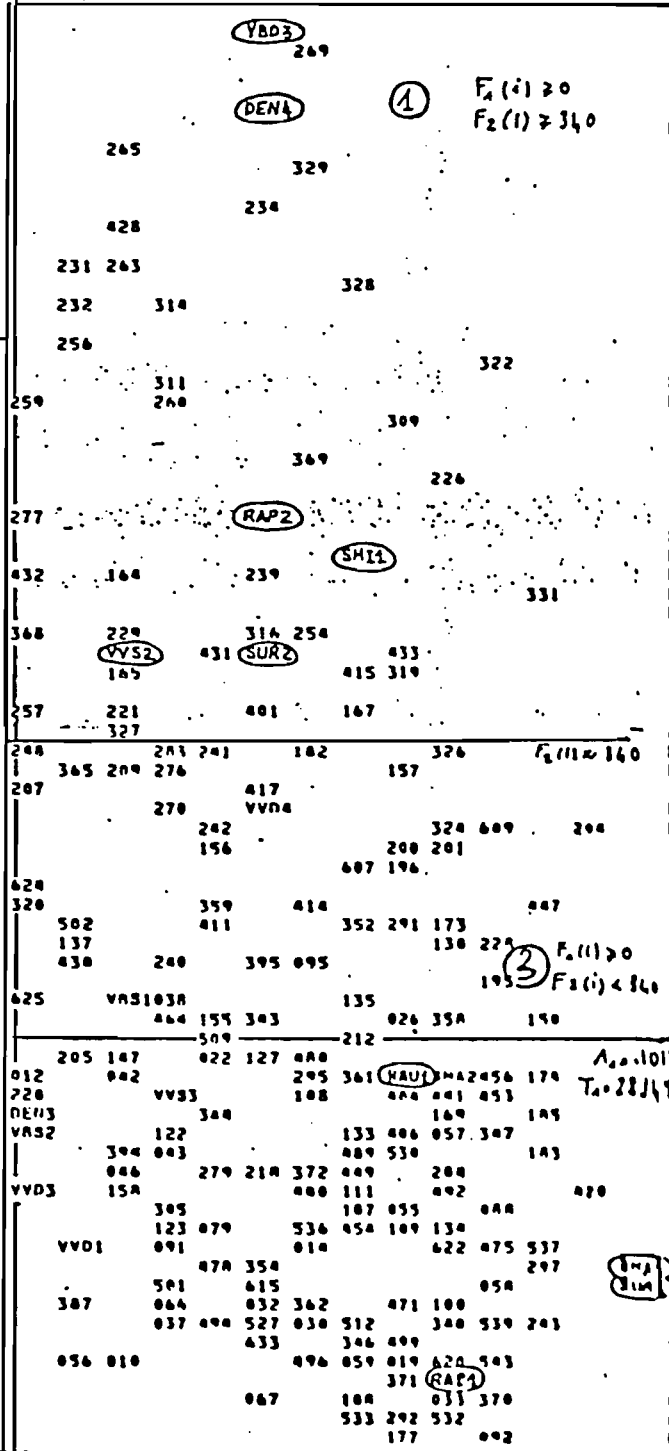
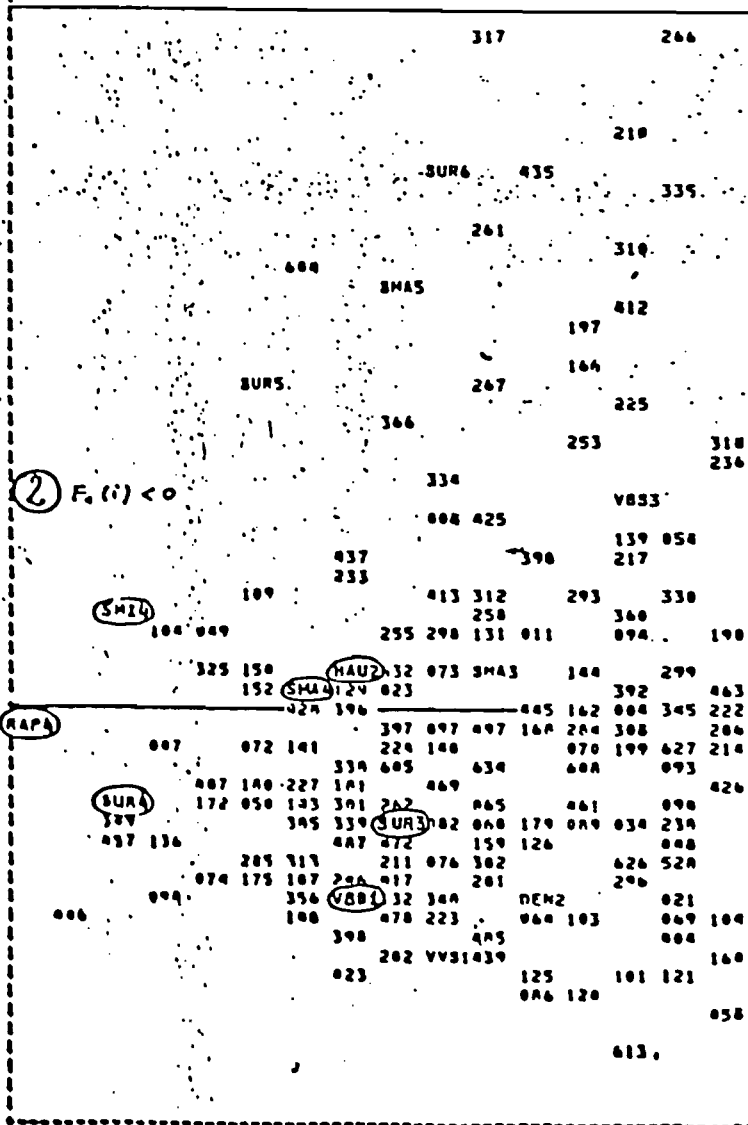


Figure 5.3 : Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2)

Représentation simultanée des lignes et des colonnes, projetées sur le plan 1 x 2 ($T_{12} \approx 40\%$).

DENS
 $A_2 \approx 0.463$
 $T_2 \approx 12.86\%$



Plan factoriel 1x2 :

En examinant, sur la figure 5.3, la projection simultanée des zones et des modalités, on peut déterminer principalement trois régions dans le plan factoriel.

1 : La région supérieure droite du plan, caractérisée par des valeurs fortement positives du second facteur ($F_2(i) > 340$) et positives du premier. Y sont regroupées les zones à forte densité de bâti (DEN4 et DEN5), la surface des constructions y est petite (SMI1, SUR2), leur valeur foncée (VBD3) et leur forme peu allongée (RAP2).

2 : La région gauche du plan (premier facteur négatif) est caractérisée par les modalités de rang élevé des variables : "surface de l'élément majoritaire" (SUR3, SUR4, SUR5, SUR6), "surface maximum de l'élément bâti" (SMA3, SMA4, SMA5) et "rapport longueur/largeur de l'élément majoritaire" (RAP3, RAP4). Les bâtiments y sont majoritairement élevés (HAU2) et de valeur claire (VBD1, VBD2).

3 : Dans la région droite du plan (premier facteur positif, deuxième facteur négatif ou faiblement positif : $-815 < F_2(i) < 560$) sont regroupées les zones à petits bâtiments (SUR1, SMA1) de formes carrées (RAP1) majoritairement bas (HAU1) ainsi que les zones peu denses (DEN1).

Ces oppositions peuvent être considérées comme une première description schématique de la géométrie du nuage des zones qui fournit une interprétation de la typologie en trois classes obtenue grâce à la C.A.H. . A elles seules, elles rendent compte d'environ 35% de l'inertie totale du nuage.

Troisième facteur :

Si l'on restreint l'ensemble des caractères morphologiques aux variables rendant compte de la taille du bâti : SURF, SMIN et SMAX, le troisième facteur apparaît comme fonction parabolique du premier. On obtient sur le plan 1x3 (fig. 5.4 et 5.5), le long de cette parabole, un classement des zones suivant la taille des éléments bâtis qu'elles contiennent. Ce facteur explique 8% de l'inertie totale du nuage et les trois variables citées y contribuent pour 52%.

Quatrième facteur :

Il rend compte principalement de deux caractéristiques du nuage:

1. L'écart important au centre de gravité des zones comportant de très grands bâtiments. La modalité SUR6 représente 12,8% de l'inertie de l'axe et SMA5 9,2%.

2. L'opposition entre, d'une part les zones où la valeur de l'environnement a une dominante foncée (VVD4: présence importante d'arbres) et une sous-dominante gris moyen (VVS3), et d'autre part celles où la valeur dominante est plus claire (VVD2 ou 3) et la sous-dominante foncée (VVS4). A ce deuxième type appartiennent en particulier les zones à vocation agricole situées à la périphérie de Marseille. Cette opposition, très visible sur le plan factoriel 1x4, peut se schématiser ainsi:

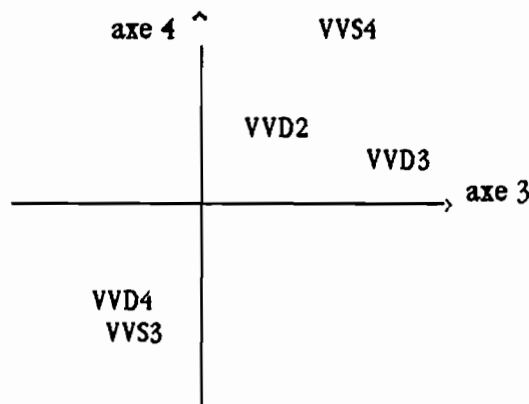


Figure 5.4 : Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3)

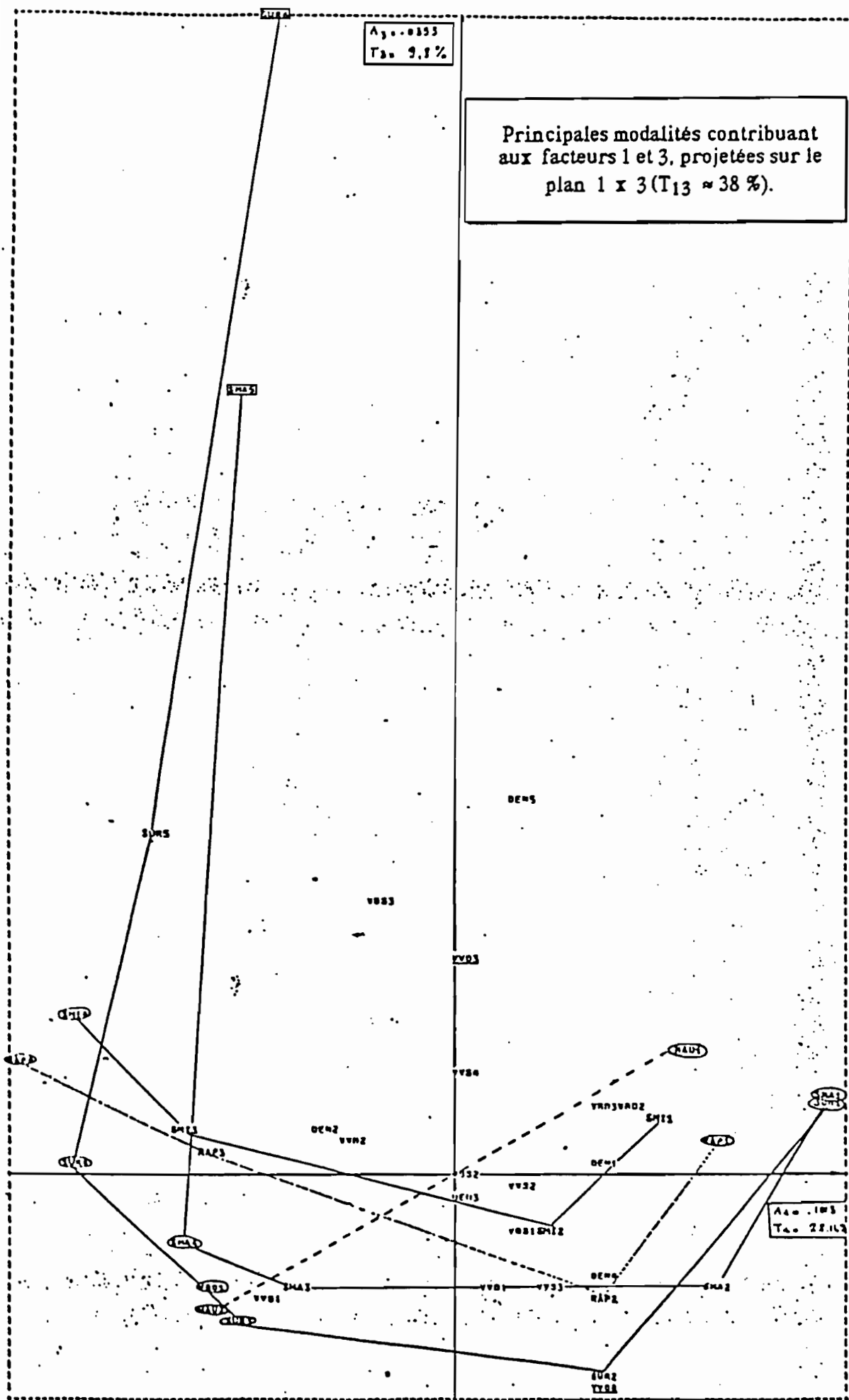
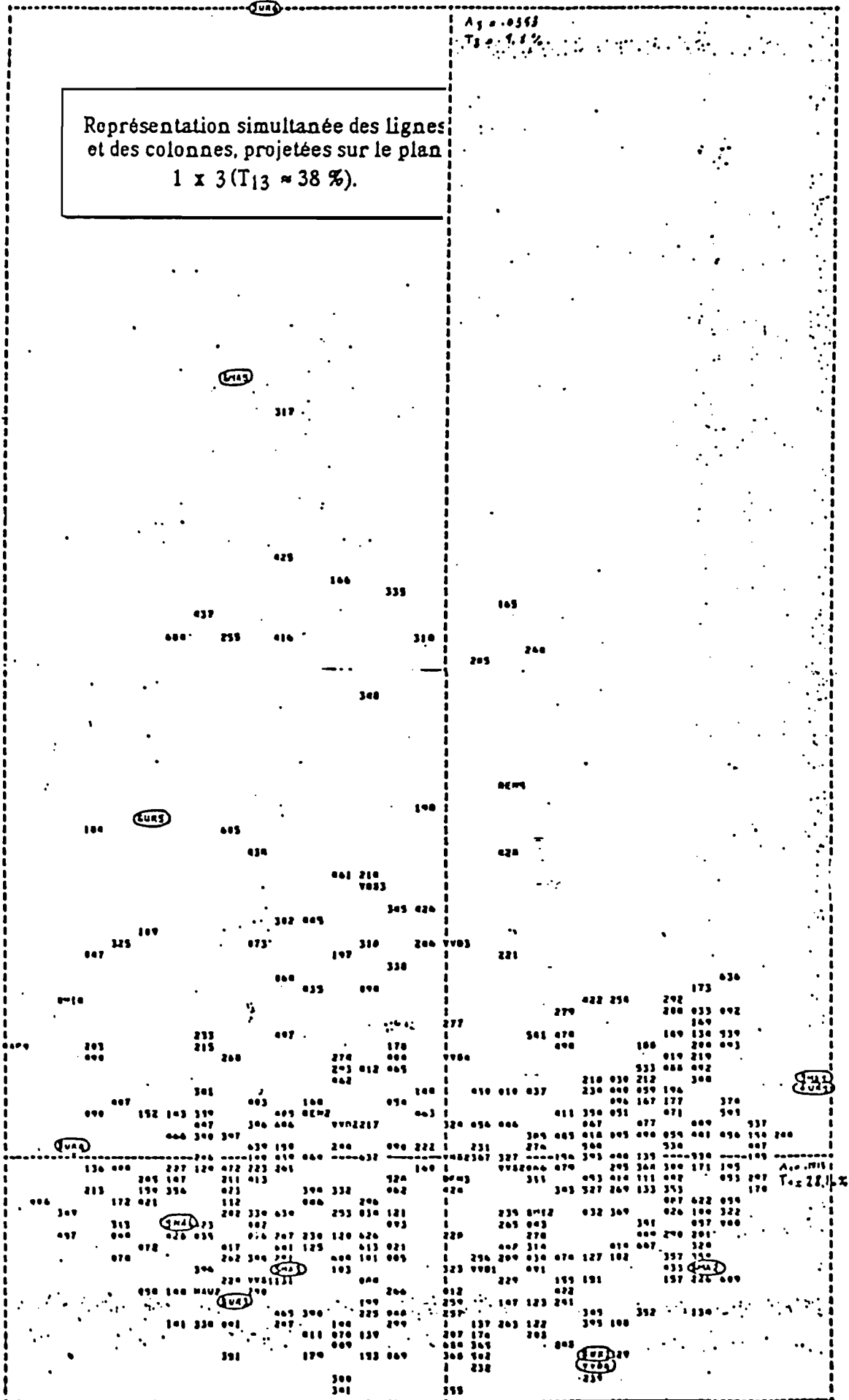


Figure 3.5 : Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3)



5.1.2 : Classification ascendante hiérarchique

Choix de la partition étudiée

Ce choix résulte d'un compromis effectué entre la logique de l'algorithme et deux contraintes imposées par les objectifs de la recherche :

1. L'histogramme des indices de niveaux de la hiérarchie présente des seuils significatifs aux passages de une à 2 classes, de 2 à 3, de 6 à 7, de 7 à 8 puis de 9 à 10 et enfin de 13 à 14 classes (figure 5.6). Après l'examen des partitions en deux et trois classes, on a donc le choix ensuite de poursuivre l'interprétation jusqu'à 7, 8, 10 ou 14 classes.
2. L'objectif stratificateur de l'analyse doit conduire à la représentativité morphologique d'un échantillon de travail tiré dans la base de sondage des districts de recensement de l'INSEE. Ceux-ci sont au nombre d'environ 4000 et le taux de sondage pressenti est de 2%. Pour avoir au minimum deux districts par strate il faut donc au moins cent districts dans chacune des strates, ce qui compte tenu de la taille moyenne des zones homogènes définies lors de la photo-interprétation, nous contraint à n'étudier que les partitions où toute strate regroupe un minimum de 25 à 30 zones. La partition en 14 classes apparaît d'ores et déjà trop fine pour satisfaire cet impératif.
3. La hiérarchie de partitions proposée par la C.A.H peut facilement être mise en relation avec celle qu'établit l'analyse factorielle dans la série des descripteurs morphologiques. C'est d'ailleurs par ce biais que l'on caractérisera en termes morphologiques chaque classe. Malheureusement, le pouvoir discriminant d'un descripteur morphologique donné, d'ailleurs en grande partie fonction des codages successifs de l'information, n'est pas toujours en rapport avec l'intérêt thématique qu'il présente du point de vue démographique. On mesurera dans la suite combien ce problème a pesé sur le choix de la partition retenue. Disons tout de suite qu'après ce constat, nous nous sommes sentis autorisés à pratiquer certains regroupements entre classes qui contredisent la logique de l'agrégation hiérarchique. Prenant le risque de scandaliser certains taxinomistes, nous espérons en revanche que le résultat sera efficace pour le télédétecteur comme pour le démographe.

Les partitions en sept et huit classes

Nous verrons ultérieurement qu'on ne peut pas retenir ces partitions comme fournissant une stratification acceptable de la base de sondage ; par contre elles servent de base à la division de l'ensemble des zones en trois grands sous-ensembles sur lesquels, moyennant la pondération de certains descripteurs, on obtiendra par de nouvelles C.A.H. un résultat convenable. Dans ce qui suit on se contente de caractériser sommairement les classes, sans retourner au tableau de données, en examinant la position des centres de classes sur les plans factoriels principaux : 1x2 et accessoirement 1x3 (figures 5.7 et 5.8).

La figure 5.7 permet de caractériser facilement les deux classes de la première partition : d'une part le sous-ensemble des zones, dont les projections sont situées dans le demi-plan droit (premier facteur positif), caractérisé globalement par les modalités HAU1, SUR1, SUR2, SMI1, RAP1, RAP2 (classe 879), d'autre part le sous-ensemble situé à gauche du même plan (premier facteur négatif), groupé autour des modalités HAU2, VBD1, SUR3, SUR4, SMA4, SMA5, SMI4 et RAP4 (classe 880). Ce deuxième ensemble est ensuite scindé en deux classes : 878 et 876. La classe 876 conserve plus de 90% des zones qui composaient la classe 880 et sa caractérisation n'en diffère guère en moyenne. Néanmoins elle est plus homogène du point de vue de la taille des éléments bâtis et de leur densité puisque la classe 878 regroupe les 20 zones ayant, soit une très grande taille de bâti (SUR6), soit la densité de bâti maximale (DEN5).

Figure 5.6 : Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie

Histogramme des indices de niveau de la hiérarchie des partitions.

SOMME DES INOICES DE NIVEAU 0.25000E+01

J	I(J)	A(J)	B(J)	IT(J)	IT(O)	HISTOGRAMME DES INOICES DE NIVEAU DE LA HIERARCHIE
1	3951	2361	3931	3941	941	941*****
1	3941	1181	3911	3921	471	1421*****
1	3931	1041	3881	3901	421	1831*****
1	3921	901	3891	3771	361	2191*****
1	3911	711	3611	3331	291	2481*****
1	3901	681	3841	3731	271	2751*****
1	3891	671	3801	3871	271	3021*****
1	3881	661	3641	3861	261	3281*****
1	3871	651	3581	3791	261	3541*****
1	3861	511	3781	3851	201	3751*****
1	3851	421	3721	3831	171	3911*****
1	3841	401	3541	3621	161	4071*****
1	3831	391	3811	3821	161	4231*****
1	3821	371	3601	3761	151	4381*****
1	3811	341	3521	3751	141	4511*****
1	3801	341	3671	3741	141	4651*****
1	3791	311	3561	3681	121	4771*****
1	3781	281	3401	2921	111	4881*****
1	3771	271	3571	3141	111	4991*****
1	3761	261	3591	3701	101	5091*****
1	3751	221	3501	3631	91	5181*****
1	3741	211	3691	3711	91	5271*****
1	3731	211	2901	3661	81	5351*****
1	3721	211	2961	3651	81	5431*****
1	3711	211	3291	2981	81	5521*****
1	3701	201	3221	3461	81	5601*****
1	3691	201	3451	2911	81	5671*****
1	3681	191	3061	2991	71	5751*****
1	3671	181	3351	3431	71	5821*****
1	3661	181	3371	3441	71	5891*****
1	3651	181	3551	3281	71	5971*****
1	3641	181	3391	3491	71	6041*****
1	3631	171	3301	3091	71	6111*****
1	3621	161	3131	3511	71	6171****
1	3611	151	2451	3341	61	6231****
1	3601	151	3241	3021	61	6291****
1	3591	141	3311	3361	61	6351****
1	3581	141	61	3531	61	6411****
1	3571	141	2821	3471	61	6471****
1	3561	141	2491	2861	61	6521****
1	3551	141	3181	3201	61	6581****
1	3541	141	941	3411	51	6631****
1	3531	141	11	3321	51	6691****
1	3521	141	3121	3081	51	6741****
1	3511	131	3381	3481	51	6791****
1	3501	131	3151	3041	51	6841****
1	3491	131	3271	3421	51	6901****
1	3481	121	3171	2931	51	6951****
1	3471	121	261	571	51	6991****
1	3461	121	2891	3071	51	7041****
1	3451	121	3031	3231	51	7091****
1	3441	121	3011	901	51	7131****
1	3431	111	281	3251	41	7181****
1	3421	111	2721	2571	41	7221****
1	3411	111	2871	2701	41	7271****
1	3401	111	2531	2791	41	7311****
1	3391	111	411	2881	41	7351****

Figure 5.7 : Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie

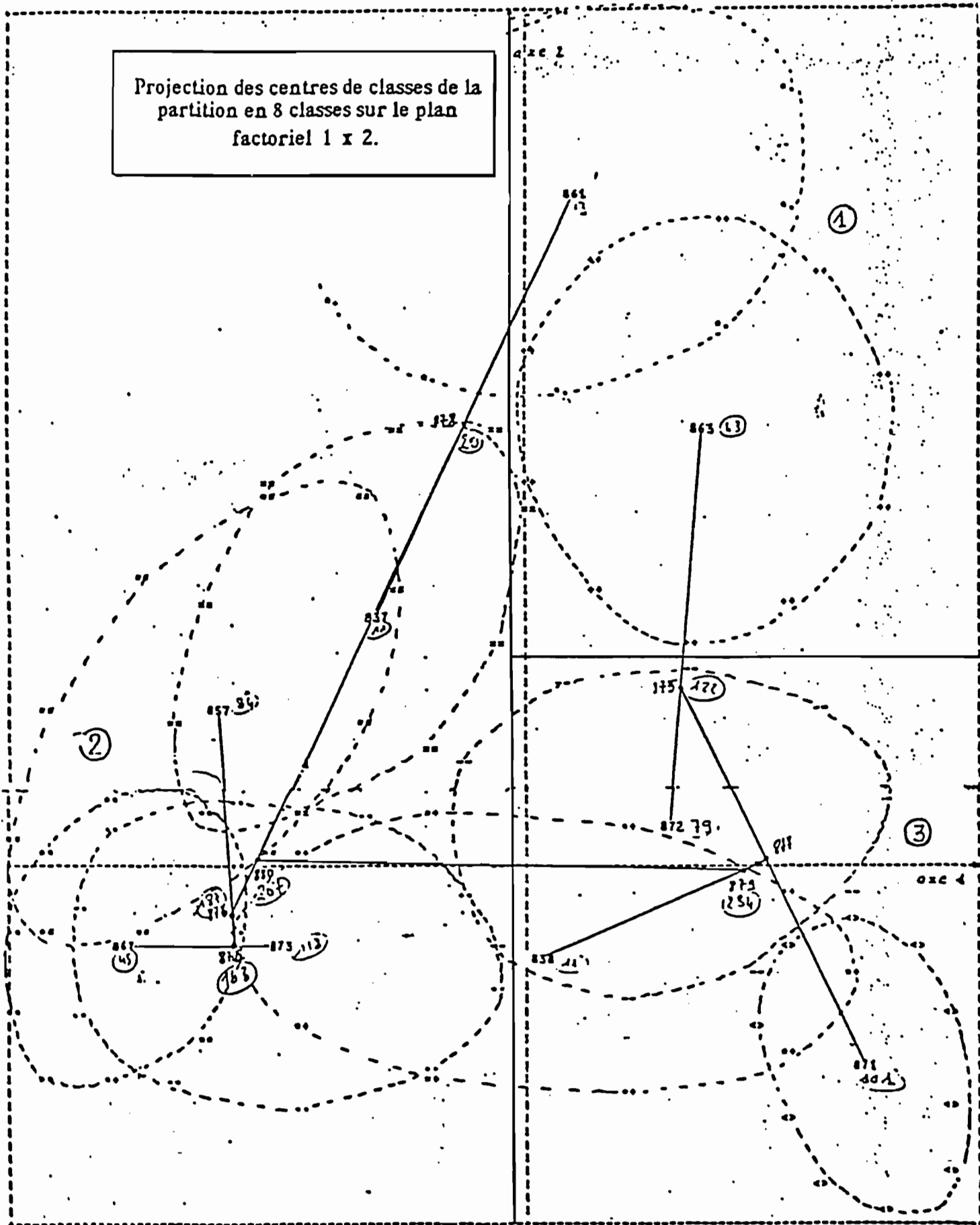
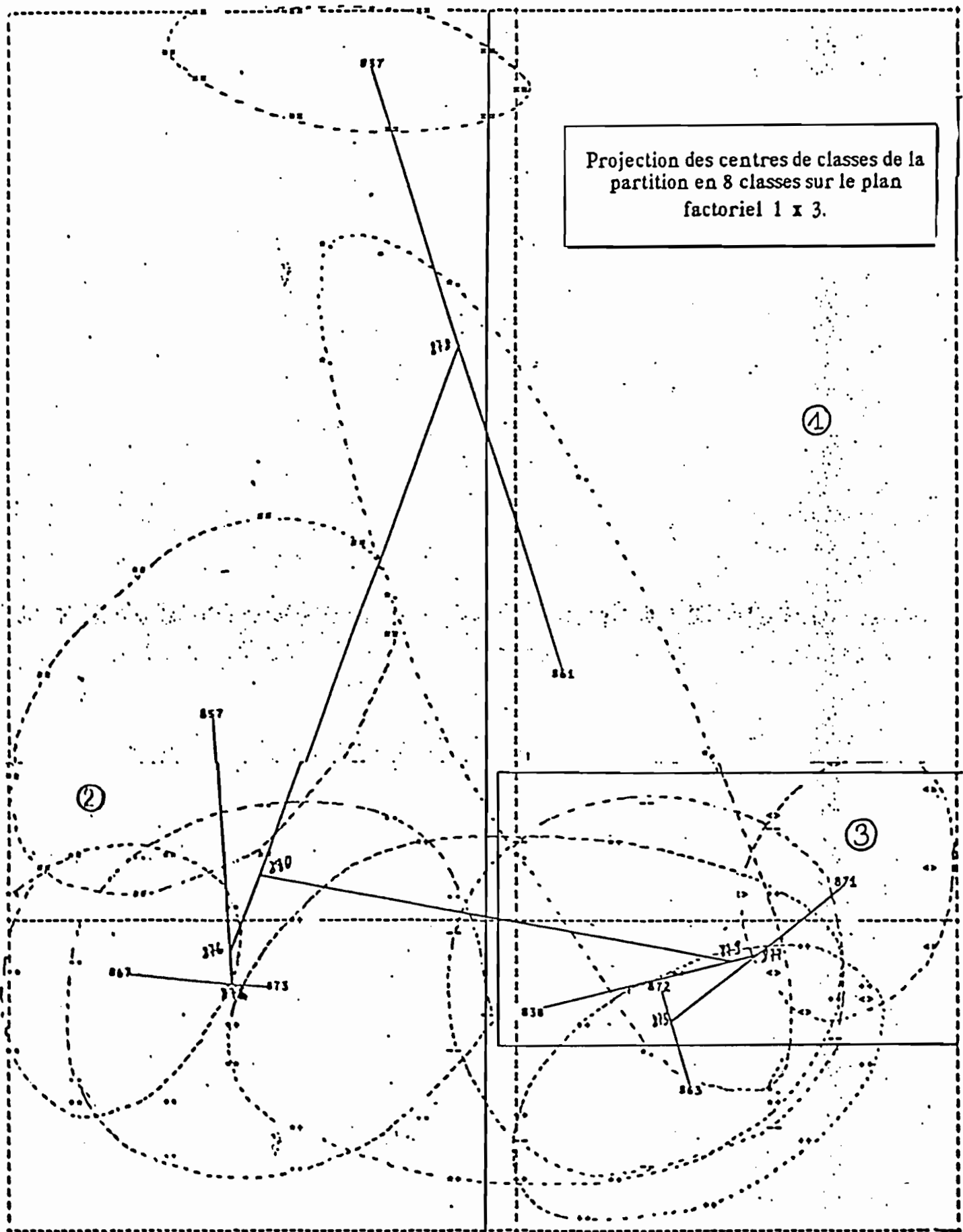


Figure 5.8 : Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie



Projection des centres de classes de la partition en 8 classes sur le plan factoriel 1 x 3.

①

②

③

857

879

861

857

870

876

873

862

874

875

871

872

875

838

863

871

La partition en sept classes divise chacune des trois classes 879, 878 et 876.

La classe 879 produit les classes 871, 875 et 838. La classe 871 est facilement caractérisable sur le plan 1x2 par les modalités HAU1, SUR1, SMA1 ou SMA2 et RAP1. Elle regroupe 101 zones et sa petite ellipse d'inertie témoigne de son homogénéité. La classe 875, qui éclatera à l'étape suivante, est hétérogène quant à la densité et à la taille du bâti. Enfin la classe 838 présente une ellipse d'inertie très allongée le long de son axe principal qui a, en gros, la direction de l'axe 1. Forte variabilité donc, dans cette classe, de la hauteur et de la taille du bâti. Par contre la densité de bâti y est assez homogène dans les modalités de petit rang (DEN1, 2 ou 3).

La classe 878 éclate en 861 et 837. On caractérise facilement ces deux classes sur les plans 1x2 et 1x3 : 837 regroupe les zones où les bâtiments ont la taille maximale (SUR6 et SMA5), tandis que 861 contient celles où la densité du bâti est maximale. Soulignons au passage la variance très élevée du troisième facteur dans cette classe : la taille des bâtiments y est donc très variable.

La classe 876 fournit les classes 857 et 874. Sur les plans factoriels 1x2 et 1x3, 857 apparaît comme groupant les zones où le bâti est de grande taille : SUR4 et plus encore SUR5, SMI3 ou SMI4, RAP4, mais de densité de bâti variable (forte variance du second facteur dans la classe). 874 contient le restant des zones situées dans la moitié gauche du plan 1x2 (163 zones) caractérisé par SUR3 ou SUR4, RAP3 ou RAP4, SMI3 ou SMI4, HAU2 et VBD1.

Enfin la partition en huit classes présente l'intérêt de couper la classe 875, dont on soulignait plus haut l'hétérogénéité, en deux : on obtient les classes 863 et 872.

Dans 863 sont groupées 43 zones caractérisées par les modalités DEN4 et VBD3 (valeur sombre du bâti dominant), SUR2, RAP2 et SMI1. Les 79 zones restantes forment la classe 872 qui présente une ellipse d'inertie légèrement allongée le long de l'axe 1 : la surface des bâtiments y est variable de SUR1 à SUR3 et la densité également de DEN1 à DEN3.

5.1.3 : Vérifications sur les photographies aériennes et sur le terrain : critique thématique de la partition en huit classes

Deux types de vérification de la typologie obtenue à l'issue de l'analyse du nuage des 441 zones ont été pratiqués :

En premier lieu, une mission à Marseille¹ a permis de vérifier, sur un échantillon représentatif de chaque classe de la partition, la signification réelle de l'information recueillie sur les photographies aériennes. On disposait d'une centaine de zones-test à partir desquelles on a pu étalonner plus précisément le codage de l'information visuelle. Ceci a naturellement conduit à une phase de correction du fichier de données et l'information morphologique relative à une quarantaine de zones a été rectifiée. Bien sûr les résultats présentés jusqu'à maintenant intègrent ces corrections.

La deuxième vérification consiste à cartographier le résultat de la classification pour qu'il puisse être superposé à la mosaïque photographique et soumis à critique visuelle, ce qui nous a conduit aux conclusions suivantes :

1. Les classes 861 et 863, bien qu'issues chacune d'une des deux classes de la première partition, peuvent être réunies sans inconvénients thématiques : elles représentent les quartiers les plus denses, probablement très spécifiques du point de vue démographique. Le seuil entre les modalités 4 et 5 de la variable "densité du bâti", qui crée la principale différence entre les profils des deux classes, n'est pas apparu

¹ : Voir au paragraphe 1.2 de l'annexe A.2.1, la méthode de collecte et la fiche de relevé terrain employées.

suffisamment robuste lors du codage pour justifier la séparation, dans cet ensemble, des zones où la densité serait maximale (DEN5) des zones où elle est simplement très forte (DEN4).

Par contre, la forte contribution au deuxième facteur des deux modalités a toujours permis de différencier clairement ce sous-ensemble dense du reste du nuage des zones : il s'agit bien, grossièrement, du centre ville plus ou moins ancien de Marseille.

2. La partition en huit classes propose une division de la classe 880, ensemble de zones caractérisées par les modalités de rangs élevés des variables décrivant les surfaces bâties, essentiellement fondée sur la taille et la densité du bâti. Hormis la classe 861 évoquée ci-dessus, apparaissent les classes 837, 857 et 874. La vérification visuelle permet d'affirmer que la classe 837 regroupe les zones où dominent les plus grands bâtiments, toutes à vocation industrielle. Mais ces onze zones ne représentent qu'une minorité des quartiers industriels qui couvrent, au total une quarantaine de zones. Dans les classes 857 et 874, les trente zones industrielles restantes ne sont pas distinguées des zones où les immeubles d'habitation dominent : dans la classe 857 il y a 11 zones à dominante industrielle sur 24 et dans 874, 19 sur 163.

3. Une fois exclues les zones densément bâties regroupées dans 863, la classe où dominent les zones à petits bâtiments (classe 879) est coupée en trois : 838 ne compte que onze zones caractérisées par la valeur claire de l'environnement du bâti. La séparation en deux classes des quelques 180 zones restantes s'effectue de façon à peu près égale (101 zones dans 871, 79 dans 875), sur le critère principal de la taille du bâti. Dans ces deux classes, la densité du bâti est variable, sans excéder toutefois le rang 3; d'autre part on trouve, dans chacune d'elles, une minorité de zones où le bâti majoritaire est constitué de petits immeubles d'habitation élevés (10 zones dans 871 et 19 dans 872).

En résumé, disons que de l'analyse du tableau global, confrontée aux objectifs thématiques et aux vérifications, se dégagent les conclusions suivantes :

Premièrement : Le principal mérite de la typologie obtenue sur l'ensemble des zones, du point de vue du thème démographique de la recherche, est d'opérer un classement de cet ensemble sur des critères de taille des éléments bâtis d'une part et de densité de l'espace bâti d'autre part. En particulier les zones denses du centre ville plus ou moins ancien sont regroupées de manière satisfaisante.

Deuxièmement : à l'intérieur des grands groupes définis ci-dessus, la partition exploitée pêche par deux points aux yeux des thématiciens.

- Dans l'ensemble des zones majoritairement occupées par les grands bâtiments, elle ne distingue pas de façon certaine celles à vocation industrielle.

- Le sous-groupe des zones, généralement moins denses, occupées par les petits bâtiments, contient des zones où l'habitat collectif domine sous forme de petits immeubles à étages et des zones à dominante pavillonnaire. De plus, il existe également à l'intérieur de cet ensemble une différence importante dans la densité du bâti entre les zones considérées comme urbaine à part entière à cause de leur densité (DEN2 et DEN3) et des zones, généralement situées aux marges de l'agglomération, où l'espace bâti est marginal (DEN1).

Pour expliquer les caractéristiques de cette première typologie, on peut avancer l'hypothèse suivante : lors du codage de l'information morphologique, on a nettement privilégié les descripteurs de la surface et de la forme des éléments bâtis : ils regroupent à eux seuls (variables SURF, SMAX, SMIN et RAPT) 19 modalités sur 40. A l'inverse, le codage de l'information relative à la hauteur des bâtiments a été fait de manière dichotomique : bâtiments majoritairement bas (un ou deux niveaux) : HAU1 ou majoritairement hauts (plus de deux niveaux) : HAU2. Ce descripteur opère une séparation de l'ensemble des zones en deux parties d'effectifs voisins et, sous cette

forme, il ne peut jouer, en l'absence de pondération, qu'un rôle mineur dans l'analyse. Or l'élévation du bâti apparaît bien déterminante pour séparer des zones où les bâtiments ont une forme, une surface au sol et une densité comparables, mais des vocations différentes : qu'il s'agisse de distinguer le bâtiment industriel du grand bâtiment d'habitation ou le petit immeuble du pavillon, la hauteur du bâtiment intervient.

Pour valider cette hypothèse et améliorer la pertinence démographique de la typologie, on a donc pris le parti suivant :

Premièrement, conserver l'acquis de cette première analyse en séparant, d'après ces résultats, le fichier de départ en trois sous-ensembles :

- I. L'ensemble des zones où la densité du bâti est forte ou très forte (DEN4 et DEN5), constitué des classes numéros 861 et 863 de la partition en huit classes. Cet ensemble contient 52 zones qui constituent dès maintenant la première strate.
- II. L'ensemble des zones où dominent les grands bâtiments, du moins en surface occupée. Constitué des classes numéros 837, 857 et 874, cet ensemble compte 198 zones.
- III. Les 191 zones où le bâti est majoritairement de petite taille contenues dans les classes numéros 871, 872 et 838.

Deuxièmement, sur les deux derniers sous-ensembles ainsi définis (II et III), orienter, grâce à différentes pondérations de certaines modalités des variables morphologiques, les résultats de l'analyse vers les impératifs fixés par le cadre thématique de la recherche. On verra ci-dessous que pour le sous-ensemble II, c'est une pondération de facteur 3 des deux modalités de la variable "hauteur majoritaire du bâti" qui permet d'en obtenir une stratification thématiquement pertinente; en ce qui concerne le sous-ensemble III, nous retiendrons la même pondération de la hauteur du bâti, mais également la multiplication par trois des colonnes DEN1 et DEN4 qui sont les modalités extrêmes de la densité du bâti sur ce sous-ensemble puisque la modalité DEN5 n'a plus de représentants ici.

5.2 : Analyse des sous-fichiers pondérés issus de la première classification

5.2.1 : Le sous-fichier des zones à grands bâtiments

Analyse factorielle

- *Premier facteur* : En privilégiant, par multiplication par un facteur 3, la variable "élévation du bâti", on a évidemment attiré artificiellement le premier facteur vers le dipôle HAUT1, HAUT2, qui explique 66% de l'inertie du facteur. Assez bien corrélées à la modalité HAUT1, les modalités SUR6 et SMA5 caractérisant les plus grands bâtiments, ont à elles deux, une contribution de 11% à l'inertie de l'axe. Ceci est dû à la co-occurrence presque systématique dans le tableau de données, entre les plus grandes tailles de bâti et la faible élévation. L'axe peut donc être interprété comme le facteur qui permettra d'extraire les zones à dominante industrielle (pour le détail des contributions, voir le tableau 5.9).

- *Deuxième facteur* : Les cinq premières contributions à ce facteur sont le fait des modalités de rang 1 des variables suivantes : "surface de l'élément bâti majoritaire" (13,4% de l'inertie du facteur), "rapport longueur/largeur de l'élément bâti majoritaire" (12,2%), "valeur sous dominante dans l'environnement du bâti" (9,1%), "surface maximum de l'élément bâti" (8,3%). Enfin la modalité de rang 3 caractérisant la valeur dominante du bâti, a une contribution de 6,1% à l'inertie de l'axe. S'agissant des quatre premières, ce sont des modalités de poids relativement faible, qui caractérisent une minorité de zones où dominent les petits bâtiments à plusieurs étages, de forme carrée. Outre qu'il met en évidence ce trait assez marginal du sous-ensemble, le second facteur est intéressant lorsque l'on considère l'ensemble des

modalités liées à la taille du bâti : pour les variables "surface de l'élément bâti majoritaire" et "surface maximum de l'élément bâti", le second facteur est une fonction monotone croissante du rang des modalités. C'est également vrai pour les modalités 2,3 et 4 de la surface minimum de l'élément bâti, mais la modalité SMI1 a une coordonnée nettement positive sur le deuxième axe, très proche de celle de SMA5. Ce phénomène n'apparaissait pas aussi clairement dans l'analyse du nuage complet.

- Plan factoriel 1-2 : (figures 5.10 et 5.11)

Deux faits marquants permettront, à partir de l'examen de la projection des centres de classes sur ce plan, de caractériser facilement les partitions obtenues par C.A.H. : (figure 5.10)

1. L'axe 1 est quasiment aligné sur le dipôle HAU1, HAU2. La moitié droite du plan est le lieu exclusif des zones où le bâti est majoritairement bas. Soulignons également la proximité, dans le quart supérieur droit du plan, des modalités SUR5, SUR6, SMA5, VBD3 et SMI1 qui caractériseront les zones où dominent les bâtiments industriels.

2. Les modalités des variables qui décrivent la taille moyenne et maximum des éléments bâtis, forment sur le plan factoriel 1x2, une parabole axée sur le dipôle HAU1, HAU2. Cette situation est à rapprocher de celle décrite lors de l'examen du plan factoriel 1x3 dans l'analyse du nuage global. Elle lui est bien sûr comparable mais en observant la projection du nuage des zones sur le plan factoriel 1x2 (figure 5.11), on constate que la pondération des modalités HAUT1 et HAUT2 fait éclater le nuage en deux parties grossièrement paraboliques et précisément localisées dans les deux moitiés du plan : bâtiments bas à droite, élevés à gauche. L'axe 2, lui, apparaît comme une échelle de mesure de la taille du bâti, petits bâtiments du côté des coordonnées négatives, grands bâtiments du côté positif.

Classification hiérarchique

Au vu de l'histogramme des indices de niveau de la hiérarchie de partitions (figure 5.12), les partitions en 2,3,4 et 5 classes représentent des étapes importantes (le seuil significatif suivant est celui de la partition en 10 classes, trop fine pour nos besoins). Nous allons nous intéresser maintenant à décrire la formation de la partition en cinq classes qui nous servira à établir la typologie définitive de ce sous-ensemble.

Comme on pouvait le prévoir, la partition en deux classes rend compte de la dichotomie introduite entre les modalités HAU1 et HAU2 (figure 5.13). La classe 392 regroupe les 60 zones situées dans la partie droite du plan factoriel 1x2. Nous caractériserons précisément les classes de la typologie définitive au chapitre suivant. Remarquons simplement que cette classe contient l'essentiel des zones où dominent les bâtiments industriels qui constitueront la strate numéro 2. Plus de 90% des zones de cette classe possèdent la modalité HAU1 et il s'agit en majorité de zones regroupant de très grands bâtiments (SUR5 ou SUR6). Le reste de l'ensemble (138 zones), dont 90% ont une majorité de bâtiments élevés, constitue la classe 393 qui éclate ensuite, donnant les classes 390 et 388. Cette séparation s'effectue principalement sur la taille des éléments bâtis comme le montre la projection des centres de classe sur le plan 1x2 : 390 regroupe 27 zones contenant des bâtiments élevés de petite taille qui ont, pour la plupart, leur coordonnée sur le deuxième axe négative et constituent une partie de la strate numéro quatre, celle des petits immeubles d'habitation. La majorité du sous-ensemble étudié se trouve donc dans la classe 388, dont le centre est le plus proche du centre de gravité de l'ensemble. Ces 111 zones où dominent les grands bâtiments élevés du type grands ensembles d'habitat collectif, constituera la strate numéro 3 des grands ensembles d'habitation.

Dans la partition en quatre classes, la classe 394 a donné les classes 392 et 391. Les 11 zones de la classe 391, situées dans la partie inférieure droite du plan 1x2, sont des zones relativement atypiques dans le sous-fichier puisque les petits bâtiments bas y dominent. Cette classe sera d'ailleurs intégrée, après vérification, dans la strate numéro cinq, majoritairement issue du sous-fichier III, dite strate du pavillonnaire.

Figure 5.10 : Analyse factorielle du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2)

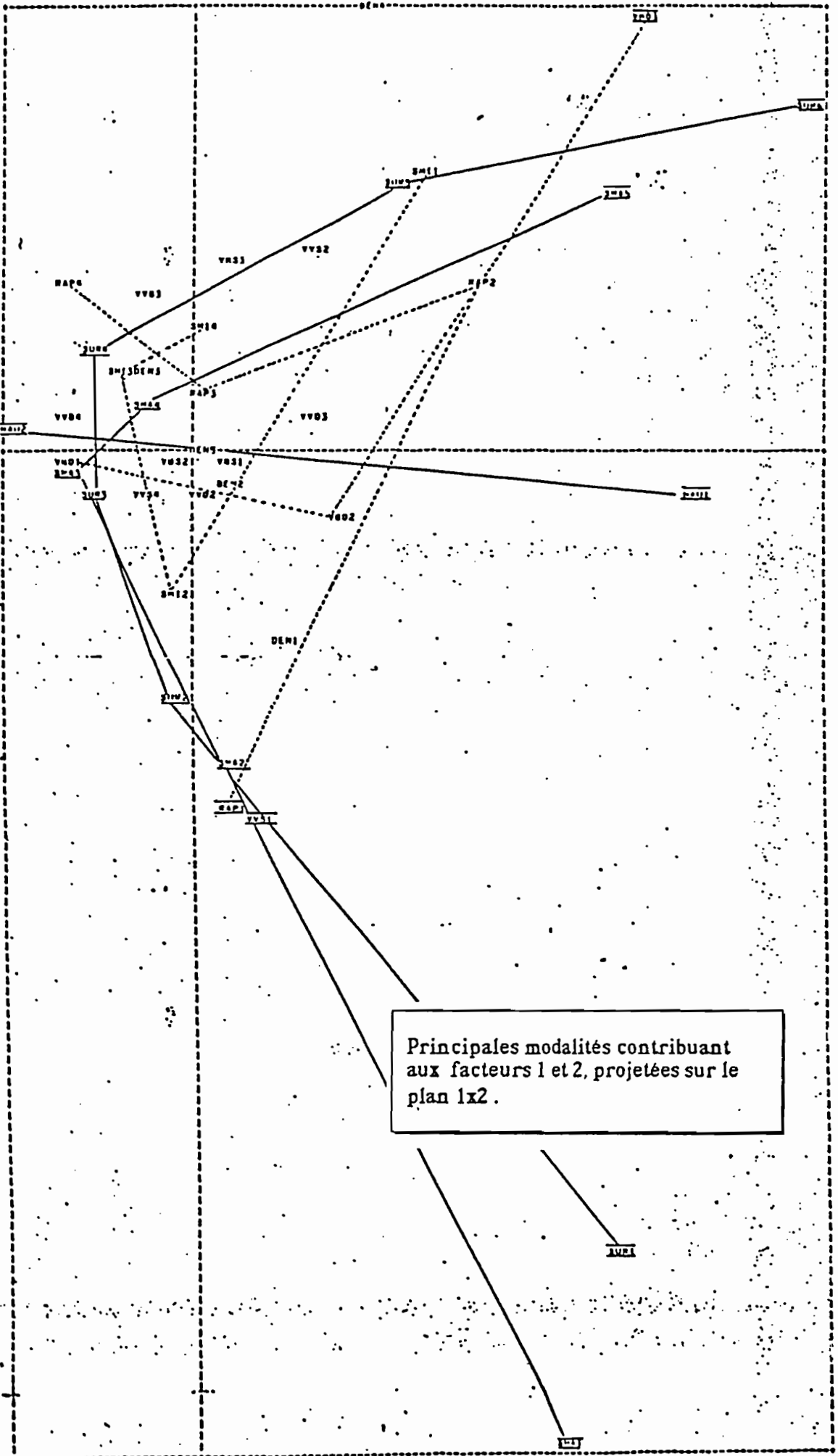


Figure 5.11 : Analyse factorielle du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2)

Représentation des zones projetées sur le plan 1x2.

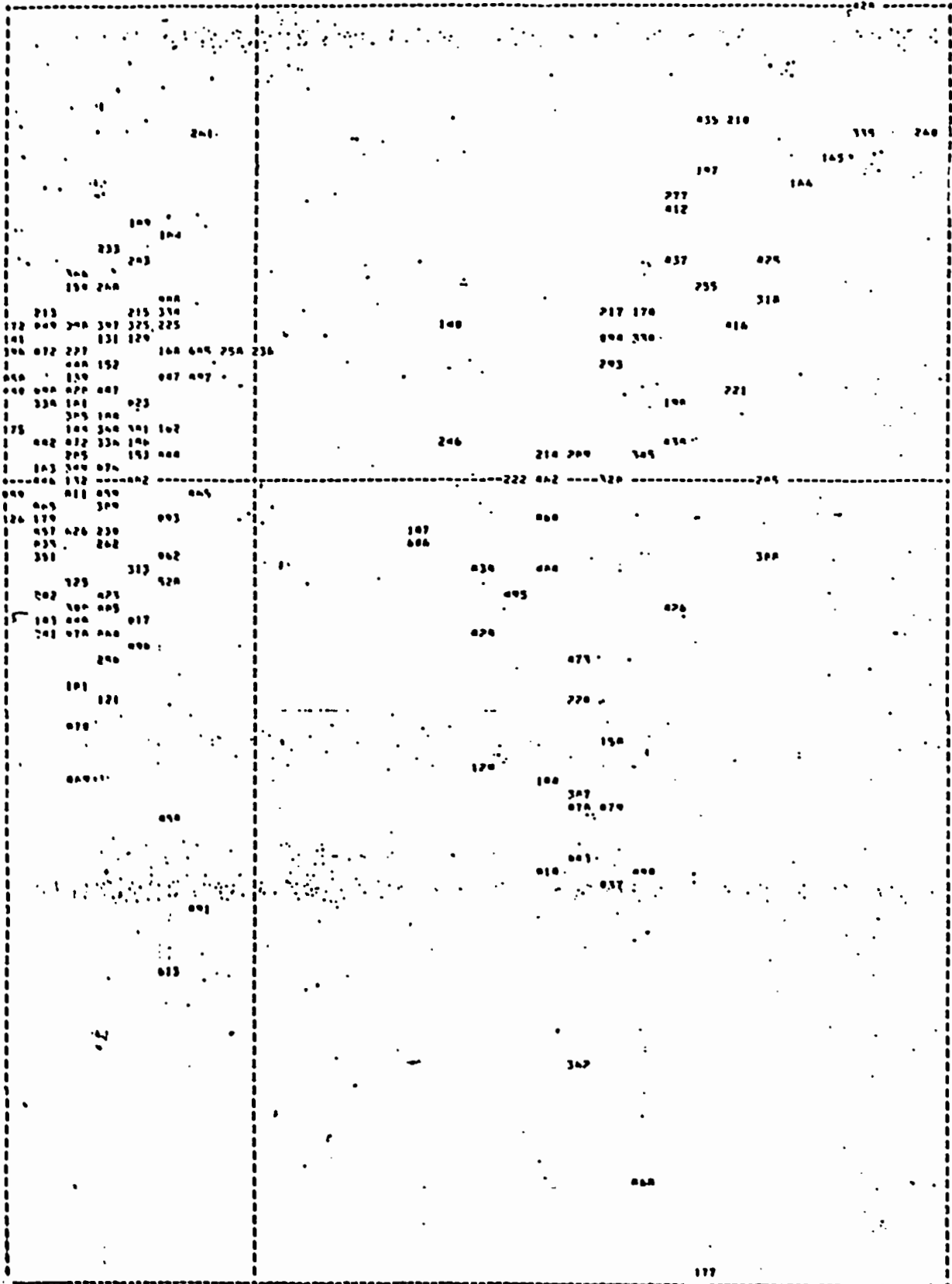


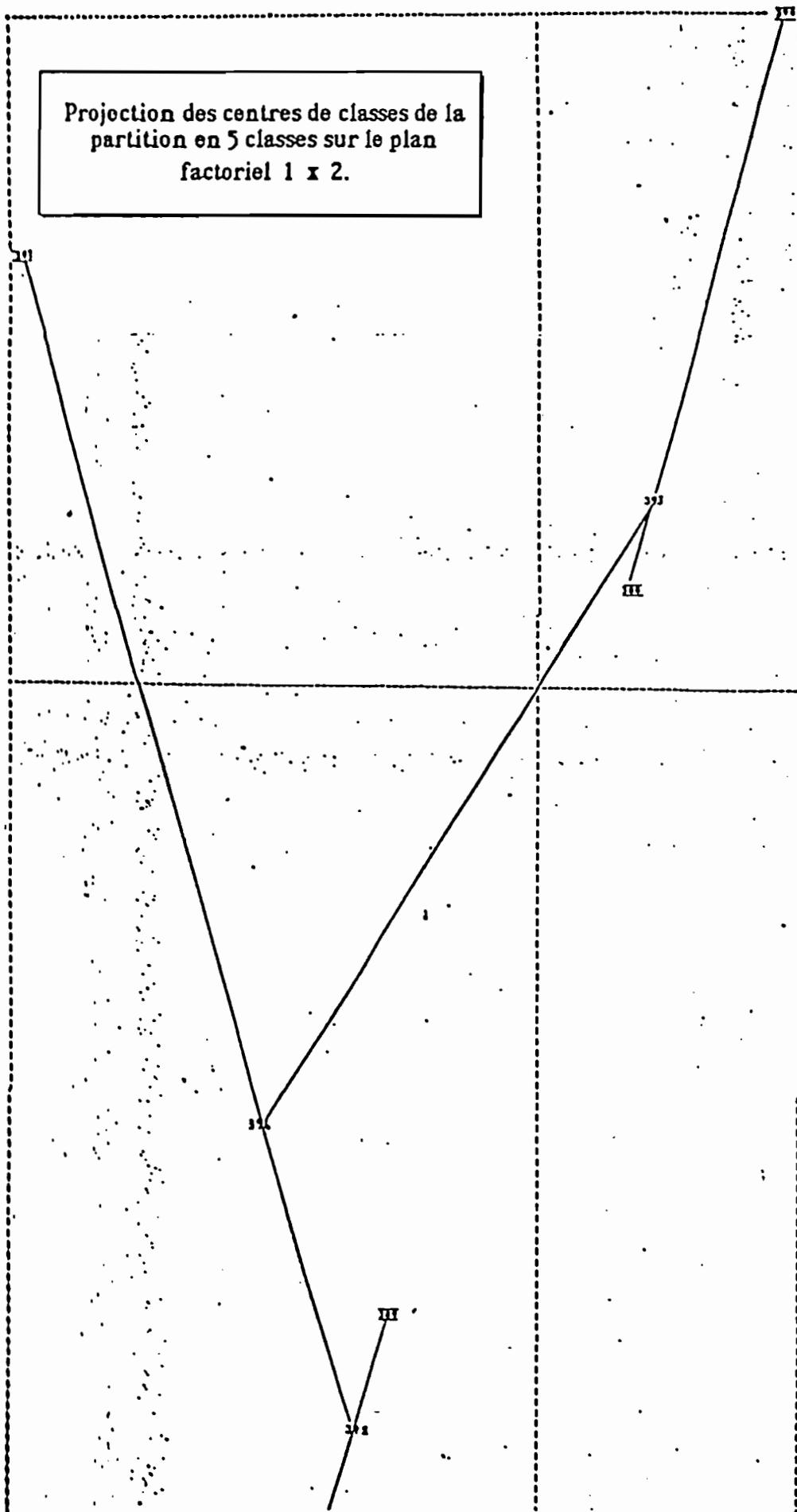
Figure 5.12 : Classification ascendante hiérarchique du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie

Histogramme des indices de niveau de la hiérarchie de partitions.

J	I(J)	A(J)	W(J)	U(J)	Q(J)	HISTOGRAMME DES INDICES DE NIVEAU DE LA HIERARCHIE

3951	2361	3931	3901	901	901
3941	1101	3911	3921	071	1021
3931	1001	3801	3901	021	1031
3921	901	3091	3771	301	2191
3911	711	3611	3331	201	201
3901	601	3001	3731	271	2751
3891	671	3001	3071	271	3021
3881	661	3601	3061	261	3201
3871	651	3501	3791	261	3501
3861	511	3701	3751	201	3751
3851	021	3721	3031	171	3911
3841	001	3501	3621	161	0071
3831	391	3011	3021	161	0231
3821	371	3601	3701	151	0301
3811	301	3521	3751	101	0511
3801	301	3671	3701	101	0651
3791	311	3561	3601	121	0771
3781	201	3001	2921	111	0001
3771	271	3571	3101	111	0991
3761	261	3591	3701	101	5091
3751	221	3501	3631	91	5101
3741	211	3691	3711	91	5271
3731	211	2901	3601	01	5551
3721	211	2961	3651	01	5031
3711	211	3291	2901	01	5521
3701	201	3221	3001	01	5601
3691	201	3051	2911	01	5671
3681	191	3001	2991	71	5751
3671	101	3351	3031	71	5021
3661	101	3371	3001	71	5691
3651	101	3551	3201	71	5071
3641	101	3391	3091	71	0001
3631	171	3301	3091	71	0111
3621	161	3131	3511	71	0171
3611	151	2051	3301	61	0231
3601	151	3201	3021	61	0291
3591	101	3311	3301	61	0351
3581	101	01	3531	61	0011
3571	101	2021	3071	61	0071
3561	101	2091	2061	61	0521
3551	101	3101	3201	61	0501
3541	101	901	3011	51	0631
3531	101	11	3321	51	0691
3521	101	3121	3001	51	0701
3511	131	3301	3001	51	0791
3501	131	3151	3001	51	0001
3491	131	3271	3021	51	0901
3481	121	3171	2931	51	0951
3471	121	261	571	51	0991
3461	121	2091	3071	51	7001
3451	121	3031	3231	51	7091
3441	121	3011	901	51	7131
3431	111	201	3251	01	7101
3421	111	2721	2571	01	7221
3411	111	2071	2701	01	7271
3401	111	2531	2791	01	7311
3391	111	011	2001	01	7351
3381	111	2651	2501	01	7001
3371	101	2771	2601	01	7001
3361	101	3211	2751	01	7001
3351	101	191	2691	01	7521
3341	101	1311	1021	01	7561
3331	101	2901	3161	01	7601
3321	91	01	3261	01	7601
3311	91	2761	2301	01	7671
3301	91	3111	2711	01	7711
3291	91	1271	3191	01	7751
3281	91	2971	2601	01	7701
3271	91	3101	711	01	7821
3261	91	2701	2731	01	7051
3251	91	2701	301	31	7091
3241	01	2201	2621	31	7921
3231	01	2251	2521	31	7951
3221	01	2611	1091	31	7901
3211	01	2291	3001	31	0021
3201	01	2231	3051	31	0051
3191	01	001	2011	31	0001
3181	71	2361	2671	31	0111
3171	71	1011	1261	31	0101
3161	71	1211	2631	31	0161
3151	71	2161	2951	31	0191
3141	71	321	331	31	0221
3131	71	2171	1731	31	0251
3121	71	2521	2001	31	0201
3111	71	2551	2371	31	0301
3101	71	2511	1931	31	0331
3091	71	2501	2001	31	0361
3081	71	2001	2031	31	0301
3071	71	001	2501	31	0011

Figure 5.13 : Classification ascendante hiérarchique du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie



Enfin, dans la partition en cinq classes, c'est à nouveau la classe 392 qui éclate en deux classes : 389 et 377. La classe 377 est constituée de sept zones que singularisent leur forte densité de bâti (DEN4). Pour cinq d'entre elles, il s'agit d'immeubles élevés, les deux autres contenant des bâtiments bas. Ces deux zones seront concervées dans la strate 2, tandis que les cinq autres constituent pour l'instant une classe à part, dont le faible effectif pourrait être compensé par un nombre de districts élevé dans chaque zone.

5.2.2 : Le sous-fichier des zones à petits bâtiments

Analyse factorielle

La pondération par un facteur 3 des modalités HAU1 et HAU2, DEN1 et DEN4, explique en grande partie le positionnement des trois premiers axes qui vont nous intéresser ici (voir le tableau 5.14 des contribution des modalités aux facteurs).

- *Premier facteur* : La modalité HAU2, située du côté positif de l'axe et qui ne concerne que 29 des zones du sous-ensemble à l'étude, contribue pour 22,4% à l'inertie du premier facteur. Les autres modalités contributives sont également de poids faibles : on trouve, dans l'ordre des contributions décroissantes, et également du côté positif de l'axe, DEN4 (8 zones, CTR=14,4%), SMA4 (8 zones, 10%), RAP3 (7 zones, 7,9%), SUR3 (10 zones, 5,8%) et VBD3 (11 zones 5%). A l'opposé, le gros du nuage ne fournit qu'une modeste contribution au positionnement de l'axe, qui est le fait de modalités "lourdes" ayant des coordonnées légèrement négatives : HAU1 (147 zones, 4,9% de l'inertie de l'axe), SUR1 (51 zones, 4,5%) et RAP1 (57 zones, 3%).

En résumé, près de 80% de l'inertie sur le premier facteur est due à l'opposition suivante : d'une part les modalités, minoritaires dans ce sous-ensemble, qui caractérisent les zones à bâtiments élevés (HAU2) ou de grande taille (SUR3, SMA4) et de formes allongées (RAP3) ou encore les zones à forte densité de bâti (DEN4), et d'autre part, la majorité du nuage, regroupé autour de HAU1, SUR1 et RAP1; les sous groupes minoritaires expliquent environ 65% de l'inertie de l'axe.

- *Deuxième facteur* : Près de 50% de l'inertie de l'axe est due à l'opposition entre les huit zones prenant la modalité DEN4 (coordonnées positives, 36% de l'inertie de l'axe) et les 29 zones comprenant une majorité de bâtiments élevés (coordonnées négatives, 14% de l'inertie de l'axe). C'est d'ailleurs cette unique caractéristique que l'on utilisera pour décrire la partition retenue.

- *Troisième facteur* : Les faits marquants et utiles par la suite sont la forte contribution des zones très peu denses (DEN1, 22% de l'inertie sur le facteur) et l'opposition entre les environnements à dominante claire (VVD1, coordonnée positive, 11,4%) et les environnements à dominante foncée (VVD4, coordonnée négative, 7%).

- *Plan factoriel 1-2* : Sur la figure 5.15, on distingue principalement trois régions sur le plan factoriel 1-2 :

1. La partie supérieure droite où figurent 9 zones, atypiques dans le sous-fichier, caractérisées par des densités de bâti élevées (DEN4) ou des grandes tailles de bâtiments (SUR4, SMI4).
2. La partie inférieure droite regroupe, autour de la modalité HAU2, les zones où les bâtiments sont majoritairement élevés, de taille moyenne et de forme plus ou moins allongée (SUR3, RAP3)
3. Enfin, autour du centre de gravité, se trouve la majorité des zones à petits bâtiments bas.

Ce plan factoriel permet de repérer et de caractériser facilement les petit sous-ensembles originaux. On y projettera avec profit les centres des classes de la partition en trois classes.

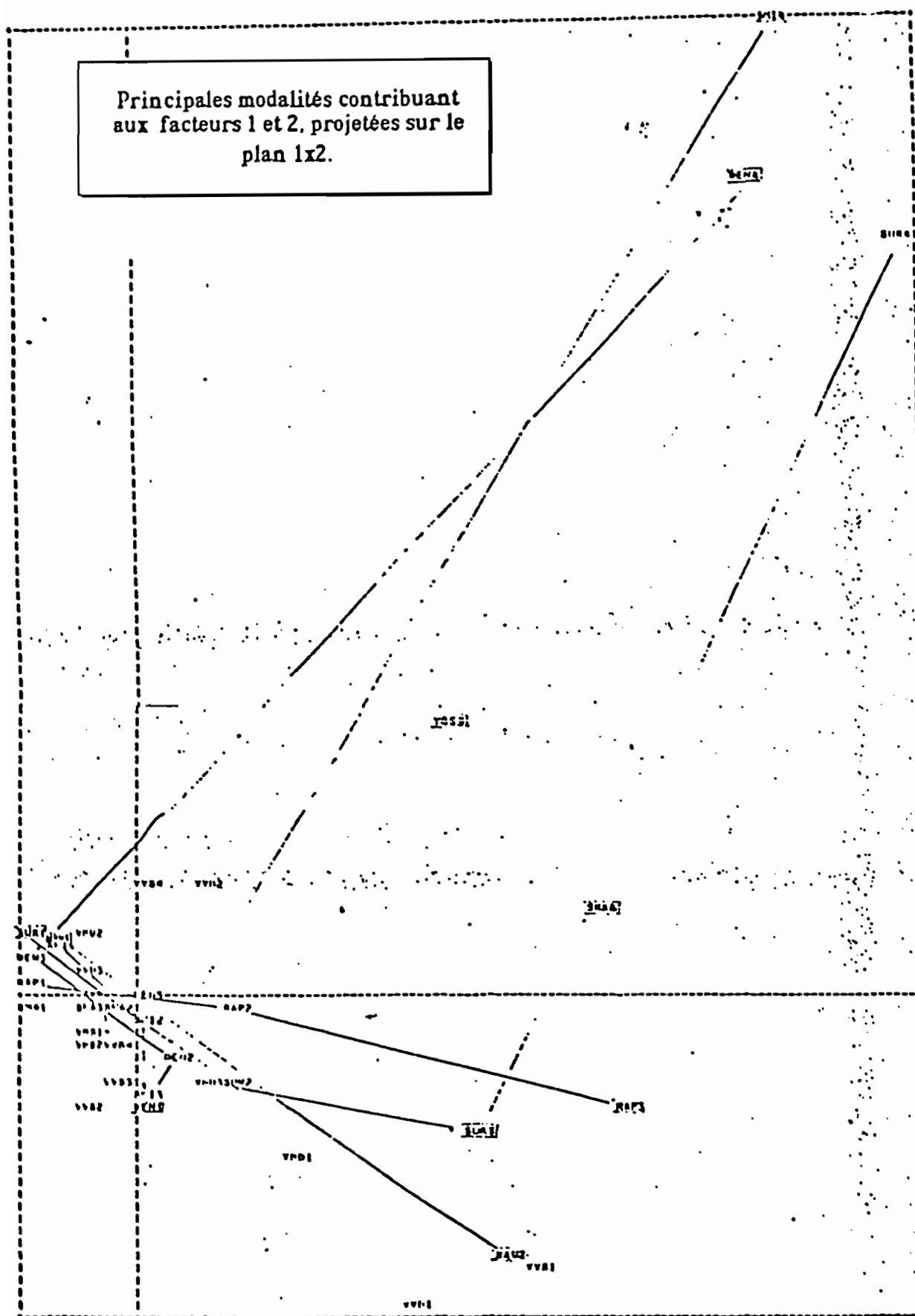
Tableau 5.14 : Analyse factorielle du sous-fichier des zones à petits bâtiments : 191 zones x 40 modalités descriptives de la morphologie

Contribution des modalités des descripteurs morphologiques aux 7 premiers facteurs

I	J1	CLT	R310	J001	10F	CUR	CTRI	20F	COP	CTRI	30F	CUR	CTRI	40F	COP	CTRI	50F	COP	CTRI	60F	COP	CTRI	70F	COP	CTRI	80F	COP	CTRI
101	422	59	45	74	3	1	-436	165	52	153	398	2191	-240	32	20	310	53	41	-674	252	2201	378	79	781				
102	356	20	35	178	8	21	-246	17	51	-601	100	37	81	2	11	-292	22	111	211	12	71	-838	195	1141				
103	495	35	231	-842	121	241	168	17	4	-812	105	301	133	11	4	-151	18	61	595	220	1001	112	0	41				
104	790	0	66	2388	260	1491	3225	491	3591	-307	4	41	377	7	41	192	2	21	-702	24	301	102	0	11				
105	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0		
106	481	20	241	-301	26	61	101	10	31	-454	60	211	296	25	101	-786	176	891	330	31	171	738	153	881				
107	438	23	101	30	8	01	-01	21	21	-08	6	11	-102	93	141	234	138	251	-65	11	21	-256	166	301				
108	278	8	301	32	0	01	-381	8	31	1836	175	231	1257	82	321	-355	7	41	-246	3	21	-45	0	01				
109	367	1	341	2530	73	211	3688	170	641	2265	54	251	1526	27	131	1883	24	151	710	6	41	808	8	61				
110	220	30	241	-450	107	221	-35	1	01	238	29	91	-273	38	131	-31	1	01	-68	2	11	-288	42	201				
111	56	81	201	-32	1	01	-36	1	01	-55	3	11	78	5	11	189	31	111	86	2	11	-122	13	51				
112	352	9	331	-188	8	11	-32	0	01	-752	63	261	457	23	111	-217	5	31	545	35	231	1011	221	1471				
113	882	6	331	1883	365	1001	337	12	41	-28	0	01	159	3	11	-680	50	291	-335	12	81	-16	0	01				
114	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0		
115	249	21	281	599	111	261	-628	122	371	-75	2	11	109	4	11	55	1	01	8	0	01	185	11	61				
116	386	82	111	-237	179	121	242	135	161	21	1	01	-23	1	01	-75	13	31	11	0	01	-169	66	151				
117	102	6	341	210	3	11	-323	7	31	-341	8	31	-58	0	01	882	15	91	217	3	21	980	66	451				
118	145	38	211	-179	28	41	-126	12	31	-237	43	111	328	79	231	-108	17	61	-151	17	71	-68	3	11				
119	263	34	201	-172	28	41	-176	25	61	278	82	151	-310	77	221	85	2	11	237	45	181	-198	30	121				
120	441	11	321	1158	140	601	1088	168	591	-360	19	81	28	0	01	278	11	61	-172	2	21	838	100	631				
121	687	4	341	1088	57	161	-1278	78	291	2282	275	1141	970	50	241	-1862	112	661	756	30	201	-282	8	31				
122	882	20	281	288	20	51	454	61	191	14	0	01	-800	141	561	-175	0	41	-168	8	01	-918	242	1361				
123	883	24	241	-221	20	51	118	7	21	334	55	161	-656	210	721	200	28	121	479	112	531	586	167	811				
124	884	35	221	-122	10	21	-208	27	71	-617	250	691	822	883	1001	52	2	11	-330	71	311	39	1	01				
125	886	1	351	1524	38	111	-1090	20	81	3893	228	991	2300	88	441	-3866	250	1541	1588	82	291	-1077	19	101				
126	221	14	311	-264	18	31	-428	35	121	178	6	21	376	26	111	-144	4	21	146	4	21	833	133	811				
127	730	32	231	-81	8	11	-315	58	151	-464	125	361	679	268	881	109	23	91	-591	203	921	-289	49	221				
128	883	40	201	81	5	11	847	166	371	129	14	31	-750	867	1321	4	0	01	424	149	581	-68	2	11				
129	799	187	121	-310	475	491	203	204	281	89	39	61	111	61	111	88	11	21	-33	5	11	-23	3	11				
130	810	29	611	1484	442	2241	-1017	206	1381	-588	71	531	-522	54	471	-301	18	191	307	19	221	42	0	01				
131	723	51	161	-503	341	451	232	72	131	-73	7	11	-105	51	111	-406	221	611	-149	30	91	-1	0	01				
132	724	26	261	331	48	101	-355	53	151	70	2	11	365	56	201	1022	436	1971	549	126	431	-138	8	81				
133	363	10	321	1272	218	581	-514	36	131	-432	25	101	-100	5	21	-488	29	161	-533	38	241	301	12	81				
134	613	1	301	3031	105	301	2967	106	301	4828	224	451	3245	120	581	-2453	88	291	1038	12	81	-583	8	31				
135	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0		
136	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0		
137	691	57	131	-397	286	311	88	8	11	-79	11	21	-179	58	111	-350	222	511	-221	89	231	94	18	81				
138	717	24	261	303	56	121	-37	1	01	200	15	51	672	174	651	808	248	1151	717	198	1021	-255	25	131				
139	341	7	331	1860	288	791	-836	16	61	-300	12	51	-608	62	291	-45	1	01	-838	16	101	-26	0	01				
140	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0	01	0	0		

I I 10001 10002 10001 10001 10001 10001 10001 10001 10001 10001 10001

Figure 5.15 : Analyse factorielle du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2)



- *Plan factoriel 1-3* (fig 5.16) : Deux remarques permettront d'utiliser ce plan factoriel pour caractériser la partition en sept classes :

1. A droite du plan, les modalités SUR4 et SMI4, de coordonnées positives sur le troisième facteur, s'opposent à la modalité DEN4, de coordonnée négative.
2. Dans la partie centrale du nuage (faible valeur absolue du premier facteur), le troisième facteur oppose les zones peu denses (DEN1) et de valeur claire (VVD1, VVS1) aux zones plus denses (DEN2 et 3) où l'environnement du bâti est foncé (VVD4).

Classification hiérarchique

L'histogramme des indices de niveau (fig 5.17) présente des seuils significatifs pour les partitions en 2, 3, 4 et 5 classes puis pour la partition en 7 classes. Ensuite la partition en 11 classes présenterait trop de singletons. Les conclusions tirées de l'analyse factorielle permettent de caractériser clairement ces différentes partitions à partir de leur projection sur les plans factoriels 1x2 et 1x3.

Sur le plan 1x2 (fig 5.18), la position des centres des classes 372 et 369, issues de la première partition, est éloquent : les 9 zones de la classe 369 apparaissent très éloignées du centre de gravité, caractérisées par les modalités SMI4, DEN4 et SUR4. De même, les 25 zones de la classe 371, issues d'une nouvelle "ponction" dans 372 (le gros du nuage étant maintenant regroupé dans 370), sont facilement caractérisées par les modalités HAU2, SUR3 et RAP3.

Les séparations qui s'opèrent ensuite à l'intérieur des classes 371, 370 et 369, doivent être observées sur le plan 1x3 (fig 5.19). En premier lieu, la classe 371 est scindée quasi parallèlement au troisième axe, en 363 (coordonnée négative) et 361 (coordonnée fortement positive). La classe 363 regroupe 22 zones où dominent les bâtiments élevés de taille moyenne aux valeurs de toit parfois foncées. Cette classe ira rejoindre dans la strate numéro 4 les plus petits des immeubles élevés issus du sous-ensemble des zones à grands bâtiments. Située à l'opposé du troisième facteur par rapport à 363, la classe 361 contient 8 zones à bâtiments élevés, caractérisées par les modalités DEN1, VVD1 et VVS1. Il s'agit de zones très peu denses à valeur très claire, où figurent quelques bâtiments isolés. Etant donné la faible densité du bâti, elles iront grossir la strate 6 dont il sera question ci-dessous.

La partition en cinq classes permet ensuite la subdivision de la classe 370 en 366 et 368. Comme en témoigne la position des deux centres de classes le long de l'axe 3, la raison principale de cette séparation est la densité du bâti dans les zones : les 92 zones qui constituent 366 sont regroupées autour des modalités SUR1, HAU1, VVD4, DEN2, DEN3, SMI1, SMA1, SMA2, etc... Il s'agit des zones, majoritaires dans le sous-ensemble à l'étude, occupées par de petits bâtiments bas, où la densité du bâti est moyenne et la valeur de l'environnement foncée. Elles constitueront la strate 5 qui rassemble ce qu'on peut appeler les zones pavillonnaires. La classe 368 est groupée autour de la modalité DEN1. La majorité des zones qui la constitue formeront la majorité de la strate 6 : il s'agira des zones à faible densité de bâti, situées pour la plupart à la périphérie de l'agglomération ; le plus souvent l'environnement des rares bâtiments qui s'y trouvent est constitué de terrains à usage agricoles dont la valeur est assez claire.

Enfin, dans la partition en sept classes, ce sont les classes 369 et 368 qui éclatent : la classe 369, qui contenait des zones atypiques, soit par la taille des bâtiments, soit par leur densité, fournit les classes 360 et 357. La classe 360 est constituée de deux zones à grands bâtiment élevés, qui seront reversées dans la strate 3, provenant majoritairement du sous-ensemble des zones à grands bâtiments. Quant à 357, elle regroupe 7 zones à forte densité de bâti, mais dont les autres caractéristiques morphologiques sont variables. Elles seront versées dans la septième strate qui contient un reliquat de zones difficiles à caractériser. Dans cette strate, on trouve également la classe 352 formée de 11 zones provenant de l'éclatement de 368. Ces zones sont aussi assez atypiques mais le restant de la classe 368 (classe 367), caractérisé par la faible densité du bâti, constitue le gros de la strate 6 dite des zones péri-urbaines peu denses.

Figure 5.16 : Analyse factorielle du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3)

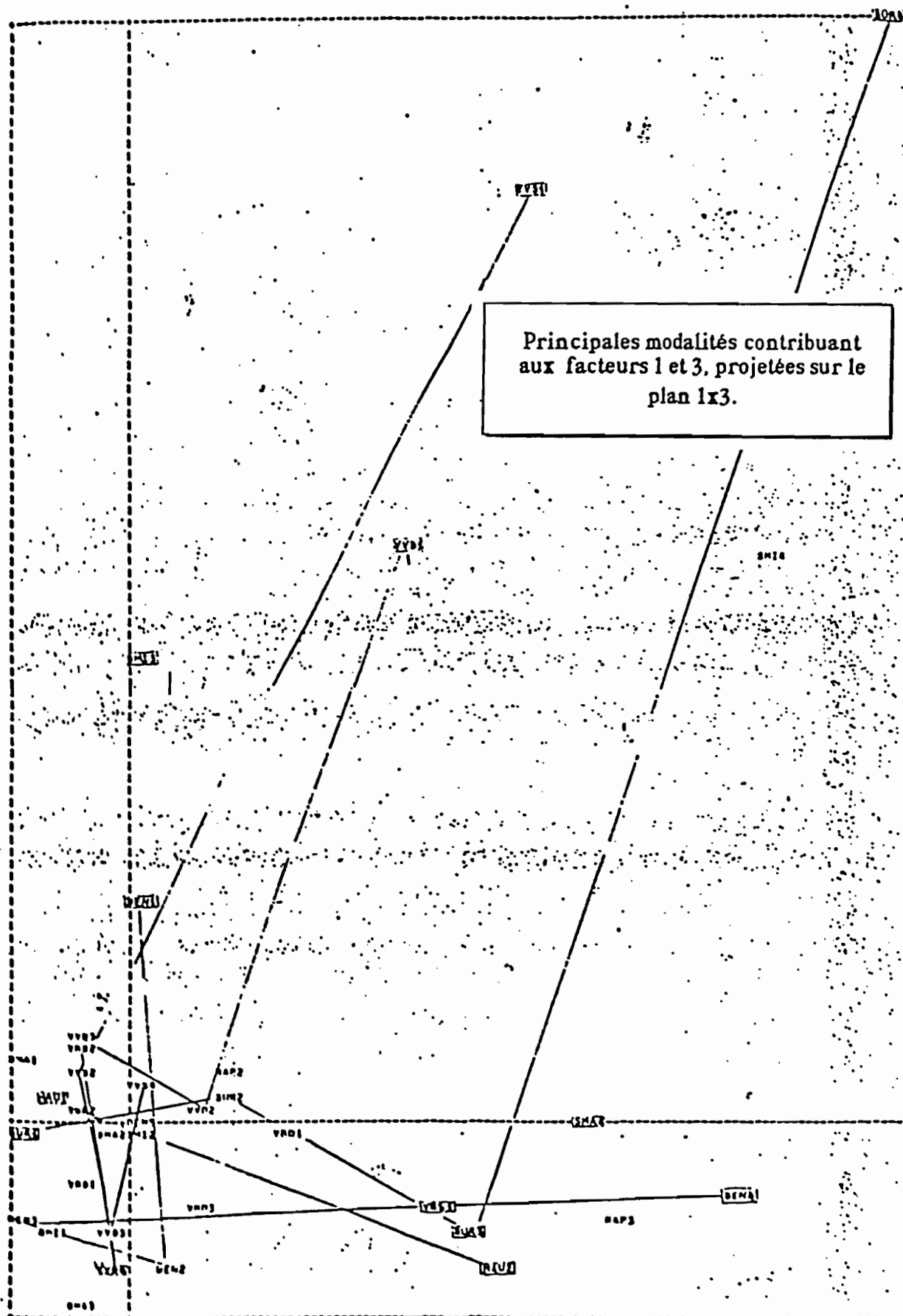


Figure 3.17 : Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie

Histogramme des indices de niveau de la hiérarchie de partitions.

HISTOGRAMME DES INDICES DE NIVEAU DE LA HIERARCHIE				
J	I	II	III	IV
3731	2011	3721	3691	601
3721	1531	3701	3711	601
3711	1321	3611	3631	551
3701	1131	3661	3691	071
3691	961	3601	3571	001
3681	921	3671	3521	101
3671	711	3501	3641	191
3661	681	3551	3651	281
3651	641	3591	3621	271
3641	601	3641	3641	271
3631	461	3691	3531	211
3621	461	3361	3641	191
3611	431	3321	01	191
3601	391	1371	1701	161
3591	371	3261	3561	161
3581	371	3001	3541	151
3571	331	2731	3101	141
3561	301	3341	3301	131
3551	201	3501	3511	121
3541	201	3021	2901	111
3531	261	3051	2161	111
3521	241	3231	3031	101
3511	231	3201	3071	101
3501	231	3041	3351	101
3491	211	2941	2561	91
3481	201	3191	3331	01
3471	201	3071	3341	01
3461	101	3121	3411	01
3451	101	3271	3391	71
3441	101	3151	2611	71
3431	171	3101	3371	71
3421	171	2531	3291	71
3411	151	41	3311	61
3401	141	3141	3241	61
3391	141	3091	3251	61
3381	141	3001	2711	61
3371	141	3161	1341	61
3361	131	2051	3221	61
3351	131	3241	3051	51
3341	131	2001	2941	51
3331	121	2721	3131	51
3321	121	51	91	51
3311	121	2991	31	51
3301	121	3021	2701	51
3291	121	2901	3011	51
3281	121	3111	2751	51
3271	111	3171	2911	51
3261	111	2931	2071	51
3251	111	3211	351	51
3241	111	3001	691	01
3231	101	1501	2921	41
3221	101	2021	3061	41
3211	91	2791	2271	41
3201	91	2571	1031	01
3191	91	2001	2591	01
3181	91	2771	3031	01
3171	91	2671	2651	01
3161	91	1051	2951	01
3151	91	2501	2971	41
3141	01	1611	2341	31
3131	01	2001	2011	31
3121	01	11	61	31
3111	01	2441	2421	31
3101	01	2461	2091	31
3091	01	2501	2001	31
3401	01	2261	2741	31
3071	01	1111	2621	31
3461	01	2491	2961	31
3051	71	2211	2641	31
3041	71	2541	2431	31
3031	71	1271	2641	31
3021	71	2601	2321	31
3011	71	2531	2031	31
3001	71	121	2741	31
2991	71	21	71	31
2981	61	1731	1101	31
2971	61	2391	1261	31
2961	61	2101	041	31
2951	61	401	1301	31
2941	61	2741	1331	31
2931	61	2351	1291	21
2921	61	1671	1651	21
2911	61	1711	1341	21
2901	61	2641	2471	21
2891	61	1661	721	21
2881	61	311	561	21

Figure 5.18 : Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie

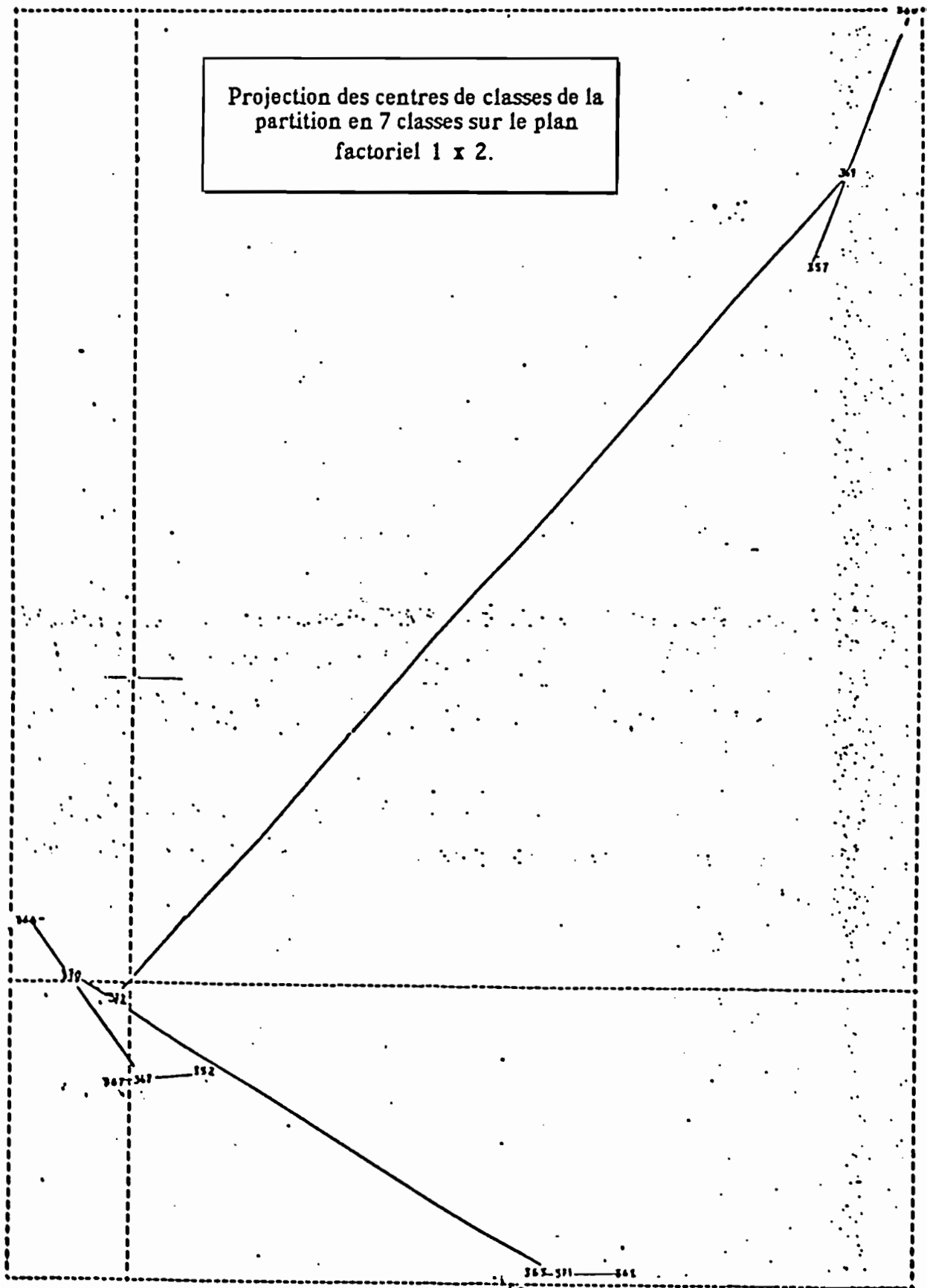
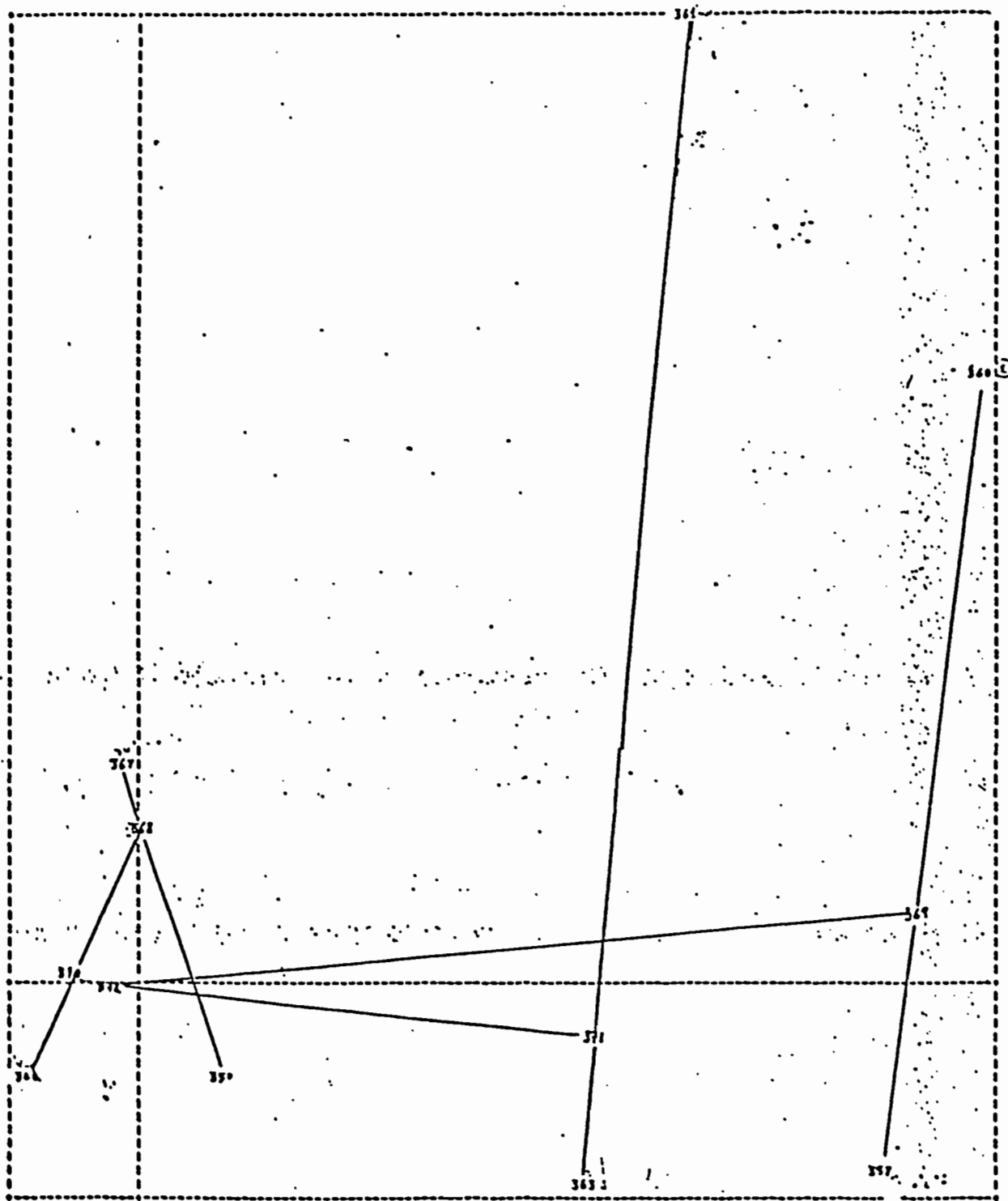


Figure 5.19 : Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie

Projection des centres de classes de la partition en 7 classes sur le plan factoriel 1 x 3.



5.3 : La typologie en huit classe de la morphologie urbaine

Le résultat de l'analyse typologique est une classification de l'ensemble des zones en huit types morphologiques. Rappelons que l'on a fixé à cette typologie un double objectif : d'une part elle doit servir de stratification pour le tirage d'un échantillon représentatif, en termes de surfaces, des types de morphologie urbaine présents sur la zone d'étude, d'autre part, elle doit permettre de tester l'efficacité d'une stratification morphologique dans l'estimation de variables démographiques par sondage. Il est donc nécessaire que la stratification remplisse deux conditions :

1 . Chaque strate doit être composée d'un nombre suffisant de districts INSEE. On a vu que cet impératif correspond à un minimum de 30 zones par strate.

2 . L'application démographique du programme de recherche impose que la typologie rende compte des distinctions majeures des types d'habitat ; par contre, les subdivisions dans les zones non habitées importent beaucoup moins.

Dans ce qui suit on décrit successivement la typologie de deux points de vue : en caractérisant la morphologie de chaque strate à partir des tableaux donnant les fréquences des modalités descriptives, puis en examinant la répartition spatiale des types à partir de leur cartographie.

5.3.1 : caractérisation morphologique des strates

La synthèse des résultats de l'analyse est présentée sous forme de tableaux croisant les huit types morphologiques et les modalités d'un descripteur donné.

On aura donc dix tableaux de ce type, dont la case courante (i,j) contient le nombre de zones de la strate j, possédant la modalité i du descripteur ainsi que les fréquences relatives en % par rapport au total général, par rapport au total de la modalité et par rapport au total de la strate (figure 5.20 à la fin du paragraphe 5.3).

Strate 1 : Elle regroupe les 49 zones les plus denses du domaine d'étude. Ses principales caractéristiques morphologiques sont les suivantes :

- 85% des zones sont concentrées sur les modalités de rang 4 ou 5 de la variable "densité du bâti" (DEN4, DEN5).

- Les éléments bâtis sont en majorité petits ou moyens et de forme peu allongée : SUR1, 2 ou 3 (98%), SMI1 ou 2 (100%) et la valeur du rapport longueur/largeur qui n'excède pas 5 est majoritairement comprise entre 1 et 2 (67%).

- La valeur du bâti est à dominante sombre : VBD3 (65%) et VBD2 (35%).

- L'environnement du bâti, ici très minoritaire en surface, a une valeur moyenne (31%) ou très foncée (63%).

Pour cette strate, on peut donc formuler une légende synthétique assez précise :

Strate 1 ("centre ville dense") : zones à densité de bâti forte ou très forte où les petits et moyens bâtiments de forme peu allongée à toits sombres dominant.

Strate 2 : Il s'agit des 48 zones, où dominant les bâtiments industriels, qui allient la grande taille des éléments bâtis et la faible élévation. On y observe les fréquences suivantes :

- SUR4, SUR5 et SUR6 : 75% des zones. Au minimum, la taille des bâtiments correspond à la modalité SUR3 (25%).

- HAUT1 : 96% des zones.

- On trouve à peu près toutes les formes de bâtiments avec toutefois une majorité de formes allongées : RAP3 et RAP4 (63%).

- La densité du bâti est moyenne : DEN2 et DEN3 (71%).

- La valeur de l'environnement est également moyenne : VVD2 et VVD3 (90%).

- Toutes les valeurs du bâti sont représentées : VBD1 (25%), VBD2 (44%), VBD3 (31%).

Légende de la strate 2 ("quartiers industriels") : zones à grands et très grands bâtiments bas. Forme majoritairement allongée et environnement clair ou gris moyen.

Strate 3 : 116 zones à grands bâtiments que l'on peut caractériser de manière assez précise :

- Grande taille du bâti : SUR3 (54%), SUR4 (35%), SUR5 (10%).
- HAU2 : 97%
- Deux modes de formes pour le bâti majoritaire : carrée (RAP1, 15%) ou allongée (RAP3 et RAP4, 78%).
- Valeur claire ou moyenne du bâti : VBD1 (68%), VBD2 (31%).
- Valeur variable de l'environnement : VVD2 (44%), VVD3 (34%), VVD4 (21%).
- Densité moyenne du bâti : DEN2 ou DEN3 (85%).

Légende de la strate 3 ("grands ensembles d'habitation") : zones à grands bâtiments hauts de forme carrée ou très allongée, la valeur du bâti est majoritairement claire.

Strate 4 : 49 zones où dominant les bâtiments de taille moyenne et majoritairement hauts :

- Taille moyenne du bâti : SUR1 (16%), SUR2 (41%), SUR3 (39%). La taille du bâti est assez homogène à l'intérieur des zones : SMI1 ou SMI2 (74%), SMA1 ou SMA2 (63%).
- Les constructions hautes dominant : HAU2 (86%), HAU1 (14%).
- Forme carrée (RAP1 : 49%) ou peu allongées (RAP2 ou RAP3 : 49%).
- Densité du bâti plutôt faible : DEN1 (37%), DEN2 (39%) ou DEN3 (24%).
- Valeur claire du bâti : VBD1 (63%), VBD2 (37%).
- Valeur foncée de l'environnement : VVD3 et VVD4 (65%).

Légende de la strate 4 ("Petits immeubles d'habitation") : zones où dominant les petits et moyens bâtiments élevés, de forme peu allongée, aux valeurs de toits claires ou moyennes. Densité du bâti plutôt faible, valeur de l'environnement plutôt foncée.

Strate 5 : 103 zones avec un bâti de petite taille et uniformément bas. Les modalités majoritaires sont les suivantes

- Petite taille du bâti : SUR1 (69%), SUR2 (25%).
- HAU1 : 98%
- Forme majoritairement carrée : RAP1 (72%), RAP2 (24%).
- Densité plutôt forte du bâti : DEN2 (35%), DEN3 (61%).
- Environnement assez foncé : VVD3 (33%), VVD4 (42%).
- Valeur gris moyen des toits : VBD2 (78%).

Légende de la strate 5 ("Zones pavillonnaires") : zones à petits bâtiments majoritairement bas (un ou deux niveaux), de forme généralement carrée. Densité assez forte du bâti, valeur plutôt foncée de l'environnement.

Strate 6 : 53 zones à faible densité de bâti généralement situées à la périphérie de l'agglomération.

- 90% des zones prennent la modalité DEN1.
- Petite taille du bâti : SUR1 et SUR2 (83%).
- Forme carrée ou peu allongée : RAP1 (70%), RAP2 et RAP3 (30%).
- Constructions basses : HAU1 (92%).
- Valeur gris moyen du bâti : VBD2 (77%).
- Hétérogénéité de la valeur de l'environnement : VVD1 (17%), VVD2 (17%), VVD3 (38%), VVD4 (28%).

Légende de la strate 6 ("Zones péri-urbaines peu denses") : zones à très faible densité de bâti, à petit bâtiments bas de forme peu allongée, valeur moyenne des toits et variable de l'environnement.

Strate 7 : 18 zones où la taille du bâti est petite ou moyenne mais plus difficiles à caractériser par ailleurs.

- Taille du bâti : SUR1 (50%), SUR2 (33%), SUR3 (17%). On trouve dans toutes ces zones de petits bâtiments (SMI1 : 39%, SMI2 : 61%), mais dans certaines d'elles il s'en trouve aussi d'assez grands (SMA2 : 50%, SMA4 : 28%).
- Les formes sont carrées ou peu allongées : RAP1 (44%), RAP2 (44%).
- La valeur du bâti est plutôt sombre : VBD2 (33%), VBD3 (61%).
- Les autres caractères morphologiques ont une forte variabilité et particulièrement la densité du bâti : DEN1 (22%), DEN2 (17%), DEN3 (22%), DEN4 (39%).

Légende de la strate 7 ("Reliquat morphologique") : Strate hétérogène où dominant les petits bâtiments aux toits sombres, de forme peu allongée. Les densités de bâti sont très variables.

Strate 8 : Cinq zones de grands bâtiments hauts à forte densité de bâti : SUR3 (60%), SUR4 (40%), HAU2 (100%), DEN4 (100%), SMA3 (100%), RAP3 (80%), RAP4 (20%).

Légende de la strate 8 : 5 zones densément bâties de grands bâtiments élevés.

5.3.2 : Répartition spatiale des strates

La cartographie des strates et le tableau 5.21 (système TIGRE : requête sur la relation morphologie, attribut strate, pour l'ensemble de la zone d'étude) précisent la répartition spatiale de chacune des strates dans la zone d'étude. Trois phénomènes majeurs apparaissent :

- *les strates occupent des surfaces variables* : alors que les strates "grands ensembles d'habitation" et "zones pavillonnaires" occupent des surfaces très importantes (environ 20 % de l'ensemble de la zone d'étude chacune), les strates 7 ("reliquat") et 8 ("zones densément bâties de grands bâtiments élevés") n'ont qu'une superficie très réduite (respectivement 3 % et 1 %), les strates "centre ville dense", "quartiers industriels" et "petits immeubles d'habitation" se trouvant en position intermédiaire (6 à 10 %).
- *le degré de connexité des strates est, lui aussi, variable* : à la concentration de la strate "centre ville dense", et à un moindre degré de la strate "quartiers industriels", s'oppose la dispersion géographique des éléments composant la strate "petits immeubles d'habitation". On observe que certaines zones connexes distinguées visuellement sont regroupées dans une même strate.
- *la répartition géographique des strates rend compte des traits majeurs de l'organisation de l'espace à Marseille*. opposition nord/sud, et disposition concentrique. La partie centrale de Marseille autour du vieux port correspond à la strate 1, "centre ville dense". Au nord-ouest de Marseille, la strate industrielle (2) jouxte les grands ensembles de la strate 3. Le croissant sud et est se caractérise par la juxtaposition des éléments appartenant aux strates 3 ("grands ensembles d'habitation"), 4 ("petits immeubles d'habitation") et 5 ("zones pavillonnaires"). Les zones peu denses de la strate 6 sont toutes à la périphérie de l'agglomération.

Figure 5.20 : fréquence des différentes modalités des descripteurs morphologiques selon les strates (1)

TABLE OF DENSK BY STRATE

DENSK	STRATE									TOTAL
	1	2	3	4	5	6	7	8		
1	0	9	16	18	4	48	4	0	99	
	0.00	2.04	3.63	4.08	0.91	10.88	0.91	0.00	27.45	
	0.00	9.09	16.16	18.18	4.04	48.48	4.04	0.00		
	0.00	18.75	13.79	36.73	3.88	90.57	22.22	0.00		
2	0	20	47	19	36	2	3	0	122	
	0.00	4.54	9.52	4.31	8.16	0.45	0.68	0.00	27.66	
	0.00	16.39	38.43	15.57	29.51	1.64	2.46	0.00		
	0.00	41.67	36.21	38.79	39.95	3.77	16.67	0.00		
3	8	14	54	12	63	3	4	0	159	
	1.81	3.17	12.47	2.72	14.29	0.68	0.91	0.00	36.05	
	5.03	8.81	38.59	7.55	39.62	1.89	2.52	0.00		
	16.33	29.17	47.41	24.49	61.17	5.66	22.22	0.00		
4	34	4	2	0	0	0	7	5	52	
	7.71	0.91	0.45	0.00	0.00	0.00	1.59	1.13	11.79	
	65.38	7.69	3.85	0.00	0.00	0.00	13.46	9.67		
	69.39	8.35	1.72	0.00	0.00	0.00	38.89	100.00		
5	7	1	1	0	0	0	0	0	9	
	1.59	0.23	0.23	0.00	0.00	0.00	0.00	0.00	2.04	
	77.78	11.11	11.11	0.00	0.00	0.00	0.00	0.00		
	14.29	2.08	9.86	0.00	0.00	0.00	0.00	0.00		
TOTAL	49	48	116	49	103	53	18	5	441	
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00	

MAUTH

FREQUENCY	STRATE									TOTAL
	1	2	3	4	5	6	7	8		
1	21	46	4	7	101	49	13	0	241	
	4.76	10.43	0.91	1.59	22.90	11.11	2.95	0.00	54.65	
	8.71	19.09	1.66	2.90	41.91	20.33	5.39	0.00		
	42.86	45.83	3.45	14.29	98.06	92.45	72.22	0.00		
2	28	2	112	42	2	4	5	5	200	
	6.35	0.45	25.40	9.52	0.45	0.91	1.13	1.13	45.35	
	14.00	1.00	56.00	21.90	1.00	2.00	2.50	2.50		
	57.14	4.17	96.55	85.71	1.94	7.55	27.78	100.00		
TOTAL	49	48	116	49	103	53	18	5	441	
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00	

VDOR STRATE

FREQUENCY	STRATE									TOTAL
	1	2	3	4	5	6	7	8		
1	0	12	79	31	23	11	1	2	159	
	0.00	2.72	17.91	7.03	5.22	2.49	0.23	0.45	36.05	
	0.00	7.55	49.69	19.50	14.47	6.92	0.63	1.24		
	0.00	25.00	68.10	63.27	22.33	20.75	5.56	40.00		
2	17	21	16	18	80	41	6	2	221	
	3.85	4.76	4.16	4.08	18.18	9.30	1.36	0.45	50.11	
	7.69	9.50	16.29	8.18	36.20	18.55	2.71	0.90		
	34.69	43.75	31.03	36.73	77.67	77.36	33.33	40.00		
3	32	15	1	0	0	1	11	1	61	
	7.26	3.40	0.23	0.00	0.00	0.23	2.49	0.23	13.83	
	32.46	24.59	1.64	0.00	0.00	1.64	18.03	1.64		
	65.31	31.25	0.86	0.00	0.00	1.89	61.11	20.00		
TOTAL	49	48	116	49	103	53	18	5	441	
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00	

Figure 5.20 : fréquence des différentes modalités des descripteurs morphologiques selon les strates (2)

SURF	STRATE								TOTAL
	1	2	3	4	5	6	7	8	
FREQUENCY									
PERCENT									
NUM PCT									
COL PCT									
1	18	0	0	8	71	30	9	0	136
	8.00	0.00	0.00	1.81	16.10	6.80	2.00	0.00	30.80
	13.20	0.00	0.00	5.80	52.21	22.00	6.62	0.00	
	36.73	0.00	0.00	16.33	68.93	56.60	50.00	0.00	
2	24	0	2	20	26	18	6	0	92
	5.40	0.00	0.45	4.50	5.90	3.17	1.36	0.00	20.86
	26.09	0.00	2.17	21.74	24.24	15.22	6.52	0.00	
	48.98	0.00	1.72	40.82	25.20	26.42	33.33	0.00	
3	6	12	60	19	6	8	3	3	117
	1.36	2.72	13.61	4.31	1.36	1.81	0.68	0.68	26.53
	5.13	10.26	51.20	16.20	5.13	6.80	2.56	2.56	
	12.24	25.00	51.72	38.78	5.83	15.09	16.67	60.00	
4	1	10	41	2	0	1	0	2	57
	0.23	2.27	9.30	0.45	0.00	0.23	0.00	0.45	12.93
	1.75	17.54	71.93	3.51	0.00	1.75	0.00	3.51	
	2.00	20.83	35.34	4.08	0.00	1.80	0.00	40.00	
5	0	13	12	0	0	0	0	0	25
	0.00	2.95	2.72	0.00	0.00	0.00	0.00	0.00	5.67
	0.00	52.00	48.00	0.00	0.00	0.00	0.00	0.00	
	0.00	27.00	10.34	0.00	0.00	0.00	0.00	0.00	
6	0	13	1	0	0	0	0	0	14
	0.00	2.95	0.23	0.00	0.00	0.00	0.00	0.00	3.17
	0.00	92.86	7.14	0.00	0.00	0.00	0.00	0.00	
	0.00	27.00	0.86	0.00	0.00	0.00	0.00	0.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

SURF	STRATE								TOTAL
	1	2	3	4	5	6	7	8	
FREQUENCY									
PERCENT									
NUM PCT									
COL PCT									
1	10	7	17	24	74	37	8	0	177
	2.27	1.59	3.85	5.44	16.70	8.39	1.81	0.00	40.14
	5.65	3.95	9.60	13.56	41.81	20.90	4.52	0.00	
	20.41	10.58	14.66	48.98	71.84	69.81	44.44	0.00	
2	33	11	8	10	25	12	8	0	107
	7.40	2.49	1.81	2.27	5.67	2.72	1.81	0.00	24.26
	30.84	10.20	7.40	9.35	23.36	11.21	7.40	0.00	
	67.35	22.92	6.90	20.41	20.27	22.64	44.44	0.00	
3	6	21	48	14	4	4	2	4	103
	1.36	4.76	10.80	3.17	0.91	0.91	0.45	0.91	23.36
	5.83	20.39	48.60	13.54	3.88	3.88	1.94	3.88	
	12.24	43.75	41.38	28.57	3.88	7.55	11.11	80.00	
4	0	9	43	1	0	0	0	1	54
	0.00	2.04	9.75	0.23	0.00	0.00	0.00	0.23	12.24
	0.00	14.67	79.63	1.85	0.00	0.00	0.00	1.85	
	0.00	18.75	37.07	2.04	0.00	0.00	0.00	20.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

Figure 5.20 : fréquence des différentes modalités des descripteurs morphologiques selon les strates (3)

MAXI		STRATE								TOTAL
FREQUENCY	PERCENT	1	2	3	4	5	6	7	8	
ROW PCT	COL PCT									
1	8	0	0	5	35	24	3	0		75
	1.81	0.00	0.00	1.13	7.94	5.44	0.68	0.00		17.01
	10.67	0.00	0.00	6.67	46.67	32.00	4.00	0.00		
	16.33	0.00	0.00	10.20	33.98	45.28	16.67	0.00		
2	17	2	1	26	45	22	9	0		122
	3.85	0.45	0.23	5.90	10.20	4.99	2.04	0.00		27.66
	13.93	1.64	0.82	21.31	36.89	18.03	7.38	0.00		
	14.69	4.17	0.86	53.06	43.69	41.51	50.00	0.00		
3	13	8	52	7	17	2	1	5		105
	2.95	1.81	11.79	1.59	3.85	0.45	0.23	1.13		23.81
	12.38	7.62	49.52	6.67	16.14	1.90	0.95	4.76		
	26.53	16.67	40.83	14.29	16.50	3.77	5.56	100.00		
4	11	14	52	10	6	5	5	0		103
	2.49	3.17	11.79	2.27	1.36	1.13	1.13	0.00		23.36
	10.68	13.59	50.49	9.71	5.83	4.85	4.85	0.00		
	22.45	29.17	44.83	20.41	5.83	9.43	27.78	0.00		
5	0	24	11	1	0	0	0	0		36
	0.00	5.44	2.49	0.23	0.00	0.00	0.00	0.00		8.16
	0.00	66.67	30.56	2.78	0.00	0.00	0.00	0.00		
	0.00	50.00	9.88	2.04	0.00	0.00	0.00	0.00		
TOTAL	49	48	116	49	103	53	18	5		441
	11.11	10.88	26.30	11.11	23.36	12.02	4.06	1.13		100.00

MINI		STRATE								TOTAL
FREQUENCY	PERCENT	1	2	3	4	5	6	7	8	
ROW PCT	COL PCT									
1	27	11	13	6	27	5	7	1		97
	6.12	2.49	2.95	1.36	6.12	1.13	1.59	0.23		22.00
	27.84	11.34	13.40	6.19	27.84	5.15	7.22	1.03		
	55.10	22.92	11.21	12.24	26.21	9.43	38.89	20.00		
2	22	17	56	30	75	39	11	2		252
	4.99	3.85	12.70	6.80	17.01	8.84	2.49	0.45		57.14
	4.73	6.75	22.22	11.40	29.76	15.48	4.37	0.79		
	44.90	35.42	48.28	61.22	72.82	73.58	61.11	40.00		
3	0	10	23	5	1	9	0	0		48
	0.00	2.27	5.22	1.13	0.23	2.04	0.00	0.00		10.88
	0.00	20.83	47.42	10.42	2.08	18.75	0.00	0.00		
	0.00	20.83	19.83	10.20	0.97	16.98	0.00	0.00		
4	0	10	24	8	0	0	0	2		44
	0.00	2.27	5.44	1.81	0.00	0.00	0.00	0.45		9.98
	0.00	22.73	54.55	18.18	0.00	0.00	0.00	4.55		
	0.00	20.83	20.69	16.33	0.00	0.00	0.00	40.00		
TOTAL	49	48	116	49	103	53	18	5		441
	11.11	10.88	26.30	11.11	23.36	12.02	4.06	1.13		100.00

Figure 5.20 : fréquence des différentes modalités des descripteurs morphologiques selon les strates (4)

SAS

10139 FRIDAY, MARCH 26, 1966

TABLE OF VVDOR BY STRATE

MODAL	STRATE										TOTAL
	1	2	3	4	5	6	7	8	9	10	
FREQUENCY	0	1	1	0	0	9	0	0	0	0	11
PERCENT	0.00	0.23	0.23	0.00	0.00	2.04	0.00	0.00	0.00	0.00	2.49
ROW PCT	0.00	9.09	9.09	0.00	0.00	11.82	0.00	0.00	0.00	0.00	29.00
COL PCT	0.00	2.04	0.46	0.00	0.00	16.98	0.00	0.00	0.00	0.00	100.00
1	15	18	51	17	26	9	5	3			144
	3.40	4.08	11.56	3.85	5.90	2.04	1.13	0.68			32.65
	10.42	17.50	35.42	11.81	18.06	6.25	3.47	2.04			60.00
	30.61	37.50	43.97	34.64	25.24	16.98	27.78	60.00			100.00
2	3	24	39	12	34	20	4	0			137
	0.68	5.67	8.84	2.72	7.71	4.54	0.91	0.00			31.07
	2.19	18.25	28.47	8.76	28.82	18.60	2.92	0.00			60.00
	6.12	52.08	33.62	24.49	33.01	37.74	27.22	0.00			100.00
3	31	4	25	20	43	15	9	2			149
	7.03	0.91	5.67	4.54	9.75	3.40	2.04	0.45			33.79
	20.81	2.68	16.78	13.42	28.86	10.07	6.04	1.34			60.00
	63.27	4.33	21.55	40.42	41.75	28.30	50.00	40.00			100.00
TOTAL	49	48	116	49	103	53	18	5			441
VSDR	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13			100.00

MODAL	STRATE										TOTAL
	1	2	3	4	5	6	7	8	9	10	
FREQUENCY	3	6	13	10	3	3	0	0			38
PERCENT	0.68	1.36	2.95	2.27	0.68	0.68	0.00	0.00			8.62
ROW PCT	7.89	15.79	34.21	26.32	7.89	7.89	0.00	0.00			60.00
COL PCT	6.12	12.50	11.21	20.41	2.91	5.66	0.00	0.00			100.00
1	26	19	26	3	14	12	2	7			104
	5.90	4.31	5.90	0.68	3.17	2.72	0.45	0.45			23.58
	25.00	18.27	25.00	2.88	13.46	11.54	1.92	1.92			60.00
	53.06	39.58	22.41	6.12	13.59	22.64	11.11	40.00			100.00
2	3	12	24	14	33	21	6	2			115
	0.68	2.72	5.44	3.17	7.48	4.76	1.36	0.45			26.08
	2.61	18.43	20.87	12.17	28.70	18.26	5.22	1.74			60.00
	6.12	25.00	20.69	28.57	32.04	39.62	33.33	40.00			100.00
3	17	11	53	22	53	17	10	1			184
	3.85	2.49	12.02	4.99	12.02	3.85	2.27	0.23			41.72
	9.24	5.98	28.80	11.96	28.80	9.24	5.43	0.54			60.00
	30.69	22.92	45.69	44.90	51.46	32.08	55.56	20.00			100.00
TOTAL	49	48	116	49	103	53	18	5			441
VSDR	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13			100.00

MODAL	STRATE										TOTAL
	1	2	3	4	5	6	7	8	9	10	
FREQUENCY	24	24	39	26	46	25	4	2			192
PERCENT	5.90	5.44	8.84	5.90	10.43	5.67	0.91	0.45			43.58
ROW PCT	13.54	12.50	20.31	13.54	23.96	13.02	2.04	1.04			60.00
COL PCT	53.06	50.00	33.62	53.06	44.66	47.17	22.22	40.00			100.00
1	27	20	62	19	47	27	6	3			206
	4.99	4.54	14.06	4.31	10.66	6.12	1.36	0.68			46.71
	18.68	9.71	30.10	9.22	22.82	13.11	2.91	1.46			60.00
	40.90	41.67	53.45	38.78	45.63	50.98	33.33	60.00			100.00
2	1	4	15	4	10	1	8	0			43
	0.23	0.91	3.40	0.91	2.27	0.23	1.81	0.00			9.75
	2.33	9.34	34.88	9.30	23.24	2.33	18.60	0.00			60.00
	2.04	8.33	12.93	8.16	9.71	1.89	44.44	0.00			100.00
TOTAL	49	48	116	49	103	53	18	5			441
VSDR	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13			100.00

**Tableau 5.21 : surfaces occupées par les strates
(récapitulatif d'une interrogation du système SAVANNE)**

RECAPITULATIF DE LA REQUETE

Nom de la base : MARSEILLE

Point bas : Longitude 0 deg 10 min - Latitude 0 deg 2' min
 Point haut : Longitude 0 deg 17 min - Latitude 0 deg 11' min

DEFINITION DE LA REQUETE

Relation	Attribut	Classe	Valeurs
MORPHOLOGIE	Strate	1 2 3 4 5 6 7 8 non bati	

IMAGE PAR RELATION

MORPHOLOGIE Superficie totale : 185342

Code	Pixel	des carre	%	Strate
1	15482	2576.7180	0.08	1
2	19384	3226.1401	0.10	2
3	36848	6465.5938	0.21	3
4	10865	1808.2961	0.06	4
5	36014	5993.9219	0.19	5
6	14512	2415.2778	0.08	6
7	4985	829.6692	0.03	7
8	1388	231.0092	0.01	8
9	43864	7300.4219	0.24	non bati

5.4 : Enseignements méthodologiques et perspectives pour la suite de l'expérience, conclusions

Dans cette dernière partie, nous nous proposons d'évaluer l'apport d'une démarche de classification automatique des unités spatiales, par rapport aux démarches visuelles, plus intuitives mais probablement plus rapides. Nous examinerons les contraintes qu'impose une telle méthode et les améliorations dont elle pourrait bénéficier, avant de conclure sur les possibilités qu'elle ouvre pour la suite de l'expérience à Marseille.

Dans un premier temps, il est nécessaire de rappeler dans quel cadre, objectifs et nature de l'information, s'inscrit cette évaluation.

Les objectifs

Le premier objectif est la représentativité de l'échantillon de travail qui doit permettre la mise au point de méthodes d'extraction de l'information morphologique à partir de l'image satellite. Mais le travail d'analyse typologique fournit également une stratification dont nous évaluerons l'intérêt en termes de sondages démographiques. Cette évaluation, on l'a vu, ne s'arrêtera pas à la mesure du gain de précision apporté par la stratification synthétique ; nous la répèterons pour chacun des descripteurs morphologiques les plus importants pris isolément. Pour ce dernier point, et afin d'opérer la sélection des descripteurs à tester, il est indispensable de quantifier aussi précisément que possible les contributions des différents descripteurs à la différenciation des strates synthétiques. Un des principaux avantages de la méthode d'analyse typologique appliquée est précisément de permettre cette quantification.

De l'information morphologique aux données traitées

Pour obtenir une information morphologique exhaustive et homogène sur l'ensemble de la zone d'étude, le support que l'on sait pour l'instant utiliser est la photographie aérienne. Ce document correspond déjà à une première transformation de l'information dépendante de différents paramètres comme la visée verticale, l'émulsion utilisée, l'échelle de prise de vue et de restitution. La médiatisation se poursuit avec l'intervention du photo-analyste : privilégier des indicateurs visuels analytiques précis permet de mieux contrôler les effets de cette deuxième transformation. Mais quelque soit le mode de traitement de l'information extraite des photographies, ces deux transformations successives sont incontournables.

Une méthode visuelle de regroupement des zones d'égale apparence en strates, outre sa rapidité, n'exige pas d'autre transformation de l'information, mais elle ne permet pas de quantifier précisément les contributions de chaque descripteur au résultat final. Les méthodes d'analyse typologique, qui permettent cette quantification nécessitent d'autres transformations : élaboration du tableau de données codant la description des zones, homogénéisation des variables métriques et qualitatives et éventuellement, rééquilibrage de l'importance relative des descripteurs par pondération. Nous développons successivement chacun de ces points.

Elaboration du tableau de données

Nous exposons en détail dans l'annexe A.2.1 les principes de la codification de l'information en un tableau descriptif de la morphologie des zones. La liste des descripteurs correspond à un choix subjectif mais qui respecte certaines règles. Pour résumer, les caractères morphologiques, définis rigoureusement, doivent permettre de discriminer des zones voisines différentes et de regrouper des zones semblables non adjacentes.

C'est au cours de cette étape qu'est fixée la précision de l'information morphologique qui va être traitée : opter pour certains critères qualitatifs conduira à une dégradation des descripteurs métriques afin d'obtenir un tableau de données homogènes. Cette dégradation peut être de deux types : passage de la métrique à une échelle de notes discrètes par seuillage des distributions ou passage au codage disjonctif où l'on ne rend plus compte du caractère ordonné des modalités.

Comme nous l'avons souligné lors de l'examen des distributions des caractères métriques, on ne perd pas beaucoup d'information en agglomérant les valeurs à la valeur centrale de chaque mode, du moins quand cette dernière est unique. Si l'on disposait par ailleurs d'échelles de notes permettant de coder les autres descripteurs jusqu'alors purement qualitatifs, on pourrait homogénéiser le tableau de description morphologique à un niveau d'information quasi-optimal compte tenu de la nature du phénomène étudié et de l'échelle d'observation.

Pour y parvenir, il faudra améliorer la méthode de codification adoptée pour la densité du bâti, sa hauteur et les valeurs de gris. En effectuant les mesures de chaque descripteur pour un certain nombre de zones-étalons représentant toutes ses modalités, on pourrait associer à chaque modalité une valeur moyenne qui deviendrait la note attribuée à la zone. Grâce à cette amélioration, on disposerait donc d'un tableau de notes qui, une fois dédoublé, se prête de façon classique à la même chaîne de traitements (AFC, CAH).

Seuillage des variables métriques et pondérations, effet Gutman et équilibre entre descripteurs

On a dit plus haut avoir préféré, pour décrire la taille et la forme de l'élément bâti majoritaire, les indicateurs synthétiques calculés "surface" et "rapport", aux variables originales "longueur" et "largeur". Ce choix n'a pas été fait sans avoir préalablement effectué l'analyse factorielle du tableau disjonctif associé aux descripteurs originalement saisis. Au cours de ce premier essai, les variables longueur et largeur étaient recodées respectivement en dix et huit modalités. La quasi-totalité de l'information apportée par les deux premiers facteurs tient alors en un classement des zones le long d'une parabole définie par les dix-huit modalités de ces deux variables. Ce phénomène, bien connu en analyse des données sous le nom d'effet Gutman, se produit souvent lorsqu'un grand nombre de modalités sont "naturellement" ordonnées (on le retrouve d'ailleurs, à un degré moindre, dans les résultats finaux exposés plus haut, avec les variables SURF, RAPT, SMIN et SMAX). Trop présent dans cette première analyse, il empêche entre autre l'émergence claire d'un classement des zones suivant la densité du bâti ; une fois calculées les variables SURF et RAPT et réduit le nombre de modalités qui leur sont affectées, la variable 'densité du bâti' prend en revanche une contribution importante au second facteur.

C'est un phénomène de même nature, mais incontournable étant donnée la nature actuelle des données, qui explique que l'on ait dû multiplier par trois l'importance relative de la variable "hauteur du bâti" pour séparer les bâtiments industriels des grands ensembles.

A l'issue de l'analyse, on peut donc souligner que le recodage disjonctif, suivant le nombre de modalités attribuées à tel ou tel descripteur, perturbe l'équilibre informationnel entre variables. Par la suite, le rétablissement de cet équilibre ou plus exactement l'établissement d'un équilibre qui satisfasse les objectifs thématiques fixés à la typologie, nécessite un ajustement par pondération qui ne peut s'effectuer que par tatonnements successifs, ce qui le rend assez laborieux.

La proposition d'amélioration de la méthode faite ci-dessus, alliée à une meilleure précision de certains descripteurs, devrait permettre de réduire de beaucoup ces tatonnements : on partirait d'un tableau où la qualité de l'information apportée serait

moins variable d'un descripteur à l'autre et on supprimerait l'arbitraire lié au nombre de modalités disjonctives créées lors du recodage. L'établissement de la typologie devenant ainsi plus "automatique", la méthode serait plus facilement transposable à d'autres sites.

Conclusions

Au terme de cette expérience, il importe de résumer les problèmes rencontrés, les solutions qui leur ont été apportées et de formuler quelques propositions qui nous semblent pouvoir améliorer la méthode. Dans cette optique, on peut souligner les points suivants :

1) L'analyse statistique de chacun des descripteurs morphologiques permet d'affirmer que, pour l'objectif poursuivi, un niveau quasi-optimal de précision de l'information extraite des photographies aériennes serait atteint si l'on codait les descripteurs "densité du bâti", "hauteur du bâti", "valeur de gris du bâti" et "valeur de gris de l'espace intersticiel" selon des **échelles de notes**, après avoir mesuré les valeurs des caractères sur des zones étalons de chaque modalité. Pour les autres caractères, quantitatifs, il semble bien que le passage du métrique à l'échelle de notes par agglomération des valeurs aux moyennes des principaux modes, ne dégrade pas trop l'information qu'ils apportent.

En adoptant ces améliorations, on homogénéiserait le tableau de description morphologique à un niveau d'information optimal, compte tenu de la nature du phénomène étudié, des objectifs poursuivis et de l'échelle d'observation.

2) Le problème majeur rencontré au cours de l'analyse est celui de **l'équilibre informationnel entre les divers descripteurs morphologiques**. Lorsqu'un descripteur, du fait du codage original ou de son recodage qualitatif, présente un grand nombre de modalités "naturellement ordonnées" (par exemple les descripteurs LONG, LARG, SURF, RAPT, SMIN et SMAX), l'effet Gutman sur ces variables empêche l'émergence claire d'un classement des zones suivant d'autres descripteurs thématiquement importants mais moins précisément observés comme l'élévation moyenne ou la densité du bâti.

Le palliatif retenu pour rétablir un équilibre qui satisfasse nos objectifs thématiques a été un ajustement par pondération de certains descripteurs, lequel peut choquer les théoriciens de l'analyse des données.

L'amélioration du codage des descripteurs devrait également permettre, en appliquant la même chaîne de traitement à l'ensemble du tableau (il s'agirait alors d'un tableau de notes dédoublées), d'obtenir la typologie sans recourir aux étapes d'analyses des sous-fichiers pondérés. Outre qu'elle serait plus élégante, la méthode, devenant aussi plus "automatique", serait certainement plus facilement transposable à d'autres sites.

Si la démarche présentée peut paraître, d'une part hétérodoxe du point de vue de la taxinomie mathématique, et d'autre part inutilement lourde aux yeux du thématicien urbaniste ou géographe, il nous semble donc qu'elle peut être assez facilement améliorée des deux points de vue grâce aux propositions faites ci-dessus. Ceci constitue à nos yeux l'acquis méthodologique principal de l'expérience.

D'autre part, quant aux objectifs même du programme de recherche, cette méthode d'analyse typologique permet d'obtenir les résultats suivants :

1) La mise au point d'une **stratification** de l'ensemble des districts de recensement de l'INSEE qui, comme on le verra par la suite, s'avère **efficace, tant du point de vue de la représentativité morphologique de l'échantillon de travail que**

travail que pour l'estimation par sondage de l'effectif de population et d'autres indicateurs démographiques.

2) Le cadre méthodologique de l'analyse typologique permet d'affirmer que **le niveau de finesse de la partition obtenue reflète à peu près exactement le niveau de précision de l'information qui a servi à son élaboration.** En d'autres termes, on ne risque pas d'avoir opéré des regroupements ou des séparations visuellement justifiables mais que l'on ne pourrait pas caractériser à partir du tableau de données. On sait donc précisément sur quels critères morphologiques repose la stratification que nous allons tester dans les chapitres suivants.

3) Cette méthode, qui permet une quantification facile et précise de la contribution de chaque caractère à la constitution de la typologie, fournit à l'équipe **un moyen de comparaison des gains de précision qu'apporte cette classification synthétique, et de ceux qu'on peut attendre d'une autre démarche de classification, fondée sur un ou plusieurs des caractères morphologique.** C'est cette quantification, alliée à la prise en compte du degré de facilité d'extraction de l'information morphologique de l'image satellite, qui déterminera le choix, parmi l'ensemble des descripteurs morphologiques, de ceux sur lesquels peuvent se fonder des stratifications praticables et efficaces. La mesure complète de cette efficacité ne peut cependant être obtenue qu'après le test de chaque stratification sur les données issues du recensement ; c'est à ce test que nous allons nous intéresser maintenant.

Chapitre 6

MISE AU POINT DE LA METHODE D'EVALUATION DES PLANS DE SONDAGE PREMIERE EXPERIENCE, FORMULAIRE GENERAL ET PROGRAMME INFORMATIQUE

Introduction

Le premier résultat concret de l'expérience : la stratification morphologique synthétique, doit maintenant démontrer son efficacité à deux niveaux :

- pour garantir la représentativité de l'échantillon de travail qui doit servir à la mise au point des méthodes en télédétection urbaine ;
- pour améliorer la précision des estimations par sondage des variables démographiques.

Le premier point n'étant pas central étant donné l'objectif de ce travail, nous ne l'abordons pas ici¹.

Quant au second point, il n'est en fait que le point de départ de l'étude que nous allons présenter maintenant. En effet l'évaluation de l'efficacité de la stratification morphologique synthétique n'est évidemment pas un but en soi ; d'une part, ce n'est pas elle, ou une stratification qui lui serait semblable, que nous envisageons d'utiliser dans la phase d'application, d'autre part, et cela quelle que soit la stratification testée, l'amélioration permise par la stratification est indissociable de celle qu'apportent les autres caractéristiques du plan de sondage : technique de sélection, probabilité d'inclusion des unités dans l'échantillon, nombre de degrés du sondage, estimateurs employés, etc.

Ce qu'il faut maintenant mettre au point, c'est une méthode permettant, à partir des données démographiques exhaustives que fournit le R.G.P., de mesurer l'impact qu'auront sur la précision des estimations l'ensemble des choix techniques liés à la définition d'un plan de sondage stratifié. On peut dès maintenant dresser la liste de ces choix, du moins en ce qui concerne le premier degré du sondage :

1. Choix du ou des critères de stratification de la base de sondage
2. Choix de la technique de sélection des unités primaires de l'échantillon : sélections systématiques ou aléatoires
3. Choix de la probabilité d'inclusion des unités dans l'échantillon : plan équiprobables ou à probabilités inégales
4. Choix de l'estimateur à employer : estimation sans biais ou estimation par le ratio etc...

¹ : Le lecteur intéressé par cet aspect peut consulter l'annexe 10 du rapport final rédigé avec F. DUREAU pour l'ATP "Télédétection" du CNRS : " Tirage de l'échantillon de travail sur Marseille", PP 159 à 166, rapport disponible au CNRS ou aux archives du département SDU à l'ORSTOM.

Cette liste n'est évidemment pas limitative et surtout ses termes ne sont pas indépendants les uns des autres ; par exemple le choix d'une technique de sélection systématique des unités primaires, à partir d'une grille de points, a des répercussions sur la probabilité d'inclusion de ces unités dans l'échantillon et sur la formule d'estimation à appliquer, et toutes les combinaisons des différentes alternatives ne seront pas possibles. De plus un choix donné d'estimateur, par exemple, n'a pas nécessairement les mêmes conséquences sur la précision des estimations suivant la stratification ou la probabilité d'appartenance retenue, etc. On le voit, la solution au problème du choix d'un plan de sondage stratifié adapté et efficace, passe par la mesure de l'influence d'un grand nombre de paramètres, et donc, conformément au principe empirique que nous nous sommes fixé, par la répétition des calculs de variance dans chacun des cas. Seul un programme informatique effectuant ces calculs et permettant de modifier à la demande l'ensemble de ces paramètres autorisera une telle démarche. Ce chapitre traite de la mise au point de cet outil.

Le travail qui a conduit à ce programme d'évaluation de techniques de sondage s'est déroulé en trois étapes qui correspondent aux trois parties de l'exposé.

La première partie rend compte de l'expérience effectuée sur le cas simple du sondage équiprobable sans remise correspondant à l'échantillon de travail. Elle conduit à une première évaluation de la stratification morphologique synthétique et des possibilités offertes par ce principe de stratification et surtout, elle permet la mise au point de la séquence du test. La seconde partie étend le formulaire à l'ensemble des cas envisagés : plan de sondage équiprobable ou plan de sondage à probabilités inégales ; dans chacun des cas on applique les méthodes d'estimation sans biais ou d'estimation par le ratio. La dernière partie présente le programme informatique mis au point (on expose l'architecture générale du programme et son fonctionnement), et son environnement matériel et logiciel : micro-ordinateur et système gestionnaire de base de données associé. Enfin, en conclusion, nous revenons sur le domaine d'application de l'outil statistique et informatique mis au point au cours de cette phase de l'expérience.

6.1 : Mise au point de la séquence du test sur un cas simple : plan équiprobable sans remise, estimation sans biais d'effectif et de densité de population, stratification morphologique synthétique

Comme on l'a déjà dit, l'échantillon de travail que nous sélectionnons à Marseille à partir de la stratification synthétique n'a pas pour but de produire des estimations démographiques. Son mode de tirage (systématique sur liste) ne correspond en rien à la méthode d'échantillonnage que nous cherchons à développer. Pourtant il sera le point de départ de l'évaluation des techniques de sondage et des stratifications dont il est maintenant question. D'une part, le sondage équiprobable et la stratification morphologique synthétique qui fournissent cet échantillon vont servir de référence pour l'évaluation des autres techniques ; il fallait donc étudier en premier lieu cette technique de base. D'autre part, à l'occasion de ce tirage, donc dans le cas simple du sondage équiprobable sans remise, on a posé les bases de la méthode d'évaluation d'un sondage à partir de données exhaustives et, à l'issue de ce travail, mené une première expérience à partir des données du recensement à Marseille. Au terme de cette expérience, la séquence du test sera établie et restera la même pour toute la suite de l'évaluation.

C'est pourquoi, avant d'élargir le champ de l'étude en modifiant les plans de sondages, les estimateurs et les caractères démographiques et en considérant d'autres stratifications, nous rendons compte, dans cette première partie, de cette expérience : il sera question successivement du plan de sondage, des taux, puis nous donnerons le formulaire simple correspondant à l'estimation sans biais dans le cas du plan

équiprobable, enfin nous examinerons les premiers résultats obtenus à partir des chiffres du recensement général de la population à Marseille en 1982.

A propos de ces données démographiques, des définitions utilisées par l'INSEE et de leur variation au sein de l'agglomération marseillaise selon les types morphologiques synthétiques retenus comme strates, on se reportera à l'annexe A.2.2 qui contient également une étude statistique préliminaire de trois indicateurs attachés aux districts de recensement : surface, population sans double compte et densité de cette même population. On y étudie rapidement les distributions de ces trois variables sur l'ensemble de la zone d'étude et dans chacune des strates morphologiques, puis on tente de prévoir le gain apporté par la stratification en se basant sur les variances inter et intra-strate des caractères. Nous avons préféré, pour ne pas alourdir l'exposé de la partie centrale, présenter ces résultats en annexe, bien qu'ils aient été connus avant ceux de l'expérience dont nous allons rendre compte ici et qu'ils soient totalement concordants avec eux.

6.1.1 : Base et plan de sondage, caractéristiques des échantillons

La base de sondage est la réunion des neuf strates morphologiques. Au cours de l'exposé, on considère des échantillons obtenus par tirage systématique dans la liste des districts INSEE. Ce mode de tirage sera considéré comme une approximation d'un tirage équiprobable sans remise¹.

Dans le plan de sondage stratifié utilisé pour le tirage de l'échantillon de travail², la même procédure a été appliquée pour chaque strate : la base de la suite arithmétique de raison 44, correspondant à un taux approximatif de 2,31%, est choisie aléatoirement et indépendamment dans chaque strate.

Soulignons enfin que, lors de ce tirage, la liste était triée sur l'identifiant alphanumérique attribué par l'INSEE à chaque district. Comme le remarque F. DUREAU dans une note sur le tirage de cet échantillon³, ce tri assure une bonne répartition géographique des districts tirés au sein de l'aire urbaine étudiée. En effet, l'identifiant INSEE regroupe les districts, quartier par quartier, puis arrondissement par arrondissement. Il est évident, pour les urbanistes et démographes de l'équipe, que ce classement n'est "neutre", ni du point de vue morphologique, ni du point de vue démographique ; des deux points de vue, les districts qui sont de rangs voisins dans la liste, auront tendance à se ressembler, autrement dit, ce classement apporte certainement une information sur les caractères que nous cherchons à estimer. Dans ce cas, on sait que l'assimilation du tirage systématique à un tirage aléatoire équiprobable sans remise, conduit à une sur-estimation des variances des estimateurs qui peut être importante (voir DESABIE : théorie et pratique des sondages, p 99, qui prend d'ailleurs le même exemple, ou GOURIEROUX, p 115). L'évaluation de cette sur-estimation est possible par assimilation du tirage systématique, dans l'ensemble de la base ou à l'intérieur de chaque strate, à un tirage stratifié de paires (voir DESABIE p 171 ou GOURIEROUX p 117). Mais, comme nous le verrons plus en détail par la suite, une procédure de tirage sur liste ne peut pas être à la base de la méthode applicable aux cas concrets de sondage aréolaire effectués sur images satellite. Il ne nous semble

¹ : Pour une justification de cette approximation, voir par exemple, J. DESABIE : Théorie et pratique des sondages (DUNOD 1966), ou C. GOURIEROUX : Théorie des sondages (ECOMICA 1981, P. 115).

² : A propos de l'échantillon de travail, son utilité dans le cadre du programme et la technique de tirage utilisée, voir l'annexe 10 du rapport ATP CNRS/CNES d'Août 1987.

³ : Voir F. DUREAU 1986 - Marseille : tirage de l'échantillon de travail. Quito, 9 p.

donc pas utile de mener ici plus avant l'évaluation du gain de précision qu'apporte le tri sur l'identifiant INSEE. Nous nous contenterons de deux observations :

1. l'information due à ce tri est certainement en partie redondante avec celle qu'apporte la stratification morphologique. On peut donc raisonnablement supposer que si les variances des estimations non stratifiées sont largement sur-estimées, au contraire, les variances des estimations stratifiées sont, elles, à peu de chose près réalistes. Il se pourrait toutefois que certaines strates géographiquement dispersées et démographiquement hétérogènes, profitent également du tri.

2. Dans le cas concret d'application, un problème semblable se posera pour évaluer l'apport du tirage aréolaire sur grille de points, qui, lui aussi, assure une bonne répartition géographique de l'échantillon. Il sera donc de nouveau question de ce problème par la suite (voir chapitre 8).

Remarque : Le plan de sondage P, défini ci-dessus, "respecte" les surfaces en espérance, c'est-à-dire au sens suivant :

$$E_P\left(\sum_{i=1}^n s_i\right) = t.S$$

où E_P désigne l'espérance mathématique calculée sous la loi de probabilité définie par P, s_i les surfaces des districts de l'échantillon, t le taux de sondage, n le nombre de districts tirés et S la surface totale de la base de sondage.

En effet,
$$E_P\left(\sum_{i=1}^n s_i\right) = E_P\left(\sum_{i=1}^n s_i \cdot 1(i \in e)\right)$$

$$= \sum_{i=1}^n s_i \cdot E_P(1(i \in e)) = \sum_{i=1}^n s_i \cdot z_i$$

Où N est le cardinal de la base de sondage, $1(i \in e)$ désigne la fonction indicatrice de l'appartenance du district i à l'échantillon (noté ici e) et $z_i = E_P(1(i \in e))$, la probabilité d'inclusion du district i dans l'échantillon (toujours sous la loi du plan P).

Le plan étant équiprobable sans remise, on a $z_i = n/N = t$.

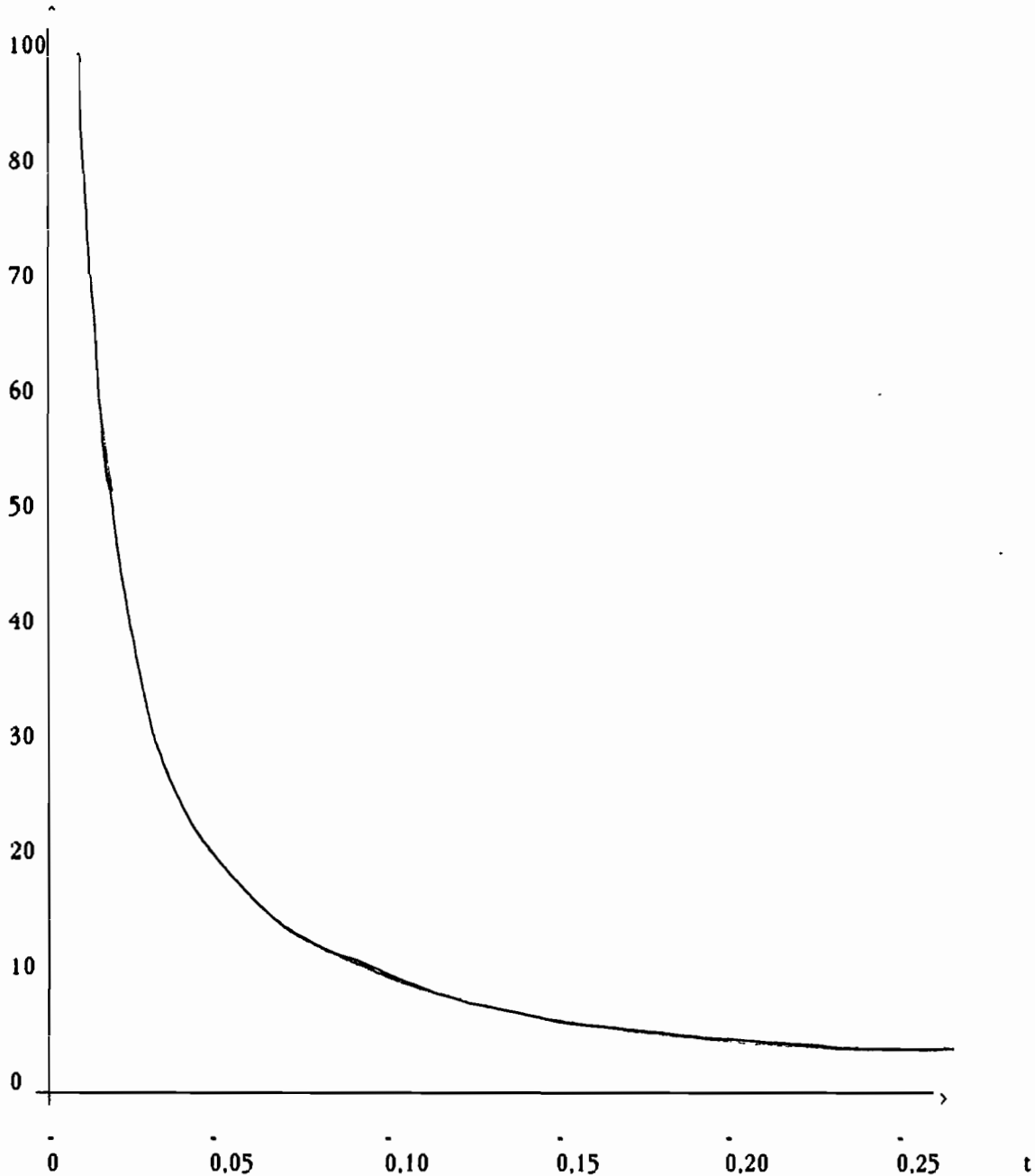
D'où
$$E_P\left(\sum_{i=1}^n s_i\right) = \sum_{i=1}^n s_i \cdot t = t \cdot \sum_{i=1}^n s_i = t.S$$

6.1.2. : Taux de sondage

Comme on le vérifiera tout au long des formulaires qui suivent, pour l'ensemble des estimateurs étudiés, la variance comprend toujours, dans son expression, un facteur $(N-n)/n$ (ou $(N_h - n_h)/n_h$ dans les expressions stratifiées), qui représente l'influence du taux de sondage sur les variances d'estimation. Il est donc très simple d'étudier les gains de précision dus aux augmentations du taux pour chacun des estimateurs que nous proposons dans ce chapitre, la multiplication par le facteur $t_1 \cdot (1-t_2)/t_2 \cdot (1-t_1)$ permettant de passer de la variance d'une estimation au taux $t_1 = n_1/N$ à la variance au taux $t_2 = n_2/N$.

L'hyperbole de la figure 6.1 donne la valeur du facteur $(N-n)/n$ en fonction du taux n/N (valable pour toute valeur de N) ; elle permet d'évaluer comment la variance des estimations va décroître dans la région de variation du taux qui nous intéresse. Sur cette même figure, le tableau donne la variance pour les taux de 4, 5, 6% en fonction de la variance V obtenue au taux de 2,31% (80/3461) qui correspond à l'échantillon de travail.

Fig. 6.1 : $f(n/N)=(N-n)/n$ c'est-à-dire $g(t)= (1/t) - 1$



Variance à 4,5 et 6% en fonction de la variance à 2.31%.

taux	2.31%	4%	5%	6%
variances	V	0,5675.V	0,4493.V	0,3705.V

6.1.3. : Formulaire du plan équiprobable sans remise

Les résultats suivants, classiques en théorie des sondages (voir par exemple DESABIE ou GOURIEROUX) sont rappelés sans démonstration. Les notations sont les suivantes : la variable estimée est toujours notée y , les valeurs qu'elle prend sur les unités statistiques y_i , son total Y , la variable définissant la probabilité z et ses valeurs z_i .

6.1.3.1. : Sans stratification

Le résultat général utilisé ici est le suivant : y étant le caractère à estimer et les probabilités d'inclusion z_i étant fixées, les seuls estimateurs sans biais du total et de la moyenne de Y , de la forme : $\sum_{i \in e} a_i y_i$ sont ceux dit d'Horwitz -Thompson et les coefficients a_i valent $1/z_i$.

Par exemple, l'estimateur du total de y , est donné par :

$\hat{T}(Y) = N \sum_{i \in e} y_i/z_i = N \sum_{i=1}^n (y_i/z_i) (1(i \in e))$, où e est l'échantillon, N le cardinal de la base de sondage et z_i , la probabilité d'appartenance du district i à l'échantillon. Le plan étant pour l'instant équiprobable, on a $z_i = n/N$ pour tout i et par conséquent, les estimateurs et leurs variances ont les expressions suivantes¹:

- estimation du total

$$\hat{T}(Y) = N \sum_{i=1}^n y_i/n = N \cdot \bar{y}$$

$$V(\hat{T}(Y)) = \frac{N(N-n)}{n} S^2$$

- estimation de la moyenne

$$\hat{M}(Y) = \sum_{i=1}^n y_i/n = \bar{y}$$

$$V(\hat{M}(Y)) = \frac{N-n}{Nn} S^2$$

où \bar{y} désigne la moyenne empirique du caractère sur l'échantillon, $V()$ la variance sous la loi du plan P , S^2 la variance à $N-1$ degrés de liberté sur l'ensemble de la base :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 = \frac{N-1}{N} \sigma^2,$$

σ^2 étant la variance classique sur l'ensemble de la base.

¹ : Voir, par exemple, C. GOURIEROUX, Théorie des sondages, chap. III, P. 77 à 81.

Application : estimation de l'effectif total de population et de la moyenne des densités.

Les quantités à estimer sont donc : $P = \sum_{i=1}^N p_i$, total de population sur l'ensemble de la

base et $\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i$, moyenne sur l'ensemble de la base des densités de districts.

- Le total des effectifs P est estimé par :

$$\hat{T}(P) = N \sum_{i=1}^n p_i/n = N\bar{p}, \text{ où } p_i \text{ désigne l'effectif de population sans double}$$

compte du district courant i et \bar{p} la moyenne empirique de cet effectif sur l'échantillon.

$N\bar{p}$ estimera donc sans biais l'effectif total et ceci avec la variance :

$$V(\hat{T}(P)) = \frac{N^2 (N-n)}{(N-1) n} \sigma_p^2$$

où σ_p^2 est la variance de l'effectif de population sans double compte sur l'ensemble de la base :

$$\sigma_p^2 = \frac{1}{N} \sum_{i=1}^N (p_i - \bar{p})^2, \text{ où } \bar{p} = \frac{1}{N} \sum_{i=1}^N p_i$$

Le plan étant équiprobable sans remise et l'estimateur celui d'Horwitz-Thompson, la loi de l'estimateur est asymptotiquement normale (théorème central limite) ce qui permet de construire l'intervalle de confiance à 95% de l'estimation du total :

$$I_p = \left[T(P) - 2N \left(\frac{N-n}{n(N-1)} \sigma_p^2 \right)^{1/2}, T(P) + 2N \left(\frac{N-n}{n(N-1)} \sigma_p^2 \right)^{1/2} \right]$$

- La moyenne de densités des districts est estimée par :

$$\hat{M}(\bar{d}) = \frac{1}{n} \sum_{i=1}^n d_i$$

La variance sera :

$$V(\hat{M}(\bar{d})) = \frac{N-n}{(N-1) n} \sigma_d^2, \text{ où } \sigma_d^2 = \frac{1}{N} \sum_{i=1}^N (d_i - \bar{d})^2$$

D'où l'intervalle de confiance à 95% :

$$I_d = \left[\bar{d} - 2 \left(\frac{N-n}{n(N-1)} \sigma_d^2 \right)^{1/2}, \bar{d} + 2 \left(\frac{N-n}{n(N-1)} \sigma_d^2 \right)^{1/2} \right]$$

6.1.3.2 : Avec stratification

- **Effectifs**

L'estimateur du total s'écrit :

$$\hat{T}(P) = \sum_{h=1}^K \frac{N_h}{n_h} \left(\sum_{i=1}^{n_h} p_{hi} \right), \text{ où } K \text{ est le nombre de strates, } N_h \text{ le cardinal de la}$$

strate h et n_h le nombre de districts tirés dans la strate h.

Cet estimateur a pour variance :

$$V(\hat{T}(P)) = \sum_{h=1}^K \frac{N_h^2 (N_h - n_h)}{(N_h - 1) n_h} \sigma_{ph}^2, \text{ où } \sigma_{ph}^2 = \frac{1}{N_h} \sum_{i=1}^{n_h} (p_{hi} - \bar{p}_h)^2, \bar{p}_h \text{ étant la}$$

moyenne du caractère "population sans doubles comptes" dans la strate h, σ_{ph}^2 est donc sa variance sur la strate h.

L'intervalle de confiance à 95% s'écrit :

$$I_p = \left[T(P) - 2(V(\hat{T}(P)))^{1/2}, T(P) + 2(V(\hat{T}(P)))^{1/2} \right]$$

- **Moyenne des densités**

Estimateur

$$\hat{M}(\bar{d}) = \sum_{h=1}^K \frac{N_h}{N n_h} \left(\sum_{i=1}^{n_h} d_{hi} \right)$$

Variance de l'estimateur

$$V(\hat{M}(\bar{d})) = \frac{1}{N^2} \sum_{h=1}^K \frac{N_h^2 (N_h - n_h)}{(N_h - 1) n_h} \sigma_{dh}^2$$

Intervalle de confiance à 95%

$$I_d = \left[\bar{d} - 2(V(\hat{M}(\bar{d})))^{1/2}, \bar{d} + 2(V(\hat{M}(\bar{d})))^{1/2} \right]$$

6.1.3.3. : Allocation optimale de NEYMAN

L'allocation proportionnelle retenue lors du tirage de l'échantillon de travail (taux uniforme dans toutes les strates) a été choisie pour que, dans l'échantillon, tous les types morphologiques figurent avec une surface approximativement proportionnelle à celle qu'ils occupent dans l'univers¹. Cet objectif n'est pas compatible avec l'optimisation de l'estimation des caractères démographiques.

En effet, pour une taille donnée de l'échantillon, la répartition optimale s'obtient en appliquant la méthode des multiplicateurs de LAGRANGE au problème de la minimisation de $V(\hat{T}_{H,T}(y))$, où y est le caractère à estimer, sous la contrainte :

$$\sum_{h=1}^K n_h = n$$

On aboutit facilement à la condition : $\frac{n_h}{N_h S_h} = a$, où a est une constante.

C'est-à-dire que les taux $\frac{n_h}{N_h}$ dans chaque strate, doivent être proportionnels à l'écart-type S_h .

On en déduit que le nombre de districts sondés doit être , dans la strate h :

$$n_h = \frac{n N_h S_h}{\sum_{h=1}^K N_h S_h}$$

6.1.4. : Application et résultats, commentaires

Sur la base de sondage, les quantités que nous voulons estimer sont :

- **La population totale sans doubles comptes** : $T(P) = 728\ 872$
- **La moyenne de la densité sur les districts** : $d = 2,4256$ h/are (soit 24256 habitants au km^2)

Ces deux caractères ont pour variance sur l'ensemble de la base :

- **variance de la population** : $\sigma_p^2 = 77250$

écart-type : $\sigma_p = 277,94$

coefficient de variation : $131,72 \%$

¹ : Ce coix correspond aux besoins de l'exploitation faite de cet échantillon en télédétection

² : Voir J. DESABIE, Théorie et pratique des sondages, P. 146 ou C. GOURIEROUX, Théorie des sondages, P. 102.

- **variance de la densité** : $\sigma_d^2 = 30,8989$

écart-type : $\sigma_d = 5,5587$

coefficient de variation : 229,0 %

Pour l'ensemble de la base de sondage et pour chacune des strates, les résultats du test mené sont consignés dans les tableaux 6.2 (estimation du total de population sans doubles comptes) et 6.3 (estimation de la moyenne du caractère "densité de la population s.d.c."). Le commentaire qui suit s'appuie sur ces tableaux que l'on peut consulter à la fin du paragraphe 6.1.

Résultats sur l'ensemble de la base

Les résultats sur l'ensemble de la base appellent les commentaires suivants.

1. Pour l'ensemble des plans testés, stratifiés ou non, l'estimateur du total de population est toujours nettement plus précis que celui de la moyenne des densités. Ce fait est en accord avec la différence importante de variabilité des deux caractères sur l'ensemble de la base : la population sans doubles comptes des districts a un coefficient de variation de 132 % alors que pour la densité, il est de 229 %. Nous reviendrons sur cette question lors de l'examen des résultats par strate, mais on peut d'ores et déjà affirmer que l'impératif qu'a l'INSEE, de constituer des districts dont la population ne dépasse pas un certain seuil, a tendance à diminuer artificiellement la variance de l'effectif de population des districts, surtout dans les endroits très densément peuplés. On ne peut donc probablement pas conclure de ce résultat que l'estimateur du total de population soit toujours plus précis que l'estimateur de la densité moyenne par district.

2. L'apport de la stratification est intéressant pour les deux caractères étudiés : au taux de 5 %, il permet de faire diminuer la variance de l'estimateur du total de population de 16,6 % en adoptant une allocation proportionnelle, et de 31,8 % en optimisant l'allocation. Quant à l'estimation de la densité moyenne, elle profite beaucoup moins de l'allocation proportionnelle (-5,6 %) mais beaucoup plus de l'allocation optimale (-38,9 %). Au total, les gains sont donc comparables, dans leur ordre de grandeur, pour les deux caractères, mais différemment répartis entre allocation proportionnelle et optimale. Reprenons plus en détail l'examen des gains dus à la stratification et à l'augmentation du taux pour chacun des caractères.

3. **Estimation du total de population** : Quelle que soit la fraction de sondage, la variance de l'estimation obtenue sans stratification est multipliée par un facteur 0,834 lors de l'introduction de la stratification avec l'allocation proportionnelle, par un facteur 0,682 si l'on choisit l'allocation de NEYMAN. Ceci permet de ramener le coefficient de variation de 10,0 % (sans stratification) à 8,3 % (allocation optimale) au taux de 5%. L'intervalle de confiance est ramené de +/- 146270 à +/- 120759, soit un gain de 17 %.

Les gains dus à la variation du taux de sondage sont, eux, beaucoup plus importants, surtout dans l'intervalle 2,31 % à 5 % (voir fig. 6.1). Par contre, au passage de 5 à 6 %, le gain est du même ordre que celui dû à la stratification.

4. Estimation de la moyenne des densités : Ici le gain qu'apporte la stratification est presque entièrement dû à l'introduction de l'allocation optimale : La variance est multipliée par 0,929 quand on emploie l'allocation proportionnelle, puis par 0,615 dans le cas de l'allocation optimale, soit 80 % du gain lors de son introduction. Au taux de 5 %, le coefficient de variation passe de 17,42 % sans stratification à 13,66 % avec l'allocation optimale.

En conclusion, l'estimation de la moyenne des densités profite donc un peu plus de la stratification que celle de l'effectif total.

5. Au taux où s'est réellement fait le tirage (80/3461), l'échantillon obtenu fournit des estimations beaucoup plus précises que ne laissent prévoir les intervalles de confiance à 95 %. L'estimation de la densité ne s'écarte que de 21,4 % de la vraie valeur tandis que le total de population est obtenu à 0,073 % près (estimations stratifiées). On a donc eu beaucoup de chance lors du tirage, en particulier du point de vue de l'estimation du total de population sans doubles comptes. Il est, bien sûr, exclu d'en tirer des conclusions quant à l'efficacité de la méthode, puisque seules les variances des estimateurs sont réellement significatives. Notons tout de même qu'une part de l'explication de ce résultat "inespéré" peut résider dans le fait que le calcul de la variance ne prend pas en compte le tri sur l'identifiant INSEE, mais nous verrons plus tard que le gain que l'on peut en attendre n'est pas de cet ordre de grandeur.

Résultats par strates

Les commentaires des résultats par strates tiennent en quatre points principaux .

1. Naturellement, la précision des estimations à l'intérieur des strates est partout très nettement inférieure à celle des estimations sur l'ensemble de la base. Le tableau ci-dessous compare, pour les deux caractères et dans le cas de l'allocation proportionnelle, la précision des estimations obtenues sur l'ensemble de la base à celle à l'intérieur des strates : le coefficient de variation à l'intérieur des strates est compris entre 1,1 fois (estimation de la moyenne de densité dans la strate pavillonnaire) et 9,7 fois (estimation du total de population dans la strate non bâti) le coefficient de variation sur l'ensemble de la base.

Coefficients de variation pour les estimations globales et les estimations par strate.

Estimateurs	Estimat- ions globales	Estimation par strates	
		Mini.	Maxi.
1. Du total de population sans doubles comptes	9,17 %	11,25 %	109,88 %
2. De la moyenne de densité de population	16,79 %	18,62 %	100,47 %

2. Ci dessus, nous remarquons que l'estimateur du total était, sur l'ensemble de la base, plus précis que l'estimateur de la moyenne de la densité. Précisons maintenant ce qu'il en est au niveau de chaque strate. Le tableau suivant donne, pour chaque strate, la précision (coefficient de variation) des estimations pour les deux caractères.

Coefficients de variation des deux caractères estimés à l'intérieur de chaque strate (au taux de 5%, estimation stratifiée à allocation proportionnelle)

Strates	estimation du total de population (%)	estimation de la moyenne de densité (%)
1 : Centre ville dense	11,25	26,29
2 : Quartiers industriels	47,50	51,12
3 : Grands ensembles d'habitation	17,43	36,48
4 : Petits immeubles d'habitation	54,34	72,55
5 : Pavillonnaire	21,33	18,62
6 : Péri-urbain peu dense	76,34	68,83
7 : Reliquat morphologique	36,79	32,74
8 : Dense : grands bâtiments élevés	45,53	32,54
9 : Non bâti	109,88	100,47

L'estimateur du total est beaucoup plus précis (environ deux fois plus) dans les strates 1 et 3, "centre ville dense" et "grands ensembles d'habitation" ; il est également plus précis dans la strate 4, "petits immeubles d'habitation" (1,34 fois).

Les deux estimateurs sont à peu près équivalents dans les strates 2 ("quartiers industriels"), 5 ("pavillonnaire"), 6 ("péri-urbain peu dense"), 7 ("reliquat morphologique") et 9 ("non bâti"). Pour ces quatre dernières, l'estimateur de la densité est même légèrement meilleur : 1,1 fois plus précis.

Enfin, dans la strate 8, l'estimateur de la moyenne des densités est 1,4 fois plus précis que celui du total de population. La précision globalement meilleure des estimations du total de population s'explique donc par les meilleurs résultats qu'elle fournit sur des sous-ensembles importants, en population et en surface, de l'agglomération marseillaise : centre ville, grands ensembles et, dans une moindre mesure, petits immeubles. Par ailleurs, la différence est peu importante, entre les deux estimateurs dans la troisième strate dominante, le pavillonnaire.

3. Examinons maintenant la précision des estimations obtenues pour chaque strate. Les figures 6.4 et 6.5 présentent, par strate et pour chacun des deux estimateurs (du total et de la moyenne), les intervalles de confiance en pourcentage de l'espérance de l'estimateur dans la strate. Les deux types d'allocations, proportionnelle et optimale, y sont considérés.

Figure 6.4 : Précision des estimations partielles du total de population sans doubles comptes au taux de 5 % .

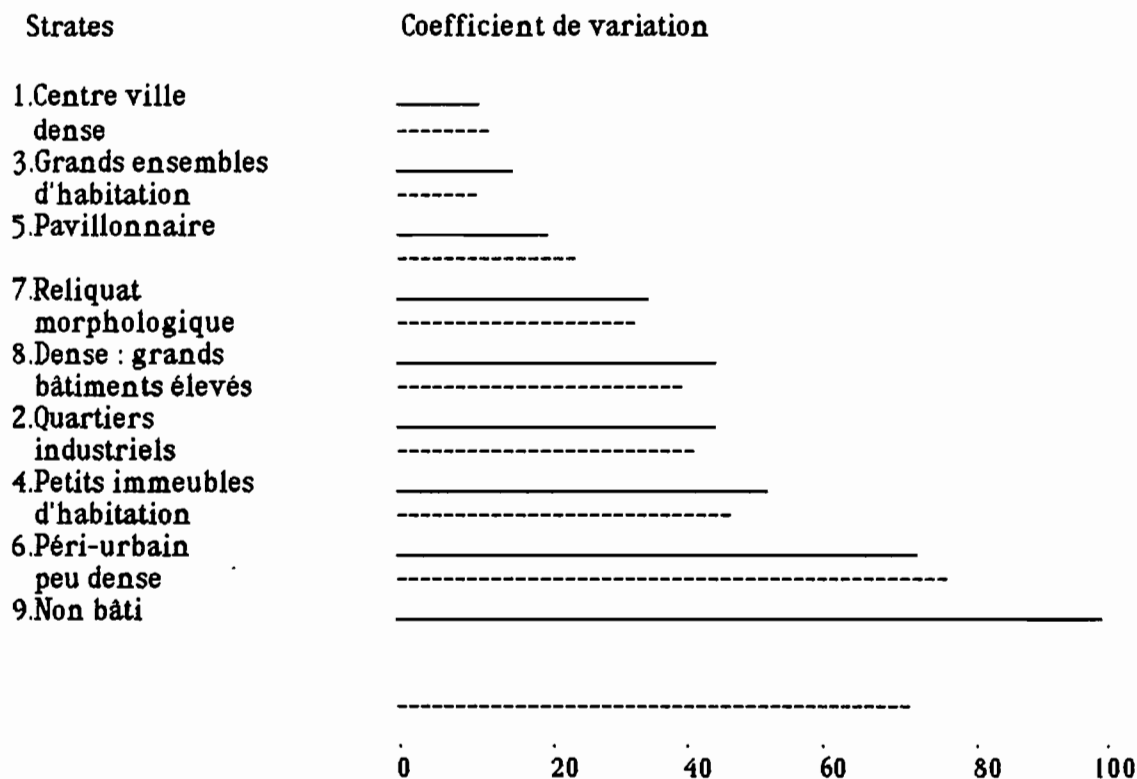
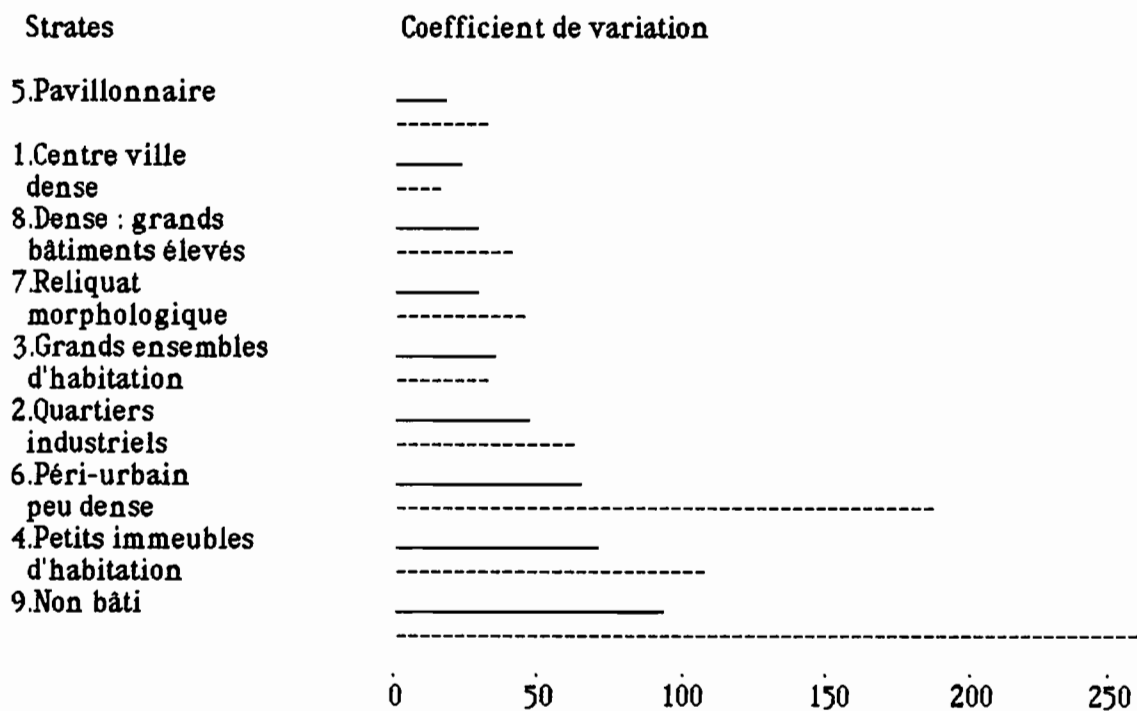


Figure 6.5 : précision des estimations partielles de la moyenne de la densité de population sans doubles comptes au taux de 5 %



figures 6.4 et 6.5 :

————— : allocation proportionnelle
 - - - - - : allocation optimale

Il se dégage clairement, de l'examen des deux figures, trois groupes de strates :

a/ Les strates dont la morphologie est homogène ("centre ville", "grands ensembles" et "pavillonnaire") bénéficient des meilleures précisions, tant pour l'estimation de l'effectif que pour celle de la densité : l'estimation du total de population sans doubles comptes se fait dans un intervalle de confiance compris entre +/- 20 % et +/- 50 % de la vraie valeur; pour l'estimation de la moyenne de densité des districts, cet intervalle varie entre +/- 40 % et +/- 70 %. Les meilleurs résultats sont obtenus, pour l'estimation du total, dans les strates 1 (+/- 21%) et 3 (+/- 34 %), et pour l'estimation de la moyenne de densité, dans les strates 5 (+/- 36,5%) et 1 (+/-51%).

b/ Les strates marginales (très minoritaires en surface), "reliquat morphologique" et "dense : grands bâtiments élevés". Pour ces deux strates, la précision des estimations est moyenne dans le cas du total et assez bonne pour la densité du moins si l'on s'en tient à l'allocation proportionnelle (intervalles de confiance de l'ordre de +/- 75 % pour l'estimation du total et +/- 65 % pour celle de la moyenne des densités).

c/ Les strates globalement peu denses et hétérogènes du point de vue morphologique et démographique : "quartiers industriels", "petits immeubles d'habitation", "péri-urbain peu dense" et "non bâti", pour lesquelles les estimations des caractères démographiques sont très imprécises : les intervalles de confiance se situent entre +/- 100 % (industriel) et +/- 220 % (non bâti) pour l'estimation du total, et entre +/- 100 % (industriel) et +/- 200 % (non bâti), pour l'estimation de la densité.

Ces résultats ne surprendront ni les urbanistes ni les démographes; ils justifient à nos yeux l'emploi des critères morphologiques pour la stratification et, a posteriori, la méthode d'analyse typologique employée pour sa mise au point. En effet, il est clair d'une part, que cette stratification a permis d'isoler des sous-ensembles urbains où les caractères démographiques ont des valeurs et des variances différentes, et d'autre part, que les zones les plus homogènes morphologiquement bénéficient de la meilleure précision d'estimation. Rappelons, à ce propos, que les groupes "centre ville", "grands ensembles" et "pavillonnaire" s'isolaient en premier lieu lors des classifications et analyses factorielles (voir chapitre précédent).

4. Enfin, il nous faut examiner dans quelle mesure l'optimisation des résultats globaux grâce à l'allocation de NEYMAN est compatible avec la précision des estimations partielles.

Les figures 6.4 et 6.5 sont également très éloquentes sur ce point :

- dans le cas de l'estimation du total de population, l'allocation optimale n'affecte qu'assez peu la précision des estimations du total de chaque strate. Elle l'améliore même sensiblement dans les strates 3 : "grands ensembles" et 9 : "non bâti".

- il en va tout autrement pour l'estimation de la moyenne des densités : la précision des estimations diminue de manière très sensible pour l'ensemble des strates, à la seule exception de la strate "centre ville dense". En particulier, les deux groupes de strates globalement peu denses (4, 5, 6 et 9) et morphologiquement marginales (7 et 8), pâtissent beaucoup de ce type d'allocation. Cela constitue évidemment un argument supplémentaire contre cet estimateur.

6.1.5. : Conclusions perspectives

Précision des estimations, ordre de grandeur des intervalles de confiance, effet de la répartition géographique de l'échantillon

En assimilant le tirage systématique pratiqué à un tirage équiprobable sans remise, une fois introduite la stratification et optimisée l'allocation aux strates, on obtient les précisions suivantes (mesurées par l'intervalle de confiance à 95 %) pour les estimations, sur l'ensemble de la base de sondage, du total de population sans doubles comptes et de la moyenne de densité des districts.

Total de population s. d. c. : +/- 16,6 % au taux de 5 %

Moyenne des densités : +/- 27,3 % au taux de 5 %

Comme nous le remarquons déjà ci dessus, une bonne répartition géographique des unités statistiques de l'échantillon au sein de l'agglomération étudiée ainsi qu'à l'intérieur de chaque strate, est assurée, dans le cas de l'expérience marseillaise, par le tri de la base de sondage sur l'identifiant INSEE. Cette répartition sera encore meilleure dans les cas concrets d'application, grâce au tirage d'unités aréolaires sur grille de points.

Cette répartition géographique apporte certainement un gain de précision notable à l'estimation des caractères démographiques. Ce gain n'est pas pris en compte dans les chiffres donnés ci-dessous : on peut donc les considérer comme des évaluations pessimistes des précisions réelles.

Le premier objectif qu'on peut fixer à la suite de ces tests est donc la quantification de ce gain et son introduction dans les calculs de précision des estimations (voir chapitre 8).

Apport de la stratification, estimations globale et estimations partielles, domaines géographiques de précision

Pour les estimations globales - on entend par là, celles qui s'étendent à l'ensemble de la base de sondage considérée - l'introduction de la stratification et de son corollaire, l'allocation optimale, permet un gain sur la variance des estimations compris, suivant les cas, entre 32 et 39 % de la variance non stratifiée.

D'autre part, en ce qui concerne les estimations partielles (estimation des caractères démographiques à l'intérieur de chaque strate), les résultats obtenus dessinent dans l'espace urbain marseillais certains domaines de bonne précision et d'autres d'imprécision (par "bonne précision" ou "imprécision", il faut entendre des chiffres comparables ou non, dans leur ordre de grandeur, à ceux obtenus au même taux sur l'ensemble de la base) :

Domaine de bonne précision : strates "centre ville", "grands ensembles d'habitation" et "pavillonnaire". Les estimations se font, au taux de 5%, avec des intervalles de confiance compris entre +/- 20 et +/- 50 %, pour le total de population s.d.c., entre +/- 40 et +/- 70 %, pour la moyenne de la densité. Rappelons que près de 80 % de la population totale de la zone d'étude est regroupée dans ces trois strates.

Domaine d'imprécision : strates "quartiers industriels", "petits immeubles d'habitation", péri-urbain peu dense et "non bâti"; ici les intervalles de confiance, compris entre +/- 100 % et +/- 170 %, ne permettent que des estimations très imprécises des caractères démographiques.

En résumé, *le principe d'une stratification de la base de sondage à l'aide de critères morphologiques ainsi que la méthode d'analyse typologique qui a servi à son élaboration sont à retenir pour deux raisons :*

- 1. parce qu'ils permettent une amélioration significative des estimations globales.*
- 2. parce qu'ils autorisent des estimations partielles avec une précision correcte dans les strates de caractéristiques morphologiques homogènes et, ce qui n'est pas moins important, permettent de quantifier l'imprécision qui affecte ces mêmes estimations dans les strates plus hétérogènes. Ce résultat représente, à notre avis, l'acquis majeur de cette première phase de test.*

Les paragraphes suivants s'attacheront à tester des méthodes de stratification plus proches de celle qu'on pense pouvoir appliquer à l'information satellitaire, en considérant comme acquise l'utilisation des caractères morphologiques pour la stratification de la base de sondage. En procédant analytiquement cette fois (par opposition au caractère synthétique des résultats d'analyse typologique), on s'attachera à évaluer séparément, l'apport des principaux caractères morphologiques que l'on pense pouvoir observer sur l'image (densité du bâti, etc.), en stratifiant la base de sondage d'après eux. Bien entendu, on continuera à mesurer le gain apporté par la stratification synthétique. On peut d'ailleurs la considérer comme une référence à laquelle on peut comparer ce qu'il sera possible de faire d'après l'image satellite (voir chapitre 7).

Comparaison entre les estimateurs de l'effectif et de la densité.

Tout au long du test mené, la comparaison entre l'estimateur du total de population sans doubles comptes et celui de la moyenne de la densité des districts s'est conclu en faveur du premier. Rappelons rapidement les avantages qu'il possède dans le cadre, qui fut le nôtre ici, du plan équiprobable .

1. Il est nettement plus précis que celui de la densité quand on s'intéresse à l'estimation globale; le gain sur l'amplitude de l'intervalle de confiance, dans le cas d'estimations stratifiées à allocation optimale, est de 40 %.

2. Il est aussi précis, et parfois plus, si l'on considère les estimations partielles dans le sous-ensemble où elles sont les meilleures : nettement plus précis dans le centre ville et les grands ensemble (gain approximatif de 55 % dans les deux cas), et à peu près équivalent à l'estimateur de la moyenne dans le pavillonnaire.

3. L'allocation optimale ne perturbe que faiblement les résultats par strate obtenus avec cet estimateur, ce qui permet une bonne précision tant de l'estimation globale que des estimations partielles. Dans le cas de l'estimation de la moyenne des densité, ces deux objectifs apparaissent difficilement conciliables.

Il nous reste à faire plusieurs remarques qui relativisent beaucoup l'intérêt de ces résultats.

Premièrement, ils sont probablement dus en partie à la réduction artificielle de la variance de l'effectif de population induite par le découpage des districts de recensement de l'INSEE. A l'examen des fonds de cartes utilisés lors du recensement, il apparaît que ce facteur a pu jouer dans les strates où le réseau de voirie étant assez lâche, un certain nombre de grands îlots ont été partagés en plusieurs districts de recensement, ayant chacun des effectifs de population "raisonnables". En revanche, ce ne peut pas être le cas dans le centre ville dense, où la taille des îlots ne justifie jamais un tel découpage.

Deuxièmement, s'il est vrai, au moins dans des sous-ensembles importants de l'agglomération, que l'effectif de population des districts est moins variable que leur densité, il ne faut pas perdre de vue que la mesure de cette variabilité qui permet de déterminer la précision de l'estimation est la variance du caractère, calculée sous la loi de probabilité correspondante au tirage. Ceci signifie que cette mesure varie suivant le plan de sondage et, avec elle, la variance de l'estimateur. Il n'est donc pas acquis que l'estimation du total de population soit toujours plus précise que celle de la densité pour un plan de sondage à probabilités inégales ; seule la suite du test nous renseignera à ce sujet.

Troisièmement, ce résultat n'est valable que tant que l'unité statistique enquêtée reste le pâté de maisons ou peut lui être comparé. Il faut également que dans la ville considérée, existent des sous-ensembles urbains, du type "centre ville dense" ou "grands ensembles d'habitation", semblables à ceux rencontrés à Marseille et qui soient prédominants en effectifs de population.

Enfin et surtout, l'intérêt de la comparaison pratiquée jusqu'à présent est très limité puisque ses deux termes n'estiment pas la même quantité : le premier fournit les totaux de population dans les strates ou l'ensemble de la base, alors que le second est la moyenne arithmétique des densités des districts, quantité qui n'a d'ailleurs, que peu d'intérêt pratique.

Conclusion

Cette étape aura donc essentiellement permis de mettre au point la séquence d'un test que nous allons étendre maintenant à des plans de sondage à probabilités inégales, à des estimateurs par le ratio faisant intervenir des variables exogènes, à d'autres caractères démographiques et enfin à d'autres stratifications. Les paragraphes 6.2 et 6.3 de ce chapitre présentent l'outil statistique et informatique mis au point pour cette généralisation du test.

Tableau 6.2 : Précision de l'estimation du total de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates	Total	Vari- -ance x 10 ⁻⁸	Coef. de variation	Intervalle de confiance à 95 % (%)	Evolution de la var. avec l'al- -location optimale (%)
1 * Centre ville **	207739	5.458	11.25	+/- 46724.79	0.00
2 * Quartiers Indus- -triels **	40680	3.734	47.50	+/- 38648.41	0.00
3 * Grands ensembles **	264764	21.292	17.43	+/- 92286.60	0.00
4 * Petits immeubles **	28750	2.441	54.34	+/- 31246.44	0.00
5 * Pavil- -lonnaire **	93937	4.016	21.33	+/- 40077.76	0.00
6 * Peri- urbain **	11648	0.791	76.34	+/- 17784.44	0.00
	11648	0.861	79.66	+/- 18557.84	8.85

* : Allocation proportionnelle

** : Allocation optimale

Tableau 6.2 (suite) : Précision de l'estimation du total de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates	Total	Vari- -ance x 10 ⁻⁸	Coef. de variation	Intervalle de confiance à 95 % (%)	Evolution de la var. avec l'al- -location optimale (%)
7 *	40534	2.224	36.79	+/- 29825.73	0.00
Reliquat morpho. **	40534	1.996	34.85	+/- 28253.18	-10.25
8 *	23788	1.173	45.53	+/- 21659.08	0.00
Dense, Grands batiments **	23788	0.930	40.54	+/- 19286.52	-20.72
9 *	17032	3.502	109.88	+/- 37429.50	0.00
Non bâti **	17032	1.450	70.70	+/- 24084.38	-58.60
Ensemble (non strati- -fié)	728872	53.488	10.03	+/-146270.83	0.00
Ensemble (stratifié, allocation prop.)	728872	44.631	9.17	+/-133612.26	-16.56
Ensemble (stratifié, allocation optimale)	728872	36.457	8.28	+/-120759.19	-31.84

* : Allocation proportionnelle

** : Allocation optimale

Tableau 6.3 : Précision de l'estimation de la moyenne de densité de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Moyenne	Vari- -ance $\times 10^{-8}$	Coef. de variation	Intervalle de confiance à 95 % (%)	Evolution de la var. avec l'al- -location optimale (%)
1	*	4.338	1.30	26.29	+/- 2.28	0.00
Centre ville	**	4.338	0.67	18.81	+/- 1.63	-48.46
2	*	1.291	0.44	51.12	+/- 1.32	0.00
Quartiers Indus- -triels	**	1.291	0.78	68.43	+/- 1.77	77.30
3	*	2.313	0.71	36.48	+/- 1.69	0.00
Grands ensembles	**	2.313	0.66	35.22	+/- 1.63	-7.00
4	*	1.050	0.58	72.55	+/- 1.52	0.00
Petits immeubles	**	1.050	1.22	105.38	+/- 2.21	110.30
5	*	0.944	0.03	18.62	+/- 0.35	0.00
Pavil- -lonnaire	**	0.944	0.11	35.29	+/- 0.67	266.60
6	*	0.416	0.08	68.83	+/- 0.57	0.00
Peri- urbain	**	0.416	0.58	183.78	+/- 1.53	625.00

* : Allocation proportionnelle

** : Allocation optimale

Tableau 6.3 (suite) : Précision de l'estimation de la moyenne de densité de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Moyenne	Vari- -ance x 10 ⁻⁸	Coef. de variation	Intervalle de confiance à 95 % (%)	Evolution de la var. avec l'al- -location optimale (%)
7	*	2.172	0.51	32.74	+/- 1.42	0.00
Reliquat Morpho.	**	2.172	1.07	47.58	+/- 2.07	109.80
8	*	4.695	2.33	32.54	+/- 3.06	0.00
Dense, Grands batiments	**	4.695	3.56	40.18	+/- 3.77	52.80
9	*	0.342	0.12	100.47	+/- 0.69	0.00
Non bâti	**	0.342	0.89	276.22	+/- 1.89	641.67
Ensemble (non strati- -fié)		2.426	0.18	17.42	+/- 0.85	0.00
Ensemble (stratifié, allocation prop.)		2.426	0.17	16.79	+/- 0.81	-5.56
Ensemble (stratifié, allocation optimale)		2.426	0.11	13.66	+/- 0.66	-38.89

* : Allocation proportionnelle

** : Allocation optimale

6.2. : Généralisation du test : plans équiprobables et plans à probabilités inégales, estimateurs sans biais et estimateurs par le ratio

6.2.1 : Position du problème

On a déjà souligné précédemment que le tirage d'échantillons d'après des documents satellitaires se fera de manière systématique, à l'aide d'une grille de points placée sur la base de sondage spatiale que constitue l'image. Cette contrainte définit un plan de sondage où la probabilité de sélection d'une unité statistique dans l'échantillon est proportionnelle à une mesure de la taille de cette unité : surface totale de l'unité si la procédure sélectionne tous les points de la grille, surface bâtie si elle ne retient que les points situés au-dessus d'une surface bâtie. Outre qu'il s'agit pour nous d'un impératif pratique¹, *de tels plans de sondage modifient la précision des estimations ; mais cette précision dépend également des estimateurs et des stratifications employées.*

COCHRAN², dans "Sampling technics" et DESABIE³ dans "Théorie et pratique des sondages", recensent les différentes possibilités d'utiliser une information exogène relative à la taille des unités statistiques ; COCHRAN écrit, par exemple : "Quand les unités sont de taille inégale, il y a le choix entre aux moins quatre techniques (en supposant connues la taille de chaque unité si ces techniques le nécessitent).", et dresse la liste suivante :

1. Plan équiprobable, estimation linéaire sans biais (bien que figurant dans la liste de COCHRAN, cette option n'utilise pas d'information exogène ; elle sert en fait de référence).
2. Plan équiprobable, estimation par le ratio relatif à la taille des unités.
3. Plan à probabilités proportionnelles à la taille des unités, estimation sans biais.
4. Plan équiprobable, stratification de la base suivant la taille des unités, estimation stratifiée usuelle (sans biais).

En fait, seule la dernière de ces techniques nécessite, avant le tirage, la connaissance sur l'ensemble de la base de la taille de toutes les unités. Elle est donc pratiquement difficile à mettre en oeuvre ; de plus notre recherche repose sur l'hypothèse qu'une stratification sur des critères morphologiques lui est préférable. Il ne sera donc plus question de cette technique dans la suite de l'exposé.

Dans la suite, on tente de mesurer l'effet sur la précision des estimations, des différentes possibilités d'intégration d'informations exogènes dans la technique de sondage : plans à probabilités inégales, estimation par le ratio et stratification. Pour cela, nous reprenons les trois premières idées exposées par COCHRAN, en leur ajoutant l'estimation par le ratio sous plan à probabilité proportionnelle à la taille des unités.

¹ : Rappelons que, dans la situation typique d'application de la méthode, nous ne pensons pas pouvoir disposer d'information qui permettent de constituer une autre base de sondage que celle, spatiale, issue de l'image satellite. Dans une telle situation, la constitution d'une liste d'unités statistiques spatiales, qui seule, permet d'envisager des tirages équiprobables et, plus généralement de contrôler complètement la probabilité de sortie des unités, n'est pas toujours possible. Elle réclamerait, de toute façon, un important travail d'identification et de vérification des unités spatiales, préjudiciable à la rapidité de mise en oeuvre de l'enquête.

² : W.G. COCHRAN : Sampling technics, WILEY 1977, troisième édition, chap. 9, P. 255.

³ : J. DESABIE : Théorie et pratique des sondages, DONOD 1966, Chap. 12, P. 241.

Nous allons maintenant décrire les outils statistiques et informatiques qui vont permettre d'effectuer le test, présenter le formulaire correspondant aux quatre possibilités d'estimation, enfin nous exposerons l'architecture et le mode d'emploi du programme informatique.

6.2.2. : Formulaire général

6.2.2.1. : Préliminaire

Lors de la phase exploratoire exposée dans la première partie de ce chapitre, le formulaire correspondait à un plan de sondage sans remise, dont la théorie, développée en 1952 par Horwitz et Thompson trouve une application simple dans le cas de plans équiprobables. Lorsque les probabilités de sélection sont inégales, le tirage sans remise, qui reste bien entendu préférable, conduit à des calculs complexes qui supposent d'explicitier les probabilités d'appartenance à l'échantillon, celles-ci se modifiant au fur et à mesure du tirage. Or, le but de l'exposé qui va suivre est, rappelons-le, triple :

- *comparer des plans de sondages équiprobables à des plans où la probabilité est proportionnelle à la taille des unités d'échantillonnage ;*
- *comparer les estimateurs sans biais aux estimateurs par le quotient ;*
- *évaluer l'apport d'une stratification morphologique de la base de sondage à la précision d'estimations démographiques.*

Relativement à ces objectifs, s'il importe que la procédure de tirage considérée soit la même dans tous les cas, on peut par contre utiliser indifféremment un tirage avec ou sans remise. L'hypothèse qui est alors faite est que le tirage sans remise améliore chaque estimation approximativement dans les mêmes proportions. C'est la raison pour laquelle on donnera ici les formules simples correspondant à des tirages avec remise. La même option étant prise au sein du programme de calcul des variances d'estimation, ces dernières sont donc surévaluées dans une proportion approximativement égale au taux de sondage : 5 à 7 % suivant la base de sondage et le taux retenus¹.

6.2.2.2. : Conventions et notations

1. Dans ce qui suit, on se limite la plupart du temps, à énoncer les expressions des estimateurs, de leur variance, etc., sans démonstrations. Ces résultats sont en effet désormais classiques et leurs démonstrations figurent dans la plupart des ouvrages généraux cités dans la bibliographie. Seule l'estimation par le ratio sous plan à probabilités inégales est exposée en détail, cet exposé étant beaucoup moins fréquent dans la littérature sur les sondages.

2. Comme on l'a constaté dans le paragraphe précédent, les formules stratifiées s'obtiennent sans difficultés par sommation sur l'ensemble des strates des formules établies pour l'ensemble de la base. Elles n'apparaîtront généralement pas ici ; seul le raisonnement menant à l'allocation optimale part nécessairement de l'expression explicite de la variance dans le cas stratifié.

3. Les notations seront identiques à celles que nous avons adoptées plus haut : la variable estimée est toujours notée y , les valeurs prises sur les unités statistiques y_i ,

¹ : Cette estimation du gain dû au tirage sans remise provient de la comparaison des expressions des variances des estimateurs sans biais obtenues avec et sans remise dans le cas d'un plan de sondage équiprobable : la variance sans remise est alors inférieure à la variance avec remise d'un facteur $(N-n)/N$ égal au taux.

son total Y, etc. Lorsqu'il est fait référence à la variable exogène utilisée comme dénominateur dans les formules relatives à l'estimation par le ratio, celle-ci est notée x, ses valeurs x_i, son total X, etc. La variable servant à définir la probabilité dans les plans à probabilités inégales est notée z, sa valeur z_i et son total Z.

6.2.2.3. : Plan équiprobable, estimateur sans biais

On a, de même que précédemment, les estimateurs :

$$\hat{T}_1(Y) = N\bar{y} = N \sum_{i=1}^n y_i/n \quad (62.1 a) \text{ et,}$$

$$\hat{M}_1(\bar{y}) = \bar{y} = \sum_{i=1}^n y_i/n \quad (62.1 b), \text{ mais les variances ont des expressions}$$

plus simples :

$$V(\hat{T}_1(Y)) = \frac{N^2 \sigma_y^2}{n} \quad \text{et} \quad V(\hat{M}_1(\bar{y})) = \frac{\sigma_y^2}{n} \quad (62.2 a \& b).$$

La détermination de l'allocation optimale se fait par minimisation de la variance stratifiée sous la contrainte $\sum_{h=1}^K n_h = n$. La solution du problème est obtenue grâce à l'utilisation des multiplicateurs de Lagrange.

on recherche $\text{Min}_{(n_h)} \left(\sum_{h=1}^K \frac{N_h^2 \sigma_h^2}{n_h} \right)$, dont le Lagrangien

$$\text{s'écrit : } L = \sum_{h=1}^K \left(\frac{N_h^2 \sigma_h^2}{n_h} \right) + (n - \sum_{h=1}^K n_h)\beta,$$

ou β est un multiplicateur de Lagrange.

Les conditions du premier ordre sont : $\frac{\partial L}{\partial n_h} = 0$ pour tout h,

$$\text{on a donc : } \frac{N_h^2 \sigma_h^2}{n_h^2} = \beta \quad (\Leftrightarrow) \quad n_h = (1/\sqrt{\beta}) N_h \sigma_h$$

En écrivant la contrainte $\sum_{h=1}^K n_h = n$, on trouve :

$$1/\sqrt{\beta} = \frac{n}{\sum_{h=1}^K n_h \sigma_h}$$

D'ou l' allocation pour la strate h :

$$n_h = n \left(\frac{n_h \sigma_h}{\sum_{h=1}^K n_h \sigma_h} \right) \quad (62.3).$$

6.2.2.4. : Plan équiprobable, estimateur par le ratio.

La connaissance du total, sur l'ensemble de la base, d'une variable exogène jamais nulle (dans nos applications, il s'agira de la taille des districts : surface totale dans un premier temps, surface bâtie par la suite), permet d'utiliser, pour estimer le total Y de y ou pour estimer directement le quotient Y/X, l'estimateur naturel du ratio Y/X = R qui s'écrit :

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \quad (62.4)$$

L'estimateur du total Y s'écrit alors :

$$\hat{T}_2(Y) = X \hat{R}$$

Celui de la moyenne :

$$\hat{M}_2(\bar{y}) = \bar{x} \hat{R}$$

Ces estimateurs sont légèrement biaisés, le calcul de ce biais se fait en utilisant le développement limité d'ordre un à l'origine de $1/(1 + \epsilon)$, où ϵ représente la quantité $(\bar{x} - \bar{X})/\bar{X}$, proche de zéro dès que n, le cardinal de l'échantillon, est assez grand. On obtient, pour un plan équiprobable avec remise, l'expression suivante pour le biais sur l'estimation du ratio R¹.

$$E(\hat{R} - R) \approx - \frac{N^2}{n\bar{X}^2} (\sigma_{xy}^2 - R\sigma_x^2) \quad (62.4 a)$$

¹ : Pour un exposé complet du calcul, voir par exemple J.M. GROBRAS : Méthodes statistiques des sondages, Pages 131 & 132.

Le biais sur l'estimation d'un total est :

$$E(\hat{T}_2(Y) - Y) \approx -\frac{N^2}{n\bar{X}} (\sigma_{xy}^2 - R\sigma_x^2) \quad (62.4 b)$$

celui sur l'estimation d'une moyenne :

$$E(\hat{M}_2(\bar{y}) - \bar{y}) \approx -\frac{N}{n\bar{X}} (\sigma_{xy}^2 - R\sigma_x^2) \quad (62.4 c)$$

La variance de l'estimateur $\hat{T}_2(Y)$ s'écrit :

$$V(\hat{T}_2(Y)) = \left(\frac{N^2}{n}\right) (\sigma_y^2 + R^2 \sigma_x^2 - 2R \sigma_{xy}^2) \quad (62.5 a)$$

Celle de l'estimateur $\hat{M}_2(\bar{y})$:

$$V(\hat{M}_2(\bar{y})) = \left(\frac{1}{n}\right) (\sigma_y^2 + R^2 \sigma_x^2 - 2R \sigma_{xy}^2) \quad (62.5 b)$$

Le risque associé aux deux estimateurs s'obtient ensuite en sommant la variance et le carré du biais.

Pour déterminer l'allocation optimale, il faut minimiser la variance totale stratifiée, dont l'expression est :

$$V = \sum_{h=1}^K \left(\frac{N_h^2}{n_h}\right) (\sigma_{h y}^2 + R_h^2 \sigma_{h x}^2 - 2R_h \sigma_{h xy}^2) + \sum_{h=1}^K \left(\frac{N_h^4}{n_h^2 \bar{X}_h^2}\right) (\sigma_{h xy}^2 - R_h^2 \sigma_{h x}^2),$$

où K est le nombre de strates, $\sigma_{h y}^2$ et $\sigma_{h x}^2$ les variances respectives des caractères y et

x dans la strate h , enfin $\sigma_{h xy}^2$ la covariance de x et y dans cette même strate. En

appliquant la méthode des multiplicateurs de Lagrange, on aboutit au système d'équations en n_h :

$$\forall h, \quad \left(\frac{N_h^2}{n_h}\right) (\sigma_{h y}^2 + R_h^2 \sigma_{h x}^2 - 2R_h \sigma_{h xy}^2) + \left(\frac{2N_h^4}{n_h^3 \bar{X}_h^4}\right) (\sigma_{h xy}^2 - R_h^2 \sigma_{h x}^2) = C$$

En négligeant le second terme d'ordre $1/n_h^3$ devant le premier d'ordre $1/n_h^2$, les conditions s'écrivent :

$$\forall h, \frac{N_h}{n_h} = C. (\sigma_h^2 y + R_h^2 \sigma_h^2 x - 2R_h \sigma_h^2 xy)^{1/2}$$

Puis, en écrivant la contrainte sur la taille de l'échantillon : $\sum_{h=1}^K n_h = n$, on obtient l'expression des n_h :

$$n_h = n \cdot \frac{N_h (\sigma_h^2 y + R_h^2 \sigma_h^2 x - 2R_h \sigma_h^2 xy)^{1/2}}{\sum_{h=1}^K N_h (\sigma_h^2 y + R_h^2 \sigma_h^2 x - 2R_h \sigma_h^2 xy)^{1/2}} \quad (62.6)$$

6.2.2.5 : Plan à probabilités inégales, estimateur sans biais.

Les probabilités d'inclusion π_i sont définies proportionnelles à une variable exogène z de total Z : il s'agit pour nous de la mesure de la taille associée à chaque unité statistique de la base (z_i). On a donc $\pi_i = nz_i/Z$ pour tout i et l'estimateur sans biais du total Y s'écrit¹ :

$$\hat{T}_3(Y) = \frac{Z}{n} \cdot \sum_{i=1}^n \frac{y_i}{z_i} \quad (62.7 a)$$

Celui de la moyenne :

$$\hat{M}_3(\bar{y}) = \frac{Z}{nN} \cdot \sum_{i=1}^n \frac{y_i}{z_i} \quad (62.7 b)$$

Pour les variances, on a :

$$V(\hat{T}_3(Y)) = \frac{1}{n} \cdot (Z \sum_{i=1}^n \left(\frac{y_i}{z_i}\right)^2 - Y^2) \quad (62.8 a)$$

$$V(\hat{M}_3(\bar{y})) = \frac{1}{nN^2} \cdot (Z \sum_{i=1}^n \left(\frac{y_i}{z_i}\right)^2 - Y^2) \quad (62.8 b)$$

¹ : Voir DESABIE : Théorie et pratique des sondages, chap. 11, P. 227.

A nouveau le calcul de l'allocation optimale se fait à partir de la variance stratifiée :

$$V(\hat{T}_3(Y)) = \sum_{h=1}^K \frac{1}{n_h} \cdot (Z_h \sum_{i=1}^{N_h} (\frac{y_i}{z_i})^2 - Y_h^2)$$

En appliquant à cette fonction la même méthode qu'au paragraphe précédent, on obtient les conditions :

$$\forall h, \quad \frac{1}{n_h} \cdot (Z_h \sum_{i=1}^{N_h} (\frac{y_i}{z_i})^2 - Y_h^2) = C$$

D'où $n_h = C (Z_h \sum_{i=1}^{N_h} (\frac{y_i}{z_i})^2 - Y_h^2)^{1/2}$.et, puisque $\sum_{h=1}^K n_h = n$,

$$n_h = n \cdot \frac{(Z_h \sum_{i=1}^{N_h} (\frac{y_i}{z_i})^2 - Y_h^2)^{1/2}}{\sum_{h=1}^K (Z_h \sum_{i=1}^{N_h} (\frac{y_i}{z_i})^2 - Y_h^2)^{1/2}} \quad (62.9)$$

6.2.2.6. : Plan à probabilités inégales, estimation d'un ratio ou d'un total par le ratio

L'estimateur

De nouveau, les probabilités de tirage π_i sont définies par les z_i , avec : $\pi_i = nz_i/Z$; x étant une variable exogène (en principe distincte de z), de total X , on s'intéresse à l'estimation du ratio $R = Y/X$, ainsi qu'à celle du total de y par le truchement de R puisque $Y = RX$.

Sous le plan π ainsi défini, les estimateurs sans biais des totaux X et Y sont connus :

$$\hat{T}_3(Y) = \frac{Z}{n} \cdot \sum_{i=1}^n \frac{y_i}{z_i} \quad \hat{T}_3(X) = \frac{Z}{n} \cdot \sum_{i=1}^n \frac{x_i}{z_i}$$

L'estimateur naturel du ratio $R = Y/X$ est donc :

$$\hat{R}_4 = \frac{\sum_{i=1}^n \frac{y_i}{z_i}}{\sum_{i=1}^n \frac{x_i}{z_i}} \quad (62.10)$$

On estimera Y par : $\hat{T}_4(Y) = X\hat{R}_4$, et \bar{y} par : $\hat{M}_4(\bar{y}) = (X/N)\hat{R}_4$. Mais ces estimateurs sont légèrement biaisés.

Expression des biais

La différence entre le ratio R et son estimateur s'écrit :

$$\hat{R} - R = \frac{\sum_{i=1}^n \frac{y_i}{z_i}}{\sum_{i=1}^n \frac{x_i}{z_i}} - \frac{Y}{X} = \frac{\hat{Y}}{\hat{X}} - \frac{Y}{X}$$

ou encore : $\hat{R} - R = \frac{\hat{Y} - R\hat{X}}{\hat{X}} = \frac{\hat{Y} - R\hat{X}}{X(1 + \frac{\hat{X}-X}{X})}$, où la quantité $\frac{\hat{X}-X}{X}$ est proche de zéro,

surtout lorsque n grandit puisque \hat{X} est un estimateur sans biais de X . On peut donc obtenir, par développement limité du premier ordre, une expression approchée de cette différence :

$\hat{R} - R = \left(\frac{\hat{Y} - R\hat{X}}{X}\right) \left(1 - \frac{\hat{X}-X}{X}\right)$, à partir de laquelle on calcule la valeur approchée du biais sur l'estimation du ratio :

$$E_{\pi}(\hat{R} - R) \approx E_{\pi}\left(\frac{\hat{Y} - R\hat{X}}{X}\right) - \frac{1}{X^2} E_{\pi}((\hat{Y} - R\hat{X})(\hat{X} - X)), \text{ où } E_{\pi} \text{ est l'espérance}$$

mathématique prise sous la loi de probabilité du plan π .

$$\text{Or, } E_{\pi}(\hat{Y} - R\hat{X}) = E_{\pi}(\hat{Y}) - RE_{\pi}(\hat{X}) = Y - RX = 0, \text{ par définition de } \hat{X} \text{ et } \hat{Y} \text{ donc,}$$

$E_{\pi}(\hat{R} - R) \approx -\frac{1}{X^2} E_{\pi}(\hat{X}(\hat{Y} - R\hat{X})) = -\frac{1}{X^2} (E_{\pi}(\hat{X}\hat{Y}) - RE_{\pi}(\hat{X}^2))$, ou encore, grâce à des relations connues :

$$E_{\pi}(\hat{R} - R) \approx -\frac{1}{X^2} (Cov_{\pi}(\hat{X}, \hat{Y}) - R \cdot Var_{\pi}(\hat{X}))$$

Exprimons $\text{Cov}_{\pi}(\hat{X}, \hat{Y})$ en fonction des variables aléatoires x, y et z :

$$\text{Cov}_{\pi}(\hat{X}, \hat{Y}) = \text{Cov}_{\pi} \left(\frac{Z}{n} \sum_{i=1}^n \frac{x_i}{z_i}, \frac{Z}{n} \sum_{i=1}^n \frac{y_i}{z_i} \right) = \frac{Z^2}{n^2} \text{Cov}_{\pi} \left(\sum_{i=1}^n \frac{x_i}{z_i}, \sum_{i=1}^n \frac{y_i}{z_i} \right)$$

Si l'on pose, pour la commodité de l'écriture, $t_1 = x/z$ et $t_2 = y/z$, on a :

$$\text{Cov}_{\pi}(\hat{X}, \hat{Y}) = \frac{Z^2}{n^2} \text{Cov}_{\pi} \left(\sum_{i=1}^n t_{1i}, \sum_{i=1}^n t_{2i} \right) \text{ et, comme les}$$

tirages sont indépendants avec remise, les variables aléatoires t_{1i} et t_{2i} sont égales à t_1 et t_2 quel que soit i .

On obtient alors facilement :

$$\text{Cov}_{\pi} \left(\sum_{i=1}^n t_{1i}, \sum_{i=1}^n t_{2i} \right) = n \text{Cov}_{\pi}(t_1, t_2) = n \text{Cov}_{\pi}(x/z, y/z), \text{ d'ou :}$$

$$\boxed{\text{Cov}_{\pi}(\hat{X}, \hat{Y}) = \frac{Z^2}{n} \text{Cov}_{\pi}(x/z, y/z)} \quad (62.11)$$

De même,

$$\text{Var}_{\pi}(\hat{X}) = \text{Var}_{\pi} \left(\frac{Z}{n} \sum_{i=1}^n \frac{x_i}{z_i} \right) = \frac{Z^2}{n^2} \text{Var}_{\pi} \left(\sum_{i=1}^n t_{1i} \right) = \frac{Z^2}{n^2} n \cdot \text{Var}_{\pi}(t_1), \text{ et on obtient :}$$

$$\boxed{\text{Var}_{\pi}(\hat{X}) = \frac{Z^2}{n} \text{Var}_{\pi}(x/z)} \quad (62.12)$$

D'où l'expression du biais en fonction des variances, sous le plan π des aléatoires x/z et y/z :

$$\boxed{E_{\pi}(\hat{R} - R) \approx - \frac{Z^2}{NX^2} \left(\text{Cov}_{\pi}(x/z, y/z) - R \cdot \text{Var}_{\pi}(x/z) \right)} \quad (62.13)$$

La formule (62.13) montre que si n tend vers l'infini, $E_{\pi}(\hat{R} - R)$ tend vers zéro. \hat{R} est donc un estimateur asymptotiquement sans biais de R . A distance finie, le biais est un infiniment petit d'ordre $1/n$. Les biais sur les estimations du total et de la moyenne de y s'obtiennent en multipliant respectivement par X et par X/N , le biais sur l'estimation du ratio.

Variance des estimations

La variance de l'estimateur du ratio s'écrit :

$$V_{\pi}(\hat{R}) = E_{\pi}((\hat{R} - R)^2) \approx E_{\pi} \left(\left(\frac{\hat{Y} - R\hat{X}}{X} \right) \left(1 - \frac{\hat{X} - X}{X} \right) \right)^2$$

Si l'on néglige, comme infiniment petit d'ordre 2, le terme en $((\hat{X} - X)/X)^2$, et comme le terme provenant du double produit s'annule, il reste à évaluer :

$$V_{\pi}(\hat{R}) \approx E_{\pi} \left(\frac{\hat{Y} - R\hat{X}}{X} \right)^2 = \left(\frac{1}{X^2} \right) E_{\pi} (\hat{Y}^2 + R^2 \hat{X}^2 - 2R\hat{X}\hat{Y})$$

D'où :
$$V_{\pi}(\hat{R}) \approx \left(\frac{1}{X^2} \right) \left(\text{Var}_{\pi}(\hat{Y}) + R^2 \text{Var}_{\pi}(\hat{X}) - 2R \text{Cov}_{\pi}(\hat{X}, \hat{Y}) \right)$$

En utilisant les relations 62.11 et 62.12 établies au paragraphe précédent , on obtient :

$$V_{\pi}(\hat{R}) \approx \left(\frac{Z^2}{nX^2} \right) \left(\text{Var}_{\pi}(y/z) + R^2 \text{Var}_{\pi}(x/z) - 2R \text{Cov}_{\pi}(x/z, y/z) \right) \tag{62.14}$$

Puis, pour l'estimation du total Y et de la moyenne y, respectivement :

$$V_{\pi}(\hat{T}_4(Y)) \approx \left(\frac{Z^2}{n} \right) \left(\text{Var}_{\pi}(y/z) + R^2 \text{Var}_{\pi}(x/z) - 2R \text{Cov}_{\pi}(x/z, y/z) \right) \tag{62.15}$$

$$V_{\pi}(\hat{M}_4(\bar{y})) \approx \left(\frac{Z^2}{Nn} \right) \left(\text{Var}_{\pi}(y/z) + R^2 \text{Var}_{\pi}(x/z) - 2R \text{Cov}_{\pi}(x/z, y/z) \right) \tag{62.16}$$

Allocation optimale

La fonction à minimiser est le risque quadratique (incluant le carré du biais) de l'estimateur stratifié qui s'écrit (pour l'estimation du total de y par exemple):

$$R_{\pi}(\hat{T}_4(Y)) \approx \sum_{h=1}^K \left(\frac{Z_h^2}{n_h} \right) \left(\text{Var}_{\pi_h}(y/z) + R_h^2 \text{Var}_{\pi_h}(x/z) - 2R_h \text{Cov}_{\pi_h}(x/z, y/z) \right) + \sum_{h=1}^K \left(\left(\frac{Z_h^2}{n_h X_h} \right) \left(\text{Cov}_{\pi_h}(x/z, y/z) - R_h \text{Var}_{\pi_h}(x/z) \right) \right)^2$$

La dérivation, par rapport aux n_h , de cette expression fournit les conditions suivantes :

$$\forall h, \left(\frac{Z_h^2}{n_h} \right) \left(\text{Var}_h(y/z) + R_h^2 \text{Var}_h(x/z) - 2R_h \text{Cov}_h(x/z, y/z) \right) - 2 \left(\frac{Z_h^4}{n_h^3 X_h^2} \right) \left((\text{Cov}_h(x/z, y/z) - R_h \text{Var}_h(x/z)) \right)^2$$

Le second terme d'ordre $1/n_h^3$ peut être négligé par rapport au premier d'ordre $1/n_h^2$.

On obtient alors que, pour tout h , n_h doit être proportionnel à :

$$Z_h \left(\text{Var}_h(y/z) + R_h^2 \text{Var}_h(x/z) - 2R_h \text{Cov}_h(x/z, y/z) \right)^{1/2}, \text{ puis en écrivant la}$$

contrainte $\sum_{h=1}^K n_h = n$, l'expression des n_h :

$$n_h = n \frac{Z_h \left(\text{Var}_h(y/z) + R_h^2 \text{Var}_h(x/z) - 2R_h \text{Cov}_h(x/z, y/z) \right)^{1/2}}{\sum_{h=1}^K Z_h \left(\text{Var}_h(y/z) + R_h^2 \text{Var}_h(x/z) - 2R_h \text{Cov}_h(x/z, y/z) \right)^{1/2}}$$

(62.17)

Remarque

A partir de l'expression 62.15, on constate que, dans le cas particulier où les variables x et z sont les mêmes, l'estimateur par le ratio a la même variance que l'estimateur sans biais. En effet, si $x/z=1$, on a immédiatement $\text{Var}_h(x/z) = 0$ et $\text{Cov}_h(x/z, y/z) = 0$ et

nous allons montrer que : $Z^2 \text{Var}_\pi(y/z) = Z \sum_{i=1}^N \left(\frac{y_i}{z_i} \right)^2 - Y^2 = V(\hat{T}_3(Y))$ (voir l'expression 62.8 a).

$$\begin{aligned} \text{Var}_\pi(y/z) &= \sum_{i=1}^N \frac{z_i}{Z} \left(\frac{y_i}{z_i} - \frac{\sum_{i=1}^N \frac{z_i y_i}{Z z_i}}{Z} \right)^2 = \sum_{i=1}^N \frac{z_i}{Z} \left(\frac{y_i}{z_i} - \frac{Y}{Z} \right)^2 = \frac{1}{Z} \sum_{i=1}^N \left(\frac{y_i^2}{z_i} - 2 \frac{Y y_i}{Z} + \frac{z_i Y^2}{Z^2} \right) \\ &= \frac{1}{Z} \left(\sum_{i=1}^N \frac{y_i^2}{z_i} - 2 \frac{Y^2}{Z} + \frac{Y^2}{Z} \right) = \frac{1}{Z} \left(\sum_{i=1}^N \frac{y_i^2}{z_i} - \frac{Y^2}{Z} \right) \end{aligned}$$

$$\text{Donc } Z^2 \text{Var}(y/z) = Z \sum_{i=1}^N \left(\frac{y_i^2}{z_i} \right) - Y^2, \text{ ce que nous devons démontrer.}$$

Ainsi, sous un plan de sondage à probabilité proportionnelle à une mesure donnée de la taille des unités primaires (leur surface par exemple), il est inutile de chercher à améliorer les estimations de totaux en employant des estimateurs par le ratio utilisant cette même mesure de la taille au dénominateur. Le formulaire développé ici reste cependant utile pour l'estimation directe des quotients comme nous le verrons par la suite.

6.3. : Le programme VARECH associé au logiciel dbaseIII

Pour obtenir, avec les plans de sondage et les estimateurs qui viennent d'être décrits, les résultats souhaités, il nous fallait réitérer les calculs de variance, pour l'ensemble du formulaire, sur plusieurs variables démographiques, plusieurs définitions de la probabilité et plusieurs types de stratifications. Il était donc nécessaire de disposer d'un outil informatique comprenant d'une part, un logiciel gestionnaire de fichiers, au sein duquel puisse être intégré et géré l'ensemble des données nécessaires et, d'autre part, un programme de calcul des variances des différents estimateurs, dans lequel soient paramétrés les choix des estimateurs, des probabilités et des stratifications. Nous décrivons ici rapidement, les caractéristiques du matériel, de la base de données et du logiciel qui permet sa gestion, ainsi que l'architecture générale du programme, l'ensemble formant un outil applicable sur d'autres sites ou à d'autres données. Le système demandera ensuite à être complété pour fournir, d'après les données recueillies par sondage, des estimations et leur précision.

6.3.1. : Matériel, système gestionnaire de fichiers et base de données Dbase III sur Marseille.

Le matériel utilisé depuis notre arrivée à Quito est un micro ordinateur IBM PC-AT, équipé d'un disque dur de 20 Mega octets, l'ensemble géré par le système d'exploitation DOS (version 3.1). Sur ce système a été implanté le S.G.B.D. dbase III qui offre, outre de multiples facilités pour la gestion des fichiers, un langage de programmation intégré, à partir duquel a été développé le programme de calcul.

Pour disposer, sous dbase III, de la base de données qui sert de support aux tests, nous avons constitué un ensemble de fichiers correspondant à des données de provenances diverses : recensement général de la population de 1982 à Marseille, photo-interprétation d'une mosaïque de photographies aériennes couvrant la zone d'étude, données de surface et de localisation des districts INSEE de la base, obtenues grâce au système d'information géographique SAVANE, développé par l'unité d'infographie de l'ORSTOM¹. Par la suite, nous y avons également intégré un quatrième type de données : celles issues de l'interprétation des images satellites fournissant une caractérisation morphologique et des stratifications de la base de sondage. La phase d'intégration et de croisement de ces données est résumée dans la figure 6.6.

¹ : Voir, à propos des divers types de données utilisées, l'annexe 7 du rapport ATP CNRS/CNES d'Août 1987.

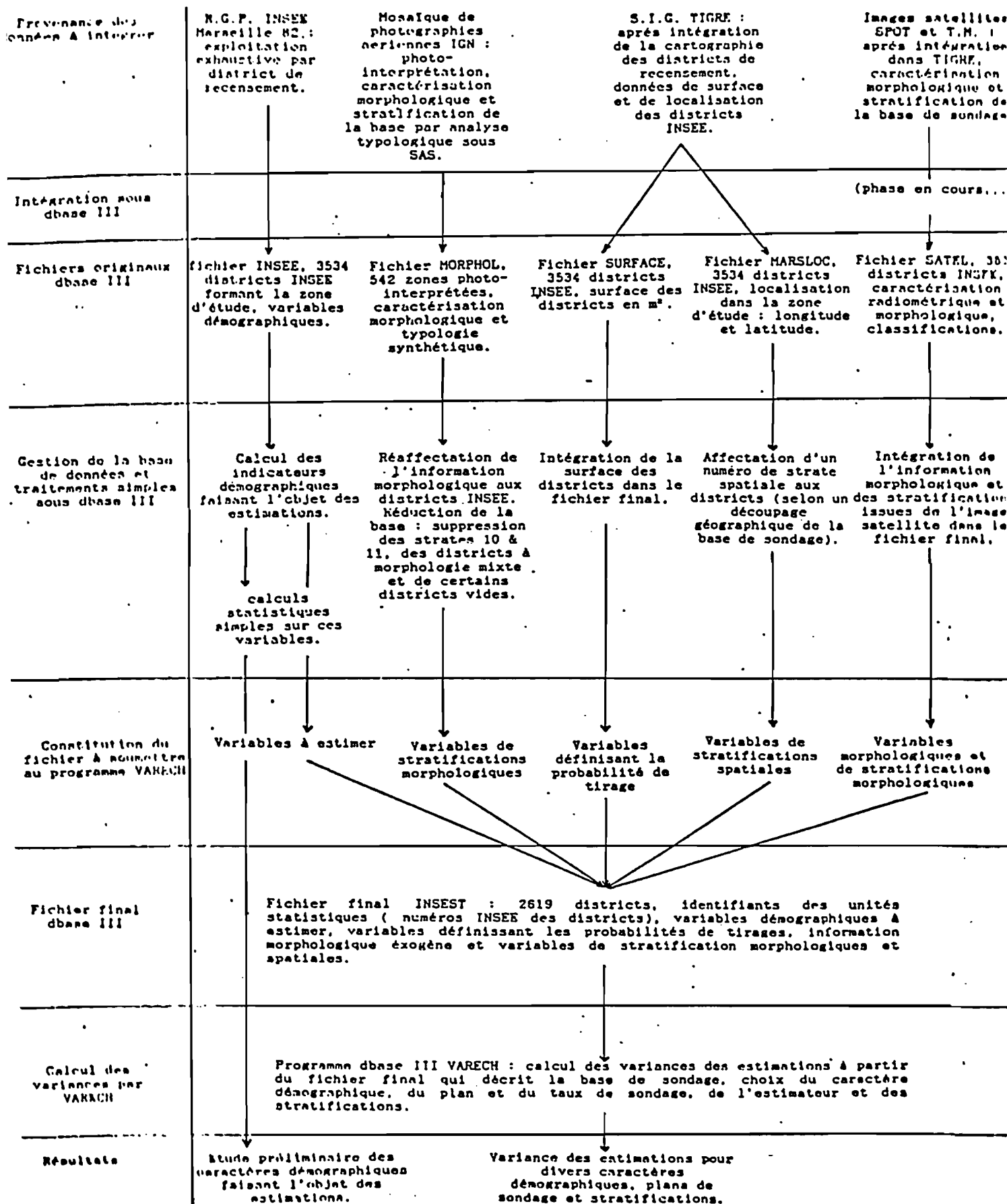
6.3.2. : Le programme VARECH (VARIance d'EChantillonnage), architecture générale et fonctionnement.

Une fois constitué, sous dbase III, le fichier contenant les informations nécessaires, le programme VARECH permet d'effectuer les calculs de variance correspondant à l'estimateur, au plan de sondage et à la stratification choisis. Ecrit sous dbase III, ce programme se présente comme un ensemble de commandes dbase, organisées en procédures qui calculent, pour un plan de sondage, un estimateur, une stratification et un taux donnés, la variance qu'aura l'estimation du total (de la moyenne ou du quotient) du caractère démographique considéré. L'organigramme général du programme est donné dans l'annexe A.2.3, ainsi que la liste des procédures dont il est constitué. On trouvera également dans cet annexe le listage complet du programme.

L'ensemble des informations nécessaires au programme doivent se trouver dans un même fichier dbase III, appelé fichier de données. Après avoir effectué les calculs, le programme écrit les résultats dans un fichier que l'on devra éditer ensuite.

Après avoir entré le nom du fichier résultat, l'utilisateur choisit, grâce au menu principal, le plan de sondage et l'estimateur; puis, dans le menu secondaire, il choisit l'estimation, stratifiée ou non, d'un total ou d'une moyenne. L'utilisateur indique ensuite au programme, le nom du fichier de données et les noms des champs contenant les différentes variables : variable à estimer, variable exogène figurant au dénominateur (dans le cas de l'estimation d'un ratio), variable définissant la probabilité (dans le cas d'un plan à probabilités inégales), variable de stratification s'il y a lieu. Il doit également choisir un taux de sondage et préciser le nombre de strates que contient la base de sondage. Dans sa version finale, le programme comprendra une procédure d'édition du fichier résultat qui permettra de sélectionner et d'imprimer les résultats. Des exemples de déroulement du programme sont présentés sous forme de copies d'écrans dans l'annexe A.2.3.

Figure 6.6 : l'intégration et les croisements de données sous dbase III



Conclusion

Soulignons tout d'abord que le programme actuel est à l'état de prototype. Pourtant on peut, dès maintenant, donner le domaine d'application qu'auront les différentes versions.

Dans son état de développement actuel, VARECH permet un travail d'évaluation, à partir de données exhaustives existant sur la base de sondage choisie (celles-ci proviennent généralement de recensements), des possibilités d'estimation par sondage de ces données. Intégrant diverses possibilités quant aux choix de plans de sondage, estimateurs et stratifications, il fournit la précision réelle des estimations et conduit ainsi à la sélection d'une ou plusieurs techniques de sondage adaptées au site et au type des données recherchées.

Il s'agit donc, à l'heure actuelle, d'un outil de mise au point méthodologique, qui trouve sa principale utilité lors d'opérations d'évaluation des possibilités d'utilisation des techniques de sondage.

Les développements futurs devront conduire à des versions assurant ou facilitant la mise en pratique de la méthode d'échantillonnage et d'estimation sélectionnée. Ces versions comprendront des procédures de sélection (ou d'aide à la sélection) des unités de l'échantillon, d'extrapolation des résultats du sondage aux domaines d'étude et d'estimation des précisions.

Le domaine d'utilisation de la version finale s'étendra donc de l'évaluation à la mise en oeuvre de méthodes de sondage, sur des sites où existent des données permettant une approche rigoureuse des choix techniques à opérer.

C'est précisément pour la mise au point, à partir des données du recensement de population de 1982, d'une méthode de sondage applicable dans l'agglomération Marseillaise à l'estimation de variables démographiques, que nous allons utiliser le programme dans les deux chapitres suivants. Par la suite il servira également à l'adaptation et au perfectionnement de cette méthode sur le site de Quito en Equateur, mais cette fois il s'appliquera à des données recueillies lors d'enquêtes préliminaires sur échantillons.

Chapitre 7

EVALUATION DE PLANS DE SONDAGE STRATIFIES POUR L'ESTIMATION DE CARACTERES DEMOGRAPHIQUES

*Résultats des test menés grace au programme VARECH sur la base de données constituée
à Marseille*

Introduction

Ce chapitre présente les résultats de l'application du programme VARECH à une série d'indicateurs démographiques issus des chiffres du recensement général de la population de 1982 à Marseille.

Dans le chapitre précédent, pour généraliser le test et mettre au point le programme¹, nous dressions la liste des possibilités existant pour le plan de sondage au premier degré :

- 1 Plan équiprobable, estimation sans biais
- 2 Plan équiprobable, estimation par le ratio
- 3 Plan à probabilités proportionnelles à la surface des unités primaires, estimation sans biais
- 4 Plan à probabilités proportionnelles à la surface des unités primaires, estimation par le ratio.

L'objectif du travail présenté ici est de déterminer, en évaluant à partir des données exhaustives ces différentes possibilités, les techniques de stratification, de sélection des unités primaires et d'estimation qui définiront un plan de sondage au premier degré adapté et efficace pour l'estimation des caractères démographiques.

L'évaluation se déroule en trois temps. Tout d'abord on compare les quatre techniques d'estimation pour une même variable : le total de population sans doubles comptes. Le test est ensuite étendu à d'autres caractères démographiques. Enfin on applique le programme à l'évaluation des stratifications construites d'après le critère de la densité du bâti qui, nous le verrons, joue un rôle important pour définir des zones où la population présente certaines caractéristiques démographiques homogènes. On s'intéresse d'abord aux stratifications établies grâce à l'analyse typologique des données issues de la photo-interprétation puis à la stratification issue de la classification de l'image satellite sur ce thème.

En conclusion, après avoir rappelé les principales questions qui se posent, on effectue une synthèse des résultats concernant les performances respectives des plans de sondages équiprobables et à probabilités inégales, le choix de l'estimateur à employer

¹ : Voir chapitre 6, paragraphe 6.2 : Généralisation du test : plans équiprobables et plans à probabilités inégales, estimateurs sans biais et estimateurs par le ratio, 6.2.1 : Position du problème.

(sans biais ou par le ratio) et les gains de précision permis par les différentes stratifications étudiées. Enfin, on examine la stabilité de ces différents résultats lorsque varie le caractère démographique à estimer.

7.1 : Comparaison de quatre techniques d'estimation pour une même variable

- Préliminaire

Comme on l'a vu au chapitre précédent (remarque du paragraphe 6.2.2.6, chapitre 6), lorsque les probabilités d'inclusion dans l'échantillon sont définies proportionnelles aux surfaces des unités primaires, l'estimateur d'un total Y par le ratio Y/S (où S est précisément la surface totale de la base) a la même variance que l'estimateur sans biais de Y . La surface des UP étant pour l'instant la seule variable exogène que nous puissions utiliser, l'estimateur par le ratio n'offre donc aucun intérêt dans le cadre du plan à probabilités inégales que nous considérons ici. La comparaison des estimations de totaux s'effectue donc entre les possibilités suivantes :

1. Plan équiprobable, estimateur sans biais
2. Plan équiprobable, estimateur par le ratio Y/S
3. Plan à probabilités proportionnelles à la surface des UP, estimateur sans biais (ou estimateur par le ratio Y/S , ce qui revient au même)

En revanche, pour l'estimation directe de quotients (voir paragraphe suivant), nous utilisons le même estimateur par le ratio, mais son dénominateur n'est plus la surface. On dispose alors dans VARECH de deux solutions :

1. Plan équiprobable (estimation d'un ratio)
2. Plan à probabilités proportionnelles à la surface des UP (estimation d'un ratio)

- Résultats

Le tableau 7.1.a fournit les coefficients de variation des trois estimateurs dans les cas stratifiés et non stratifiés, pour l'ensemble de la base et pour chacune des neuf strates morphologiques. Le graphique 7.1.b visualise ces résultats pour l'estimateur sans biais sous plan à probabilités inégales, dans le cas de l'allocation optimale.

Il y a quatre conclusions à tirer de la comparaison des trois estimateurs.

1. La surface des unités statistiques (districts INSEE), prise ici comme mesure de leur taille, n'est pas suffisamment corrélée à l'effectif de population pour que son introduction dans la formule d'estimation par l'intermédiaire du ratio P/S , améliore la précision. Au contraire, cet estimateur s'avère moins précis que les deux autres, tant pour l'ensemble de la base, qu'au niveau de chacune des strates. *Il sera donc écarté sous cette forme et nous chercherons, pour la suite, une meilleure mesure de la taille des unités.*

2. L'estimation sans biais se fait avec une précision équivalente sous les deux plans de sondage considérés : plan équiprobable et plans à probabilités proportionnelles à la surface des districts. Pour une estimation non stratifiée, il y a un léger avantage au premier (5,75 % de la variance en moins), mais cet avantage devient négligeable dans le cas stratifié avec l'allocation optimale (1,3 % de la variance). *L'équivalence des deux plans est un résultat important pour la suite de l'étude puisqu'en pratique, le tirage de l'échantillon sur documents issus de l'image satellite nous contraint à utiliser un plan à probabilités inégales.* Pour conclure sur ce point, il reste à tester un plan où les probabilités soient déterminées par la surface bâtie dans l'unité primaire et non plus par sa surface totale. Une mesure exacte de cette surface est certainement impossible sur l'image satellite. Néanmoins l'obtention d'un coefficient qui permette de l'approcher dans chaque strate, alliée à une technique de tirage des unités de

l'échantillon qui garantisse une probabilité approximativement égale à leur surface bâtie, pourrait apporter aux estimations un gain de précision sensible. Cette question sera abordée, pour la mise au point définitive du plan de sondage à Quito, à partir des données recueillies lors des enquêtes préliminaires. Comme nous le verrons alors, la difficulté réside en fait dans la mise en œuvre pratique de la stratégie de sélection des unités primaires.

3. Les domaines de précision et d'imprécision des estimations dans l'espace urbain, tels qu'ils ont été identifiés lors de la mise au point du test sur le cas simple du plan équiprobable sans remise et de l'estimateur sans biais (voir paragraphes 6.14 & 6.15 du chapitre 6), demeurent identiques quels que soient le plan de sondage et l'estimateur retenus :

- centre ville, grands ensembles et habitat pavillonnaire (strates 1,3 et 5) : domaine de bonne précision où les intervalles de confiance sont compris entre +/- 20% et +/- 40% ;

- les strates 7 et 8, "reliquat morphologique" et "grands bâtiments élevés", malgré leur faibles effectifs, conservent une précision acceptable ;

- le domaine d'imprécision couvre les strates 2, 4, 6 et 9, c'est-à-dire les quartiers industriels, les petits immeubles d'habitation et les zones périphériques peu denses ou non bâties ; ici les intervalles de confiances dépassent +/- 100%, ce qui rend très hasardeux toute estimation. Notons que le résultat catastrophique obtenu dans la strate 9 s'explique en partie par son effectif trop faible : 7 districts ; d'autre part ne perdons pas de vue le fait que ces strates sont peu peuplées dans leur ensemble.

4. Les gains qu'apportent la stratification morphologique puis, l'allocation optimale, sont également toujours du même ordre. Ils sont consignés dans le tableau suivant qui donne la diminution de la variance en pourcentage de cette dernière dans le cas non stratifié.

	Plan équiprob. est. sans biais	Plan équiprob. est. du ratio	Plan à proba. inégale est. sans biais
estimation stratifiée à allocation proport.	19,4 %	31,2	19,7 %
estimation stratifiée à allocation optimale	40,0 %	57,5 %	42,6 %

Soulignons, au vu du tableau, l'équivalence des gains dans les deux cas : équiprobabilité et probabilités inégales. *Ce résultat confirme l'intérêt de la stratification morphologique dans le cadre d'une technique d'échantillonnage d'après document satellitaire, technique qui contraint à utiliser un plan à probabilités inégales.*

Il importe maintenant de vérifier que les résultats obtenus jusqu'à présent s'appliquent non seulement à l'estimation de la population sans doubles comptes mais aussi à d'autres indicateurs démographiques.

Tableau 7.1.a : Résultats du test mené avec le programme VARECH. Estimation du total de population sans doubles comptes : deux plans de sondage, trois estimateurs, neuf strates morphologiques synthétiques, taux de 6,6% appliqué a une base de 2619 districts (R.G.P. Marseille 1982).

Estimations :	Plan équiprob. Estimateur sans biais		Plan équiprob. estimateur par le ratio		Plan à proba. proportion. à la surface Estim. sans biais	
	(*)	(**)	(*)	(**)	(*)	(**)
Strates						
1.Centre ville dense	9.90	10.85	10.42	11.96	9.71	10.79
2.Quartiers industriels	55.71	49.66	70.51	60.43	76.04	58.21
3.Grands ensembles d'habitation	21.06	13.92	24.46	16.07	21.61	14.15
4.Petits immeubles d'habitation	69.30	69.91	82.45	84.86	57.64	63.98
5.Pavillonnaire	16.02	23.17	21.85	29.40	14.25	21.93
6 Peri-urbain peu dense	59.81	107.45	125.34	214.74	113.11	148.27
7.Reliquat morphologique	37.14	32.91	33.38	33.56	31.70	30.51
8.Dense : grands bâtiments élevés	38.04	33.24	69.21	48.49	34.41	31.72
9.Non bâti	370.09	598.17	454.80	912.63	320.98	559.00
Ensemble de la base stratifiée	8.83	7.62	11.35	8.92	9.08	7.67
Ensemble de la base non stratifiée.	9.84		13.67		10.13	

(*) : Allocation proportionnelle

(**) : Allocation optimale

Graphique 7.1.b : Résultats du test mené avec le programme VARECH

**Estimation du total de population sans doubles comptes.
Comparaison graphique des coefficients de variation pour les trois techniques de sondages (allocation optimale).**



_____ : Estimation sans biais, plan équiprobable.
 - - - - - : Estimation sans biais, plan à probabilités inégales.
 : Estimation par le ratio, plan équiprobable.

7.2 : Extension du test à d'autres caractères démographiques

Disposant d'un outil de test qui fournit les précisions associées à différents plans de sondage, estimateurs et stratifications, nous l'appliquons maintenant à divers indicateurs démographiques. Deux questions principales se posent alors :

1. Existe-t-il des types de variables démographiques qui, du fait des caractéristiques de leur distribution (au sens statistique comme au sens spatial), se prêtent mieux que d'autres aux estimations par sondage ?

2. Les choix des probabilités de tirage, des formules d'estimation et des stratifications, qui définissent le sondage, ont-ils les mêmes conséquences sur la précision des estimations, quelles que soient les variables considérées ?

Pour tenter d'y répondre, nous avons soumis au programme le fichier contenant, pour les 2619 districts de la base, les données de 28 indicateurs démographiques, dont la liste figure à la fin du paragraphe 7.2, dans le tableau 7.2. Parmi eux, deux catégories principales sont à distinguer : d'une part les mesures d'effectifs comme le nombre d'individus ou de ménages du district répondant à une certaine caractéristique (15 variables), d'autre part les quotients mesurant l'importance relative au sein du district de catégories données d'individus ou de ménages (14 variables).

Nous nous intéresserons, tout d'abord, aux estimations concernant l'ensemble de la base, successivement, estimations d'effectifs et de ratios, puis, après sélection de 15 caractères parmi les 28 de départ, aux estimations partielles dans chacune des strates. Enfin, en conclusion, nous tenterons d'apporter des éléments de réponse aux deux questions formulées ci-dessus.

7.2.1 : Résultats sur l'ensemble de la base de sondage, estimateurs et caractères démographiques

1- Estimation d'effectifs

Le tableau 7.3.a donne, pour les 15 variables retenues, les coefficients de variation ($C = \text{écart-type}/\text{quantité à estimer}$, en %) des trois estimateurs non stratifiés et stratifiés de l'effectif total sur l'ensemble de la base (dans le cas stratifié, c'est l'allocation optimale qui est choisie). La dernière colonne mesure le gain de précision dû à la stratification par le pourcentage de variance qu'elle permet d'éliminer. Enfin le graphique 7.3.b fournit une traduction visuelle des précisions obtenues avec chacun des trois sondages, dans le cas de l'allocation optimale. De l'ensemble de ces résultats se dégagent quatre conclusions :

1. La précision des estimations est du même ordre pour la plupart des variables considérées, avec des coefficients de variation de 7 à 8% lorsque l'on choisit le meilleur estimateur et que l'allocation aux strates est optimale. Il n'y a d'ailleurs rien d'étonnant à cela puisque les variables considérées sont, pour la plupart d'entre elles, fortement corrélées. Notons tout de même que, lorsqu'il s'agit d'estimer les totaux de catégories nettement minoritaires, la précision se dégrade assez sensiblement. C'est le cas pour la population de nationalité étrangère, celle des hommes de plus de 19 ans au chômage ou encore pour le nombre de ménages de plus de six personnes et, dans une moindre mesure, pour la population des moins de 5 ans. Les coefficients de variation les meilleurs sont alors compris entre 11 et 21%.

2. Comme le montre le graphique 7.3.b, en général, les deux estimateurs sans biais ont une précision équivalente alors que l'estimation de ces totaux par le ratio faisant intervenir la surface des districts est moins précise. Ceci confirme l'idée déjà

exprimée selon laquelle la surface totale des districts ne constitue pas une mesure efficace de la taille des unités lorsqu'on s'intéresse à l'estimation de caractères démographiques.

3. On constate également sur le graphique 7.3.b que *lorsque deux catégories de population sont complémentaires au sein d'une population totale (populations française et étrangère par exemple), les estimations concernant la catégorie la plus nombreuse sont généralement les plus précises*. L'exception est ici la population des plus de 60 ans, dont l'estimation est plus précise que celle des moins de 60 ans. Nous verrons plus loin que, dans le cas des quotients, ce phénomène devient systématique. Mais il trouve alors une explication fort simple et nous aboutirons en fait à la conclusion que les deux estimations sont équivalentes.

4. Enfin, l'examen de la dernière colonne du tableau 7.3 a permis de *chiffrer à environ 40 % la diminution moyenne de la variance due à l'introduction de la stratification morphologique*. Ce gain peut atteindre 50 à 55 % dans le cas des caractères dont la variance inter-strate est importante : population étrangère, hommes de plus de 19 ans au chômage, nombre de ménages d'une personne. A l'inverse, il n'est que de 20 % lors de l'estimation de la population âgée de 60 ans et plus.

Tableau 7.3.a : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 15 caractères démographiques (effectifs). Mesure du gain dû à la stratification.

Estimations :	Plan équiprob. Estimateur sans biais		Plan équiprob. estimateur par le ratio		Plan à proba. proportion. à la surface Estim. sans biais ou par le ratio		Gain de Précision dû à la stratif. (***)
	(*)	(**)	(*)	(**)	(*)	(**)	
Strates							
Population totale	9.83	7.62	13.67	9.95	10.15	7.67	42.9
Population totale des ménages ordinaires	9.88	7.64	13.77	9.02	10.25	7.76	42.7
Population de nationalité étrangère	30.81	21.39	32.62	22.32	37.83	26.58	50.6
Population de nationalité française	9.59	7.59	13.39	8.78	9.43	7.18	42.0
Population des hommes de plus de 19 ans au chômage	14.92	12.34	18.25	14.71	22.04	14.64	55.9
Population des hommes de plus de 19 ans ayant un emploi	9.19	7.61	13.56	9.01	10.87	8.03	45.4
Population âgée de moins de cinq ans	16.27	10.73	18.12	11.68	14.76	11.67	37.5

(*) : Estimation non stratifiée.

(**) : Estimation stratifiée, allocation aux strates optimale.

(***) : Gain de précision dû à la stratification, en % de la variance non stratifiée (Plan à probabilités proportionnelles aux surfaces, estimateurs sans biais).

Tableau 7.3.a (suite) : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 15 caractères démographiques (effectifs). Mesure du gain dû à la stratification.

Estimations :	Plan équiprob. Estimateur sans biais		Plan équiprob. estimateur par le ratio		Plan à proba. proportion. à la surface Estim. sans biais ou par le ratio		Gain de Précision dû à la stratif. (***)
	(*)	(**)	(*)	(**)	(*)	(**)	
Strates							
Population âgée de cinq ans et plus	9.66	7.56	13.60	8.87	10.07	7.59	43.3
Population âgée de soixante ans et plus	8.57	7.45	14.12	9.07	8.75	7.76	21.3
Population âgée de moins de soixante ans	11.02	8.15	14.15	9.30	10.66	8.21	40.7
Nombre de ménages ordinaires	8.70	7.20	13.66	8.74	10.30	7.43	48.0
Nombre de ménages d'une personne	9.44	7.98	15.21	9.51	12.81	8.70	53.9
Nombre de ménages de deux personnes ou plus	9.26	7.43	13.50	8.89	9.97	7.53	43.0
Nombre de ménages de six personnes ou plus	29.98	17.59	30.40	18.11	24.43	18.88	40.5
Nombre de ménages de moins de six personnes	8.65	7.22	13.73	8.78	10.36	7.46	48.1

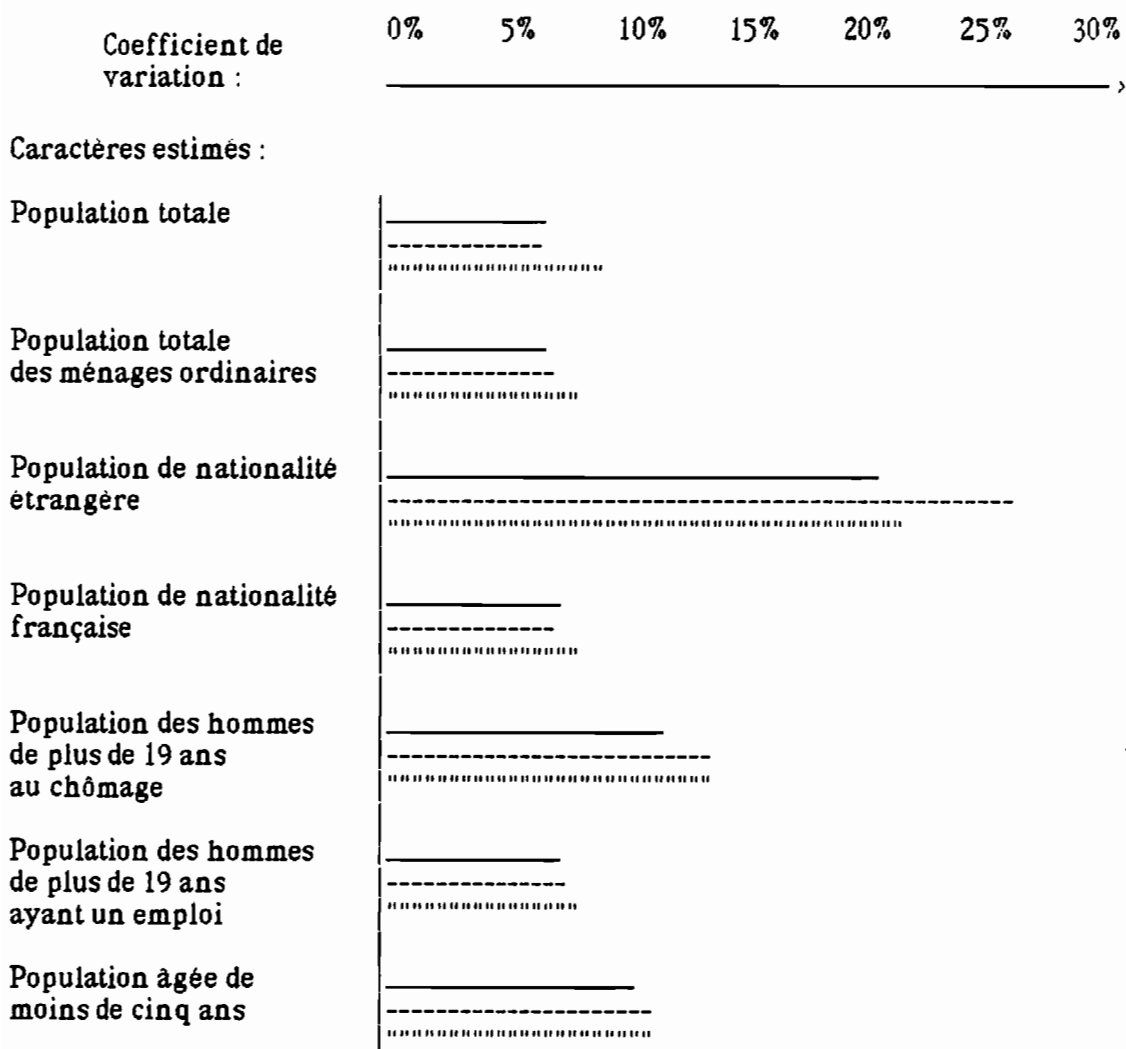
(*) : Estimation non stratifiée.

(**) : Estimation stratifiée, allocation aux strates optimale.

(***) : Gain de précision dû à la stratification, en % de la variance non stratifiée (Plan à probabilités proportionnelles aux surfaces, estimateurs sans biais).

Graphique 7.3.b : Résultats du test mené avec le programme VARECH

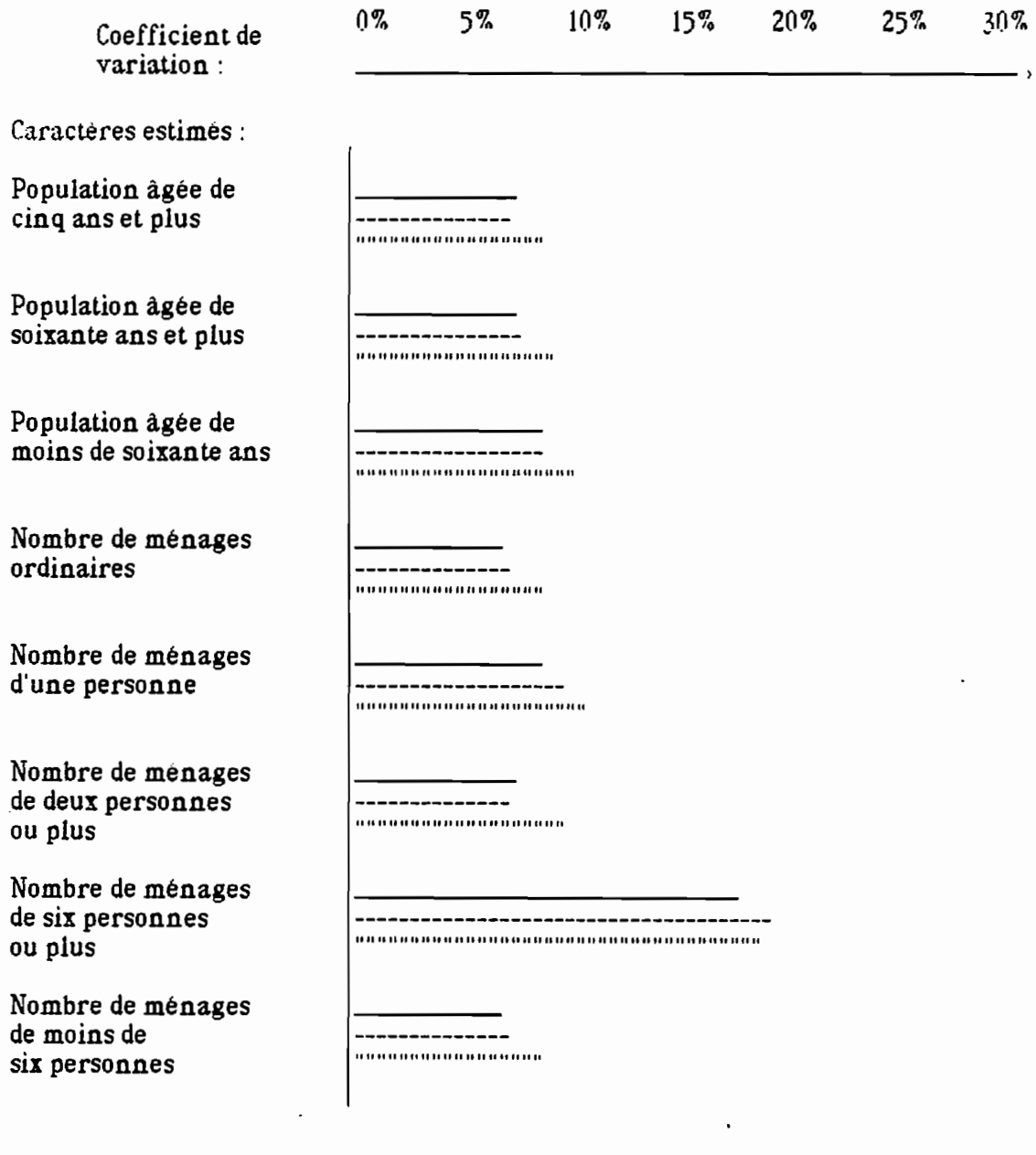
Comparaison, pour trois sondages, des coefficients de variation des estimateurs stratifiés de 15 mesures d'effectifs. Allocation optimale aux strates.



————— : Estimation sans biais, plan équiprobable.
 - - - - - : Estimation sans biais, plan à probabilités inégales.
 : Estimation par le ratio, plan équiprobable.

Graphique 7.3.b (suite) : Résultats du test mené avec le programme VARECH

Comparaison, pour trois sondages, des coefficients de variation des estimateurs stratifiés de 15 mesures d'effectifs. Allocation optimale aux strates.



————— : Estimation sans biais, plan équiprobable.
 - - - - - : Estimation sans biais, plan à probabilités inégales.
 : Estimation par le ratio, plan équiprobable.

2. Estimations de ratios

Si l'estimateur par le ratio n'a pas, comme on l'a montré, amélioré les estimations d'effectifs, son intégration dans le programme conserve néanmoins un grand intérêt puisqu'il permet d'estimer directement les valeurs des quotients décrivant les structures de la population. Le tableau 7.4 présente les coefficients de variation de ces estimations et une mesure du gain de précision obtenu grâce à la stratification pour 13 de ces quotients. Son examen conduit à quatre remarques :

1. Ce qui frappe tout d'abord, c'est que dans l'ensemble, les coefficients de variation sont nettement inférieurs à ceux qu'on obtenait lors de l'estimation d'effectifs : sept des quotients sont estimés avec des coefficients inférieurs à 2,5 % et on atteint même, dans le cas du quotient population de 5 ans et plus / population totale, 0,29 %. Ce résultat n'est pas surprenant : ces ratios ont, par nature, une variabilité bien moindre que celle des effectifs. Nous verrons d'ailleurs plus loin que, *même avec des précisions de cet ordre, le sondage ne fournit qu'une image floue des différences qui existent, à l'intérieur de l'agglomération, entre les strates morphologiques.* Il apparaît donc que la mesure de la précision des estimations par le calcul de la variance de l'estimateur est très relative : *comparer entre elles les variances d'estimation de différents caractères démographiques n'a pas grand sens. Ce qui importe surtout, c'est de vérifier qu'une modification donnée du sondage a bien les mêmes conséquences sur l'ensemble des estimations auxquelles ont s'intéresse.*

2. De nouveau, et systématiquement cette fois, *si le numérateur est une partie minoritaire de la population décrite au dénominateur, l'estimation du quotient complémentaire a un coefficient de variation moindre.*

L'explication du phénomène est, en fait, triviale : la somme des deux quotients que l'on cherche à estimer étant constante (elle vaut 1 ou 100 si l'on transforme en pourcentage), on montre immédiatement que les variances des deux estimations sont égales. Il en est évidemment de même pour les écarts-type et les intervalles de confiance, en revanche le coefficient de variation, proportionnel à l'écart-type et inversement proportionnel à la quantité à estimer, sera bien sûr d'autant plus grand que le ratio lui-même est petit. Ainsi par exemple, le quotient "population de nationalité française / population totale", est estimé avec un coefficient de variation plus faible que le quotient "population étrangère / population totale".

Le coefficient de variation, rappelons-le, a été choisi pour mesurer la précision des estimations parce qu'il permet d'évaluer l'effet d'une modification du plan de sondage : changement de probabilité d'inclusion, d'estimateur ou de stratification lorsque l'on s'intéresse à une même variable. En revanche, on a déjà constaté qu'il n'est pas un bon indicateur pour comparer, dans un même cadre de plan de sondage, les précisions d'estimation de plusieurs caractères. *Il devient même incohérent lorsque l'on s'intéresse comme ici à l'estimation de plusieurs quantités linéairement dépendantes : les variances d'estimation sont alors égales tandis que les coefficients de variation sont eux extrêmement variables.*

Il faut donc, dans le cas des quotients complémentaires s'en tenir à l'observation des variances des estimateurs et l'on conclut qu'il est toujours équivalent d'estimer l'une ou l'autre des quantités.

Pour les habitués des techniques de sondages, ces deux premières remarques vont de soi, mais étant donné leur importance pratique, nous avons considéré que leur rappel était utile.

3. Comme on l'a déjà constaté lors de l'estimation de la population sans doubles comptes et pour l'ensemble des estimations d'effectifs, les deux plans de sondages, équiprobable et à probabilités proportionnelles aux surfaces, conduisent à des précisions à peu près équivalentes pour l'estimation des quotients. Il est même parfois avantageux d'utiliser le plan de sondage à probabilités inégales dans le cas non stratifié.

Tableau 7.4 : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 13 caractères démographiques (ratios). Mesure du gain dû à la stratification.

Quotients estimés :	Estimation non stratifiée		Estimation stratifiée allocation optimale		Gain de précision en % de la var. non stratifiée	
	(*)	(**)	(*)	(**)	(*)	(**)
Population de nationalité étrangère/ Population totale	27.02	32,91	19.02	22,92	50.0	51,5
Population de nationalité française/ Population totale	3.01	3,67	2.12	2,55	50.4	51,7
Population des hommes de plus de 19 ans au chômage/ Population des hommes de plus de 19 ans	9.89	14,48	8.31	9,60	29.4	56,0
Population des hommes de plus de 19 ans ayant un emploi/ Population des hommes de plus de 19 ans	0.99	1,45	0.82	0,96	31.4	56,2
Population âgée de moins de cinq ans/ Population totale	9.71	8,90	6.06	6,93	61.1	39,4
Population âgée de cinq ans ou plus/ Population totale	0.47	0,43	0.29	0,33	61.9	41,1
Population âgée de soixante ans ou plus/ Population totale	6.33	5,92	4.10	4,56	58.0	40,7

(*) : Plan équiprobable

(**) : Plan à probabilités proportionnelles aux surfaces des unités primaires (districts de recensement INSEE)

Tableau 7.4 (suite) : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 13 caractères démographiques (ratios). Mesure du gain dû à la stratification.

Quotients estimés :	Estimation non stratifiée		Estimation stratifiée allocation optimale		Gain de précision en % de la var. non stratifiée	
	(*)	(**)	(*)	(**)	(*)	(**)
Population âgée de moins de soixante ans/ Population totale	2.06	1,93	1.34	1,49	57.7	40,4
Population des ménages ordinaires/ Nombre de ménages ordinaires	3.72	3,23	2.28	2,50	62.3	40,1
Nombre de ménages d'une personne/ Nombre de ménages ordinaires	4.90	5,44	3.59	4,12	46.3	42,6
Nombre de ménages de plus d'une personne/ Nombre de ménages ordinaires	2.28	2,53	1.67	1,92	46.4	42,4
Nombre de ménages de six personnes ou plus/ Nombre de ménages ordinaires	27,76	22,64	16,09	17,82	66,4	38,0
Nombre de ménages de moins de six personnes/ Nombre de ménages ordinaires	0.86	0,70	0.50	0,55	66.2	38,3

(*) : Plan équiprobable

(**) : Plan à probabilités proportionnelles aux surfaces des unités primaires (districts de recensement INSEE)

En revanche lorsque l'on stratifie la base de sondage et que l'allocation aux strates est optimale, l'avantage revient au plan équiprobable.

4. Que la sélection soit équiprobable ou qu'elle s'effectue avec probabilité proportionnelle à la surface, la diminution de variance due à la stratification est du même ordre que celle obtenue lors des estimations d'effectifs (parfois même légèrement supérieure) : entre 30 et 60 % de la variance non stratifiée.

7.2.2 : Résultats par strate : variabilité des caractères démographiques et précision des estimations partielles

Les conclusions de l'examen des résultats sur l'ensemble de la base permettent, en écartant les variables décrivant des catégories très minoritaires au profit de leurs complémentaires, de ne retenir, pour l'examen plus détaillé des résultats par strate, que huit effectifs et six ratios, qui constituent un échantillon restreint mais significatif de caractères démographiques. A travers cet échantillon, outre la mesure de l'effectif total de population (au sens de la population totale¹), quatre thèmes seront abordés.

- **La population étrangère** : effectif et ratio population française / population totale.

- **Le chômage** : effectif de la population des hommes de plus de 19 ans ayant un emploi et ratio par rapport à la population totale.

- **La structure par âge de la population totale** : effectifs des populations de 5 ans et plus et de moins de 60 ans, ainsi que les ratios correspondant par rapport à la population totale.

- **La taille des ménages** : nombre de ménages ordinaires, nombre de ménages de plus d'une personne et nombre de ménages de moins de six personnes, taille moyenne des ménages, c'est-à-dire le quotient population des ménages ordinaires / nombre de ménages ordinaires et enfin les deux ratios donnant la proportion de ménages de plus d'une personne et de moins de six personnes.

Le tableau 7.5 donne, pour chaque caractère démographique, le coefficient de variation sur l'ensemble de la base et dans chacune des neuf strates morphologiques (on a choisi ici le plan de sondage à probabilités proportionnelles à la surface des districts de recensement). L'examen des chiffres conduit aux conclusions suivantes :

1. Les domaines de précision (strates 1,3,5,7 et 8) et d'imprécision (strates 2, 4, 6 et 9), sont les mêmes quel que soit le caractère considéré.

2. L'ensemble des effectifs soumis au test, sont estimés avec des précisions très voisines tant pour la totalité de la base de sondage qu'au niveau de chaque strate. Par contre, **en ce qui concerne les ratios, la précision varie beaucoup plus d'un caractère à l'autre.** Ce phénomène trouve son explication dans les différences que présentent les distributions des caractères : une différence de variance entre deux variables se répercute évidemment dans la variance de leurs estimations. Et on observe en effet, une plus grande diversité de distributions au sein de la famille des variables de type ratios que pour celle des effectifs.

¹ : Voir la définition retenue par l'INSEE dans l'annexe A.2.2.

Tableau 7.5 : Résultats du test mené avec le programme VARECH. Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques (première partie, estimation sans biais d'effectifs).

Caractères : Strates	Pop. totale	Pop. de nat. française	Pop. des hommes de + de 19 ans ayant un emploi	Pop. de plus de cinq ans	Pop. de moins de 60 ans	Nb de ménages ordinaires	Nb de ménages de plus d'une pers.	Nb de ménages de moins de 6 pers.
1. Centre ville	10.80	9.73	11.70	10.66	12.26	9.46	10.48	9.33
2. Quartiers industriels	58.05	55.78	58.79	57.68	59.14	59.23	57.45	59.93
3. Grands ensembles	14.17	13.50	15.16	14.10	14.17	15.46	14.24	15.99
4. Petits immeubles	63.85	59.92	67.72	63.60	65.68	69.32	62.50	69.65
5. Zones pavillonnaires	21.61	20.15	21.52	21.33	23.20	22.50	20.84	22.51
6. Peri-urbain peu dense	149.05	138.78	152.23	149.22	155.79	156.01	141.40	156.98
7. Reliquat morphologique	30.53	28.71	30.55	30.14	32.61	30.53	30.37	30.36
8. Dense, grands bâtiments	31.68	29.37	32.63	31.31	34.74	29.31	30.69	29.02
9. Non bâti	561.80	518.17	582.91	550.01	527.03	639.32	576.92	641.42
Total (non stratifié)	10.15	9.43	10.87	10.07	10.66	10.30	9.97	10.36
Total (stratifié, alloc. optim.)	7.67	7.18	8.03	7.59	8.21	7.43	7.53	7.46

Tableau 7.5 : Résultats du test mené avec le programme VARECH. Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (seconde partie, estimation de ratios)

Quotients :	Pop. Français	Pop. ayant un emp.	Pop. + 5 ans	Pop. de moins de 60 ans	Pop. des mén. ord.	Nb. de mén. de + de 1 p.	Nb de mén. de + de 1 p.
	Pop. totale	Pop. des hommes	Pop. totale	Pop. totale	Nb des mén. ord.	Nb. de mén. ordinaires	Nb de mén. ordinaires
1. Centre ville	4,66	1,70	0,47	2,40	3,37	2,93	0,66
2. Quartiers industriels	14,26	5,13	2,40	10,61	18,53	13,67	4,25
3. Grands ensembles	5,02	1,68	0,70	2,28	5,23	3,23	1,54
4. Petits immeubles	13,11	6,36	2,40	11,47	15,78	13,14	3,83
5. Zones pavillonnaires	2,98	1,94	0,79	4,27	5,61	5,14	1,04
6. Péri-urbain peu dense	25,38	18,13	6,41	29,61	57,46	50,14	10,79
7. Reliquat morphologique	4,41	2,59	1,01	6,52	9,17	9,08	1,02
8. Dense, grands bâtiments	6,83	2,51	0,92	6,53	6,76	6,79	1,25
9. Non bâti	577,04	252,23	80,72	305,91	627,82	521,77	187,44
Total (non stratifié)	3,67	1,45	0,43	1,93	3,23	2,53	0,70
Total (stratifié, alloc. optim.)	2,55	0,96	0,33	1,49	2,50	1,92	0,55

3. Le tableau 7.6, reprend les résultats du test pour trois caractères démographiques en donnant les quantités à estimer dans chaque strate et les intervalles de confiance des estimations correspondantes. En l'examinant, on prend conscience que les mesures de précision fournies par les coefficients de variation sont très relatives. De manière générale, les degrés de précision atteints sont encore insuffisants pour obtenir des images nettes des variations démographiques à l'intérieur de l'agglomération. De plus, dans le cas des ratios caractérisant la structure de la population, des coefficients de variation de l'ordre de 2 ou 3%, ou même inférieurs à 1%, ne doivent pas faire illusion : il s'agit en effet de caractères dont les variations d'une strate à l'autre se situent précisément dans cet ordre de grandeur (2 ou 3%, parfois moins, autour de la moyenne générale). Les intervalles de confiance associés aux estimations dans chaque strate se chevauchent alors beaucoup trop pour que le sondage à ce taux renseigne de façon fiable sur les variations inter-strates de ces indicateurs.

7.2.3 : Premières conclusions

Au terme des comparaisons entre variables démographiques que l'on vient d'effectuer, on peut répondre aux deux questions formulées au début du paragraphe 7.2. Rappelons-les tout d'abord.

Questions :

1. Existe-t-il des types de variables démographiques qui, du fait des caractéristiques de leur distribution (au sens statistique comme au sens spatial), se prêtent mieux que d'autres aux estimations par sondage ?
2. Les choix des probabilités de tirage, des formules d'estimation et des stratifications, qui définissent le sondage, ont-ils les mêmes conséquences sur la précision des estimations, quelles que soient les variables considérées ?

Réponses :

1. L'estimation des variables de type ratio se fait dans l'ensemble avec des coefficients de variation sensiblement plus faibles que ceux correspondant aux effectifs. Ce résultat ne doit pas surprendre puisque, par nature, les premières ont des variances nettement inférieures aux secondes. C'est également ce qui explique que ces coefficients de variation plus faibles ne permettent pas pour autant de saisir de manière précise les variations des ratio au sein de l'agglomération marseillaise. *Qu'il s'agisse d'estimation d'effectifs ou de quotients, il reste donc beaucoup à faire pour atteindre une précision satisfaisante de ce type de sondage.*

2. *Les conclusions établies lors de l'examen des résultats de l'estimation de l'effectif de population sans doubles comptes, qui concernent le plan de sondage et la formule d'estimation, demeurent valables pour l'ensemble des caractères démographiques considérés par la suite. Il en est de même pour l'ordre de grandeur du gain apporté par la stratification ainsi que pour la définition des domaines de "bonne ou de mauvaise précision". La démarche d'amélioration du sondage présentée jusqu'ici semble donc bien robuste lorsqu'on change la variable à estimer. Ce résultat important ne peut que nous encourager à la perfectionner.*

À ce propos, et afin de relativiser l'impression négative que peut produire l'examen des intervalles de confiance rapportés dans le tableau 7.6., dont les amplitudes restent élevées, soulignons que la technique de sondage à laquelle ils correspondent est encore très frustrante puisqu'elle ne comprend qu'un seul degré. On perd évidemment

beaucoup à enquêter exhaustivement les districts de l'échantillon, mais l'évaluation du gain possible grâce au deuxième degré de sondage ne sera possible qu'à partir des données recueillies à Quito. Comme on l'a déjà dit, à Marseille, ce qui nous intéresse, plus que l'ordre de grandeur des précisions obtenues, c'est l'évaluation des possibilités d'amélioration de la technique de sondage au premier degré grâce aux choix d'estimateurs, de probabilités d'inclusion et de stratifications qui peuvent être fait. De ce point de vue, ces premiers résultats sont encourageants.

Tableau 7.6 : Résultats du test mené avec le programme VARECH.

Quantités à estimer et intervalles de confiance à 95% par strate pour trois caractères démographiques.

Caractère estimé:	POPULATION TOTALE		RATIO POP. DE PLUS DE CINQ ANS / POPULATION TOTALE		RATIO POP. FRANCAISE / POPULATION TOTALE	
	Total	Intervalle de confiance à 95 % (*)	Valeur en %	Intervalle de confiance à 95 % (**)	Valeur en %	Intervalle de confiance à 95 % (**)
Strates						
1. Centre ville	181508	[142302;220714]	96.03	[95,12;96,93]	86.67	[78,59;94,75]
2. Quartiers industriels	20830	[0;45014]	94.78	[90,28;99,28]	89.42	[63,92;100]
3. Grands ensembles	142480	[102101;182859]	93.91	[92,60;95,21]	89.36	[80,40;98,32]
4. Petits immeubles	8581	[0;19539]	95.41	[90,83;99,99]	94.70	[69,88;100]
5. Zones pavillonnaires	54298	[30830;77766]	96.49	[94,97;98,01]	96.50	[90,76;100]
6. Péri urbain peu dense	2287	[0;9105]	95.10	[83,98;100]	96.24	[71,81;100]
7. Reliquat morphologique	26679	[10389;42969]	96.50	[97,55;98,46]	96.50	[87,99;100]
8. Dense, grands bâtim.	18282	[6699;29865]	96.50	[94,72;98,28]	96.22	[83,08;100]
9 Non bâti	183	[0;2239]	98.36	[0;100]	98.91	[0;100]
Total non stratifié	455128	[356820;553436]	95.39	[94,57;96,21]	89.97	[83,37;96,57]
Total stratifié alloc. optim.	455128	[385311;524945]	95.39	[94,75;96,03]	89.97	[85,37;94,57]

(*) : Plan de sondage à probabilités proportionnelles à la surface des districts, estimation sans biais

(**) : Plan de sondage à probabilités proportionnelles à la surface des districts, estimation de ratios.

Tableau 7.2 : liste des caractères démographiques ayant fait l'objet du test.

EFFECTIFS	QUOTIENTS
1. Population totale	1. Population de nationalité étrangère / Population totale
2. Population des ménages ordinaires	2. Population de nationalité française / Population totale
3. Population de nationalité étrangère	3. Population des hommes de plus de 19 ans au chômage / Population des hommes de plus de 19 ans
4. Population de nationalité française	4. Population des hommes de plus de 19 ans ayant un emploi / Population des hommes de plus de 19 ans
5. Population des hommes de plus de 19 ans au chômage	5. Population âgée de moins de cinq ans / Population totale
6. Population des hommes de plus de 19 ans ayant un emploi	6. Population âgée de cinq ans ou plus / Population totale
7. Population âgée de moins de cinq ans	7. Population âgée de soixante ans ou plus / Population totale
8. Population âgée de cinq ans ou plus	8. Population âgée de moins de soixante ans / Population totale
9. Population âgée de soixante ans ou plus	9. Population des ménages ordinaires / Nombre de ménages ordinaires
10. Population âgée de moins de soixante ans	10. Nombre de ménage (ordinaires) d'une personne / Nombre de ménages ordinaires
11. Nombre de ménages ordinaires	11. Nombre de ménages (ordinaires) de plus d'une personne / Nombre de ménages ordinaires
12. Nombre de ménages (ordinaires) d'une personne	12. Nombre de ménages (ordinaires) de six personnes ou plus / Nombre de ménages ordinaires
13. Nombre de ménages (ordinaires) de plus d'une personne	13. Nombre de ménages (ordinaires) de moins de six personnes / Nombre de ménages ordinaires
14. Nombre de ménages (ordinaires) de six personnes et plus	
15. Nombre de ménages (ordinaires) de moins de six personnes	

7.3 : Stratifications construites sur le critère de la densité du bâti issu de la photo-interprétation, puis de l'image satellite

7.3.1 : Préliminaire

Nous venons de mesurer l'intérêt que peut avoir, pour améliorer la précision du sondage, la stratification morphologique de la base. Mais il ne faut pas perdre de vue que la stratification testée jusqu'à présent représente une synthèse de l'information morphologique issue des photographies aériennes et contrôlée par des vérifications au sol. Elle ne peut donc pas servir de base à la mesure qui nous intéresse réellement : le gain dû à une stratification construite d'après l'image satellite. C'est précisément cette mesure que nous abordons dans cette fin de chapitre.

Suivant en cela le travail fourni par l'équipe de télédétection, nous voudrions au moins amorcer ici une démarche analytique dont l'esprit serait le suivant : tester une à une les stratifications fondées sur les différents descripteurs de la morphologie urbaine jugés pertinents pour une observation démographique. Le premier, et jusqu'à présent l'unique descripteur morphologique que l'on peut extraire de l'information satellitaire avec fiabilité est la mesure de la densité du bâti ; c'est donc aux stratifications basées sur la seule densité du bâti que nous allons nous intéresser maintenant, stratifications construites d'abord en partant de l'information issue de la photo-interprétation, puis à partir de l'image satellite.

En premier lieu nous exposons rapidement la démarche qui a conduit aux stratifications étudiées, puis nous commentons les résultats de l'évaluation de ces stratifications, enfin nous évoquons la possibilité d'une suite dans cette recherche qui, en étendant le champ de l'étude à d'autres descripteurs morphologiques, pourrait conduire à une stratification morphologique plus efficace, issue de l'image satellite.

7.3.2 : Provenance et élaboration des stratifications

Dans un rapport remis en avril 87¹, Alain MICHEL résume ainsi le travail qu'il a effectué :

" Le descripteur densité du bâti, (...) caractérise l'ensemble de la zone à l'aide des six modalités suivantes : NON BATI, TRES FAIBLE, FAIBLE, MOYENNE, FORTE, TRES FORTE. Le but de ce travail est de générer, à partir de l'image Landsat T.M. six canaux, une image qui rende compte de la densité du bâti observée sur mosaïque de photographies aériennes et vérifiée sur le terrain. De façon plus pragmatique, il s'agit donc de classer les images Landsat T.M. afin de se rapprocher le plus possible de l'interprétation visuelle de la densité du bâti réalisée sur la mosaïque."

La démarche adoptée est donc la suivante :

1. Identification d'un descripteur morphologique et d'une stratification de référence ; ici, il s'agit de la stratification en six classes définie par le descripteur photo-interprété : "densité du bâti".
2. Recherche d'une classification de l'image satellite la plus proche possible de la stratification de référence.

¹ : Alain MICHEL, Télédétection et observation suivie de la morphologie et de la démographie des villes des pays en développement. Mise au point des méthodes d'analyse des images satellite à haute résolution et évaluation des informations fournies par ces images. Etude statistique du descripteur "densité du bâti", présentation des classifications, 99 p., Paris, 1987.

De notre côté, nous allons suivre la même démarche pour évaluer le gain de précision attaché à chaque stratification :

1. Mesure du gain apporté par la stratification de référence (six classes de densité de bâti définies lors de la photo-interprétation).

2. Mesure du gain dû à la stratification correspondante de l'image satellite.

Pratiquement, se pose alors le problème de la réaffectation à la base de sondage (ici, les 2619 districts INSEE), de stratifications définies sur d'autres unités spatiales : les zones photo-interprétées pour la première, les pixels de l'image satellite pour la seconde. La première réaffectation est facile puisque la base de sondage actuelle est constituée de districts entièrement contenus dans une même zone morphologique. Elle fournit donc une stratification de la base de sondage en six classes :

- 1 : NON BATI
- 2 : TRES FAIBLE
- 3 : FAIBLE
- 4 : MOYENNE
- 5 : FORTE
- 6 : TRES FORTE.

C'est le système SAVANE qui permet de mener à bien la seconde réaffectation en mettant en correspondance l'image issue de la classification des pixels avec celle de la base de sondage. C'est de cette stratification qu'il sera question dans le paragraphe 7.4.4. ; mais il ressort également du travail d'Alain MICHEL (rapport d'avril 87, P. 72), qu'elle peut être approchée par la stratification que l'on obtient en regroupant deux à deux les modalités du descripteur photo-interprété pour obtenir trois classes :

- 1 : NON BATI et TRES FAIBLE
- 2 : FAIBLE et MOYENNE
- 3 : FORTE et TRES FORTE.

Pour pouvoir ensuite situer la stratification issue de l'image satellite par rapport aux stratifications sur le même thème élaborées d'après l'information morphologique de référence, nous nous intéressons maintenant à l'évaluation de ces deux stratifications.

7.3.3 : Evaluation des stratifications issues de la photo-interprétation

Pour cette évaluation, nous avons retenu 11 caractères démographiques, choisis dans l'ensemble de ceux qui ont servi aux tests des estimateurs et des plans de sondage : six effectifs et cinq ratios dont la liste est donnée dans le tableau 7.7 , en fin de chapitre. Conformément aux conclusions de l'étape précédente¹, ce test a été mené avec un plan de sondage à probabilités proportionnelles à la surface des districts ; l'estimateur est sans biais pour les totaux et par le ratio pour les quotients.

Le tableau 7.8.a résume les résultats : pour chacun des caractères étudiés, il donne le coefficient de variation de l'estimation non stratifiée (col.1), puis ceux obtenus avec trois stratifications distinctes :

- la stratification morphologique synthétique en neuf classes utilisée jusqu'à présent (col.2), qui sert à apprécier le rôle joué par le descripteur "densité du bâti" dans la stratification synthétique.
- la stratification en six classes selon la densité du bâti qu'a permis d'obtenir la photo-interprétation (col.3)

¹ : voir paragraphes 7.1 et 7.2.

- la stratification en trois classes de densité de bâti, obtenue par regroupement, et considérée ici comme une approximation de celle qu'on pourra tirer de l'image satellite (col.4).

Dans le même tableau, on trouve ensuite le gain de précision apporté par chacune des stratifications par rapport à l'estimation non stratifiée (col. 5, 6 et 7). Ce gain est mesuré par la différence entre les écarts-type avant et après stratification, rapportée à l'écart-type de l'estimation non-stratifiée et exprimé en %. Il s'agit donc d'une mesure du gain obtenu sur le coefficient de variation (ou si l'on veut sur l'amplitude des intervalles de confiance à 95%), et non sur les variances des estimations, comme nous le pratiquons jusqu'à présent. Enfin, la figure 7.8.b (fin de paragraphe) donne une représentation graphique de ces gains.

Les conclusions qui suivent sont relatives à deux comparaisons distinctes effectuées sur l'ensemble des caractères :

- La stratification en trois classes de densité du bâti est confrontée à sa référence, la classification en six classes issue de la photo-interprétation, afin d'évaluer dans quelle fourchette de précision doit se situer une classification de l'image satellite sur ce thème pour approcher de manière satisfaisante une "réalité terrain" (il s'agit ici de l'information tirée de la photo-interprétation), dont on sait que la prise en compte améliore le sondage.

- Les stratifications basées sur la seule densité du bâti sont comparées à la stratification morphologique synthétique afin de mesurer l'apport spécifique de ce descripteur à une stratification synthétique (qui devrait par la suite être construite à partir de l'information satellitaire).

Tableau 7.8.a : coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.

Caractères estimés	Coef. de variation des estimations				Gain de précision(en % du coef. de variation sur l'estimation non stratifiée)		
	(0)	(1)	(2)	(3)	(1)	(2)	(3)
A) Effectifs							
1 POPSDC	10.13	7.67	8.59	8.77	24.28	15.20	13.43
2 POPFRANC	9.43	7.18	8.13	8.22	23.86	13.79	12.83
3 POPHEMP	10.87	8.03	8.52	8.93	26.13	21.62	17.85
4 POPMEN	10.25	7.76	8.71	8.89	24.29	15.02	13.27
5 NBMEN-6	10.36	7.46	8.08	8.24	27.99	22.01	20.46
6 POP+5	10.07	7.59	8.48	8.66	24.63	15.79	14.00
B) Ratios							
7 $\frac{\text{POPFRANC}}{\text{POPTOTAL}}$	3.67	2.55	2.87	3.16	30.52	21.80	13.90
8 $\frac{\text{POPHEMP}}{\text{POPH+19}}$	1.45	0.96	0.94	1.10	33.79	35.17	24.14
9 $\frac{\text{POPMEN}}{\text{NBMENAG}}$	3.23	2.50	2.79	2.83	22.60	13.62	12.38
10 $\frac{\text{NBMEN-6}}{\text{NBMENAG}}$	0.70	0.55	0.65	0.66	21.43	7.14	5.71
11 $\frac{\text{POP+5}}{\text{POPTOTAL}}$	0.43	0.33	0.39	0.39	23.26	9.30	9.30

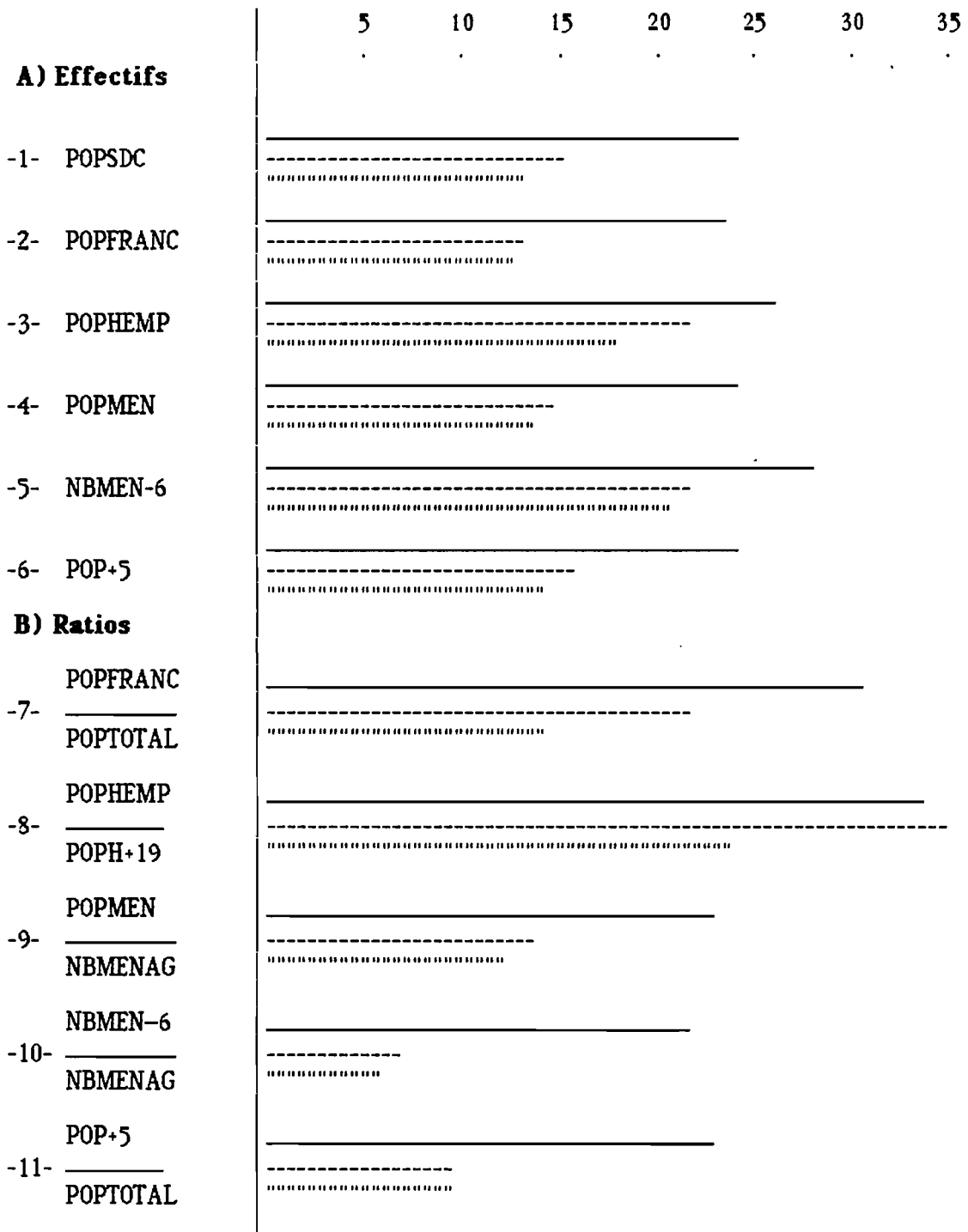
(0) : Sans stratification

(1) : Stratification morphologique synthétique en 9 classes

(2) : Stratification sur la densité du bâti en 6 classes

(3) : Stratification sur la densité du bâti en 3 classes

Graphique 7.8.b : Comparaison des gains de précision apportés par trois stratifications morphologiques à l'estimation de 11 caractères démographiques (en % du coefficient de variation non stratifiée)



Stratification synthétique morphologique (9 classes) : _____
 Stratification sur la densité du bâti (6 classes) : - - - - -
 Stratification sur la densité du bâti (3 classes) :

7.3.3.1. Comparaison des stratifications selon la densité du bâti en trois et six classes.

Le tableau suivant donne, pour chaque caractères, le gain de précision dû à la stratification en trois classes (G1) en pourcentage du gain dû à celle en six classes (G2).

Caractères :	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-
G1/G2 en % :	88,4	93,0	82,6	88,3	93,0	88,7	63,8	68,7	90,9	80,0	100

Hormis pour les deux ratios relatifs à la nationalité (population française/population totale) et à l'emploi (population des hommes de plus de 19 ans ayant un emploi/population des hommes de plus de 19 ans), **le gain apporté par la stratification en trois classes est très proche de celui induit par celle qui a servi de référence au travail sur l'image satellite**. Sous réserve que la stratification obtenue par classification des pixels de l'image T.M soit effectivement comparable au regroupement en trois classes opéré (voir paragraphe suivant), on peut d'ores et déjà conclure à la pertinence de la télédétection comme moyen de stratification d'une base de sondage urbaine sur un critère de densité du bâti. Nous allons constater maintenant qu'une telle stratification améliore déjà de façon notable la précision de l'ensemble des estimations.

7.3.3.2. Contribution du caractère morphologique "densité du bâti" à une stratification morphologique synthétique.

A nouveau, le tableau ci-dessous donne le gain dû au descripteur "densité du bâti" (G1) (stratification en six classes) par rapport à celui de la stratification synthétique (G2).

Caractères :	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-
G1/G2 en % :	62,6	57,8	82,7	61,8	78,6	64,1	71,4	104,1	60,3	33,3	40,0

A l'examen du tableau, on constate que la contribution du descripteur "densité du bâti", au gain qu'apporte une stratification morphologique synthétique varie entre 30 et 100% du gain total et se situe, pour la plupart des variables démographiques étudiées, au-dessus de 50% (les seules exceptions étant les ratios 10 et 11 concernant la taille des ménages et la structure par âge). Rappelons ici que la densité du bâti fut l'un des descripteurs morphologiques dont on a augmenté l'importance par pondération lors de l'analyse typologique. Il n'est donc pas surprenant qu'il soit responsable d'une grande part du gain de précision que permet la stratification synthétique. Il se pourrait d'ailleurs que la somme des contributions de chacun des descripteurs morphologiques pris isolément, soit supérieure à 100 % . Peut-être même obtiendrait-on, par superposition des classifications fondées sur chacun d'eux, une meilleure stratification (au sens du sondage à but démographique), que celle qui nous sert pour l'instant de référence. Mais avant de revenir sur cette perspective, examinons quelle précision permet d'atteindre la stratification de la base de sondage selon la densité du bâti d'après l'image Thématic Mapper.

7.3.4 Evaluation de la stratification issue de l'image satellite

De même que pour les deux stratifications précédentes, l'évaluation du gain de précision que permet la stratification issue de l'image T.M. est réalisée pour les 11 caractères de la liste du tableau 7.7 et bien sûr avec le même plan de sondage et les mêmes estimateurs. Le tableau 7.9.a et le graphique 7.9.b, construits d'après les mêmes principes que 7.8.a et 7.8.b, résument les résultats. Ici, les trois stratifications mises en parallèle sont, d'une part, les deux stratifications qui rendent compte de la densité du bâti observée sur les photographies aériennes (col. 2 : 6 classes et col.3 : 3 classes), et d'autre part, la stratification sur le même thème issue de l'image satellite (col. 4). On mesure également dans les colonnes 5, 6 et 7 les gains que permettent les trois stratifications par rapport aux estimations non stratifiées (col. 1), gains visualisés dans la figure 7.9.b.

Comme au paragraphe précédent, ces chiffres conduisent à des conclusions qui ont trait à deux questions :

- Quelle est l'efficacité de cette classification de l'image comme stratification par rapport à la référence que l'on s'était fixé : la stratification en six classes basée sur la densité du bâti extraite des photographies aériennes ?
- Quel gain, sur la précision des estimations, peut-on attendre de la stratification de la base de sondage sur le critère de la densité du bâti issue de l'image satellite et quelle part du gain permis par une stratification synthétique établie sur photographies aériennes représente-t-il ?

7.3.4.1 Comparaison à la référence "photographies aériennes"

Le tableau suivant donne, pour chaque caractère, le gain de précision dû à la stratification tirée de l'image (G1) en pourcentage du gain dû à celle venant de la photo-interprétation (G2) ; les deux stratifications comportent six classes. On peut également comparer l'efficacité des deux stratification sur le graphique 7.9.b.

Caractères :	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-
G1/G2 en % :	93,6	93,0	87,2	94,8	96,5	93,80	88,8	80,4	70,5	40,1	75,1

Les deux stratifications sont donc d'une efficacité tout à fait comparable pour la quasi totalité des estimations pratiquées. Il n'y a que dans le cas du ratio "nombre de ménages de moins de six personnes / nombre total de ménages" que la stratification tirée de l'image déçoit par rapport à sa référence. Notons qu'il s'agit précisément du caractère pour lequel la stratification sur le critère de la densité du bâti est inefficace. La méthode mise au point de classification de l'image satellite sur ce critère s'avère donc parfaitement efficace du point de vue de l'objectif stratificateur qui lui était fixé : il est possible de construire, à partir de l'information satellitaire, une stratification de l'agglomération qui rend compte de la densité de construction et permet l'amélioration des estimations que l'on attendait précisément de l'emploi de ce critère.

7.3.4.2 Gain permis par le critère de densité du bâti et comparaison à la stratification morphologique synthétique

Du tableau 7.9.a, il importe de retenir la mesure de l'efficacité de la stratification réalisée d'après l'image que fournit la dernière colonne. Cette efficacité s'avère variable suivant le caractère démographique estimé : le gain est égal ou supérieur à 20% sur le coefficient de variation (35 à 50% de la variance) lorsque la stratification sur ce critère est efficace, de l'ordre de 15 % (30% de la variance) lorsqu'elle est moyennement efficace et inférieur à 10% (20% de la variance) pour les caractères peu liés à la densité du bâti.

De nouveau, on peut comparer ces gains à ceux que permet l'exemple de stratification synthétique dont nous disposons : le tableau ci-dessous donne le gain dû au descripteur "densité du bâti" tel qu'il est pris en compte dans la stratification issue de l'image (G1) par rapport à celui de la stratification synthétique (G2).

Caractères :	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-
G1/G2 en % :	58,57	53,77	72,13	58,62	75,88	60,09	63,40	83,69	42,48	13,35	30,01

La conclusion est la même qu'au paragraphe précédent : le descripteur "densité du bâti", tel qu'en rend compte la classification de l'image, "explique", pour la plupart des caractères démographiques, plus de la moitié du gain que permet la stratification synthétique (à l'exception cette fois des ratios 9,10 et 11).

7.3.5 Perspectives pour la suite de la recherche

Au terme de cette évaluation des stratifications basée sur le critère de la densité du bâti, on peut conclure que ce critère doit être pris en compte prioritairement lorsque l'on cherche à stratifier l'espace urbain pour améliorer un sondage démographique. Outre que ce résultat est conforme à l'intuition que l'on pouvait avoir du problème, il est également bien venu puisque, comme on l'a vu, ce critère est le seul, dans l'état actuel des recherches, dont on sache bien rendre compte à partir de l'information satellitaire. Cependant, la comparaison effectuée avec la stratification synthétique conduit à penser que la recherche de stratifications morphologiques efficaces construites à l'aide de l'information satellitaire ne doit pas s'arrêter là. C'est pourquoi il paraît intéressant que soit poursuivi, dans des recherches ultérieures, le travail analytique qui a été commencé ici, et ce, dans la direction suivante : il faudrait soumettre au même type de test, les autres descripteurs morphologiques issus de la photo-interprétation, construire avec eux des stratifications et sélectionner les plus efficaces en mesurant les gains de précision qu'elles permettent d'obtenir.

Ce travail effectué, il appartiendrait ensuite aux spécialistes de la télédétection d'identifier sur l'image, à l'aide de l'information spectrale, chacun des thèmes correspondant aux descripteurs sélectionnés. Les résultats qu'ils obtiendraient, sous forme de classifications thématiques de l'image satellite, seraient réaffectés à la base de sondage et testés comme de nouvelles stratifications. A l'issue de la démarche, on disposerait d'un ensemble de stratifications de la même base de sondage, chacune représentant un thème morphologique ; il suffirait alors de les superposer pour obtenir la stratification finale et pouvoir en évaluer l'efficacité.

Tableau 7.9 a. : coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.

Caractères estimés	Coef. de variation des estimations				Gain de précision(en % du coef. de variation sur l'estimation non stratifiée)		
	(0)	(1)	(2)	(3)	(1)	(2)	(3)
A) Effectifs							
1 POPSDC	10.13	8.59	8.77	8.69	15.20	13.43	14.22
2 POPFRANC	9.43	8.13	8.22	8.22	13.79	12.83	12.83
3 POPHEMP	10.87	8.52	8.93	8.82	21.62	17.85	18.86
4 POPMEN	10.25	8.71	8.89	8.79	15.02	13.27	14.24
5 NBMEN-6	10.36	8.08	8.24	8.16	22.01	20.46	21.24
6 POP+5	10.07	8.48	8.66	8.58	15.79	14.00	14.80
B) Ratios							
7 $\frac{\text{POPFRANC}}{\text{POPTOTAL}}$	3,67	2,87	3,16	2,96	21,80	13,90	19,35
8 $\frac{\text{POPHEMP}}{\text{POPH+19}}$	1,45	0,94	1,10	1,04	35,17	24,14	28,28
9 $\frac{\text{POPMEN}}{\text{NBMENAG}}$	3,23	2,79	2,83	2,92	13,62	12,38	9,60
10 $\frac{\text{NBMEN-6}}{\text{NBMENAG}}$	0,70	0,65	0,66	0,68	7,14	5,71	2,86
11 $\frac{\text{POP+5}}{\text{POPTOTAL}}$	0,43	0,39	0,39	0,40	9,30	9,30	6,98

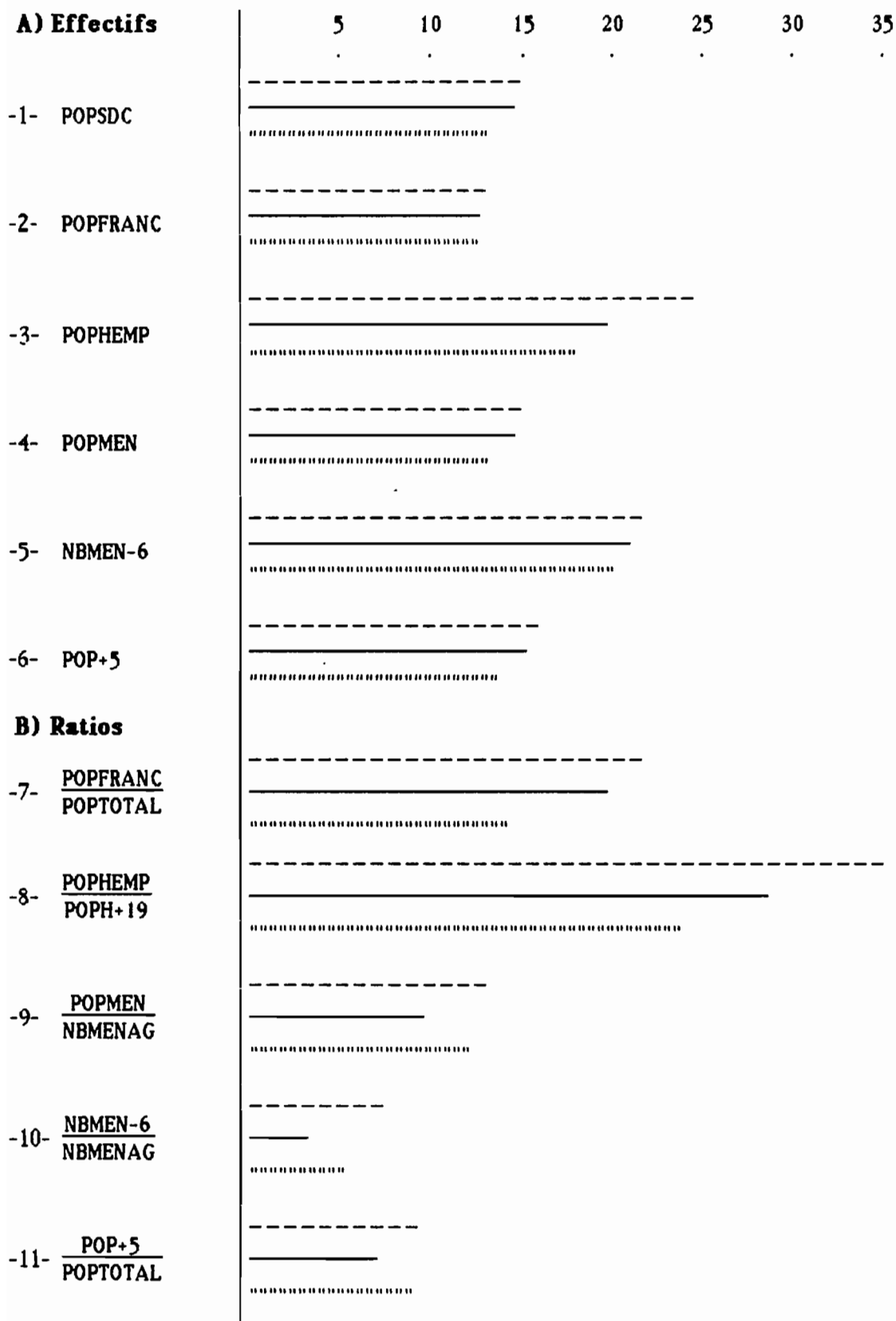
(0) : Sans stratification

(1) : Stratification sur la densité du bâti en 6 classes (photographies aériennes)

(2) : Stratification sur la densité du bâti en 3 classes (photographies aériennes)

(3) : Stratification sur la densité du bâti d'après image satellite en 6 classes.

Graphique 7.9.b : Comparaison des gains de précision apportés par trois stratifications morphologiques à l'estimation de 11 caractères démographiques (en % du coefficient de variation non stratifiée)



Stratification synthétique morphologique (9 classes) : _____
 Stratification sur la densité du bâti (6 classes) : - - - - -
 Stratification sur la densité du bâti (3 classes) :

Tableau 7.7 : liste des caractères démographiques considérés dans le chapitre

A) effectifs

N°	Nom du caractère	Abréviation utilisée
-1-	Population sans doubles comptes	POPSDC
-2-	Population totale de nationalité française	POPFRANC
-3-	Population totale des hommes de plus de 19 ans ayant un emploi	POPHEMP
-4-	Population totale des ménages ordinaires	POPMEN
-5-	Nombre de ménages ordinaires de moins de 6 personnes	NBMEN-6
-6-	Population totale de plus de cinq ans	POP+5

B) ratios

N°	Nom du caractère	Abréviation utilisée
-7-	$\frac{\text{Population totale de nationalité française}}{\text{Population totale}}$	$\frac{\text{POPFRANC}}{\text{POPTOTAL}}$
-8-	$\frac{\text{Population totale des hommes de plus de 19 ans ayant un emploi}}{\text{Population totale des hommes de plus de 19 ans}}$	$\frac{\text{POPHEMP}}{\text{POPH+19}}$
-9-	$\frac{\text{Population totale des ménages ordinaires}}{\text{Nombre de ménages ordinaires}}$	$\frac{\text{POPMEN}}{\text{NBMENAG}}$
-10-	$\frac{\text{Nombre de ménages ordinaires de moins de 6 personnes}}{\text{Nombre de ménages ordinaires}}$	$\frac{\text{NBMEN-6}}{\text{NBMENAG}}$
-11-	$\frac{\text{Population totale de plus de cinq ans}}{\text{Population totale}}$	$\frac{\text{POP+5}}{\text{POPTOTAL}}$

Conclusion

Cette étape, durant laquelle on a soumis les données du R.G.P. de 1982 à Marseille au programme VARECH afin d'évaluer l'efficacité de divers plans de sondage, a permis d'aborder plusieurs questions ; rappelons les principales :

1. Quelles performances, en terme de précision des estimations démographiques en ville, peut-on attendre d'un sondage où la probabilité d'inclusion des unités primaires (pâtes de maison) est proportionnelle à leur surface, ceci par rapport au sondage équiprobable ?

2. Dans chacun des deux cas (équiprobabilité ou probabilités inégales), a-t-on intérêt à utiliser l'information exogène concernant la surface des UP dans un estimateur par le ratio ?

3. Quelles sont les possibilités d'améliorer la précision des estimations en stratifiant la base de sondage d'après des critères de morphologie urbaine extraits de l'information satellitaire ?

4. Quel degré de stabilité offrent les résultats relatifs à l'ensemble des problèmes posés lorsque l'on fait varier le caractère démographique à estimer ?

Pour conclure, essayons de faire une synthèse des réponses apportées à chacune des questions.

1. Les deux définitions de la probabilité d'inclusion des unités primaires sont apparues équivalentes quant à leurs performances et cela sur trois plans :

- équivalence des précisions atteintes pour l'estimation sans biais des totaux et pour l'estimation biaisée des quotients ;
- équivalence quant aux gains permis par les stratifications de la base de sondage ;
- équivalence enfin quant aux domaines géographiques de bonne ou de médiocre précision des estimations.

Il s'agit là nous le verrons, d'un premier résultat important pour la définition du plan de sondage au premier degré puisqu'en pratique, la sélection des unités primaires sur document provenant d'images satellite conduit précisément à un plan à probabilités inégales.

2. La réponse à la deuxième question est négative : la surface des unités primaires, prise comme mesure de leur taille, n'est pas suffisamment corrélée aux variables démographiques que nous considérons ici pour que l'estimation d'un total Y par l'intermédiaire du ratio Y/S soit avantageuse, au contraire elle s'avère toujours moins précise que l'estimation sans biais.

3. En ce qui concerne les améliorations qu'apportent les stratifications, les conclusions sont à nuancer, d'une part en fonction des critères morphologiques qui président à la stratification, d'autre part en fonction des caractères démographiques que l'on cherche à estimer.

La stratification morphologique synthétique réalisée par analyse typologique des données provenant de la photo-interprétation permet une diminution de la variance des estimations comprise entre 20 et 60 % selon les caractères démographiques estimés.

Le premier critère morphologique à propos duquel on sache extraire de l'image satellite une information fiable est la densité du bâti. La méthode, mise au point par l'équipe de télédétection, de classification de l'image satellite Thematic Mapper sur ce critère, aboutit à une stratification tout à fait satisfaisante : le gain de précision réalisé varie alors entre 20 et 50 % de la variance non stratifiée suivant le caractère estimé et représente généralement un gain égal ou supérieur à la moitié de celui que permet la stratification synthétique.

4. Les conclusions relatives aux deux premières questions sont les mêmes quel que soit le caractère démographique estimé. En revanche, comme on l'a vu, les gains permis par les stratifications sont variables suivant que l'indicateur démographique à estimer est plus ou moins lié aux critères morphologiques pris en compte dans la stratification.

À l'issue de ce travail d'évaluation des diverses techniques de stratification, de sélection des unités et d'estimation, qui nous paraissent pouvoir être appliquées au premier degré d'un sondage aréolaire à but démographique en milieu urbain, certains résultats sont donc acquis. Ils permettront, en conclusion de l'expérience menée à Marseille, de définir presque complètement le plan de sondage au premier degré. Néanmoins il demeure un point sur lequel les tests menés jusqu'à présent ne permettent pas de conclure : il s'agit des conséquences qu'aurait sur la précision des estimations un mode de sélection systématique des unités primaires. C'est ce point que nous abordons dans le chapitre suivant.

Chapitre 8

APPORT DU TIRAGE SYSTEMATIQUE AREOLAIRE

Introduction

Dans la méthode que nous tentons de mettre au point, la collecte se fera par sondage aréolaire systématique à partir de documents satellitaires ; il s'agit d'une contrainte pratique dont on a déjà abordé les conséquences sur les probabilités de sélection des districts ou, plus généralement, des unités d'échantillonnage, et donc, sur la définition du plan de sondage. Mais ce type de sondage possède une autre caractéristique importante : il fournit toujours un échantillon géographiquement bien réparti au sein de la base de sondage. Or, il est clair que dans la plupart des villes, les caractéristiques démographiques des quartiers ne sont pas indépendantes de la localisation. Marseille et Quito, nos deux sites d'expérience, n'échappent pas à cette règle, même si l'organisation concentrique de la première s'oppose, par certains aspects, à la structure linéaire de la seconde. C'est sur cette évidence que repose l'idée directrice du travail dont l'exposé va suivre : les caractères démographiques que nous cherchons à estimer n'étant pas indépendants de la localisation, comment évaluer le gain de précision qu'apporte la bonne répartition spatiale de l'échantillon obtenue quand la technique de sondage prend en compte la localisation des unités statistiques ?

Par ailleurs, il n'y a pas non plus indépendance entre les spécificités morphologiques et la localisation : à un type morphologique donné correspond une localisation plus ou moins précise. Donc, si l'on considère, comme c'est le cas ici, morphologie et localisation comme deux variables "explicatives" de la démographie, il y a très probablement une certaine redondance dans l'information apportée par les deux variables et les gains permis par les stratifications construites à partir de chacune d'elles ne s'additionnent pas totalement lorsqu'on les prend en compte simultanément. Pourtant cette redondance n'est sûrement pas totale dans la mesure où, au sein d'une même strate morphologique, les différences de localisation peuvent à nouveau "expliquer" une part de la variance intra-strate des caractères démographiques. Dans le cadre de notre sondage stratifié, peut-on chiffrer l'apport, non redondant avec la morphologie, de la localisation telle qu'elle est prise en compte par un tirage spatial systématique ?

Pour répondre à ces questions, la première partie de ce chapitre propose une méthode d'approximation de la précision du sondage aréolaire systématique par assimilation à un sondage stratifié sur un critère de localisation. Puis, on dresse la liste des données nécessaires au test et on expose comment sont obtenues les stratifications employées. Enfin, après avoir présenté les résultats, on tire les conclusions de cette expérience.

8.1 La méthode : approximation du tirage systématique aréolaire par stratification de la base de sondage sur un critère de localisation.

La méthode exposée ici s'inspire très largement d'une proposition faite par DESABIE dans "Théorie et pratique des sondages" à propos du tirage systématique sur liste¹, dont nous reprenons l'idée centrale pour l'adapter au tirage aréolaire utilisant une grille de points. DESABIE assimile le tirage systématique sur liste au tirage suivant (M est le cardinal de la base de sondage, m celui de l'échantillon).

" L'univers est découpé en $m/2$ strates égales, chacune d'entre elles contenant $2(M/m)$ unités consécutives de la population. Dans chaque strate on tire au sort deux unités" et l'estimateur de la moyenne du caractère Y est :

$$\bar{Y} = (2/m) \cdot \sum_{h=1}^{m/2} Y_h, \text{ où}$$

$$Y_h = 1/2 (Y_{2h-1} + Y_{2h}).$$

L'idée est en fait d'utiliser les formules d'estimations et de variances liées au sondage stratifié, en stratifiant la base de sondage à l'aide de l'ordre de la liste, et cela en créant le nombre maximal de strates. Ce nombre n'est limité que par la contrainte qu'on se donne de pouvoir calculer un estimateur de la variance de l'estimation dans chaque strate ; l'échantillon doit donc contenir au moins deux individus de chaque strate.

L'adaptation au tirage d'unités aréolaires à l'aide d'une grille de points se fait en considérant comme strates les mailles, carrées et de surface égale, d'une seconde grille placée sur le document satellitaire : chaque strate est définie par la longueur de son côté et les coordonnées géographiques d'un de ses sommets.

Ici apparaît l'unique problème posé par l'adaptation de la méthode au sondage aréolaire : si la grille de tirage a une maille constante sur l'ensemble de la base, chaque strate définie par la grille de stratification contient le même nombre de districts échantillonnés, mais le taux de sondage varie alors suivant les strates puisqu'il est fonction du rapport entre la surface de la maille qui sert au tirage et la surface moyenne des unités de la strate. L'approximation, par cette méthode, du tirage aréolaire à l'aide d'une grille uniforme nécessite donc le calcul de taux spécifiques à chaque strate. Le programme actuel de calcul des variances d'estimation ne le permettant pas, nous avons préféré procéder, dans un premier temps, à l'approximation de la variance qu'aurait un sondage aréolaire à taux constant sur l'ensemble de la base. Ce sondage correspond à la procédure de tirage suivante : la taille de la grille, redéfinie pour chaque sous-ensemble de la base au sein duquel les districts ont une taille voisine, est proportionnelle à la surface moyenne des unités du sous-ensemble, ce qui permet d'obtenir un taux à peu près constant. Une telle procédure nécessite en revanche un premier zonage de la base de sondage, effectué d'après la taille des unités. Les résultats qui vont suivre sont donc une approximation de la précision attachée à une telle procédure.

Enfin, la méthode appelle une dernière remarque : pour optimiser l'approximation du sondage systématique, si l'on suit l'idée de DESABIE, le nombre de strates et donc leur surface, devrait en principe être définis de manière à ce que chaque strate contienne deux points de la grille de tirage. En pratique, aux taux de sondages que nous envisageons, ce nombre de strates (environ 90) ralentit énormément l'exécution du

¹ : J. DESABIE, "Théorie et pratique des sondages" (DUNOD 1966) chapitre 7.7.7 : Le tirage systématique envisagé comme un sondage stratifié, P. 171.

programme de calcul des variances d'estimation. Nous avons donc retenu des maillages de stratification plus larges, comportant suivant les cas 6 ou 15 strates, ce qui permet des temps de calculs raisonnables tout en gardant un découpage spatial suffisamment fin.

8.2 Les données nécessaires au test : élaboration des stratifications et remarques sur l'interprétation des résultats

L'information permettant l'élaboration des stratifications spatiales dont il va être question provient du système de gestion de base de données localisées SAVANE. Les fonds cartographiques concernant l'ensemble des districts de la base de sondage ayant été intégrés dans ce système, nous disposons, grâce à lui, d'un fichier contenant la latitude et la longitude des centres de ces districts. A partir de ces coordonnées géographiques, un programme affecte chacun des 2619 districts de la base de sondage à une strate définie par un intervalle de longitude et de latitude. Trois stratifications ont été produites qui contiennent respectivement six, huit et quinze classes. Les résultats du test ayant fait apparaître une très faible différence entre les stratifications en six et huit classes, nous n'avons finalement retenu que les stratifications en six et quinze classes schématisées sur la figure 8.1.

Les stratifications spatiales permettent, lorsqu'on les utilise seules, d'approcher la précision du sondage systématique non stratifié. Pour approcher celle d'un sondage stratifié selon un critère morphologique, nous croisons deux à deux les stratifications morphologique¹ et spatiale pour obtenir les six stratifications du tableau 8.2.

Figure 8.1 : schéma du nombre de districts par strate dans les deux stratifications spatiales.

	Stratification en 15 classes		stratification en 6 classes	
		↑ (Nord)		↑ (Nord)
159	86	43		
			386	275
222	306	108		
			1138	458
592	506	81		
			155	206
100	207	20		
73	97	18		

¹ : Au moment où nous réalisons cette expérience, la stratification issue de l'image satellite en six classes de densité du bâti n'était pas disponible. Elle n'est donc pas intégrée dans ce test mais nous avons constaté depuis qu'elle conduit aux mêmes résultats que les stratifications sur ce même thème étudiées ici.

Tableau 8.2. : croisement des stratifications spatiales et morphologiques, nombre de classes des stratifications résultantes.

Localisation : Morphologie :	Grille 6 classes	grille 15 classes
Stratif. synthétique 9 classes	43 classes	80 classes
Stratif. sur densité du bâti en 6 classes (photo. aériennes)	31 classes	62 classes
Stratif. sur densité du bâti en 3 classes (photo. aériennes)	18 classes	38 classes

Ce principe des stratifications croisées appelle deux remarques :

1. Dans certaines strates contenant un très petit nombre d'unités, l'échantillon ne contient qu'une fraction de district. De ce fait, un tel tirage est impossible à mettre en oeuvre pratiquement. Ce n'est d'ailleurs pas l'objectif de la méthode qui se propose simplement de fournir une approximation de la précision du tirage systématique qui est, lui, praticable.

2. Au sein du programme de calcul des variances d'estimation, la détermination de l'allocation optimale se fait en minimisant l'expression de la variance stratifiée : il s'agit d'une somme sur l'ensemble des strates issues du croisement des stratifications spatiales et morphologiques. En toute rigueur, ce calcul n'est pas légitime puisqu'il conduit à faire varier le taux de sondage selon la localisation alors que la seule "vraie" stratification est la stratification morphologique. On aurait donc dû déterminer l'allocation optimale uniquement par rapport à elle. Là encore, l'état actuel du programme de calcul nous en empêchait et les résultats relatifs à l'allocation optimale ne sont que des indications sur l'ordre de grandeur du gain amené par la procédure de tirage sur grille.

8.3 Résultats

Pour ce test, nous n'avons retenu que trois caractères démographiques qui sont le total de population sans doubles comptes et deux ratios par rapport à la population totale : population de nationalité française et population de 5 ans et plus.

Le tableau 8.3 présente les gains de précision obtenus grâce aux deux stratifications spatiales et les compare aux gains dus à la stratification morphologique en 9 classes. Sont donnés successivement pour chacun des caractères, le coefficient de variation de l'estimation correspondant à l'allocation proportionnelle, puis celui correspondant à

l'allocation optimale, enfin le gain réalisé sur le coefficient de variation (ou si l'on préfère sur l'amplitude de l'intervalle de confiance) lorsque l'on passe de l'estimation non stratifiée à l'estimation stratifiée avec allocation optimale.

Tableau 8.3 : coefficients de variation¹ et gain de précision obtenus avec les stratifications morphologiques et spatiales pour trois caractères démographiques

Caractères	POPSDC (*)			POP+5 POPTOTAL (**)			POPFRANC POPTOTAL (**)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Non stratifié	10.13	..	0,0	0,43	..	0,0	3,67	..	0,0
Stratif. morpho. synthétique (9 classes)	9.08	7,67	24,3	0,40	0,33	23,3	3,33	2,55	30,5
Stratif. sur la localisation (6 classes)	9,70	9,42	7,01	0,40	0,34	20,9	3,44	2,82	23,2
Stratif. sur la localisation (15 classes)	9.25	8,87	12,44	0,40	0,34	20,9	3,15	2,34	36,2

(*) : Estimateur sans biais

(**) : Estimateur biaisé

(1) : Coefficient de variation de l'estimation avec allocation proportionnelle (%)

(2) : Coefficient de variation de l'estimation avec allocation optimale (%)

(3) : Gain de précision, en % du coefficient de variation non stratifié, une fois optimisée l'allocation

On remarque d'abord, à l'examen du tableau, que le gain permis par la bonne répartition spatiale de l'échantillon peut être égal et même supérieur à celui dû à la stratification morphologique mais qu'il varie suivant le caractère démographique auquel on s'intéresse. L'effectif de population sans doubles comptes ne profite que peu de cette bonne répartition tandis que pour les deux quotients étudiés les gains sont compris entre 20 et 40 % du coefficient de variation c'est-à-dire entre 30 et 60 % de la variance, une fois optimisée l'allocation aux strates. L'extension du test à d'autres caractères confirmerait que le gain permis par les stratifications "spatiales" est variable suivant les cas comme l'était celui dû aux stratifications morphologiques.

¹ : Plan à probabilités proportionnelles aux surfaces des districts, estimation sans biais des totaux et estimation biaisée des quotients (allocation proportionnelle et optimale).

L'amélioration obtenue en passant de 6 à 15 strates est, elle aussi, variable : gain important pour l'estimation de la proportion de population française ; pas d'effet en revanche sur celle de la population de plus de cinq ans.

Deux questions se posent maintenant :

1. Du point de vue de l'information qu'ils apportent, les deux types de stratifications, morphologique et spatiale, sont en partie redondants. Il est donc probable, comme on l'a déjà dit, que les gains liés à chacun d'eux ne s'additionnent pas entièrement lorsqu'on construit les stratifications croisées. On peut, en plus, se demander si l'amélioration permise par le tirage systématique, représenté ici par la stratification spatiale, est la même quelle que soit la stratification morphologique employée.

2. La variabilité, selon les caractères estimés, des gains permis par les stratifications spatiales et par le passage de six à quinze classes peut avoir, à notre sens, deux explications :

- premièrement, comme pour la morphologie, certains caractères démographiques sont plus liés que d'autres à la localisation des districts.

- deuxièmement, le niveau de finesse de la partition spatiale retenue comme stratification peut suffire à rendre compte des variations dans l'espace urbain de certains caractères, alors que pour d'autres, ces variations ont lieu pour une part importante à l'intérieur de la maille employée. Dans ce cas, on devrait en stratifiant plus, augmenter le gain de précision de manière sensible.

On le voit, le gain dû au sondage systématique peut varier suivant le caractère estimé mais peut être également fonction du degré de finesse de la stratification morphologique de départ. En examinant les résultats obtenus lorsqu'on croise les deux types de stratification, nous tentons maintenant de préciser ce qui vient d'être dit. Le tableau 8.4 donne les coefficients de variation correspondant aux six croisements de stratifications possibles (3 stratifications morphologiques de départ et 2 niveaux de stratification spatiale).

Tout d'abord, on constate que le gain apporté par la stratification spatiale est effectivement du même ordre quelle que soit la stratification morphologique¹. Par contre, comme précédemment, il n'est pas le même pour tous les caractères démographiques estimés. L'effectif de population sans doubles comptes profite moins de la stratification spatiale que l'estimation des deux ratios pour lesquels l'amélioration est sensible (20 à 35 %). Pour ces variables, on constate un premier gain lors de l'introduction d'une stratification morphologique, puis un second gain du même ordre que le premier lorsqu'on stratifie spatialement.

Il semble donc que certains caractères démographiques présentent plusieurs niveaux de variation spatiale appréhendables, au niveau de l'amélioration d'un sondage, soit par la stratification morphologique, soit par le tirage systématique aréolaire. Le découpage de l'espace urbain en sous-ensembles morphologiquement homogènes, rend compte d'une partie de la variation, tandis qu'à l'intérieur de ces strates, le tirage systématique à l'aide d'une grille de points permet d'appréhender les variations plus locales.

¹ : Il s'agit ici des gains mesurés en % du coefficient de variation atteint avec la stratification morphologique de départ. On obtient donc malgré tout, après croisement des stratifications morphologiques avec les stratifications spatiales, un certain tassement des écarts de précision existant entre les différentes stratifications morphologiques étudiées. Il reste qu'une stratification morphologique meilleure que les autres lorsqu'on considère son effet "isolé" demeure meilleure lorsque cet effet est cumulé avec celui du tirage systématique.

Tableau 8.4 : coefficients de variation¹ et gains de précision obtenus sur l'estimation de trois caractères démographiques à partir de stratifications croisant un critère morphologique et un critère de localisation.

Caractères :	POPSDC (*)			$\frac{POP+5}{POPTOTAL}$ (**)			$\frac{POPFRANC}{POPTOTAL}$ (**)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Non stratifié	10.13	..	0,0	0,43	..	0,0	3,67	..	0,0
1) Croisements avec la stratification morphologique synthétique en neuf classes.									
Strat. morph.(9 cl.)	9.08	7.67	24.3	0,40	0,33	23,3	3,33	2,55	30,5
morph.(9 cl.) x loc.(6 cl.)	8.96	7.16	29.3	0,41	0,29	32,6	3,11	2,10	42,8
morph.(9 cl.)x loc.(15 cl.)	8.93	6.83	32.6	0,38	0,27	37,2	2,91	1,72	53,1
2) Croisements avec la stratification en six classes de densité du bâti									
Strat. dens.(6 cl.)	8,81	8,59	15,2	0,41	0,39	9,3	2,92	2,87	21,8
Dens.(6 cl.) x loc.(6 cl.)	8,57	7,98	21,2	0,42	0,32	25,6	3,00	2,16	41,1
Dens.(6 cl.)x loc.(15 cl.)	8,37	7,48	26,2	0,39	0,31	27,9	2,84	1,91	48,0
2) Croisements avec la stratification en trois classes de densité du bâti									
Strat. dens.(3 cl.)	8,90	8,77	13,43	0,41	0,39	9,3	3,19	3,16	13,9
Dens.(3 cl.) x loc.(6 cl.)	8,74	8,29	18,16	0,41	0,33	23,3	3,22	2,51	31,6
Dens.(3 cl.) x loc.(15 cl.)	8,85	7,91	21,90	0,41	0,32	25,6	3,23	2,16	41,1

(*) : Estimateur sans biais

(**) : Estimateur biaisé

(1) : Coefficient de variation de l'estimation avec allocation proportionnelle (%)

(2) : Coefficient de variation de l'estimation avec allocation optimale (%)

(3) : Gain de précision, en % du coefficient de variation non stratifié, une fois optimisée l'allocation

¹ : Plan à probabilités proportionnelles aux surfaces des districts, estimation sans biais des totaux et estimation biaisée des quotients (six croisements, allocations proportionnelles et optimales).

La dernière constatation à faire sur le tableau 8.4 est que le gain de précision obtenu au passage de la division spatiale en six classes à celle en quinze classes demeure relativement important. On peut donc penser que l'on parvient, grâce à la méthode retenue, à une approximation légèrement pessimiste du gain que permet le tirage systématique. Il faudrait probablement stratifier plus finement la base de sondage pour améliorer l'approximation.

Conclusion

Pour conclure, résumons les leçons que l'on peut tirer du test. En premier lieu, il confirme qu'un tirage aréolaire systématique, en assurant une bonne répartition géographique de l'échantillon, permet une amélioration sensible des estimations. Le gain se situe pour certaines estimations, entre 20 et 40 % du coefficient de variation correspondant au tirage strictement aléatoire mais il peut être assez variable suivant le type de caractère que l'on cherche à estimer. Par contre, pour un caractère démographique donné, le gain dû au tirage systématique a été le même pour les trois stratifications morphologiques testées. On parvient ainsi, si l'on considère l'effet cumulé du critère morphologique et du critère spatial, à une amélioration comprise entre 20 et 30 % du coefficient de variation pour le total de population sans double compte et entre 25 et 55% pour les deux quotients retenus.

On peut, nous semble-t-il, conclure à une certaine robustesse de cette technique quand on change de critère de stratification, du moins en gardant un niveau de finesse des partitions comparable.

Enfin, la méthode d'assimilation à un tirage stratifié d'après critère spatial semble efficace pour évaluer le gain apporté par le tirage systématique que nous pratiquerons à la condition, nous l'avons vu, de retenir une stratification géographique suffisamment fine de l'agglomération. Pour une ville de la taille de Marseille, il semble qu'un découpage de l'agglomération en une dizaine de zones géographiques soit un minimum pour obtenir une approximation satisfaisante de ce gain. La méthode testée ici ayant d'une certaine façon fait ses preuves, nous l'emploierons à nouveaux pour approximer la précision du sondage systématique que nous mettrons en œuvre à Quito (voir chapitre 13).

CONCLUSION

Le but principal de l'expérience menée sur le site de Marseille était, comme on l'a vu, de tester et d'évaluer des méthodes de sondages spatiaux à but démographique en milieu urbain. En conclusion de ce travail, il nous semble utile tout d'abord de rappeler ses principaux objectifs et de résumer les étapes de son déroulement. Ensuite, pour faire la synthèse de l'expérience marseillaise, nous tenterons de proposer des réponses claires, reposant sur les résultats acquis, aux questions que nous nous posions à son début. Enfin, pour finir, nous essaierons d'identifier les principaux problèmes restant posés et nous donnerons la liste des travaux qui devront être réalisés à Quito pour essayer de les résoudre.

1. Bilan de l'expérience menée à Marseille : objectifs, déroulement, résultats

Pour la recherche que nous menons actuellement sur les sondages spatiaux en milieu urbain, l'expérience réalisée à Marseille avait trois objectifs principaux qui d'ailleurs se conditionnent mutuellement dans une large mesure :

1. Vérifier l'hypothèse fondamentale selon laquelle une stratification morphologique de la base de sondage issue d'une image satellite améliore de façon significative la précision de l'estimation des informations démographiques et quantifier ce gain de précision.
2. Sélectionner, après évaluation des différentes possibilités, un plan de sondage efficace et adapté aux contraintes qui surgissent de son application à la collecte de données démographiques en milieu urbain et de l'utilisation de l'image satellite comme base de sondage.
3. Mettre en évidence les principales spécificités du plan de sondage défini et analyser les conséquences qu'elles auront sur la précision des estimations et sur la mise en oeuvre pratique de l'enquête.

Dans le déroulement de l'expérience on peut distinguer quatre étapes dont les trois premières fournissent des résultats principalement méthodologiques, lesquels peuvent d'ailleurs intéresser d'autres domaines que celui de la mise au point de sondages spatiaux, tandis que la dernière permet de préciser ce que pourrait être un sondage à but démographique en milieu urbain utilisant comme base une image satellite stratifiée.

Première étape : constitution et intégration dans le système d'information géographique SAVANE, d'une base de données sur la ville de Marseille. Dans le chapitre 5, on a abordé l'utilisation que nous faisons, dans notre propre travail, de cette base et du système permettant sa gestion ; mais cette étape a également permis d'importantes avancées dans le domaine de l'intégration dans un système de gestion de

bases de données d'informations géographiques et satellitaires, résultats qui intéressent plus spécialement les géographes et les informaticiens¹.

Deuxième étape : établissement, d'après des informations issues de photographies aériennes, d'une stratification morphologique de l'espace urbain marseillais. Comme on l'a vu au chapitre 5, le résultat principal de cette étape est bien entendu la stratification elle-même et son utilisation pour la sélection d'un échantillon de travail indispensable au travail des spécialistes en télédétection. Elle laisse également certains acquis dans le domaine de l'application des techniques d'analyse des données aux problèmes d'analyse typologique et de classification automatique des données de morphologie urbaine, ses possibilités et ses limites.

Troisième étape : développement du programme VARECH qui permet l'évaluation, d'après des données exhaustives, des performances de différents plans de sondage et de différentes stratifications. L'outil statistique et informatique présenté au chapitre 6 est, bien sûr, applicable à d'autres sites géographiques urbains ou non et pour répondre à d'autres demandes thématiques, démographiques ou autres.

Quatrième étape : application du programme VARECH aux données du recensement général de la population de 1982 à Marseille pour l'évaluation de plans de sondage et de stratifications, matière des chapitres 7 et 8.

Nous ne répèterons pas ici les résultats obtenus dans cette quatrième étape puisqu'ils figurent, exposés en détail, dans les conclusions des chapitres en question. Il nous semble néanmoins nécessaire d'en faire maintenant une synthèse et de montrer en quoi ils ont permis d'atteindre les objectifs fixés au début de l'expérience.

La vérification de l'hypothèse fondamentale

Il est acquis, à l'issue de cette expérience, qu'une stratification de l'espace urbain construite par synthèse d'un ensemble de critères morphologiques permet une amélioration significative de la précision de l'estimation par sondage des caractéristiques démographiques des habitants. On peut quantifier cette amélioration en disant qu'elle représente une diminution de 40 %, en moyenne, de la variance des estimateurs de totaux et de 50 % de celle des estimateurs de ratios.

D'autre part le critère de densité de l'espace bâti s'avère jouer un rôle déterminant dans ce gain de précision : à lui seul, il permet une amélioration égale à 50 %, voir 75 % pour certaines estimations, de celle due à la stratification synthétique. Ce résultat ne doit pas surprendre dans la mesure où ce critère a justement été privilégié lors de l'élaboration de la stratification synthétique.

L'autre résultat fondamental, à porter au crédit de l'équipe de télédétection, est que l'on peut produire des classifications de l'image satellite qui rendent compte de façon parfaitement satisfaisante de la densité du bâti en milieu urbain.

On dispose donc, à l'issue de ce travail, de stratifications des images satellite qui permettent de réduire au moins de moitié la variance des estimations par sondage de caractères démographiques et cela quel que soit le plan de sondage considéré.

Sélection d'un plan de sondage efficace et adapté

Comme nous l'avons déjà signalé, l'utilisation de l'image satellite comme base de sondage spatiale implique presque nécessairement un plan de sondage à probabilités inégales. La technique de sélection des unités spatiales à enquêter la plus simple à mettre en oeuvre est celle du tirage systématique à l'aide d'une grille de points

¹ : Les personnes intéressées peuvent se reporter aux deux rapports rédigés par l'équipe de recherche en Avril 86 et Août 87 pour l'ATP Télédétection spatiale CNRS/CNES, lesquels sont disponibles au CNRS ou à l'ORSTOM.

plaquée sur l'image. La probabilité d'inclusion est alors proportionnelle à la surface de chaque unité spatiale.

Les tests réalisés à partir des données de Marseille permettent d'affirmer que les estimateurs sans biais ont à peu près la même précision sous ce plan de sondage que sous un plan équiprobable. De plus, il s'avère que l'introduction des stratifications améliore les précisions dans les mêmes proportions pour les deux plans.

D'autre part il apparaît que la surface totale des districts n'est pas suffisamment corrélée aux variables que nous cherchons à estimer pour que les estimateurs par le ratio faisant intervenir cette surface dans leurs dénominateurs soient meilleurs que les estimateurs sans biais d'Horwitz-Thompson.

Enfin, comme l'ensemble de ces résultats est quasiment invariant lorsqu'on change le caractère à estimer, on peut considérer, au vu de ces conclusions, que le sondage que nous cherchons à définir doit avoir les caractéristiques suivantes :

1. La base de sondage est l'ensemble de l'aire métropolitaine matérialisée sur l'image satellite par la limite de la zone considérée comme urbaine.

2. Cet espace pourra être stratifié suivant un ensemble de critères morphologiques représentés par diverses classifications de l'image satellite. Pour l'instant nous sommes en mesure, quel que soit le site, de construire d'après l'image satellite une stratification efficace sur un critère de densité du bâti. Il est d'ailleurs certain que dans toute stratification du milieu urbain à but démographique, ce critère doit jouer un rôle très important.

3. La sélection de l'échantillon s'effectue grâce à un tirage systématique où la probabilité d'appartenance à l'échantillon d'une unité spatiale donnée est proportionnelle à une mesure de sa taille. La plus "naturelle" de ces mesures est évidemment la surface totale de l'unité, c'est aussi la plus facile à mettre en oeuvre lors du tirage mais, comme nous le verrons plus tard, ce n'est bien sûr pas la seule possible.

4. Les estimateurs sont sans biais lorsqu'il s'agit d'estimer des totaux (estimateurs classiques d'Horwitz-Thompson) mais les estimateurs par le ratios (biaisés) restent utiles pour estimer directement des quotients de totaux.

2. Analyse des spécificités du plan défini

Par rapport aux sondages les plus couramment pratiqués, par exemple les sondages aléatoires équiprobables, le sondage que nous proposons présente deux particularités principales :

1. En tant que sondage stratifié à probabilités d'inclusion inégales, il nécessite, pour que l'on puisse extrapoler correctement les données recueillies sur l'échantillon au domaine d'estimation choisi, la connaissance d'informations exogènes :

- Sur chacune des unités spatiales enquêtées, outre l'ensemble de l'information correspondant aux variables que l'on cherche à estimer, il faudra connaître la taille de l'unité puisque cette grandeur entre dans le calcul de l'estimateur. Il s'agira très probablement de sa surface totale et il faut donc disposer d'un moyen fiable de la mesurer.

- Il faut également connaître la taille de chaque strate, quand cette taille est la surface totale, sa mesure s'effectuera sans problèmes, on le verra, sur l'image satellite.

2. Le caractère systématique du sondage, en assurant une bonne répartition géographique de l'échantillon, améliore la précision des estimations. Il est malheureusement impossible de mesurer rigoureusement ce gain puisqu'il n'existe pas, dans le strict cadre de la théorie des sondages, d'expression de la variance des estimateurs sous un tel plan. Les tests dont il a été question au chapitre 8 ont

néanmoins permis de le quantifier approximativement. En tant que tel, le tirage systématique fait diminuer la variance de 20 à 30 % par rapport au tirage strictement aléatoire au même taux. Si l'on considère l'effet cumulé du tirage systématique et de la stratification, l'allocation aux strates étant optimale, le gain total se situe, suivant les caractères estimés, entre 45 et 75% de la variance de départ (tirage aléatoire non stratifié).

3. Questions en suspens, perspectives de travail à Quito

L'expérience menée à Marseille aura donc permis de préciser un certain nombre des caractéristiques que doit avoir un sondage à but démographique en milieu urbain utilisant comme base de sondage une image satellite stratifiée. Cependant, d'une part la description de ce plan de sondage reste encore incomplète, et d'autre part il reste à vérifier la stabilité des résultats acquis à Marseille lorsqu'on change de site d'application. En conséquence, trois questions principales encore en suspens peuvent être formulées :

1. L'usage de l'image satellite comme base de sondage aréolaire et support de la sélection de l'échantillon implique, on l'a vu, un plan de sondage où la probabilité d'appartenance d'une unité à l'échantillon est proportionnelle à la taille de cette unité. Reste à définir quelle est la mesure de cette taille la plus efficace au sens de la précision du sondage. Concrètement, compte tenu de l'information actuellement disponible grâce à l'interprétation des images satellite, l'alternative est la suivante : soit la mesure de la taille des unités spatiales est leur surface totale, soit il s'agit d'une approximation de leur surface bâtie.

La question est donc : a-t-on intérêt, du point de vue de la précision des estimations produites, à choisir une probabilité de sélection proportionnelle à la surface bâtie des unités plutôt qu'à leur surface totale? Dans l'hypothèse affirmative, existe-t-il une stratégie de sélection de ces unités praticable sur l'image satellite qui garantisse une telle probabilité?

2. L'expérience marseillaise a permis de définir presque totalement le plan de sondage applicable à des unités spatiales du type districts de recensement INSEE. Ces unités correspondent en gros aux pâtés de maison physiques identifiables sur l'image satellite. Or, l'unité statistique prise en compte lors des enquêtes démographiques est soit l'individu, soit, plus généralement, le ménage. Le plan de sondage complètement défini doit donc comprendre deux degrés, le premier degré correspond à la sélection des pâtés de maison et le second à celle de l'unité statistique à enquêter : le ménage. On s'attend d'ailleurs, dans la mesure où la composante intra-unités primaires (c'est-à-dire la somme des variances au sein des îlots) est probablement faible devant la composante inter-unités primaires (variance entre les moyennes des divers îlots), à ce qu'une procédure à deux degrés permette d'améliorer les précisions des estimations pour un même nombre total de ménages enquêtés.

Il reste donc à fixer les caractéristiques du second degré du sondage, le taux et la technique de sélection à pratiquer. Après avoir évalué la précision que permet d'atteindre une telle procédure, les choix techniques seront faits de manière à optimiser la précision du sondage pour un coût de mise en oeuvre donné c'est-à-dire pour un nombre de ménages à enquêter fixé à l'avance.

3. Quel degré de stabilité offrent les résultats acquis à Marseille lorsque l'on change de site? La méthode définie ici nécessite-t-elle des adaptations pour être efficace ailleurs?

A l'issue de l'expérience menée à Marseille, trouver des réponses aux questions formulées ci-dessus constitue l'objectif principal de la phase suivante que nous avons réalisée à Quito. Comme on le verra plus en détail dans la troisième partie, son déroulement comprendra les étapes suivantes :

1. Sélection d'un échantillon de "manzanas" (pâtés de maisons à Quito) et collecte d'informations morphologiques et démographiques sur cet échantillon.
2. Application des outils statistiques et informatiques mis au point à Marseille à ce corpus de données et réponse aux questions restant posées.
3. Définition complète du plan de sondage à mettre en oeuvre lors de l'enquête démographique finale à Quito et des techniques adéquates pour sa réalisation sur document satellitaire.
4. Sélection de l'échantillon et réalisation de l'opération de collecte, contrôle, codification et saisie des données collectées.
5. Développement du programme informatique permettant l'extrapolation des données de l'échantillon aux domaines d'estimations retenus : domaines, strates, ensemble de l'agglomération.
6. Exploitation de l'enquête, production d'estimations de caractères démographiques et d'indicateurs de migrations sur la ville de Quito, estimation de la précision des données produites, conclusion sur l'efficacité réelle de la méthode mise au point.

Ce sont ces différentes phases du travail réalisé à Quito, qui doivent conduire au résultat final de cette recherche, qui font l'objet des troisième et quatrième parties.

TROISIEME PARTIE

DEFINITION ET MISE EN ŒUVRE D'UN PLAN DE SONDAGE AREOLAIRE

SUR IMAGE SPOT POUR UNE ENQUETE DEMOGRAPHIQUE A QUITO

INTRODUCTION

Nous abordons, dans cette troisième partie, la phase décisive de notre recherche : l'application de la méthode à une ville d'un pays en développement, la ville de Quito. Toutefois, à l'issue de l'expérience Marseillaise que nous avons décrite dans la deuxième partie, cette méthode est loin d'être entièrement définie. La phase d'application à Quito va donc débiter par la réalisation d'enquêtes préliminaires dont l'exploitation nous permettra de parvenir à la définition complète du plan de sondage. En conclusion de la seconde partie, nous définissons ainsi les principales étapes du travail à réaliser à Quito :

1. Sélection d'un échantillon de pâté de maisons à Quito et collecte d'informations morphologiques et démographiques sur cet échantillon.
2. Application des outils statistiques et informatiques mis au point à Marseille à ce corpus de données pour fournir des réponses aux questions restant posées.
3. Définition complète du plan de sondage à mettre en oeuvre lors de l'enquête démographique finale à Quito et des techniques adéquates pour sa réalisation sur document satellitaire.
4. Sélection de l'échantillon et réalisation de l'opération de collecte, codification, saisie et contrôle des données collectées.
5. Développement du programme informatique permettant l'extrapolation des données de l'échantillon aux domaines d'estimations retenus : domaines d'étude, strates, ensemble de l'agglomération.
6. Exploitation de l'enquête, production d'estimations de caractères démographiques et d'indicateurs de migrations sur la ville de Quito, estimation de la précision des données produites, conclusion sur l'efficacité réelle de la méthode mise au point."

Ce sont les quatre premiers points que nous aborderons dans cette troisième partie et les annexes lui correspondant. Comme dans la seconde partie, l'exposé suivra la chronologie définie par les étapes du travail.

Le chapitre 9, après avoir détaillé les objectifs et le plan d'expérience à Quito, décrit les opérations de collecte correspondant aux deux enquêtes préliminaires réalisées : l'enquête "morphologie" et l'enquête "bâti/population". Dans le chapitre 10, on rend compte de la première exploitation des résultats de l'enquête "bâti population", dont le but est la définition de la stratégie de sondage à appliquer aux unités primaires, les pâtés de maisons. Puis, dans le chapitre 11, la même enquête est utilisée pour aborder le second degré du sondage, l'échantillonnage des ménages au sein des pâtés de maisons. Enfin, le chapitre 12 présente les objectifs et le questionnaire de l'enquête finale, définit dans son ensemble le plan de sondage et décrit les différentes étapes de sa mise en oeuvre.

Chapitre 9

OBJECTIFS GENERAUX, PLAN D'EXPERIENCE, REALISATION DES ENQUETES PRELIMINAIRES

9.1. Objectifs généraux, plan d'expérience¹

Les principales étapes du travail à Quito, décrites dans la Figure 9.1, sont celles qui auraient lieu, dans une phase opérationnelle, pour toute application de la méthode :

- *Enquête sur la morphologie urbaine*, où sont collectées des informations précises sur l'occupation du sol dans un échantillon d'îlots de Quito, afin d'initialiser l'interprétation des images satellite devant servir de base de sondage pour l'enquête démographique finale.

- *Enquête démographique légère*, sur un sous-échantillon des îlots enquêtés du point de vue morphologique, afin d'analyser les relations bâti / population et de définir la stratégie de sondage à appliquer à l'enquête finale.

- *Stratification sur image satellite et tirage de l'échantillon d'enquête*

- *Collecte démographique* : mesure des effectifs de population et analyse des systèmes de résidence

- *Exploitation de l'enquête*, production des données démographiques par estimation

Jusqu'à la production de données démographiques lors de l'enquête finale, nous n'utilisons que les documents de base susceptibles d'exister dans toute ville des pays en développement : image satellite, et plans de la ville, pas nécessairement actualisés. Les autres types d'information existant à Quito, mais généralement pas dans d'autres villes, ne seront consultés qu'à la fin de l'expérimentation, afin d'examiner la précision de nos résultats.

Cette vérification ne pourra d'ailleurs pas être très précise, étant donné le décalage de temps entre le dernier recensement (1982), la dernière couverture aérienne (1984) et nos propres observations (1986-1987). Néanmoins, nous pourrons ainsi contrôler, au moins grossièrement :

- l'interprétation des images satellite, en utilisant comme référence les photographies aériennes,

- les estimations d'effectifs démographiques, par rapport à des projections sous forme de "fourchettes" des chiffres du recensement de 1982. Un deuxième passage, en Novembre 1987, sur l'échantillon de l'enquête Bâti/population nous sera d'une aide précieuse pour asseoir les projections des chiffres du recensement de 1982.

¹ : Ce paragraphe (§ 9.1) est repris intégralement de l'annexe 21 du rapport ATP : "Plan d'expérience sur Quito", F.DUREAU, décembre 1986, Paris, 3 p.

Figure 9.1 - Plan d'expérience sur Quito

Enquête Morphologie Urbaine
 Echantillon d'îlots, par tirage systématique
 sur cartes INEC 1/10 000 (taux de recensement = 3 %)
 Collecte d'informations sur l'occupation du sol
 de chacun des îlots de l'échantillon (description
 de chaque bâtiment et des portions d'îlots non
 bâties)

Exploitation des données morphologiques
 Fichier décrivant les bâtiments — Typologie de bâtiments
 Fichier décrivant les îlots — Typologie des îlots

**Collecte d'informations démographiques sur un
 sous-échantillon d'îlots déjà enquêtés d'un point
 de vue morphologique**
 Ce sous-échantillon est composé des différents types
 d'îlots identifiés précédemment

Analyse des relations bdtu / population
 sur le sous-échantillon d'îlots enquêtés sur les
 plans morphologique et démographique

Définition des unités statistiques à enquêter
 Tests menés sur le sous-échantillon d'îlots enquêtés
 sur les plans morphologiques et démographiques

*Les variables morphologiques devant servir
 de base à la stratification*

Le type d'unité statistique à enquêter

Tirage de l'échantillon d'enquête

Enquête démographique
 Objectifs : - mesure des effectifs de population
 - analyse des systèmes de résidence
 1) Collecte des données
 2) Exploitation systématique de l'enquête
 3) Production des données par estimation

**Comparaison des résultats avec les sources
 de données démographiques existantes**

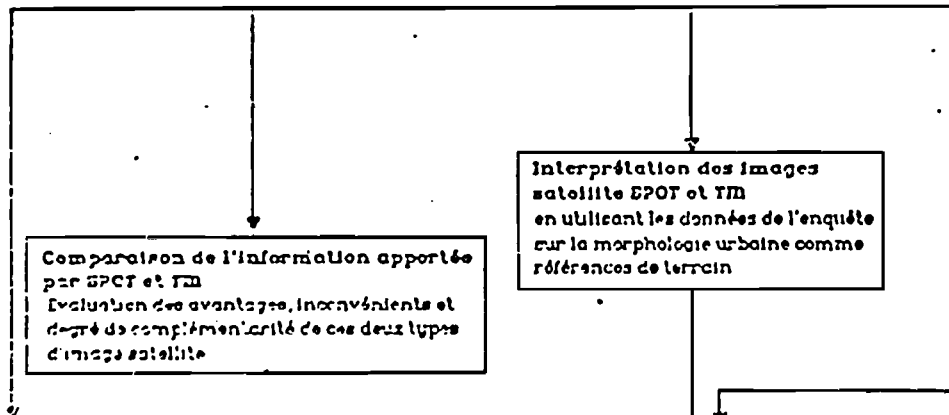
**Interprétation des images
 satellite SPOT et TM**
 en utilisant les données de l'enquête
 sur la morphologie urbaine comme
 références de terrain

**Zonage de l'agglomération en zones
 de caractéristiques morphologiques
 homogènes**

**Comparaison de l'information apportée
 par SPOT et TM**
 Evaluation des avantages, inconvénients et
 degré de complémentarité de ces deux types
 d'image satellite

Test de la stéréoscopie permis par SPOT
 Evaluation de la précision, et de l'intérêt de cette
 information sur l'élévation du bdtu

**Comparaison des résultats avec les sources de
 données sur la morphologie urbaine existantes
 (notamment photographies aériennes, et plans au
 1/1000 de l'agglomération)**



Dans la conception générale du programme de recherche, nous avons choisi de réaliser sur le site de Marseille les développements méthodologiques. Néanmoins, pour les deux axes de recherche poursuivis dans la seconde partie de l'expérience : "sondages spatiaux" et "interprétation d'images satellite", quelques points de méthode non abordés à Marseille doivent faire l'objet de travaux spécifiques à Quito :

- *la définition des unités statistiques à enquêter et plus généralement, du plan de sondage à appliquer à l'enquête finale.* A Marseille, du fait que nous utilisons des données du recensement pour calculer la variance des estimations, nous sommes dans l'obligation de travailler sur les districts de recensement définis par l'INSEE, c'est à dire un découpage spatial résultant de critères administratifs, démographiques et physiques. La prise en considération de ces différents paramètres aboutit à un découpage spatial en unités de taille très variable, et ne correspondant pas systématiquement à l'ilot urbain au sens physique du terme, c'est à dire le pâté de maisons. D'autre part, les données étant agglomérées par district, elles ne permettent pas, comme on l'a déjà dit pendant l'expérience marseillaise¹, d'évaluer les possibilités offertes par les sondages à plusieurs degrés. A Quito, les deux enquêtes préliminaires fournissent sur un échantillon d'ilots des informations morphologiques et démographiques, au niveau de chaque bâtiment pour les premières, et pour chaque ménage pour les secondes. On veut, grâce à ces données, tester des procédures de sondage à plusieurs degrés qui, tout en restant praticables et en offrant de faibles risques d'erreur lors de la collecte, améliorent la précision des estimations démographiques.

- *La stéréoscopie sur image satellite en milieu urbain :* le satellite SPOT, du fait des variations possibles de l'angle de visée, permet la vision stéréoscopique. Nous testerons sur Quito l'intérêt de cette nouvelle information en télédétection urbaine, son degré de précision pour l'appréhension des hauteurs de bâti.

- *Comparaison de l'information apportée par SPOT et par TM :* disposant de ces deux types d'image à des dates très voisines, respectivement juin, septembre et novembre 1986, nous pouvons évaluer les avantages et inconvénients de chacune de ces sources d'information et leur degré de complémentarité.

Dans les chapitres suivants, c'est le premier point qui sera développé ; outre cet aspect méthodologique, le travail préparatoire réalisé à Quito avant l'enquête finale a aussi pour objectif la vérification de la stabilité des principaux résultats obtenus à Marseille.

9. 2 Réalisation de l'enquête morphologie²

9.2.1 Sélection de l'échantillon

Pour la sélection de l'échantillon de l'enquête morphologie, comme pour toute l'expérimentation menée à Quito, nous n'utilisons que les documents de base susceptibles d'exister dans toute ville d'un pays en développement. C'est donc sur un plan de Quito non actualisé que nous avons réalisé la sélection de l'échantillon : les plans au 1/10.000 établis par l'INEC en 1982 pour le recensement général de la population.

¹ : Voir par exemple la conclusion de la deuxième partie.

² : Les paragraphes 9.2.1 à 9.2.4 sont repris de l'annexe 22 du rapport ATP : " Enquête morphologie Quito (octobre-décembre 1986) : collecte et exploitation des données", F. DUREAU et A. MICHEL, juillet 1986, Paris, 6 p.

La technique d'échantillonnage a été des plus simples ; nous avons opté pour un tirage systématique sur grille plaquée sur les plans INEC. Ayant fixé le taux de sondage à 4 %, connaissant approximativement le nombre total d'îlots de Quito et la surface totale occupée par la ville, nous en avons déduit la taille de la maille de la grille : 750 mètres de côté.

Nous avons donc tracé sur calque un carroyage de 7,5 cms de côté que nous avons superposé au plans INEC au 1/10.000. Pour chaque intersection de la grille, on incluait dans l'échantillon l'îlot qui contenait cette intersection.

Dans le cas où une intersection de la grille tombait sur une rue, une règle systématique était appliquée pour déterminer un îlot voisin à inclure dans l'échantillon ; si nous n'avions pas procédé ainsi, le taux de sondage aurait été inférieur à 4 %.

Avec ce mode de tirage, la probabilité pour un îlot d'appartenir à l'échantillon est proportionnelle à sa surface ; de ce fait, certains très grands îlots ont été tirés plusieurs fois (ils contenaient plusieurs intersections de la grille). Nous avons pratiqué un tirage "sans remise", en reportant les intersections appartenant à un même îlot sur des îlots voisins¹.

Nous avons ainsi constitué un échantillon de 195 îlots. Les cartes de l'INEC, datant de 1982, ne couvrent pas les extensions de la ville de 1982 à 1986 ; il a donc fallu compléter notre échantillon tiré systématiquement sur document cartographique par quelques îlots de la périphérie dont le type morphologique n'était pas représenté dans l'échantillon initial. Cet ajout de 4 îlots s'est fait en fin d'enquête, une fois bien connues par l'équipe les caractéristiques morphologiques de l'échantillon enquêté et celles des extensions récentes de Quito.

Au total, notre échantillon compte donc 199 îlots.

9.2.2. Elaboration des documents d'enquête

Comme l'indique le manuel d'instructions aux enquêteurs², chaque îlot de l'échantillon est décrit par trois types de documents :

- *Carte* : sur le fond de plan INEC agrandi au 1/500, sur lequel ne figurent au départ de l'enquête que les limites de chaque îlot à enquêter, les enquêteurs établissent la cartographie schématique de l'occupation du sol de l'îlot. Tous les bâtiments ainsi que tous les éléments non bâtis doivent figurer sur ces plans.

- *Fiche de description des bâtiments* : chacun des édifices dessinés sur la carte est décrit sur cette fiche. Il est repéré par un numéro de 1 à n.

- *Photographies* : pour chaque îlot, les chercheurs de l'Orstom font des photographies depuis un point surplombant l'îlot (étage élevé d'un immeuble ou relief), et au sol, en bordure et/ou à l'intérieur de l'îlot. Ces photographies sont destinées à préciser et contrôler l'information recueillie par les enquêteurs.

La connaissance du bâti et du non bâti acquise lors de diverses sorties sur le terrain effectuées à Quito nous a permis de modifier en l'adaptant la fiche d'enquête conçue et expérimentée à Marseille ; les conseils de collègues de l'équipe Atlas connaissant bien la ville nous ont aidé à préciser les termes à employer pour caractériser les différents matériaux de construction utilisés à Quito et à adapter les fiches en fonction des

¹ : De ce fait, la probabilité d'appartenance n'est plus réellement contrôlée, mais cela importe peu puisque l'objectif de l'opération n'est pas d'effectuer une inférence rigoureuse à partir de cet échantillon.

² : Voir annexe A.3.1.

possibilités et des conditions d'enquête propres à ce site. Une relecture de ces fiches a ensuite été faite par un architecte équatorien.

Nous avons testé avec les enquêteurs recrutés pour l'enquête le caractère opérationnel de ces fiches à l'occasion d'une première sortie commune sur le terrain. La méthode de collecte a été décrite avec précision dans le manuel d'instructions aux enquêteurs, afin qu'ils puissent s'y référer en cas d'ambiguïté.

9.2.3 Collecte des données

La procédure de collecte des données étant décrite précisément dans les points 3 et 4 du manuel d'instructions aux enquêteurs, nous ne donnerons ici que quelques précisions sur les temps et conditions de collecte.

Temps de collecte

La collecte des données (remplissage des questionnaires et établissement des schémas sur les plans au 1/500) a été réalisée par deux enquêteurs entre le 1er Octobre et le 9 Décembre 1986 ; la moyenne d'un îlot par demi-journée, estimée au début de l'enquête d'après notre expérience marseillaise, s'est donc avérée exacte.

Quant aux photographies prises depuis des immeubles ou des points hauts de la ville, elles ont nécessité environ 20 matinées de travail à deux personnes ; 10/12 îlots par demi-journée semble le maximum qui puisse être fait, compte tenu des temps de déplacement et de discussion pour avoir accès aux sommets des immeubles.

Conditions de collecte

Les enquêteurs n'ont pas eu de difficulté à collecter l'information : ce type d'enquête ne réclame que peu de participation de la population, puisqu'elle ne nécessite que l'autorisation d'accéder aux terrasses des maisons ou immeubles. Pour tous les îlots, cette autorisation a pu être obtenue sans problème majeur, quelque soit le type de quartier de Quito.

9.2.4 Codification, saisie, contrôles de cohérence et recodage

Codification

La codification, réalisée par F. DUREAU sur des fiches du type de celle présentée en annexe A.3.1, a été établie à partir de toutes les données collectées au cours de l'enquête : questionnaires, plans au 1/500, photographies. Ces photographies se sont révélées fondamentales pour préciser l'information collectée par les enquêteurs : ce sont des documents qu'il faut s'efforcer de recueillir lors de toute collecte de données-terrain en milieu urbain. Outre leur rôle de contrôle du travail des enquêteurs, ils apportent l'information qualitative complémentaire des données quantitatives des questionnaires simplifiés que nous utilisons.

Le principe général de la fiche de codification est le suivant : constituer deux fichiers, l'un décrivant les bâtiments, l'autre les "manzanas" (C'est à dire les pâtés de maisons). Le fichier des bâtiments ne décrit pas les édifices un par un, mais par groupes présentant des caractéristiques morphologiques semblables ou très voisines ; ce sont ces groupes que nous appelons "genres".

Saisie

Les fiches de codification ont été saisies sous DBASEIII, sur l'IBM AT, en une dizaine d'heures à deux personnes.

A l'issue de cette phase de saisie, nous disposons de deux fichiers :

QMANZANA : fichier descriptif des 199 manzanas enquêtées (recto de la fiche de codification) champs 1 à 3 : identification manzana
champ 4 : identification enquêteur
champs 5 à 7 : mesure de surface
champs 8 à 10 : typologies (sources exogènes, estimations enquêteur et codificateur)
champs 11 à 42 : description de l'espace bâti

QBATIMEN : fichier descriptif des 2134 genres de bâtiments (verso de la fiche de codification)
champ 1 : identification manzana
champ 2 : identification du genre dans la manzana
champ 3 : nombre de bâtiments du genre dans la manzana
champs 4 à 13 : caractérisation du genre

Contrôles de cohérence ; redressement des mesures de bâtiments

L'analyse des tableaux de fréquence et histogrammes nous ont permis de corriger quelques erreurs de codification ou de saisie.

Le point principal sur lequel ont porté les corrections a été la surface des bâtiments.. Il avait été demandé aux deux enquêteurs de mesurer les dimensions au sol des bâtiments en comptant leurs pas, après avoir étalonné ceux-ci en effectuant leurs premières mesures avec un décamètre. La comparaison des histogrammes des surfaces des bâtiments de chacun des enquêteurs, sur un échantillon de bâtiments censés avoir sensiblement les mêmes dimensions, nous a montré que l'un des deux enquêteurs avait sous-estimé les surfaces d'environ 20 % par rapport à l'autre enquêteur (ce dernier ayant une bonne appréciation des distances). Nous avons donc corrigé le fichier des bâtiments en augmentant de 20 % toutes les surfaces mesurées par l'enquêteur n° 1. Les surfaces des bâtiments ont ensuite été réparties en 10 modalités.

Recodages

Au vu des histogrammes, deux autres variables ont été seuillées : les surfaces de bâtiments (7 classes) et le nombre de pans du toit (5 classes).

9.2.5. Exploitation de l'enquête morphologie en vue du tirage d'un échantillon représentatif¹

Afin de pouvoir sélectionner un échantillon représentatif pour l'enquête sur les relations bâti/population, il est nécessaire de définir une typologie des bâtiments identifiés et décrits lors de l'enquête morphologie, à partir de laquelle on peut caractériser l'espace bâti des manzanas et classer ces dernières.

La figure 9.2 retrace les différentes étapes de la classification des bâtiments, puis des manzanas; les figures 9.3 et 9.4 précisent le contenu des fichiers soumis aux traitements.

¹ : Ce paragraphe est repris de l'annexe 23 du rapport ATP : "Schema de la démarche de classification des 199 Manzanas de l'enquête morphologie de Quito", F. DUREAU et O. BARBARY, mai 1987, Paris, 6 p.

La classification des bâtiments en 38 types (décrits dans le tableau 9.5) est obtenue par segmentation sur 3 variables : la surface du sol, le nombre d'étages, et le matériau du toit. Les bornes des deux premières variables ont été définies d'après les résultats d'une classification automatique réalisée, avec le logiciel SPSS, sur les 4595 bâtiments regroupés en 352 "super-genres", caractérisés par leur surface au sol (10 modalités) et leur nombre d'étages (7 modalités).

La classification des manzanas est basée sur l'analyse du tableau de contingence caractérisant l'espace bâti de chaque manzana, en fonction des 38 types définis précédemment : la case courante k_{ij} du tableau contient le pourcentage de surface bâtie dans l'îlot i occupée par les bâtiments de type j .

En combinant les résultats de la classification ascendante hiérarchique réalisée sur ce tableau de contingence (17 classes), la densité du bâti, et un indicateur de l'homogénéité des bâtiments au sein de chaque îlot, on obtient une classification finale en 9 classes, décrites dans le tableau 9.6.

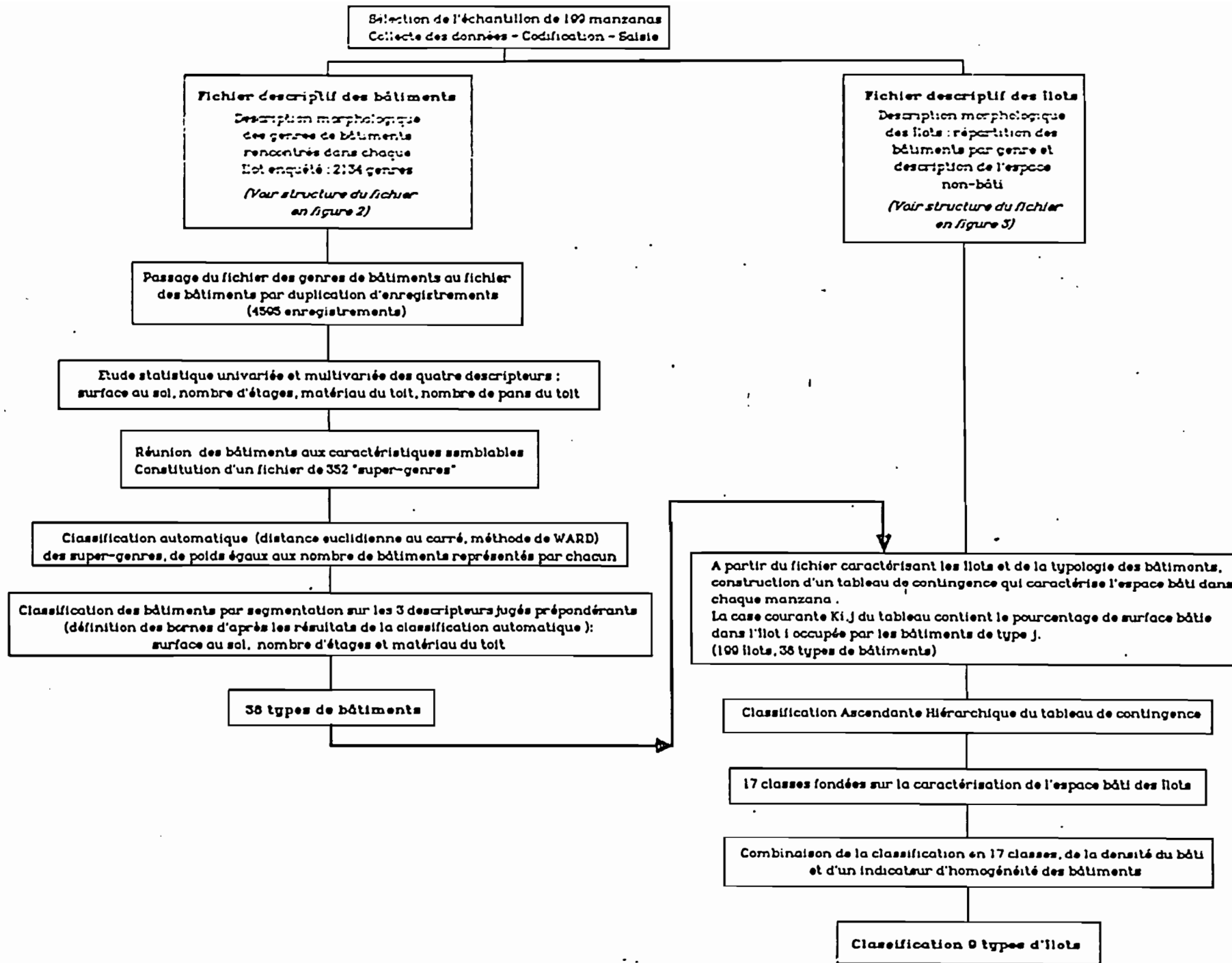


Figure 9.2 - Schéma de la démarche de classification

Figure 9.3 - Structure du fichier descriptif des bâtiments

```
Structure du fichier      : a:qbatiment.dbf
Nombre d'enregistrements: 2134
Dernière mise à jour    : 31/03/87
Champ  Nom champ      Type      Dim      Dec
  1  NOMANZANA      Numérique  3
  2  GENRE          Numérique  2
  3  NBATI          Numérique  2
  4  MATERIAU       Numérique  1
  5  FORMEMAT       Numérique  1
  6  COULEUR        Numérique  2
  7  NBPANS         Numérique  2
  8  NBETAGES       Numérique  2
  9  SURFACEBAT     Numérique  5
 10  SURFBATC       Numérique  5
 11  SURGENRE       Numérique  7
 12  TYPEBATI       Numérique  2
 13  NOMENQ         Numérique  1
 14  SGENRE         Numérique  5
** Total **                41
```

Figure 9.4 - Structure du fichier descriptif des flots

Structure du fichier : a:gmanzana.dbf
 Nombre d'enregistrements: 199
 Dernière mise à jour : 03/04/87

Champ	Nom champ	Type	Dim	Dec
1	NOMANZANA	Caractère	3	
2	NOINEC	Caractère	7	
3	PARTINEC	Caractère	1	
4	ENQETEUR	Caractère	1	
5	SURFILOT	Numérique	6	
6	SURF1_2VOI	Numérique	6	
7	SURFTOTALE	Numérique	6	
8	TYPECAZA	Caractère	2	
9	TYPEENQUET	Caractère	2	
10	TYPEPHOTO	Caractère	2	
11	SURFBATI	Numérique	6	
12	DENSBATI	Caractère	1	
13	IMPLANTBAT	Caractère	2	
14	NBBATITOTA	Numérique	3	
15	NBBATIFINI	Numérique	3	
16	NBBATICON	Numérique	3	
Appuyez sur une touche pour continuer...				
17	NBBATI1	Numérique	2	
18	NBBATI2	Numérique	2	
19	NBBATI3	Numérique	2	
20	NBBATI4	Numérique	2	
21	NBBATI5	Numérique	2	
22	NBBATI6	Numérique	2	
23	NBBATI7	Numérique	2	
24	NBBATI8	Numérique	2	
25	NBBATI9	Numérique	2	
26	NBBATI10	Numérique	2	
27	NBBATI11	Numérique	2	
28	NBBATI12	Numérique	2	
29	NBBATI13	Numérique	2	
30	NBBATI14	Numérique	2	
31	NBBATI15	Numérique	2	
32	NBBATI16	Numérique	2	
Appuyez sur une touche pour continuer...				
33	NBBATI17	Numérique	2	
34	PENTETOIT	Caractère	1	
35	NBRATIPENT	Numérique	3	
36	NBBATIFER	Numérique	3	
37	NBBATICHEM	Numérique	3	
38	VEGEBALCON	Caractère	1	
39	DIRPENTDOM	Caractère	1	
40	NIVPENTDOM	Caractère	1	
41	DIRPENTSDO	Caractère	1	
42	NIVPENTSDO	Caractère	1	
43	TYPEFEAU	Caractère	1	
44	SURFEAU	Numérique	6	
45	VEGEBHTYDO	Caractère	1	
46	VEGEBHTYSO	Caractère	1	
47	VEGEBHSURF	Numérique	6	
48	VRQKCHTYDO	Caractère	1	

Appuyez sur une touche pour continuer

49	VEGECBTYSO	Caractère	1	
50	VEGECBSURF	Numérique	6	
51	VOIRIETYDO	Caractère	1	
52	VOIRIETYSD	Caractère	1	
53	SURFVOIRIE	Numérique	6	
54	MURDSILOT	Caractère	1	
55	MURHSILOT	Numérique	3	
56	TROTTOIRDO	Caractère	1	
57	TROTTOIRSD	Caractère	1	
58	CHASSEEDO	Caractère	1	
59	CHASSEESD	Caractère	1	
60	PELOUSE	Numérique	3	
61	ARBRES	Numérique	3	
62	ARBUSTES	Numérique	3	
63	VOITURESTA	Numérique	3	
64	VOITURECIR	Numérique	3	
Appuyez sur une touche pour continuer				
65	TYPE17	Numérique	2	
** Total **			160	

Tableau 9.5 - Typologie des bâtiments

Surface au sol	Nombre D'étages	Béton ou Eternit	Tuiles vernies, Carrelage, Shingle	Zinc	Tuiles non vernies
1 - 29 m ²	1 ét.	1	1	1	1
30 - 79 m ²	1 ét.	3	12	21	30
80 - 224 m ²	1 ét.	4	13	22	31
225 - 14999 m ²	1, 2 ét.	5	14	23	32
1 - 29 m ²	2, 3 ét.	6	15	24	33
30 - 79 m ²	2, 3 ét.	7	16	25	34
80 - 224 m ²	2, 3 ét.	8	17	26	35
225 - 14999 m ²	2, 3 ét.	9	18	27	36
30 - 79 m ²	4, 5 ét.	10	19	28	37
80 - 224 m ²	4, 5 ét.	11	20	29	38
225 - 14999 m ²	5 à 15 ét.	2	2	2	2

(Dans chaque case du tableau, est noté le numéro du type de bâtiment, selon la surface au sol, le nombre d'étages et le matériau du toit)

Tableau 9.6 - Typologie des îlots

Type	Intitulé du type	Description morphologique du type	Nombre d'îlots
1	Non bâti	<i>Densité du bâti nulle</i>	6
2	Péri-urbain	<i>Densité du bâti très faible</i> (bâtiments petits ou moyens, tous matériaux sauf tuiles vernies)	59
3	Habitat populaire récent (1 : à étages)	Bâtiments de taille petite ou moyenne, <i>Rdc +1 ou 2 étages</i> , toiture en béton ou éternit, <i>peu de tuiles vernies</i> ou non (densité du bâti faible à très forte)	19
4	Habitat populaire récent (2 : sans étage)	Bâtiments de taille petite ou moyenne, <i>Rdc seul</i> très dominant, toiture en béton, éternit, zinc, <i>peu de tuiles vernies</i> ou non (densité du bâti faible à très forte)	20
5	Vieux centre	Bâtiments taille moyenne ou grande, <i>Rdc + 1 ou 2 étages</i> , <i>% important de toitures en tuiles non vernies</i> (densité du bâti faible à très forte)	29
6	Habitat de standing	Bâtiments <i>taille moyenne</i> (80, 225m ²), <i>Rdc, Rdc +1 ou 2 ét.</i> , <i>% important de toiture en carrelage ou tuiles vernies</i> , pas de zinc (densité du bâti moyenne)	22
7	Industries et équipements	<i>Très grands bâtiments bas</i> à toiture en béton, éternit ou zinc (densité du bâti faible ou moyenne)	21
8	Immeubles	<i>Grands bâtiments élevés</i> à toiture surtout en béton et éternit (densité du bâti faible ou moyenne)	12
9	Viviendas	<i>Densité du bâti forte ou très forte</i> , <i>très grande homogénéité</i> des bâtiments (bas et de grande taille)	11
TOTAL			199

9.3. Réalisation de l'enquête "Relations bâti/population"¹

9.3.1 Sélection de l'échantillon

Comme il l'a été vu dans le paragraphe précédent, les 199 îlots et les 2134 genres de bâtiments enquêtés lors de l'enquête morphologie ont fait l'objet d'une classification en, respectivement 9 types d'îlots et 38 types de bâtiments.

Etant donnés les objectifs assignés à l'enquête bâti/population, trois critères ont servi à sélectionner les îlots de l'échantillon, parmi les 199 déjà enquêtés sur le plan morphologique :

- le type de bâtiments
- le type d'îlot
- la localisation au sein de la ville de Quito.

Cette sélection s'est faite par choix raisonné dans la base des 199 îlots, stratifiée par type d'îlot. Le tirage a été réalisé à partir du listing décrivant la composition par types de bâtiments des 199 îlots et des cartes de localisation des îlots de l'enquête morphologie. Le nombre d'îlots à tirer dans chaque strate (type) a été défini par l'application d'un taux sensiblement constant (environ 25 %) ; pour chaque type d'îlot, nous nous sommes efforcés de choisir des îlots à la fois bien représentatifs des caractéristiques de leur strate et offrant la plus grande variété possible de types de bâtiments et de localisations dans Quito.

Type d'îlot	Nombre d'îlots Enq. morphologie	Nombre d'îlots Enq. bâti/population	Numéros des îlots Enq. bâti/population
1 (non bâti)	6	0	-
2	59	15	158,166,256,164,216, 146,120,142,326,362, 361,319,251,148,352,
3	19	6	147,214,228,204,345 263
4	20	7	208,245,110,329,318, 342,230
5	29	8	203,121,237,249,223, 232,156,248
6	22	6	111,266,311,324,246 347
7	21	4	130,217,304,344
8	12	4	212,250,262,307
9	11	4	143,257,258,322
TOTAL	199 (196 bâtis)	54	

¹ : Ce paragraphe est repris de l'annexe 24 au rapport ATP : "Enquête Bâti / Population (avril 1987) : collecte et exploitation des données", F. DUREAU, juillet 1986, Paris, 6 p.

9.3.2 Elaboration des documents d'enquête

Pour collecter les données de l'enquête bâti/population, chaque enquêteur devait remplir deux types de documents (voir manuel d'instructions aux enquêteurs en annexe A.3.1) :

- la fiche logement, pour décrire les personnes résidant dans les logements des bâtiments faisant partie de l'îlot ;
- la fiche récapitulative, pour rassembler les informations générales relatives aux logements appartenant à l'îlot.

Les données socio-démographiques devant être mises en relation avec les données morphologiques de l'enquête précédente, au niveau du bâtiment comme au niveau de la manzana, il importait qu'il y ait concordance entre les deux enquêtes sur l'identification des manzanas et des bâtiments au sein de chaque manzana ; c'est pourquoi chaque enquêteur recevait des copies des questionnaires et plans de l'enquête morphologie, et était tenu de conserver la même délimitation et la même numérotation de bâtiments.

Les thèmes abordés par le questionnaire de la fiche logement ont été choisis par rapport à l'objectif de l'enquête finale du dernier trimestre 1987 (estimation d'effectifs et analyse des systèmes résidentiels) : outre les caractéristiques démographiques classiques (âge, sexe ...etc), l'accent a été mis sur la saisie de la résidence, le statut d'occupation du logement et les revenus.

La formulation des questions a été soumise à l'avis de différentes personnes ayant une expérience solide de la collecte des données dans les villes équatoriennes.

9.3.3 Collecte des données

La procédure de collecte étant décrite précisément dans les points 3 et 4 du manuel d'instructions aux enquêteurs (annexe A.3.1), nous ne donnerons ici que quelques précisions sur les temps et conditions de collecte.

Temps de collecte

La collecte a été réalisée par 5 enquêteurs, encadrés par un superviseur, entre le 11 Avril et le 8 Mai (compte tenu des jours fériés, 3 semaines effectives de travail pour 4 des enquêteurs et le superviseur ; et 2 semaines de travail pour le dernier enquêteur). Environ 1700 logements ont pu être enquêtés durant cette période, soit une moyenne de 17 questionnaires remplis par enquêteur et par jour ; en tenant compte des interdictions d'accès à certains bâtiments qu'ont subies les enquêteurs, le temps d'enquête a donc été plus long que ce que nous avons estimé a priori (25 logements par jour et par enquêteur).

Conditions de collecte

Cette seconde enquête s'est déroulée dans un contexte bien particulier, lié au séisme qu'a connu l'Equateur le 5 Mars 1987. Du fait de la rupture de l'oléoduc (le pétrole est la principale source de revenus de l'Etat Equatorien), l'économie nationale s'est trouvée complètement désorganisée durant la période qui a suivi le tremblement de terre : augmentation immédiate du prix du carburant (multiplié par 2), rations de carburant, augmentation des prix pour de nombreux produits de base...etc. C'est donc dans un contexte économique et social très tendu que nous avons réalisé notre enquête.

Contrairement à l'enquête morphologie, l'enquête bâti/population supposait une participation de la population pour remplir le questionnaire logement. Le taux de non-réponse est très variable selon les îlots.

Un premier constat, bien que très classique, mérite d'être rappelé : le degré de participation à l'enquête est inversement proportionnelle aux revenus des habitants. Dans les quartiers populaires de Quito de nombreux îlots ont été enquêtés à 100 % et les taux de non-réponses se situent toujours en dessous de 10 % (nombre de logements non enquêtés/nombre de logements total de l'îlot) : ces non-réponses ne correspondent pas à des refus, mais à des logements toujours trouvés vides malgré 4 ou 5 visites de l'enquêteur, leurs occupants s'absentant longuement de leur domicile pour aller travailler souvent très loin.

Par contre, dans les quartiers de moyen et haut standing les taux de non-réponses (là, il s'agit essentiellement de refus de réponse) sont beaucoup plus élevés 15 à 30 % le plus souvent. Nous avons même dû renoncer à l'enquête dans un îlot de standing très élevé : malgré une demi-journée de discussion et avec lettres officielles à l'appui, nous nous sommes fait interdire l'accès des immeubles par les administrateurs de la quasi totalité des tours constituant l'îlot à enquêter, et, dans la seule tour à laquelle nous avons pu accéder, ce sont les résidents eux-mêmes qui ont claqué la porte au nez de l'enquêtrice ...

Il faut d'autre part noter, dans chaque îlot, une certaine homogénéité dans les comportements des résidents vis à vis des enquêteurs (en rapport, bien évidemment, avec l'homogénéité de la population elle-même) d'où des taux de non-réponses très tranchés : 10 % dans un îlot "sans problème", et 70 % dans un autre îlot peu éloigné du premier, mais ayant la particularité, par exemple, d'être majoritairement occupé par des militaires et policiers à la retraite.

Ces taux de non-réponses élevés ne nous posent pas de problèmes majeurs étant donné le type d'exploitation des données que nous réalisons pour cette enquête bâti/population. Par contre, s'il en était de même pour l'enquête finale, ce serait plus fâcheux, mais rappelons ici que c'est l'un des intérêts du sondage à deux degrés que de permettre des solutions de substitution en cas de non réponse.

L'on ne peut certainement pas mettre ces problèmes de non-réponses au compte des enquêteurs : d'une part, des passages supplémentaires que nous avons effectués nous-même ont confirmé l'efficacité de leur travail ; d'autre part, nos propres enquêteurs ont eu des taux de réponse satisfaisants sur des îlots réputés difficiles à enquêter, et ayant posé les mêmes problèmes aux enquêteurs du recensement (alors que les gens sont obligés de rester chez eux durant le recensement, et que le certificat de réponse au recensement est indispensable pour nombre de démarches administratives). Il semble donc que le climat tendu qui régnait à Quito au moment de l'enquête a joué contre les enquêteurs ; le contexte social lors de l'enquête morphologie était nettement plus favorable.

Qualité de la collecte

Les premiers contrôles des questionnaires faits dès leur remise par les enquêteurs, et ceux réalisés au cours de la codification ont montré une bonne cohérence des données collectées.

Un problème particulier, apparu au cours de la collecte, mérite toutefois d'être signalé, étant donné l'objectif de notre enquête : il s'agit de la difficulté qu'ont eue les enquêteurs dans certains quartiers construits en continu comme le centre historique de Quito, à identifier les bâtiments repérés par leurs collègues de l'enquête morphologie. Pour la première enquête, c'est une définition physique du bâtiment vu du dessus qui primait : dans un dédale de couloirs serpentant à l'intérieur de vieux bâtiments accolés, où rien ne traduit la limite des différents bâtiments, il est bien évidemment difficile de savoir dans quel bâtiment précis l'on se trouve, sous quelle toiture ! Dans ce cas précis, nous avons donc pris le parti d'affecter la population enquêtée à un groupe de bâtiments, sans chercher à identifier précisément si tel logement appartenait à un bâtiment ou à son voisin, qui a, dans la quasi-totalité des cas, les mêmes caractéristiques morphologiques.

Ces difficultés à identifier les bâtiments ne se sont produites que dans le cas du vieux centre de Quito ; dans tous les autres types d'implantation spatiale du bâti, il n'y a pas eu de problèmes à retrouver les bâtiments figurant sur les plans au 1/500 établis lors de l'enquête morphologie.

9.3.4 Codification et saisie

Deux types de codification ont été réalisées, générant deux fichiers :

- codification des caractéristiques individuelles notées sur les fiches logement =====> fichier où chaque enregistrement correspond à un individu enquêté (fichier QDEMINDI, comptant près de 5000 enregistrements) ;
- codification des caractéristiques des logements et des bâtiments notées sur les fiches logement et sur les fiches récapitulatives =====> fichier où chaque enregistrement correspond à un logement enquêté (fichier QDEMVI, comptant environ 1700 enregistrements).

Cette structuration des données en deux fichiers permet de ne pas dupliquer les informations relatives au logement pour chaque individu résidant dans celui-ci ; le numéro de logement existant dans les deux fichiers, toutes les données peuvent être mises en relation par le système de gestion de données DBASEIII, sans qu'il y ait redondance de l'information.

La codification des fiches logement a été réalisée par un étudiant, qui a entamé son travail une semaine après le début de l'enquête ; il a été aidé durant sa dernière semaine de travail par le superviseur. Ce dernier, plus qualifié, a été chargé de la codification des questions sur l'activité et les revenus. De la sorte, la codification a été terminée seulement une dizaine de jours après la fin de l'enquête.

La codification des caractéristiques individuelles a été réalisée directement sur la fiche d'enquête, où des cases avaient été prévues pour cela ; cette solution permet d'éviter le report sur une fiche de codification spécifique des données recueillies directement sous forme de codes, comme l'âge, la date de naissance, le statut de résidence...etc. Cette procédure diminue nettement les temps de codification ; elle avait été choisie parce que initialement, nous pensions effectuer la saisie nous-mêmes à Quito, sur l'IBM AT avec le logiciel DBASEIII. Pour des raisons de calendrier, nous avons finalement confié cette saisie à une société de service informatique en France, où nous étions en mission durant tout le mois de Juin 1987 : la saisie directe sur les fiches d'enquête est nettement plus complexe et longue pour les opérateurs de saisie. Lorsque les travaux de saisie sont sous-traités à des sociétés spécialisées où le travail est effectué très "mécaniquement", il nous semble bien préférable d'opter pour la codification sur bordereau spécifique.

La codification des données relatives aux logements et bâtiments, très rapide, a été réalisée sur bordereau (voir modèle en annexe A.3.1).

Au terme de ces deux opérations de collecte réalisées au début de l'expérimentation à Quito, nous disposons de l'information nécessaire d'une part, à la mise au point du processus de traitement de l'image satellite devant conduire à la stratification de l'agglomération, d'autre part, à la définition complète du plan de sondage à deux degrés que nous allons appliquer à l'enquête finale. C'est ce deuxième point que nous allons aborder dans les deux chapitres suivants.

Chapitre 10

ENQUETE PRELIMINAIRE "BATI POPULATION", EXPLOITATION POUR LA MISE AU POINT DU PREMIER DEGRE DU SONDAGE : ECHANTILLONNAGE DES PATES DE MAISONS

10.1 : Introduction : Le contexte et les questions posées

La manière dont on a abordé le problème de la définition d'une stratégie de sondage à appliquer aux unités primaires à Quito résulte d'un compromis : explorer un éventail de méthodes le plus large possible tout en restant dans le cadre des contraintes fixées au départ du projet, c'est à dire d'une part, l'utilisation de l'image satellite comme base de sondage et d'autre part, le caractère rapide et peu coûteux des solutions recherchées. Nous avons, au chapitre 1, dressé une liste des conséquences qu'ont ces contraintes sur les objectifs que nous nous fixons. Nous les évoquerons maintenant de manière plus détaillée, à chaque fois qu'elles interviennent dans une question posée, une réponse apportée ou une décision prise.

A propos de la sélection de l'échantillon des pâtés de maisons et de la démarche d'inférence qui doit suivre, quatre questions principales se posent.

1. Quelle probabilité doit avoir chaque unité d'appartenir à l'échantillon ?

L'éventail des possibilités est en fait assez restreint. Un sondage équiprobable est impraticable puisqu'on ne dispose d'aucune liste des unités et que les techniques d'interprétation de l'image ne permettent pas encore de dresser une telle liste du moins avec la rapidité souhaitée. On retiendra pourtant cette alternative dans la suite comme base de comparaison et également parce qu'il n'est pas exclu qu'elle puisse être praticable dans l'avenir.

L'idée qui vient ensuite "naturellement" à l'esprit, s'agissant d'un sondage aréolaire qui s'applique à des unités de tailles variables, est celle d'une probabilité d'appartenance à l'échantillon proportionnelle à la surface de ces unités. C'est elle qui s'applique au cas d'une sélection aléatoire de points dans l'espace géographique de la base de sondage (on retient alors dans l'échantillon l'ensemble des pâtés de maisons "touchés" par ces points) et, par une approximation admise, au cas du tirage systématique à l'aide d'une grille de points.

Mais une autre idée est de rechercher une stratégie de sélection des pâtés de maisons qui fournisse une probabilité d'appartenance proportionnelle à une meilleure mesure de la taille des unités, c'est à dire à une mesure plus corrélée aux caractères démographiques que la surface totale. Il s'agirait par exemple d'une mesure approchée de la surface bâtie ou, mieux encore, de la surface habitable.

On ne peut pas encore obtenir par interprétation d'image satellite, une mesure, même approchée, de la surface habitable; par contre des classifications de l'image existent qui fournissent un découpage de l'agglomération en zones homogènes du point de vue de leur densité de bâti ou de leur densité de végétation. A partir de ces classifications et de l'information obtenue lors de l'enquête "morphologie", on peut, par étalonnage,

obtenir une valeur approchée de la densité de bâti à l'intérieur de chaque classe et donc une approximation de la surface bâtie de chaque pâté de maisons. Mais la réalisation d'un sondage suivant une telle probabilité de sélection pose un certain nombre de problèmes; les difficultés majeures sont les suivantes.

A: Dans l'étape de sélection de l'échantillon

Pour respecter lors d'un tirage systématique à l'aide d'une grille de points une probabilité d'appartenance approximativement proportionnelle à la surface bâtie de l'unité primaire, on doit penser à une stratégie qui n'accepte un pâté de maisons dans l'échantillon que lorsque le point de la grille qu'il contient est situé au dessus d'un bâtiment; dans le cas contraire l'unité est rejetée. Une telle stratégie complique considérablement la gestion du taux de sondage à travers la taille de la maille de la grille. *Mais surtout elle nécessite que le document servant de base au tirage permette de décider si un point donné est situé au dessus d'un bâtiment ou au contraire inclus dans une surface non bâtie*; à la limite, pour être totalement rigoureux, il faudrait être en mesure d'identifier chaque bâtiment du pâté de maisons. *La résolution spatiale actuelle des images satellite est bien évidemment incompatible avec une telle contrainte*. La probabilité d'appartenance qui résulterait de l'application d'une stratégie de ce type sur l'image satellite ne serait donc au mieux que très grossièrement proportionnelle à la surface bâtie des unités et n'améliorerait sans doute pas les estimations.

B. Dans l'étape d'inférence

Pour calculer les estimateurs sans biais associés à ce plan de sondage il faut connaître, pour chaque unité de l'échantillon et pour chaque domaine d'extrapolation, leur surface bâtie. La mesure de cette surface, qu'on devrait tirer de l'image satellite en ce qui concerne les domaines d'extrapolation (univers entier, strates et éventuellement domaines), ne serait bien sûr qu'approximative ce qui pourrait avoir de fâcheuses conséquences sur la précision des estimations puisqu'il s'agit des facteurs d'extrapolation. *Enfin cette même mesure au niveau des pâtés de maisons de l'échantillon, ne peut pas être effectuée sur l'image satellite et sa collecte sur le terrain, également sujette à erreurs, allongerait énormément les temps d'enquête.*

Pour l'ensemble de ces raisons, nous n'envisageons pas de retenir pour son application à Quito un tel plan de sondage, cependant il nous semble important pour les développements futurs de la méthode d'en évaluer dès maintenant la précision.

Le test qui va suivre concernera donc trois plans :

1. Plan équiprobable
2. Plan à probabilité proportionnelle à la surface des unités
3. Plan à probabilité proportionnelle à la surface bâtie des unités.

Pour répondre complètement à la question posée, il est bien entendu indispensable de comparer également les trois plans dans le cadre stratifié. En effet, le problème du choix d'une probabilité d'appartenance est, on le verra, très lié à la stratification employée.

2. Quels estimateurs employer dans l'étape d'inférence au premier degré : de l'échantillon de pâtés de maisons au total des strates ?

Il s'agit ici de confirmer les résultats obtenus lors de l'expérience sur Marseille concernant les comparaisons entre estimation sans biais et estimation par le ratio intégrant la surface des unités, qui s'étaient conclues à l'avantage de la première¹.

3. Quel gain de précision permet la stratification de la base de sondage à partir de l'image satellite?

Au départ de notre recherche, cette question était posée différemment : il s'agissait de comparer entre elles plusieurs stratifications issues de l'image correspondant chacune à des critères ou associations de critères différents. Malheureusement, étant donné le calendrier de la phase d'application, nous ne disposions au moment du test que d'une seule classification de l'image correspondant au critère de densité du bâti. C'est donc à l'évaluation de cette stratification qu'on s'attachera et à sa validation en la comparant à une stratification sur le même critère obtenue à partir de l'enquête préliminaire "morphologie". L'évaluation d'autres stratifications reste donc à mener² et c'est d'ailleurs à notre avis une des directions prometteuses à explorer dans des recherches futures.

4. Y a-t-il des possibilités d'optimisation de l'allocation aux strates?

A t-on intérêt à faire varier le taux de sondage selon les strates de densité de bâti pour améliorer la précision des estimations sur l'ensemble des strates? La réponse à cette question dépend évidemment de l'ensemble de variables qu'on considère dans l'enquête. Pour chacune d'elles on peut essayer, à partir des données de l'enquête préliminaire "bâti/population", d'obtenir une valeur approchée de la variance du caractère sur chacune des strates et, partant, de définir une "bonne" allocation. Les différentes allocations ainsi définies seront elles compatibles entre elles ou, au contraire contradictoires? Dans la première hypothèse on pourrait adopter, pour la stratégie d'allocation de l'échantillon aux strates, un moyen terme qui serait alors profitable à l'ensemble des estimations.

Les problèmes étant posés, les paragraphes suivants sont consacrés aux données et aux traitements statistiques auxquels elles ont été soumises, puis à l'examen des différents résultats obtenus. Enfin, en conclusion, on tentera d'apporter des réponses aux questions formulées ci dessus.

10.2 : Les données et les outils statistiques utilisés

10.2.1 : Les données

Puisqu'il s'agit dans cette étape d'examiner les unités primaires du sondage, c'est-à-dire les pâtés de maisons, il nous a fallu constituer, en partant du fichier ménage, un fichier pâté de maisons contenant les informations nécessaires aux tests envisagés. Comme à Marseille, mais cette fois ci sur l'échantillon de pâtés de maisons enquêtés, on souhaite disposer d'informations fiables concernant d'une part un certain nombre de caractéristiques démographiques moyennes des ménages du pâté de maisons et d'autre part la localisation, la taille et la morphologie de chaque îlot. La liste de ces informations est donnée dans le tableau 10.1. À propos de ces données, il y a plusieurs remarques à faire.

¹ : Voir à ce sujet la deuxième partie de cette thèse : "Mise au point de la méthode à Marseille, stratifications et approche du premier degré du sondage".

² : A ce propos, A. MICHEL, dans la thèse qu'il rédige à l'issu de son travail au sein de l'équipe ORSTOM en télédétection urbaine, évalue une stratification construite sur une typologie du bâti extraite de l'image SPOT, et la compare à celle fondée sur la seule densité du bâti.

10.2.1.1 : Données démographiques

On l'a vu, durant la phase de collecte, le taux de non réponses à tout ou partie du questionnaire a été très variable suivant le quartier enquêté. Des données trop partielles risquant d'oter toute signification aux calculs de moyennes, la première étape de la constitution du fichier est la sélection, pour chaque variable, du sous ensemble de l'échantillon dans lequel l'information est suffisante pour fournir une moyenne significative ; connaissant pour tous les pâtés de maisons considérés le nombre de ménages y résidant, on a pu, lorsque, pour une variable donnée, l'information était disponible pour la grande majorité d'entre eux, étendre la donnée à l'ensemble de l'ilot. Il résulte de cette étape que, comme dans le fichier ménage, le domaine où la donnée est connue peut être différent suivant la variable considérée. Dans le tableau 10.1, on indique le nombre de pâtés de maisons pour lesquelles chaque variable est soit totalement connue, soit extrapolée avec une fiabilité acceptable.

Une fois définis ces domaines de validité, le calcul des indicateurs démographiques sur les pâtés de maisons se fait par sommation à partir du fichier ménage. Cette opération a été grandement facilitée par les possibilités de gestion relationnelle des données offerte par dbaseIII. On trouvera en annexe le listage des données correspondant aux pâtés de maison enquêtés¹.

Signalons que si les données obtenues au niveau des îlots peuvent être entachées de certaines inexactitudes dues aux non réponses, nous avons néanmoins pu vérifier à partir des chiffres du recensement de 1982 qu'au niveau de l'ensemble de l'échantillon les valeurs obtenues pour un certain nombre d'indicateurs étaient en parfaite cohérence avec la donnée "réelle". Nous pensons donc que le degré de précision des informations qui vont servir aux tests est suffisant pour que leurs résultats soient fiables.

Enfin le lecteur remarquera qu'une partie seulement des traitements s'applique à la totalité des variables citées dans le tableau 10.1. Pour les tests menés avec le programme VARECH, les variables "AGECHEF", "REVENU" et "DURESMIG" ont été supprimées d'une part pour ne pas allonger l'étape de traitement et respecter le calendrier prévu pour la réalisation de l'enquête et d'autre part parce que, lors de l'introduction des stratifications, nous aurions disposé pour ces variables d'un nombre insuffisant de données dans certaines strates.

¹ : Voir annexes A.3.4 : "Résultats numériques obtenus à Quito".

Tableau 10.1 : Liste des champs du fichier "pâté de maisons".

Champ	Nom du champ	Définition	Nombre d'îlots où la donnée existe
1) Identification			
1	NOMA	Numéro de l'îlot (manzana)	53
2) Localisation, taille, morphologie			
2	SUDNORD	Latitude	53
3	ESTOUEST	Longitude	53
4	LOC	Appartenance à une des douze zones géographiques définies dans Quito.	53
5	SURFACE	Surface de la manzana	53
6	DENSBAT	Pourcentage de la surface bâtie dans la manzana, d'après l'enquête "morphologie"	53
7	DENSBATI	Classification en cinq classes de densité de bâti : résultat de l'enquête "morphologie"	53
8	DENSBATIM	Classification en cinq classes de densité de bâti : résultat du traitement de l'image SPOT	53
9	DENSBATIME	Classification en cinq classes de densité de bâti obtenue après "redressement" de la classification DENSBATIM	53
10	DENSVEG	Densité de végétation d'après l'enquête "morphologie"	53
11	SURFBAT	Surface bâtie, mesure approchée par le pourcentage de surface bâtie.	53
12	CSURFVEG	Complémentaire à la surface en végétation : mesure approchée par l'indice de végétation tiré de l'image satellite.	53

Tableau 10.1 (suite) : Liste des champs du fichier "pâté de maisons".

Champ	Nom du champ	Définition	Nombre d'ilots ou la donnée existe
3) Démographie			
13	EFFECTRES	Population résidente ¹	46
14	NBMENAGE	Nombre de ménages	53
15	NBMEN1	Nombre de ménages d'une personne	46
16	NBMEN7	Nombre de ménages de sept personnes et plus	46
17	NBMENP1	Nombre de ménages de plus d'une personne (complémentaire de NBMEN1 ²)	46
18	NBMENM7	Nombre de ménages de moins de sept personnes (complémentaire de NBMEN7 ³)	46
19	EFFM5A	Population âgée de moins de cinq ans	44
20	EFFMIG	Population de migrants	44
21	EFFMIGM5	Population de migrants ayant moins de cinq ans de résidence à Quito	44
22	EFFECTMOY	Effectif moyen des ménages	46
23	AGECHEF	Age moyen des chefs de ménage (en années)	45
24	PCM0INS5	Proportion de la population de moins de cinq ans dans la population totale (en %)	44
25	PCMIG	Proportion de la population de migrants dans la population totale (en %)	44
26	PCMIGM5	Proportion de la population de migrants ayant moins de cinq ans de résidence à Quito dans la population totale (en %)	44
27	DURESMIG	Durée moyenne de résidence à Quito des migrants (en mois)	41
28	REVENU	Revenu moyen des résidents de l'ilot (en millier de sucres)	38
29	PCMENAG1	Proportion des ménages d'une personne dans la totalité des ménages (en %)	46
30	PCMEN7	Proportion des ménages de sept personnes et plus dans la totalité des ménages (en %)	46

¹ : Pour les définitions des différentes catégories de populations considérées ici, ménages, population résidente, population migrante, voir le manuel d'instruction au enquêteurs en annexe A.3.1.

² : NBMENP1 - NBMENAGE - NBMEN1

³ : NBMENM7 - NBMENAGE - NBMEN7

10.2.1.2 : Taille des unités primaires

On considérera dans les analyses trois mesures de taille des pâtés de maisons :

1. La surface totale de l'unité
2. La surface bâtie de l'unité (exclus la voirie interne, les jardins, les terrains vagues avec ou sans végétation, les parkings ou toute autre surface ne comportant pas de bâtiments)
3. Le complémentaire de la surface en végétation (inclus la voirie interne, les terrains vagues sans végétation, les parkings ou toute autre surface ne contenant pas de végétation). Ces données ne sont pas issues de l'interprétation de l'image satellite mais proviennent de l'enquête préliminaire "morphologie" réalisée sur l'échantillon. Le niveau de précision de la mesure de la surface totale nous paraît en accord avec celui des mesures praticables sur l'image, par contre la donnée concernant la surface bâtie est probablement plus précise que son équivalent obtenu d'après l'information satellitaire. Pour pallier à cette différence, nous avons considéré une troisième mesure : obtenue par soustraction de la surface en végétation à la surface totale, elle est sans doute plus proche de l'approximation de la surface bâtie qu'on pourrait tirer de l'indice de végétation calculé sur l'image.

10.2.1.3 : Stratifications

Dans ce travail, on ne considère qu'un seul critère de stratification qui est la densité de bâti, pourtant on s'intéresse à trois stratifications différentes construites toutes les trois sur ce critère. La première est issue des observations réalisées lors de l'enquête "morphologie" et la variable lui correspondant sera appelée DENSBATI. La seconde, DENSBATIM, est la stratification de la base de sondage spatiale obtenue par classification de l'image satellite SPOT de juin 1986¹. Quant à la troisième, DENSBATIME, elle correspond aux résultats des corrections faites sur la seconde. Nous avons en effet jugé utile de "redresser" cette seconde stratification pour deux raisons.

D'une part, pour des impératifs de calendrier, la stratification de l'image dont nous disposons (DENSBATIM) est une classification des pixels et non pas la classification d'îlots obtenue en utilisant les valeurs radiométriques moyennes des pixels de l'îlot. Les travaux de l'équipe télédétection ayant montré que cette dernière classification est nettement préférable, nous avons pris le parti de corriger la stratification dans ce sens en changeant certains îlots de strate.

D'autre part, pour alléger la collecte à réaliser dans l'enquête préliminaire démographique, lorsque qu'on pouvait identifier sans ambiguïtés plusieurs parties au sein d'un même pâté de maisons, on n'a parfois enquêté qu'une seule de ces parties. Il peut alors se produire que la densité de bâti caractérisant la partie enquêtée soit significativement différente de la densité moyenne du pâté de maisons. L'unité réellement enquêtée se voit alors affectée, si l'on s'en tient à la seconde des stratifications, à une strate qui ne correspond pas réellement à sa densité de bâti. Nous avons donc dans ce cas réaffecté ces unités aux strates présentant la même densité moyenne de bâti qu'elles.

Qu'elles aient été effectuées pour la première ou la seconde raison, ces réaffectations se font dans l'immense majorité des cas à la strate immédiatement plus dense ou moins dense que la strate d'origine (DENSBATIM).

¹ : Voir à ce sujet le chapitre 12, paragraphe 12.3.2.

10.2.2 : Les outils statistiques utilisés

Pour apporter des solutions aux problèmes posés, on a procédé aux traitements sur le fichier des pâtés de maisons en deux temps.

1) Phase exploratoire : corrélations, régressions linéaires, analyses de la variance.

Dans cette étape, en identifiant et quantifiant à l'aide d'outils statistiques simples les principales liaisons existant entre caractères démographiques à estimer et mesures de la taille des unités primaires, on poursuit un double objectif. D'une part on veut écarter, pour gagner du temps dans la suite de l'expérience, l'étude de solutions à priori envisageables mais que l'absence de liaisons statistiquement significatives font apparaître illusoirs. On pourra par exemple conclure dès cette étape à l'inutilité d'intégrer les informations sur la taille des unités primaires dans des estimateurs par le ratio. D'autre part, on cherche à formuler des hypothèses concernant l'intérêt d'adopter une probabilité de sélection des unités de l'échantillon proportionnelle aux diverses mesures possibles de la taille des pâtés de maisons. On passera ensuite, dans la seconde phase, à la vérification de ces hypothèses.

2) Phase conclusive : test des plans de sondage à l'aide du programme VARECH mis au point au cours de l'expérience menée à Marseille¹.

Le programme VARECH permet dans cette étape d'évaluer les conséquences qu'auront sur la précision des estimations les différents choix possibles quant aux probabilités d'inclusion dans l'échantillon et aux stratifications. Pour cela on considère l'échantillon des pâtés de maisons enquêtés comme une base de sondage et le programme calcule la variance des estimateurs sur cette base. Afin de pouvoir comparer entre elles l'ensemble des variances calculées, nous avons systématiquement appliqué un même taux de sondage ($2/3 = 66,6\%$) ; ce taux, par ailleurs arbitraire, a été choisi assez grand pour limiter l'ordre de grandeur des coefficients de variation. Pour vérifier les hypothèses formulées dans la phase exploratoire, les traitements effectués à l'aide du programme se sont déroulés en trois temps. Tout d'abord on a comparé les effets, sur la précision des estimations, des différentes probabilités d'appartenance à l'échantillon qu'on peut envisager. Ces effets varient évidemment suivant le fait que la base de sondage est ou non stratifiée, l'expérience est répétée dans les deux cas. Puis on a évalué les gains de précision permis par les trois stratifications. Enfin on s'est intéressé aux allocations optimales pour chacun des caractères afin d'essayer de dégager une règle d'allocation améliorant l'ensemble des estimations.

¹ : A propos du programme VARECH, son domaine d'application, le formulaire mis en œuvre et le fonctionnement du programme, voir le chapitre 7 et, plus spécialement, le paragraphe 7.3 : Le programme Variance d'Echantillonnage (VARECH) associé au logiciel Dbase III.

10.3 : Résultats

10.3.1 : Phase exploratoire

Tout d'abord on s'est intéressé aux corrélations existant entre indicateurs démographiques.

Tableau 10.2 : Matrice des corrélations entre variables démographiques. (logiciel SPSS)

COEFFICIENTS *	1	2	3	4	5	6	7	8	9	10	11
1 EFFECTRES	1	.99	.74	.76	.93	.99	.87	-.28	.23	-.10	.10
2 NBMENAGE	.99	1	.79	.69	.96	.99	.90	-.32	.24	-.08	.11
3 NBMEN1	.74	.79	1	.43	.77	.80	.74	-.48	.48	-.08	.27
4 NBMEN7	.76	.69	.43	1	.54	.72	.52	.03	.23	-.25	.10
5 EFFM5A	.93	.96	.77	.54	1	.94	.93	-.34	.19	.08	.11
6 EFFMIG	.99	.99	.80	.72	.94	1	.91	-.28	.27	-.09	.21
7 EFFMIGM5	.87	.90	.74	.52	.93	.91	1	-.15	.25	-.06	.37
8 EFFECTMOY	-.28	-.32	-.48	.03	-.34	-.28	-.15	1	-.11	-.19	.28
9 AGECHF	.23	.24	.48	.23	.19	.27	.25	-.11	1	-.48	.42
10 PCMOINS5	-.10	-.08	-.08	-.25	.08	-.09	-.06	-.19	-.48	1	-.23
11 PCMIG	.10	.11	.27	.10	.11	.21	.37	.28	.42	-.23	1
12 PCMIGM5	-.08	-.08	-.05	-.06	-.07	-.02	.26	.62	.11	-.20	.72
13 DURESMIG	.08	.07	.05	.13	.02	.02	-.22	-.29	-.03	-.01	-.65
14 REVENU	.56	.58	.45	.32	.59	.58	.62	-.22	.32	-.14	.22
15 PCMENAG1	.15	.19	.68	.07	.14	.21	.14	-.48	.50	-.18	.28
16 PCMEN7	-.20	-.23	-.26	.21	-.25	-.19	-.11	.79	.16	-.32	.32

COEFFICIENTS	12	13	14	15	16
1 EFFECTRES	-.08	.08	.56	.15	-.20
2 NBMENAGE	-.08	.07	.58	.19	-.23
3 NBMEN1	-.05	.05	.45	.68	-.26
4 NBMEN7	-.06	.12	.32	.07	.21
5 EFFM5A	-.07	.02	.60	.14	-.25
6 EFFMIG	-.02	.02	.58	.21	-.19
7 EFFMIGM5	.26	-.22	.62	.14	-.11
8 EFFECTMOY	.62	-.29	-.22	-.48	.79
9 AGECHF	.11	-.03	.32	.50	.16
10 PCMOINS5	-.20	-.01	-.14	-.18	-.32
11 PCMIG	.72	-.65	.22	.28	.32
12 PCMIGM5	1	-.65	.04	-.10	.42
13 DURESMIG	-.65	1	-.11	.07	-.21
14 REVENU	.04	-.11	1	.10	-.16
15 PCMENAG1	-.10	.07	.10	1	-.18
16 PCMEN7	.42	-.21	-.16	-.18	1

* : En gras, les coefficients supérieurs à 0,70.

De l'examen du tableau 10.2 se dégagent deux groupes de caractères.

Premièrement les effectifs (caractères 1 à 7) qui forment un ensemble au sein duquel les valeurs des coefficients sont généralement élevées et toutes positives. En particulier les trois variables "Effectif de résidents", "Nombre de ménages", "Effectif des migrants" sont fortement corrélées entre elles et à tous les autres effectifs considérés. Un tel phénomène de bloc, qui n'est d'ailleurs pas surprenant, autorise à espérer, pour la suite des traitements, un comportement semblable de l'ensemble de ces variables.

Le second groupe est formé par les variables de type quotient (caractères 8 à 16) pour lesquelles les corrélations sont en général médiocres, à deux exceptions près qui reflètent des liaisons évidentes : corrélation élevée entre l'effectif moyen du ménage et la proportion de ménages de plus de sept personnes et entre la proportion de migrants et la proportion de migrants ayant moins de cinq ans de résidence à Quito. Après ce qui vient d'être dit, on peut penser qu'il sera relativement facile de formuler, à l'issu des tests qui vont suivre, des conclusions valables pour l'estimation de l'ensemble des effectifs alors qu'à l'inverse il risque fort de ne pas se dégager de règles applicables à l'ensemble des quotients. Comme entre les deux groupes, il n'existe aucune corrélation linéaire significative, il faudra prendre garde qu'une stratégie de sondage améliorant l'estimation des effectifs ne détériore pas au contraire celle des quotients.

S'agissant de l'étude des différentes mesures de la taille des unités primaires qui, selon la stratégie de sélection de l'échantillon retenue, définissent la probabilité d'inclusion, il nous semble naturel d'examiner d'abord les corrélations entre ces mesures et les caractères démographiques. En effet, notre sondage à probabilités inégales sera bien sûr d'autant plus efficace que la variable définissant la probabilité sera mieux corrélée aux variables à estimer.

Le tableau 10.3 donne les coefficients de corrélation linéaire existant entre les indicateurs démographiques et les trois mesures de la taille des pâtés de maisons retenues.

L'examen des résultats conduit aux conclusions suivantes :

1. Il n'y a pas de corrélations significatives entre les caractères démographiques et la surface totale des unités. Il faut donc s'attendre à ce que l'introduction d'une probabilité proportionnelle à la surface des pâtés de maisons n'améliore pas la précision des estimations.

2. La plupart des effectifs, à l'exception du nombre de ménage de sept personnes et plus, sont corrélés positivement avec la surface bâtie. Les coefficients sont compris entre 0,55 et 0,61 ; de telles valeurs, si elles paraissent insuffisantes pour justifier l'emploi d'estimateurs par le ratio, devraient permettre en revanche un gain appréciable de précision si on retient une probabilité d'appartenance proportionnelle à la surface bâtie. Dans le cas des quotients au contraire, les coefficients de corrélation sont faibles et varient de -0,35 à 0,50, ce qui laisse peu d'espoir d'améliorer les estimations en adoptant une telle probabilité.

3. La troisième mesure, le complémentaire de la surface en végétation, apparaît bien, au regard des coefficients de corrélation avec les indicateurs démographiques, comme une approximation de la surface bâtie. Les corrélations, systématiquement inférieures à celles qu'on observe dans le cas de la surface bâtie, doivent néanmoins être suffisantes pour que l'introduction de cette mesure comme déterminant de la probabilité améliore la précision du sondage.

Tableau 10.3 : coefficients de corrélation entre variables démographiques et mesures de taille des unités primaires du sondage. (logiciel SPSS)

	(1) SURFACE	(2) SURFBAT	(3) CSURFVEG
	Surface totale de l'ilot	Surface bâtie de l'ilot	Complément à la surface en végétation
1) Totaux			
1 EFFECTRES	0,068	0,572	0,442
2 NBMENAGE	0,075	0,603	0,453
3 NBMEN1	0,042	0,610	0,420
4 NBMEN7	-0,070	0,308	0,239
5 EFFM5A	0,071	0,608	0,405
6 EFFMIG	0,060	0,581	0,437
7 EFFMIGM5	0,018	0,548	0,365
2) Quotients			
8 EFFECTMOY	-0,055	-0,349	-0,236
9 AGECHEF	-0,192	0,150	0,023
10 PCMOINS5	-0,012	-0,107	-0,183
11 PCMIG	-0,034	0,147	0,095
12 PCMIGM5	-0,154	-0,155	-0,112
13 DURESMIG	0,216	0,144	0,118
14 REVENU	-0,146	0,380	0,112
15 PCMENAG1	0,173	0,504	0,396
18 PCMEN7	-0,165	-0,218	-0,196

Quels modèles explicatifs des caractères démographiques peut-on maintenant construire à partir des informations relatives à la taille des unités primaires? Il serait vain, étant donné le degré de corrélation observé ci dessus, de considérer la surface totale des pâtés de maisons comme facteur explicatif. En revanche il peut être intéressant de voir quel type de modèle rend le mieux compte des liaisons existant entre indicateurs démographiques et surface bâtie. On comparera ici trois modèles linéaires :

1. Régression entre effectifs et surface bâtie
2. Régression entre densités et densité de bâti
3. Analyse de la variance pour tester l'effet du classement des observations en six niveaux de densité de bâti.

Le tableau 10.4 résume les résultats obtenus lorsqu'on applique successivement ces trois modèles aux sept variables démographiques.

Tableau 10.4 : Résultats résumés des régressions et analyses de la variance entre caractères démographiques (effectifs ou densité) et surface bâtie ou densité du bâti. (Logiciel SPSS)

	Régressions entre effectifs et surface bâtie		Régressions entre densités et densité de bâti		Analyse de la variance à un critère : densité de bâti en 6 cl.
	Coeff. de reg.	% de varian. expl.	Coeff. de reg.	% de varian. expl.	% de variance expliquée
1 EFFECTRES	0,52	26,7	0,67	44,6	52,6
2 NBMENAGE	0,55	30,2	0,67	45,0	53,8
3 NBMEN1	0,52	26,7	0,45	19,8	26,8
4 NBMEN7	0,21	4,2	0,42	17,3	20,9
5 EFFMOINS5A	0,53	27,8	0,66	43,4	51,0
6 EFFMIG	0,52	27,1	0,62	37,8	55,2
7 EFFMIGM5	0,47	21,8	0,37	13,7	29,3

Tout d'abord il faut remarquer que les parts de variance expliquées par chacun des modèles sont toutes trop faibles pour qu'on puisse espérer estimer la valeur prise par le caractère démographique à partir de celle du facteur explicatif. Cela confirme qu'il y a peu d'avantage à utiliser des estimateurs par le ratio intégrant ces mesures de taille (il est même probable que la précision des estimations en souffre). Mais l'expérience ne perd pas pour autant tout intérêt. En effet on constate que les trois modèles n'ont pas des performances équivalentes : d'une part, les régressions entre densité d'une catégorie de population ou de ménages et densité de bâti sont généralement nettement meilleures que celles entre l'effectif correspondant et la surface bâtie (sauf pour les catégories très minoritaires comme les ménages d'une personne ou les migrants ayant moins de cinq années de résidence à Quito, variables pour lesquelles il faudrait étudier les quantités complémentaires), d'autre part, le modèle d'analyse de la variance s'ajuste toujours mieux aux données que les régressions. Ces constatations conduisent à avancer l'hypothèse suivante concernant la stratégie de sondage à adopter et qu'il s'agira de vérifier dans la phase conclusive de l'expérience. Plutôt que de chercher à mettre en pratique une stratégie qui garantisse un tirage à probabilité approximativement proportionnelle à la surface bâtie des unités primaires, ce qui de toute façon sera difficile étant donné l'information disponible sur l'image base de sondage, n'a-t-on pas intérêt à pratiquer, à l'intérieur de strates offrant une densité de bâti homogène, un tirage avec probabilité proportionnelle à la surface totale des pâtés de maisons, ce qui est, par contre, facilement réalisable?

Il nous faut maintenant vérifier le bien fondé de l'ensemble des hypothèses formulées dans cette phase exploratoire. Cette vérification se fera à partir du calcul sur l'échantillon de la variance des estimateurs qu'effectuera le programme VARECH.

10.3.2 : Phase conclusive

10.3.2.1 : Probabilités d'appartenance des unités primaires à l'échantillon

Le tableau 10.5 donne les coefficients de variation calculés par le programme dans le cas non stratifié et dans le cas d'une stratification de la base de sondage en cinq classes de densité de bâti (variable DENSBATI). Pour le cas non stratifié on a retenu quatre définitions de la probabilité d'appartenance :

1. Probabilité égale pour chaque unité (colonne 1)
2. Probabilité proportionnelle à la surface totale de l'unité (colonne 2)
3. Probabilité proportionnelle à la surface bâtie de l'unité (colonne 3)
4. Probabilité proportionnelle au complémentaire de la surface en végétation (colonne 4).

Pour le cas stratifié seules les trois premières définitions ont été appliquées, les colonnes impaires donnant le coefficient obtenu avec l'allocation proportionnelle et les colonnes paires ceux correspondants à l'allocation optimale :

1. Equiprobabilité (colonnes 5 et 6)
2. Probabilité proportionnelle à la surface totale (colonnes 7 et 8)
3. Probabilité proportionnelle à la surface bâtie (colonnes 9 et 10).

Enfin les colonnes 11 et 12 mesurent le gain dû à la stratification sous le plan à probabilités proportionnelles aux surfaces totales des unités, en pourcentage du coefficient de variation non stratifié (colonne 11) et par le facteur de division de la variance (colonne 12).

L'observation du tableau fournit quatre résultats principaux.

1. Comparé à la référence équiprobable et dans le cadre non stratifié, un sondage à probabilités proportionnelles aux surfaces totales des unités primaires (sondage aréolaire aléatoire non stratifié) ne présente pas d'intérêt ; au contraire, dans la plupart des cas il augmente l'erreur d'échantillonnage (voir colonnes 1 et 2). On trouve ici confirmation d'un résultat important obtenu à Marseille.

2. Une autre hypothèse faite lors de l'expérience marseillaise se vérifie : s'il était possible de gérer correctement une probabilité de sélection proportionnelle à la surface bâtie, ou du moins à son approximation d'après l'image satellite, l'amélioration du sondage serait en revanche notable, du moins en ce qui concerne les effectifs : 30 à 40 % de réduction du coefficient de variation dans le cas de la mesure de la surface bâtie, 20 à 30 % s'il s'agit de l'approximation par la surface en végétation (voir colonnes 1,3 et 4 de la première page du tableau). Pour les quotients en revanche l'avantage ne va systématiquement à aucune des deux définitions de la probabilité.

3. Quand on s'en tient au tirage équiprobable et à l'allocation proportionnelle, la stratification en six classes sur le critère de densité du bâti n'apporte qu'un gain assez faible. Par contre l'allocation optimale permet une diminution sensible de la variance des estimations sur l'ensemble de la base (colonnes 1,5 et 6).

Tableau 10.5 : Coefficients de variation des estimations des variables démographiques avec différentes définitions des probabilités d'appartenance à l'échantillon, cas non stratifié et cas stratifié. (Programme VARECH)

	Non stratifié				Stratifié sur la densité du bâti (DENSBATI)						Gains dus à la stratif.	
	1 Equi.	2 P-su	3 P-sb	4 P-csv	5 Equi. alloc. prop.	6 Equi. alloc. opti.	7 P-su alloc. prop.	8 P-su alloc. opti.	9 P-sb alloc. prop.	10 P-sb alloc. opti.	11 ¹ Gain (8-2) 2 en %	12 ² Gain (2/8) ²
1) Totaux												
1 EFFECTRES	24,91	25,42	15,97	17,40	22,46	15,92	13,28	12,30	13,73	12,23	-51,6	4,3
2 NBMENAGES	26,72	26,26	15,58	17,75	24,15	16,95	13,27	12,19	13,03	11,59	-53,6	4,7
3 NBMEN1	34,10	40,58	24,03	27,99	31,40	27,39	26,25	21,79	24,49	21,19	-46,3	3,5
4 NBMENP1	26,79	26,25	15,83	17,88	24,19	16,88	13,36	12,33	13,15	11,70	-54,0	4,5
5 NBMEN7	20,55	28,87	23,92	24,90	19,61	18,29	23,10	21,91	23,41	21,94	-24,1	1,7
6 NBMENM7	28,17	27,14	15,92	18,33	25,50	17,74	13,63	12,40	13,18	11,64	-54,3	4,8
7 EFFM5A	25,53	26,54	16,11	18,42	23,24	17,37	14,50	13,63	12,66	12,25	-48,6	3,8
8 EFFP5A	25,34	26,45	16,64	18,10	22,87	16,14	13,92	12,77	14,57	12,85	-51,7	4,3
9 EFFMIG	28,31	29,62	17,80	19,79	25,05	17,23	13,84	12,79	14,53	12,96	-56,8	5,3
10 EFFNMIG	23,61	25,02	16,60	16,60	21,61	16,08	14,58	13,60	14,82	13,48	-45,6	3,4
11 EFFMIGM5	26,02	30,50	29,42	21,55	23,38	19,18	20,03	19,23	21,76	20,38	-37,0	2,5
12 CEFFMIGM5	25,38	26,42	16,61	18,12	22,95	16,29	14,06	12,92	14,53	12,83	-51,1	4,2

Equi. : sélection équiprobable des unités primaires dans l'échantillon.

P-su : probabilité d'appartenance à l'échantillon proportionnelle à la surface totale des unités.

P-sb : probabilité d'appartenance à l'échantillon proportionnelle à la surface bâtie des unités.

P-csv : probabilité d'appartenance à l'échantillon proportionnelle au complémentaire de la surface en végétation des unités.

¹ : Gain en % du coefficient de variation non stratifié

² : Facteur de division de la variance (par rapport à la variance non stratifiée)

Tableau 10.5 (suite) : Coefficients de variation des estimations des variables démographiques avec différentes définitions des probabilités d'appartenance à l'échantillon, cas non stratifié et cas stratifié. (Programme VARECH)

	Non stratifié				Stratifié sur la densité du bâti (DENSBATI)						Gains dus à la stratif.		
	1 Equi.	2 P-su	3 P-sb	4 P-csv	5 Equi. alloc. prop.	6 Equi. alloc. opti.	7 P-su alloc. prop.	8 P-su alloc. opti.	9 P-sb alloc. prop.	10 P-sb alloc. opti.	11 ¹ Gain (8-2) 2 en %	12 ² Gain (2/8) ²	
2) Quotients													
13	EFFECTRES / NB MENAGE	3,71	4,07	3,08	3,40	3,35	2,68	3,15	2,83	3,13	2,81	-30,5	2,1
14	EFFECTRES / SURFACE	29,92	25,42	23,77	23,48	15,64	13,96	13,28	12,30	13,79	12,49	-51,6	4,3
15	NBMEN1 / NB MENAGE	22,34	30,09	21,57	23,21	20,11	18,08	22,51	18,80	20,94	17,19	-37,5	2,6
16	NBMENP1 / NB MENAGE	1,14	1,54	1,10	1,19	1,03	0,93	1,15	0,96	1,07	0,92	-37,7	2,6
17	NBMEN7 / NB MENAGE	21,46	23,60	19,92	21,59	17,86	16,52	19,71	18,06	19,29	18,19	-23,5	1,6
18	NBMENM7 / NB MENAGE	2,18	2,40	2,02	2,19	1,81	1,68	2,0	1,89	1,96	1,85	-21,2	1,6
19	EFFM5A / EFFECTRES	9,88	10,91	9,57	9,66	9,69	8,02	9,09	8,29	9,37	8,57	-24,0	1,7
20	EFFP5A / EFFECTRES	1,02	1,12	0,99	0,95	1,0	0,83	0,94	0,85	0,97	0,88	-30,2	1,7
21	EFFMIG / EFFECTRES	5,55	7,58	6,33	6,23	3,99	3,49	5,10	4,73	5,06	4,75	-37,6	2,6
22	EFFNMIG / EFFECTRES	3,54	4,84	4,04	3,98	2,55	2,23	3,25	3,02	3,23	3,03	-37,6	2,6
23	EFFMIGM5 / EFFECTRES	12,95	18,07	24,31	14,88	13,17	12,17	17,19	16,17	19,12	17,58	-10,5	1,3
24	CEFFMIGM5 / EFFECTRES	1,06	1,48	1,99	1,22	1,08	0,99	1,41	1,32	1,52	1,44	-10,8	1,3

Equi. : sélection équiprobable des unités primaires dans l'échantillon.

P=su : probabilité d'appartenance à l'échantillon proportionnelle à la surface totale des unités.

P=sb : probabilité d'appartenance à l'échantillon proportionnelle à la surface bâtie des unités.

P=csv : probabilité d'appartenance à l'échantillon proportionnelle au complémentaire de la surface en végétation des unités.

¹ : Gain en % du coefficient de variation non stratifié

² : Facteur de division de la variance (par rapport à la variance non stratifiée)

4. L'observation comparée des colonnes 1, 2 et 5 à 10 permet d'affirmer qu' *un sondage à probabilités proportionnelles à la surface totale des unités, une fois stratifié la base de sondage selon la densité du bâti, est la solution qui s'impose*. Premièrement, il est la plupart du temps meilleur que le sondage équiprobable stratifié et a fortiori non stratifié (colonnes 7 et 8 comparées aux colonnes 1,5 et 6). *Deuxièmement avec un tel plan de sondage le gain induit par la stratification, une fois optimisée l'allocation aux strates est important*, puisqu'il s'agit d'une diminution des coefficients de variation comprise entre 25 et 55% ce qui représente une *division de la variance par un facteur compris entre 2 et 5* (voir colonnes 2 et 8 ou, plus commodément, 11 et 12). Enfin la comparaison des colonnes 7 et 8 avec les colonnes 9 et 10 permet de conclure à l'inutilité de rechercher plus avant un mode de tirage qui garantisse une probabilité approximativement proportionnelle aux surfaces bâties, cette recherche se heurtant d'ailleurs comme on l'a vu à de nombreuses difficultés pratiques. On constate en effet, en examinant conjointement les colonnes 8 et 10, qu' *une fois retenu le principe de la stratification sur la densité du bâti, le tirage avec probabilités proportionnelles aux surfaces bâties n'améliore que très peu, voir pas du tout, l'estimation par rapport au tirage à probabilités proportionnelles aux surfaces totales*. Ce résultat apparaît d'ailleurs intuitivement satisfaisant : à l'intérieur d'une même classe de densité de bâti, surface totale et surface bâtie des unités primaires sont grosso modo proportionnelles et donc présentent les mêmes corrélations avec les caractères démographiques. On peut également considérer que ce résultat valide d'une certaine manière la stratification employée : les six classes de densité du bâti ont un bon niveau d'homogénéité.

10.3.2.2 : Stratifications

Le tableau 10.6 permet de comparer en termes d'amélioration de la précision du sondage, les trois stratifications construites sur l'échantillon, toutes trois fondées, comme on l'a vu avec des degrés de précision différents, sur le critère de densité du bâti. Il contient tout d'abord, comme référence pour l'appréciation des gains, les coefficients de variation des estimations non stratifiées (colonnes 1 et 2), puis (colonnes 3 et 4) ceux obtenus à partir de la stratification de référence issue de l'exploitation de l'enquête "morphologie" (DENSBATI), sous plan de sondage à probabilité proportionnelle à la surface totale des unités (P=su). La comparaison avec les stratifications établies d'après l'image satellite, stratification "brute" ou "corrigée" (respectivement DENSBATIM et DENSBATIME), doit être faite dans le cadre de ce même plan de sondage (P=su) à l'aide des colonnes 5 et 6 ou 9 et 10. Les colonnes 7,8,11 et 12 donnent, pour les deux nouvelles stratifications, les coefficients obtenus sous plan à probabilité proportionnelle à la surface bâtie des unités (P=sb) ; elles permettent de vérifier que les conclusions de l'expérience précédente sur la probabilité d'appartenance ne sont pas affectées par ces changements de stratification.

1. Pour un même plan de sondage et une même variable de stratification on constate que de manière générale les estimations d'effectifs profitent plus de l'introduction de la stratification que les estimations de quotients. *Le gain se situe autour de 50 % du coefficient de variation* pour les premiers (variance divisée environ par 4,5) alors qu'il est beaucoup plus variable pour les seconds, s'étalant entre 10 et 40 % (division de la variance par 1,3 à 2,5).

2. *Une fois redressées les erreurs d'affectation de l'échantillon aux strates, la stratification issue de l'image remplit fort bien son rôle puisque les gains qu'elle permet sont tout à fait comparables à ceux apportés par la stratification de référence* : une fois optimisée l'allocation aux strates, la différence entre les deux stratifications n'exède pas 1,5 % du coefficient de variation non stratifié dans le cas des estimations d'effectifs. Cette différence, tout en restant faible, est plus variable en ce qui concerne l'estimation des quotients. Elle peut même être alors à l'avantage de la stratification tirée de l'image, il est vrai dans des cas où le gain dû aux stratifications est de toute manière faible.

Tableau 10.6 : Coefficients de variation des estimations des variables démographiques avec différentes stratifications.(Programme VARECH)

	Non stratifié		Stratifié DENSBATI		Stratifié DENSBATIME				Stratifié DENSBATIM			
	1	2	3	4	5	6	7	8	9	10	11	12
	P-su alloc.	P-sb	P-su alloc. prop.	P-su alloc. opti.	P-su alloc. prop.	P-su alloc. opti.	P-sb alloc. prop.	P-sb alloc. opti.	P-su alloc. prop.	P-su alloc. opti.	P-sb alloc. prop.	P-sb alloc. opti.
1) Totaux												
1 EFFECTRES	25,42	15,97	13,28	12,30	14,13	12,69	14,68	13,18	17,92	16,40	15,42	14,20
2 NBMENAGES	26,26	15,58	13,27	12,19	14,08	12,57	14,16	12,68	18,13	16,44	15,18	13,72
4 NBMENP1	26,25	15,83	13,36	12,33	14,01	12,56	14,23	12,76	18,27	16,63	15,29	13,89
6 NBMENM7	27,14	15,92	13,63	12,40	14,39	12,83	14,32	12,80	18,69	16,83	15,51	13,85
8 EFFP5A	26,45	16,64	13,92	12,77	14,71	13,17	15,46	13,79	18,66	16,92	16,08	14,69
10 EFFNMIG	25,02	16,60	14,58	13,60	14,97	13,85	15,42	14,19	18,75	17,36	15,69	14,76
12 CEFFMIGM5	26,42	16,61	14,06	12,92	14,80	13,31	15,42	13,78	18,81	17,08	16,00	14,64
2) Quotients												
13 EFFECTRES / NBMENAGE	4,07	3,08	3,15	2,83	3,14	2,86	2,96	2,73	3,18	2,87	2,82	2,57
14 EFFECTRES / SURFACE	25,42	23,77	13,28	12,30	14,13	12,69	15,04	13,51	17,92	16,40	20,35	18,09
16 NBMENP1 / NBMENAGE	1,54	1,10	1,15	0,96	1,14	1,0	1,06	0,94	1,10	0,95	0,96	0,83
18 NBMENM7 / NBMENAGE	2,40	2,02	2,0	1,89	1,96	1,84	1,89	1,80	2,07	1,95	1,92	1,80
20 EFFP5A / EFFECTRES	1,12	0,99	0,94	0,85	0,95	0,91	0,98	0,93	0,95	0,91	0,91	0,87
22 EFFNMIG / EFFECTRES	4,84	4,04	3,25	3,02	3,41	3,26	3,30	3,14	3,30	3,16	3,16	2,95
24 CEFFMIGM5 / EFFECTRES	1,48	1,99	1,41	1,32	1,40	1,28	1,55	1,40	1,44	1,31	2,03	1,69

P=su : probabilité d'appartenance à l'échantillon proportionnelle à la surface totale des unités.

P=sb : probabilité d'appartenance à l'échantillon proportionnelle à la surface bâtie des unités.

La stratification d'après l'image satellite non corrigée apparaît en revanche, étant donné les biais évoqués plus haut, beaucoup moins bonne que la stratification de référence du moins en ce qui concerne les effectifs. Pour les ratios les trois stratifications sont en fait globalement équivalentes, l'avantage revenant selon les cas à l'une ou l'autre. Ce phénomène, surprenant au premier abord, s'explique en réalité très simplement. Comme le suggère la première observation faite sur le tableau, dans le cas des quotients les liaisons entre variables à estimer et variables de stratification sont assez ténues. Il n'y a donc pas de raison pour qu'une stratification seulement légèrement plus "exacte" que les autres en ce qui concerne le critère morphologique soit systématiquement plus efficace au sens du caractère démographique. La "meilleure" stratification peut donc être tour à tour chacune des trois.

3. Comme on l'a déjà remarqué, et c'est une évidence, les conséquences sur la précision des estimations du choix de la probabilité d'appartenance à l'échantillon des unités primaires ne sont pas indépendantes de la stratification retenue. On peut à ce propos formuler la règle générale suivante : *plus la redondance entre les deux variables définissant d'une part la stratification et d'autre part la probabilité d'appartenance est forte, moins il y a d'intérêt à pratiquer avec cette stratification, une telle probabilité.*

Dans le cas qui nous intéresse, on trouve une bonne illustration de cette règle en comparant l'ensemble des colonnes "P=su" et "P=sb" dans les tableaux 10.5 et 10.6 : le gain sur le coefficient de variation des estimations obtenu en passant d'une probabilité d'inclusion proportionnelle à la surface totale des unités à une probabilité proportionnelle à leur surface bâtie est de l'ordre de 30 à 40% dans le cas non stratifié, il n'est plus que de 10 à 20 % lorsque la stratification de la base sur le critère de densité du bâti est "floue" (variable DENSBATIM) et il devient négligeable (on obtient même une détérioration de la précision pour certaines estimations) lorsque la stratification s'améliore (variables DENSBATI et DENSBATIME).

Plus que sa mise en évidence, qui peut paraître triviale, c'est la quantification du phénomène qui a eu une réelle importance pratique : *l'équivalence, en terme de précision des estimations obtenues, des deux stratégies de sélection des unités primaires (probabilités proportionnelles aux surfaces totales ou aux surfaces bâties), justifie l'emploi de la méthode de tirage la plus simple, rapide et adaptée à la base de sondage dont nous disposons : le sondage systématique aréolaire à l'aide d'une grille de points.*

10.3.2.3 : Allocations optimales

D'après les résultats précédents, pour le premier degré du sondage, on peut considérer comme acquis un plan stratifié d'après la densité du bâti observée sur l'image satellite, à probabilité proportionnelle à la surface totale des unités primaires. Les colonnes 5 et 6 du tableau 10.6 montrent que dans ce cadre, l'introduction de l'allocation optimale de l'échantillon aux strates améliore de façon significative les estimations sur l'ensemble de la base. En moyenne, passer de l'allocation proportionnelle à l'allocation optimale fait gagner 20% de la variance mais l'allocation optimale, fonction de la variance au sein de chaque strate du caractère que l'on estime, doit être redéfinie pour chaque variable. Les corrélations entre indicateurs démographiques observées lors de la phase exploratoire de l'expérience peuvent faire penser que l'évolution de la variance du caractère à estimer d'une strate à l'autre est semblable pour bon nombre des indicateurs démographiques. Si tel était le cas on pourrait construire une stratégie d'allocation "intermédiaire" c'est-à-dire qui soit un compromis entre les différentes allocations optimales, permettant d'améliorer l'ensemble des estimations. A partir du tableau 10.7 on peut comparer, pour l'ensemble des indicateurs démographiques retenus, les répartitions de l'échantillon entre les strates correspondant aux allocations optimales.

Tableau 10.7 : Comparaisons des allocations optimales obtenues pour l'ensemble des caractères démographiques considérés.(Programme VARECH)

Répartition de l'échantillon entre les strates (%)	Stratification "DENSBATIM"					Stratification "DENSBATIME"				
	Strates :					Strates :				
	1	2	3	4	5	1	2	3	4	5
A) Valeurs connues sur 46 îlots										
NOMBRE D'ILOTS	5	6	16	15	4	3	11	13	11	8
ALLOC. PROPORTION.	10,9	13,1	34,8	32,6	8,7	6,5	23,9	28,3	23,9	17,4
1) Totaux										
1 EFFECTRES	4,6	3,7	34,2	49,7	7,8	2,8	9,3	33,8	40,2	13,9
2 NBMENAGES	3,8	3,8	32,5	50,8	9,2	2,5	8,9	31,9	41,4	15,3
4 NBMENP1	3,9	4,2	32,6	50,8	8,5	2,6	9,5	31,6	41,3	14,9
6 NBMENM7	3,0	4,4	31,4	51,9	9,3	2,5	9,4	29,6	42,5	15,9
2) Quotients										
7 EFFECTRES / NBMENAGE	2,7	4,2	40,7	48,0	4,3	0,5	12,1	33,0	37,8	16,6
8 NBMENP1 / NBMENAGE	0,1	6,9	54,0	23,3	15,8	0,1	9,8	43,7	19,2	27,2
9 NBMENM7 / NBMENAGE	7,6	7,0	42,8	41,0	1,5	0,7	18,0	37,1	31,4	12,9

Tableau 10.7 (suite) : Comparaisons des allocations optimales obtenues pour l'ensemble des caractères démographiques considérés.(Programme VARECH)

Répartition de l'échantillon entre les strates (%)	Stratification "DENSBATIM"					Stratification "DENSBATIME"				
	Strates :					Strates :				
	1	2	3	4	5	1	2	3	4	5
B) Valeurs connues sur 44 îlotss										
NOMBRE D'ÎLOTS ALLOC.	5	6	15	14	4	3	11	12	11	7
PROPORTION.	11,4	13,7	34,1	31,8	9,1	6,8	25,0	27,3	25,0	15,9
1) Totaux										
8 EFFP5A	4,7	3,7	34,0	49,9	7,5	3,1	9,0	33,5	41,3	13,1
10 EFFNMIG	6,0	4,1	37,8	45,9	6,1	3,7	11,4	35,4	36,4	13,2
12 CEFFMIGM5	4,8	3,5	34,8	49,3	7,5	3,0	9,6	33,6	40,8	12,8
2) Quotients										
20 EFFP5A / EFFECTRES	7,6	6,3	34,5	44,7	6,8	6,0	14,6	30,2	35,3	14,0
22 EFFNMIG / EFFECTRES	6,9	5,5	39,8	37,7	10,1	1,5	21,5	26,1	26,4	24,5
24 CEFFMIGM5 / EFFECTRES	4,6	7,1	54,0	29,3	5,0	1,7	42,1	20,9	23,9	11,4

En premier lieu, l'observation des résultats conduit, une fois encore, à nuancer les conclusions selon que l'on s'intéresse aux effectifs ou aux quotients. Pour les premiers, toutes les répartitions correspondantes aux allocations optimales sont à peu près équivalentes. Pour les seconds, on observe une plus grande variabilité dans le détail des répartitions cependant, les strates sur et sous représentées sont toujours les mêmes : respectivement les strates 4 (sur-représentée) et les strates 1 et 2 (sous-représentée).

La seconde constatation a trait à l'effet du changement de stratification. Il modifie bien entendu la répartition optimale de l'échantillon entre les strates mais, nous semble-t-il, dans des proportions raisonnables. Les allocations "moyennes" sont les suivantes (dans le cas des effectifs qui fournissent des répartitions très régulières) :

	Strates :	1	2	3	4	5	Total
DENSBATIME							
Allocation proportionnelle :		6,5	23,9	28,3	23,9	17,4	100 %
Allocation "optimale moyenne" :		3	9	32	41	15	100 %

	Strates :	1	2	3	4	5	Total
DENSBATIM							
Allocation proportionnelle :		10,9	13,1	34,8	32,6	8,7	100 %
Allocation "optimale moyenne" :		4	4	33	50	9	100 %

Dans les deux cas les modifications apportées par rapport à l'allocation proportionnelle sont comparables : peu de changements dans les strates 3 et 5, augmentation sensible du taux dans la strate 4 au détriment des strates 1 et 2.

On peut donc conclure à une relative robustesse des stratégies d'allocation optimale vis à vis de modifications opérées soit sur la stratification soit concernant les variables à estimer. Partant de cette constatation, l'idée qui a été retenue pour construire la règle d'allocation aux strates est de moyenner les allocations optimales au sein de l'ensemble où elle apparaissent les plus stables c'est-à-dire les effectifs. Par ailleurs comme on l'a vu, DENSBATIM est la stratification qu'on appliquera à la base de sondage, c'est donc à partir des allocations optimales lui correspondant que l'on effectue le calcul. On parvient ainsi à une répartition qui correspond, par rapport à l'allocation proportionnelle, à une multiplication du taux dans chaque strate par les coefficients suivants :

Strates :	1	2	3	4	5	Total
Coefficients :	0,4	0,3	1	1,5	0,9	1

Cette allocation est surtout bénéfique à l'estimation des effectifs mais elle devrait également améliorer l'estimation de la plupart des quotients (voir tableau 10.7).

10.4 Conclusions

Au début de ce travail sur le premier degré du sondage nous avons posé quatre questions :

1. Quelle probabilité doit avoir chaque unité primaire d'appartenir à l'échantillon ?
2. Quels estimateurs employer pour l'étape d'inférence de l'échantillon de pâtés de maisons au total des strates ?
3. Quel gain de précision permet la stratification de la base de sondage à partir de l'image satellite ?
4. Y a-t-il des possibilités d'optimisation de l'allocation aux strates ?

En conclusion, en faisant la synthèse des résultats obtenus, on peut formuler pour chacune d'elle les réponses suivantes.

Définition de la probabilité d'appartenance des unités primaires à l'échantillon :

L'alternative est la suivante : la probabilité doit-elle être proportionnelle à la surface totale des unités ou à une approximation de leur surface bâtie.

Dans le cadre non stratifié, par rapport à la référence équiprobable, un sondage à probabilités proportionnelles aux surfaces totales des unités primaires ne présente pas d'intérêt ; par contre, une probabilité de sélection proportionnelle à la surface bâtie, ou à son approximation d'après l'image satellite permet une amélioration notable : 30 à 40 % de réduction du coefficient de variation dans le cas de la mesure de la surface bâtie, 20 à 30 % s'il s'agit de l'approximation par la surface en végétation.

Une fois stratifiée la base de sondage selon la densité du bâti, il faut par contre retenir le sondage à probabilités proportionnelles à la surface totale des unités. En effet, d'une part il est meilleur que le sondage équiprobable stratifié, d'autre part dans un tel cadre, le tirage avec probabilités proportionnelles aux surfaces bâties n'améliore plus l'estimation par rapport au tirage à probabilités proportionnelles aux surfaces totales. *Ce résultat, qui s'ajoute à la nécessité d'une méthode de tirage simple, rapide et adaptée à la base de sondage dont nous disposons, justifie l'emploi du sondage systématique aréolaire à l'aide d'une grille de points.*

Définition de l'estimateur à employer :

Les seules informations exogènes qu'on puisse envisager d'utiliser dans l'étape d'inférence (estimation par le ratio ou par la régression), sont à nouveau les mesures de taille des unités primaires. *Les coefficients de corrélation entre ces mesures et les indicateurs démographiques, compris entre 0,5 et 0,6, sont nettement insuffisants pour justifier l'emploi de ces estimateurs pour l'estimation des totaux, mais, comme toujours, l'estimateur du ratio reste utile pour l'estimation directe des quotients; pour le reste, on utilisera les estimateurs sans biais¹.*

Choix des critères morphologiques de stratification :

Le gain que permet la stratification en cinq niveaux de densité de bâti choisie comme référence se situe autour de 50 % du coefficient de variation (estimation des effectifs). Au terme de l'expérience, on peut penser que la stratification issue de l'image permet une amélioration comparable, du moins une fois optimisée l'allocation de l'échantillon aux strates. C'est donc cette stratification qui sera utilisée lors de l'enquête finale ; néanmoins il serait bon, si l'on prolonge ces recherches, de tester d'autres classifications de l'image satellite.

Définition d'une règle d'allocation aux strates :

En moyenne, *passer de l'allocation proportionnelle à l'allocation optimale fait gagner 20% de la variance et la robustesse des règles d'allocation optimale vis à vis des modifications opérées soit sur la stratification soit concernant les variables à estimer autorise à "moyenner ces allocations".* On parvient ainsi à une allocation qui ne modifie presque pas la part de l'échantillon affectée aux strates 3 et 5 (densités moyenne et très forte), mais qui favorise nettement la strate 4 (densité forte) au détriment des strates 1 et 2 (densités faible et très faible). Cette allocation, surtout

¹ : Le formulaire associé au plan de sondage que nous appliquerons pour l'enquête finale est donné en résumé au chapitre 12 (§ 12.2), ou plus complètement, du moins en ce qui concerne le premier degré du sondage, dans le chapitre 6 (§ 6.2).

bénéfique à l'estimation des effectifs, devrait néanmoins améliorer l'estimation de la plupart des quotients.

Au terme de cette étape on est parvenu à proposer des solutions concrètes aux problèmes de sélection et d'inférence liés à l'échantillonnage des pâtés de maisons. On sait de plus quelles conséquences aura la mise en œuvre de ces solutions sur la précision des estimations. Les contraintes, liées à l'emploi de l'image satellite comme base de sondage et aux objectifs de rapidité, simplicité et faible coût que l'on s'était fixés au départ, s'avèrent compatibles avec une amélioration sensible de la précision des estimations grâce à l'emploi de la stratification, du tirage à probabilités inégales et d'une allocation de l'échantillon aux strates optimisée. Pour aboutir à la définition complète du plan de sondage, il reste à décrire les étapes de sélection et d'inférence correspondant au second degré du sondage : l'échantillonnage des ménages au sein des pâtés de maisons. C'est l'objet du chapitre suivant.

Chapitre 11

ENQUETE PRELIMINAIRE "BATI POPULATION", EXPLOITATION POUR LA MISE AU POINT DU SECOND DEGRE DU SONDAGE : ECHANTILLONNAGE DES MENAGES DANS LES PATES DE MAISONS

11.1 : Introduction, les objectifs principaux et la démarche

Le principal résultat de l'expérience réalisée à Marseille et de la première exploitation de l'enquête "bâti population" faite à Quito est la détermination du premier degré du sondage. Dans ce chapitre on tente de compléter la définition du plan de sondage en abordant, à partir de ces mêmes données de l'enquête "bâti population", le problème de l'échantillonnage des ménages au sein des pâtés de maisons. Cette tentative comporte deux volets aux objectifs bien différents.

Le premier est l'analyse, à l'aide d'outils statistiques classiques, des possibilités "d'explication" des caractéristiques démographiques des ménages par trois facteurs : la morphologie du pâté de maisons, sa localisation dans l'agglomération et la morphologie du bâtiment occupé par le ménage. L'objectif assigné à cette analyse du fichier "ménages" est la mesure du degré d'hétérogénéité (ou d'homogénéité) que présentent les caractéristiques démographiques des ménages au sein des sous-ensembles définis par les facteurs explicatifs. *Plus concrètement, il s'agit en fait de statuer sur la possibilité et l'intérêt d'une stratification (ou post-stratification) du second degré du sondage.*

Dans le second volet, on aborde réellement le problème de la sélection des ménages dans chaque pâté de maisons de l'échantillon. Après avoir exposé, en termes "intuitifs", *comment un sondage à deux degrés peut permettre, en enquêtant un même nombre total de ménages, d'améliorer la précision des estimations par rapport au sondage où toutes les unités primaires sont enquêtées exhaustivement*, on formalise cette idée dans le formulaire. Il fournit les expressions des variances des estimateurs dans le cas d'un plan de sondage à deux degrés. A partir de ces expressions, on a construit le programme VAR2DEG qui permet de mesurer la part de variance due à chacun des deux degrés du sondage. Le principe est ensuite de considérer l'échantillon de pâtés de maisons de l'enquête "bâti population" comme une base de sondage sur laquelle le programme effectue les calculs des variances correspondant aux plans envisagés. *Les résultats permettent d'évaluer le gain de précision en fonction de la fraction de sondage au second degré et donc de déterminer la stratégie optimale de sélection des ménages dans chaque pâté de maisons.*

En conclusion de chacun des deux volets, on trouvera un résumé des principaux résultats qui débouche sur la solution pratique retenue dans le plan de sondage.

11.2 Analyses préliminaires : les possibilités "d'explication" des caractéristiques démographiques des ménages par la morphologie du bâtiment, la morphologie et la localisation de l'îlot.

11.2.1 Les questions posées, les données et les outils statistiques utilisés

11.2.1.1 : Les questions posées

La solution au problème posé passe par l'évaluation et la comparaison du pouvoir discriminant des facteurs explicatifs : ceux caractérisant les unités primaires d'une part (morphologie et localisation du pâté de maisons), celui pouvant servir de base à la stratification au sein de ces unités d'autre part (morphologie du bâtiment occupé par le ménage). Il nous faut donc répondre à plusieurs questions :

1 : Quel est le pouvoir discriminant, pour chacun des caractères démographiques considérés, de chaque facteur indépendant pris isolément : morphologie du bâtiment, morphologie de la manzana, localisation? Quelles sont les variables démographiques se prêtant réellement à une discrimination ces facteurs? Avec ceux des facteurs indépendants "naturellement ordonnés", pour lesquels un codage métrique est possible, peut-on construire des modèles de régression permettant l'estimation des variables démographiques?

2 : Lorsqu'il existe une possibilité réelle de discrimination d'une variable démographique donnée par les facteurs indépendants, les discriminations opérées par chacun des critères "explicatifs" sont-elles redondantes ou au contraire peut-on s'attendre à un effet d'addition? Au cas où un modèle intégrant l'ensemble des critères permet en effet une discrimination plus fine, ce modèle peut-il être limité aux effets additionnés ou doit-il intégrer les interactions entre critères (d'ordre 1 ou supérieures) ? Autrement dit, l'effet d'un critère "explicatif" donné est-il le même quelles que soient les valeurs ou modalités des autres critères?

3 : *En conclusion, la liaison existant entre caractéristiques démographiques des ménages et morphologie des bâtiments est elle suffisamment forte pour justifier la stratification ou la post-stratification du second degré du sondage?*

11.2.1.2 : Les données

Les analyses statistiques réalisées portent sur un fichier de 1381 enregistrements correspondant à l'ensemble des ménages enquêtés. Chaque ménage est décrit par quatre caractères démographiques et par les trois facteurs "explicatifs" :

Variables démographiques

1. EFFECT : Effectif du ménage (nombre de personnes résidentes).
2. REVENU : Revenu moyen des personnes du ménage (somme des revenus déclarés par chacune des personnes du ménage divisée par le nombre de personnes) ; et également son logarithme en base 10 : LOGREV.
3. AGECHF : Age du chef de ménage en années ou tranches d'âge quinquennales (AGECHFR).
4. DURESMIG : Durée moyenne de résidence à Quito des migrants du ménage (somme des durées de résidence déclarées par chacun des migrants du ménage divisée par le nombre des migrants) en années ou tranches d'âge quinquennales (DURESMIGR).

Facteurs explicatifs

1. TYPÉMORF : Type morphologique du bâtiment occupé par le ménage. Typologie issue de la classification en 38 types résultant de l'exploitation de l'enquête préliminaire "morphologie" après regroupement de certaines classes morphologiquement proches pour aboutir à 14 types (fig. 11.1 & 11.2).

Figure 11.1 : Regroupement des 38 types morphologiques de bâtiments en 14 nouveaux types, rappel de la typologie des bâtiments en 38 types

Surface au sol	Nombre D'étages	Béton ou Eternit	Tuiles vernies, Carrelage, Shingle	Zinc	Tuiles non vernies
1 - 29 m ²	1 ét.	1	1	1	1
30 - 79 m ²	1 ét.	3	12	21	30
80 - 224 m ²	1 ét.	4	13	22	31
225 - 14999 m ²	1, 2 ét.	5	14	23	32
1 - 29 m ²	2, 3 ét.	6	15	24	33
30 - 79 m ²	2, 3 ét.	7	16	25	34
80 - 224 m ²	2, 3 ét.	8	17	26	35
225 - 14999 m ²	2, 3 ét.	9	18	27	36
30 - 79 m ²	4, 5 ét.	10	19	28	37
80 - 224 m ²	4, 5 ét.	11	20	29	38
225 - 14999 m ²	5 à 15 ét.	2	2	2	2

(Dans chaque case du tableau, est noté le numéro du type de bâtiment, selon la surface au sol, le nombre d'étages et le matériau du toit)

Figure 11.2 : Regroupement des 38 types morphologiques de bâtiments en 14 nouveaux types, regroupements effectués pour parvenir aux 14 types

Nouveau type	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Ancien type	1	3	7	21	30 & 31	32 & 36	22,25 & 26	34 & 35	4	5 & 9	8	11	14,17 & 20	2
Nb de bâtiments	158	112	61	50	70	43	20	60	147	79	175	129	39	140

Les anciens types 6, 10, 13, 13, 15, 16, 18, 19, 23, 24, 27, 28, 29, 33, 37 et 38, qui regroupent au total 23 bâtiments, ont été exclus de la nouvelle typologie parcequ'ils contenaient des bâtiments non occupés ou que ces types concernaient trop peu de ménages dans l'échantillon.

2. TPEMANZ : Type morphologique du pâté de maisons occupé par le ménage. Classification en 9 types obtenue lors de l'exploitation de l'enquête préliminaire "morphologie". Pas de type "non bâti", 8 modalités¹.

3. LOC : Appartenance de l'îlot à un des 12 sous-ensembles géographiques définis dans la ville, issus du croisement d'une division Nord/Sud en 4 zones avec une division Est/Ouest en 3 zones.²

Cependant, du fait d'absences ou de refus de réponses, l'ensemble de ces données n'est pas disponible sur la totalité des 1381 ménages enquêtés et le nombre de ménages pour lesquels les données sont connues varie suivant les caractères considérés. Plutôt que de chercher à constituer un fichier sur lequel on dispose de la totalité de l'information, ce qui restreint beaucoup le nombre de ménages pris en compte dans la plupart des analyses, on a préféré ne conserver pour chaque analyse que l'ensemble des ménages pour lesquels les données nécessaires sont connues. Ainsi, selon les variables prises en compte dans tel ou tel traitement, l'analyse s'effectue sur un nombre plus ou moins grand de ménages. Ce nombre sera toujours indiqué pour chaque variable ou groupe de variables dans les tableaux de résultats.

¹ : Voir la description des types morphologiques d'îlots au chapitre 9, tableau 9.6.

² : Les modalités de la variable sont les suivantes :

- Modalités 1, 2 & 3 : zone sud (1: sud ouest, 2 : sud, 3 : sud est)
- Modalités 4, 5 & 6 : zone centre-sud (4: centre-sud ouest, 5 : centre-sud, 6 : centre-sud est)
- Modalités 7, 8 & 9 : zone centre-nord (7: centre-nord ouest, 8 : centre-nord, 9 : centre-nord est)
- Modalités 10, 11 & 12 : zone nord (10: nord ouest, 11 : nord, 12 : nord est)

11.2.1.3 : Les outils statistiques utilisés

Pour aborder les questions posées on a utilisé, tour à tour et selon la nature du problème et des données, des techniques d'analyse de la variance¹ à un ou plusieurs critères, avec ou sans interactions et de régression simple ou multiple².

Ainsi, pour mesurer le pouvoir discriminant de chacun des facteurs explicatifs pris isolément, on a construit à l'aide du logiciel SPSS³ les modèles d'analyse de la variance à un critère correspondant à l'ensemble des combinaisons possibles entre variables dépendantes (chacun des quatre caractères démographiques) et variables indépendantes (morphologie du bâtiment, morphologie de la manzana, localisation de la manzana). Le pourcentage de variance expliquée par la variable indépendante (variance inter-classe) permet alors d'évaluer son pouvoir discriminant sur chacune des variables démographiques.

La construction de modèles de régression linéaire qui permettent la prévision des caractéristiques démographiques des ménages à l'aide de l'information morphologique ou de la localisation dans l'agglomération ne peut être envisagée qu'à condition que les variables exogènes puissent être codées sur des échelles métriques. Les données recueillies ne permettent un tel codage que pour les deux variables décrivant la localisation de l'îlot au sein de l'agglomération : latitude et longitude d'un point choisi comme centre du pâté de maisons. Ce sont ces deux variables qui ont été prises comme facteurs explicatifs. Malheureusement, mais cela était prévisible, les coefficients de régressions obtenus, que ce soit lors de régressions simples ou multiples⁴, ne sont jamais supérieurs à 0,3. C'est évidemment très insuffisant pour espérer prévoir, avec la technique des moindres carrés, les caractéristiques des ménages. Nous n'exposerons donc pas plus avant les résultats obtenus puisqu'ils sont de toute manière dénués d'intérêt pratique.

Enfin, pour essayer de mesurer le degré de redondance existant entre les trois informations exogènes que nous considérons, cela lorsqu'au moins deux des critères explicatifs avaient un pouvoir discriminant réel sur une variable démographique, nous avons soumis les données à des analyses de la variance multi-critères⁵. En comparant les parts de variance expliquée par les modèles intégrant plusieurs facteurs explicatifs, modèles à deux puis à trois critères, à celles expliquées par les modèles à un critère, cette technique a permis de voir dans quelle mesure les effets explicatifs des différents critères s'additionnent. Dans cette étape on a également cherché à savoir s'il existe des effets d'interaction entre les trois critères. Après avoir recodé les variables pour limiter les modalités prises par les critères obtenus par croisement et l'espace mémoire nécessaire aux calculs, on construit les modèles comportant les termes d'interactions d'ordre un (interactions entre deux critères) et les résultats permettent de conclure.

¹ : sur l'analyse de la variance, voir par exemple : L. LEBART , A. MORINEAU et J.P. FENELON : Traitement des données statistiques, méthodes et programmes, 2^e édition, DUNOD, 1982, Paris.

² : sur la régression, voir E. MALINVAUD (1981, 3^e édition) : Méthodes statistiques de l'économétrie. DUNOD, Paris, 846 p.

³ : Statistical Package for Social Sciences, programme ONEWAY. Pour documentation, se reporter au manuel de référence du logiciel : SPSS/PC+ documentation (1986), Chapitre 14, One way analysis of variance, procedure ONEWAY, section B, p. 152.

⁴ : On a utilisé les programme PLOT et REGRESSION de SPSS. Voir SPSS/PC+ documentation (1986), chapitres PLOT (pp. C 114 à C 119) et REGRESSION (pp. C 123 à C 144).

⁵ : Programme ANOVA de SPSS. SPSS/PC+ Documentation (1986), chapitre 15 : Analysis of variance, procedure ANOVA, section B, p. 164.

11.2.2 : Résultats

11.2.2.1 : Analyse de la variance à un critère : pouvoir discriminant des facteurs "explicatifs" sur les différents caractères démographiques étudiés.

- Préliminaire : condition d'interprétabilité des analyses

Pour interpréter légitimement les résultats d'une analyse de variance, on doit avant tout s'assurer que les données entrent bien dans le cadre des hypothèses qui fondent la méthode. La condition principale pour cela est que la variance du caractère analysé n'apparaisse pas significativement différente d'une classe à l'autre¹. Les sorties fournies par le programme SPSS contiennent, comme indicateur d'homogénéité des variances, le rapport : variance maximale/variance minimale ; lignes 4, 8 et 12 du tableau 11.3. Pour juger une analyse "interprétable", on a considéré que la valeur de ce rapport ne doit pas dépasser 3. Les suppressions d'individus trop originaux ou le recodage des variables² ont permis dans presque tous les cas d'atteindre ce seuil d'interprétabilité. La seule analyse pour laquelle les résultats doivent être considérés prudemment est celle "expliquant" l'âge du chef de ménage par les modalités de la variable "localisation".

- Résultats

Dans le tableau 11.3, pour chacun des trois facteurs indépendants examinés, on retient cinq indications fournies par le programme d'analyse de la variance :

1. Effectif : Nombre de ménages pris en compte dans cette analyse
2. Part de la variance expliquée : quotient (exprimé en %) de la variance inter-classes sur la variance totale.
3. Rejet de l'hypothèse d'égalité des moyennes : il s'agit du résultat de l'application du test de Fisher au seuil de confiance de 5%.
 - oui : $P < 0,05$, rejet de l'hypothèse d'égalité
 - non : $P > 0,05$, pas de rejet
4. Indicateur d'homogénéité des variances selon les classes : quotient de la variance dans la classe la plus hétérogène sur la variance dans la classe la plus homogène.

¹ : Ici, comme en général en analyse de la variance, on entend par classe l'ensemble des individus ayant une modalité donnée pour le caractère exogène.

² : Les caractéristiques des variables démographiques avant et après "recodage" sont les suivantes :

Variables "brutes" :

- 1. EFFECT : De 0 à 14 personnes, 1381 ménages
- 2. REVENU : De 0 à 205 000 sucres, 751 ménages
- 3. AGECHF : En années, 1086 ménages
- 4. DURESMIG : En années, 745 ménages

Variables "recodées" :

- 1. EFFECTR : Suppression du ménage sans résident et des ménages de plus de 9 personnes (7 ménages supprimés, restent 1374 ménages).
- 2. REVENUR : suppression des revenus nuls et supérieurs à 45 000 sucres (11 ménages supprimés, restent 740 ménages).
- 3. LOGREV : Passage au logarithme en base 10 sur REVENUR (740 ménages).
- 4. AGECHFTR : Transformation en tranches d'âge quinquennales (1086 ménages).
- 5. DURESMIGR : Transformation en tranches d'âge quinquennales (745 ménages).

Tableau 11.3 : Résultats résumés des analyses de la variance à un critère
(Programme "ONEWAY", SPSS)

		EFFECT	EFFECTR	REVENU	REVENUR	LOGREV
L O C	Nombre de ménages	1381	1374	762	751	751
	Part de variance expliquée en % de la variance totale	4,3	3,7	13,8	23,5	27,9
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	oui	oui	oui
	Indicateur d'homogénéité des variances (VM/Vm)	3,64	2,21	135,94	35,29	2,79
	Nombre de classes différentes au sens de la procédure de Scheffé	2/66	1/66	9/66	25/66	21/66
T Y P E M O R F	Nombre de ménages	1283	1276	725	714	714
	Part de variance expliquée en % de la variance totale	3,4	3,1	15,0	21,6	23,6
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	oui	oui	oui
	Indicateur d'homogénéité des variances (VM/Vm)	4,80	3,04	88,06	13,14	2,91
	Nombre de classes différentes au sens de la procédure de Scheffé	1/91	0/91	11/91	22/91	20/91
T Y P M A N Z	Nombre de ménages	1381	1374	762	751	751
	Part de variance expliquée en % de la variance totale	2,2	2,0	13,4	21,5	24,1
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	oui	oui	oui
	Indicateur d'homogénéité des variances VM/Vm)	2,45	2,17	21,17	8,79	2,27
	Nombre de classes différentes au sens de la procédure de Scheffé	1/28	2/28	6/28	11/28	14/28

Tableau 11.3 (suite) : Résultats résumés des analyses de la variance à un critère (Programme "ONEWAY", SPSS)

		AGECHEF	AGECHEFR	DURES MIG	DURES MIGR
L O C	Nombre de ménages	1086	1086	745	745
	Part de variance expliquée en % de la variance totale	5,4	5,3	2,6	2,5
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	non	non
	Indicateur d'homogénéité des variances (VM/Vm)	4,48	4,52	2,1	2,2
	Nombre de classes différentes au sens de la procédure de Scheffé	5/66	6/66	0/66	0/66
T Y P E M O R F	Nombre de ménages	1017	1017	702	702
	Part de variance expliquée en % de la variance totale	2,9	2,9	2,3	2,4
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	non	non
	Indicateur d'homogénéité des variances (VM/Vm)	2,39	2,49	2,30	2,70
	Nombre de classes différentes au sens de la procédure de Scheffé	0/91	0/91	0/91	0/91
T Y P M A N Z	Nombre de ménages	1086	1086	745	745
	Part de variance expliquée en % de la variance totale	2,0	2,0	0,7	0,7
	Rejet de l'hypothèse d'égalité des moyennes	oui	oui	non	non
	Indicateur d'homogénéité des variances (VM/Vm)	2,32	2,23	2,12	2,18
	Nombre de classes différentes au sens de la procédure de Scheffé	2/28	2/28	0/28	0/28

5. Nombre de classes différentes : il s'agit des classes significativement différentes au sens de la procédure de Scheffé¹. On donnera dans le tableau le nombre de différences constatées au seuil fixé à l'avance, 5 ou 10% selon les cas, rapporté au nombre de différences N théoriquement possibles ($N = n.(n-1)/2$ où n est le nombre de classes de la variable indépendante).

- Conclusions

1 : Durée moyenne de résidence des migrants du ménage

Aucun des trois facteurs indépendants ne possède de pouvoir discriminant notable sur cette variable. L'hypothèse d'égalité des moyennes n'est jamais rejetée et aucune classe n'apparaît significativement différente d'une autre au sens de Scheffé.

2 : Effectif des résidents du ménage et âge du chef de ménage

Pour ces deux variables, certaines possibilités de discrimination existent à l'aide de la localisation et du type morphologique de la manzana, par contre le type morphologique du bâtiment n'a pas d'effet notable, aucune classe n'étant significativement différente d'une autre. Les parts de variance expliquées restent de toute manière très faibles (inférieures à 5%); les quelques différences entre moyennes à noter sont les suivantes :

Taille des ménages :

Les classes de localisation 1 (zone extrême sud-ouest, majoritairement péri-urbaine où les ménages ont des effectifs importants) et 9 (zone nord-est incluant "El Batán", quartier résidentiel de standing où les ménages sont en général de petite taille) s'opposent. De même, le type d'îlots 2 (péri-urbain), avec des ménages nombreux se différencie des types 8 (grands ensembles) et 5 (centre colonial) où les ménages sont peu nombreux.

Age du chef de ménage :

Les jeunes chefs de ménage résident surtout dans les zones 3 (extrême sud-est), 6 (sud-est) et 12 (nord-est), tandis que la zone 8 (centre nord, nouveau centre d'affaires) regroupe un nombre important de chefs de ménage âgés. Les zones 5 (centre sud colonial) et 9 (nord est résidentiel) offrent en général des valeurs intermédiaires de la variable. Quant au type morphologique de l'îlot, l'habitat de standing (classe 6) avec ses "vieux" chefs de ménage, s'oppose aux quartiers "jeunes" péri-urbain et d'habitat populaire (respectivement classes 2 et 3).

3 : Revenu moyen des personnes du ménage

Les facteurs explicatifs retenus permettent tous les trois une discrimination intéressante des valeurs prises par le logarithme en base 10 du revenu moyen des personnes du ménage. Les trois facteurs ont des pouvoirs discriminants équivalents se situant autour de 25% de la variance totale (localisation 28%, morphologie de la manzana et du bâtiment 24%). Le passage au logarithme sur cette variable améliore nettement les résultats, tant du point de vue de la variance expliquée que de celui de l'homogénéité des variances au sein des classes, cela dans la même proportion pour les trois facteurs explicatifs. Le principal résultat à noter est le suivant : le facteur

¹ : A propos de la procédure de Scheffé, voir la documentation du logiciel SPSS : SPSS/PC+ Documentation (1986), chapitre 14 (procédure ONEWAY), paragraphe 14.6, section B, p. 156.

"localisation" permet de discriminer quatre catégories de revenus décrites dans le tableau 11.4 ci-dessous. Les modalités des facteurs explicatifs regroupées pour former chaque catégorie y sont indiquées par leurs numéros. Pour interpréter ces regroupements en terme de localisation dans l'agglomération ou de morphologie des bâtiments ou des pâtés de maisons, il faut se reporter à la description des modalités¹.

Tableau 11.4 : Regroupements effectués d'après les analyses de la variance à un critère pour obtenir des catégories de revenus.

Modalités de la variable explicative et nombre de ménages concernés		Catégories de revenus et moyenne de la variable (en sucres/mois)		
Modalités	Ménages	Catégories	Moyennes	
L	1 : 1, 2, 3, 4, 7, 12	236	Petits revenus	5200
	2 : 6,10	87	Moyens revenus	7300
	3 : 5,11	247	Hauts revenus	10700
	4 : 8,9	181	Très hauts revenus	15500
	Total :	751	Moyenne d'ensemble :	9700
T	1 : 1,2,4,5,6	268	Petits revenus	6100
	2 : 3,7,9	189	Moyens revenus	9700
	3 : 8,10,11,12,13	199	Hauts revenus	12300
	4 : 14	58	Très hauts revenus	17500
	Total :	714	Moyenne d'ensemble :	9700
T	1 : 2,4,7	287	Petits revenus	5850
	2 : 3,5	186	Moyens revenus	9200
	3 : 6,9	78	Hauts revenus	11650
	4 : 8	200	Très hauts revenus	14400
	Total :	751	Moyenne d'ensemble :	9700
N				
Z				

11.2.2.2 : Analyse de la variance multi-critères : redondance/indépendance de l'information apportée par les trois facteurs "explicatifs".

D'après les résultats précédents, il est clair que seule la variable indiquant le revenu moyen des personnes du ménage, plus précisément son logarithme en base 10, se prête à une discrimination intéressante à l'aide des trois facteurs explicatifs considérés. *Dans la suite des traitements, où il s'agit d'intégrer dans une même analyse plusieurs facteurs, on n'a donc retenu comme variable à expliquer que cet indicateur du revenu.* Le tableau 11.5 donne la part de variance (en % de la variance totale) expliquée par les facteurs indépendants dans les trois analyses à deux critères possibles (croisements de deux facteurs) et dans l'analyse intégrant les trois facteurs.

¹ : Les modalités de la morphologie des pâtés de maisons sont décrites au chapitre 9 (tableau 9.6), celles de la morphologie du bâtiment au début de ce chapitre (figure 11.1). Les modalités de la localisation ont été données en note au paragraphe 11.2.1.2.

Figure également en première ligne le rappel des résultats obtenus avec les analyse à un seul critère, enfin les colonnes "gain" expriment le gain obtenu lors de l'introduction dans l'analyse du deuxième ou troisième critère (en pourcentage de la variance expliquée par l'analyse à un critère).

L'examen du tableau 11.5 montre que la redondance d'information entre les trois facteurs "explicatifs" est importante sans être totale. On peut d'ailleurs classer les combinaisons deux à deux des facteurs de la moins redondante à la plus redondante :

1 : localisation et morphologie du bâtiment

2 : morphologie du bâtiment et morphologie de l'îlot.

3 : localisation et morphologie de l'îlot.

Cependant même la dernière combinaison permet une amélioration non négligeable du pourcentage de variance expliquée si on la compare à l'analyse n'intégrant qu'un seul des deux critères.

Tableau 11.5 : Pourcentage de variance expliquée dans les analyses de variance à un, deux et trois critères.

	Localisation		Morphologie du bâtiment		Morphologie de l'îlot		
	% de var. expliquée	Gain	% de var. expliquée	Gain	% de var. expliquée	Gain	
Un seul critère	28 %	0 %	24 %	0 %	24 %	0 %	
D e u x c r i. t e r i e s	Localisation	34,5 %	43,2 %	32,2 %	34,2 %
	Morphologie du bâtiment	34,5 %	23,2 %	29,5 %	22,9 %
	Morphologie de la manzana	32,2 %	15 %	29,5 %	22,9 %
Trois critères	36,5 %	30,4 %	36,5 %	52,1 %	36,5 %	52,1 %	

Le passage de deux à trois critères apporte à nouveau un léger gain, mais on constate que ce dernier est très faible lorsque le troisième facteur introduit est la morphologie du bâtiment. *On peut donc penser qu'une fois prise en compte dans la procédure de sondage la morphologie des manzanas (par stratification du premier degré du sondage sur critère morphologique) et la localisation (par tirage systématique assurant la bonne répartition géographique de l'échantillon), il y a relativement peu d'intérêt à stratifier le second degré du sondage (le choix de l'échantillon de ménages à enquêter dans le pâté de maison) sur un critère tel que la morphologie du bâtiment.* Ce résultat important sera retenu pour l'élaboration du plan de sondage final.

Le tableau 11.6, reprenant le même principe que le tableau 11.5, fournit les chiffres permettant d'apprécier le gain qu'amène la prise en compte dans les analyses de variance, des interactions d'ordre 1. Le test a été mené après un recodage des variables

indépendantes qui réduit le nombre de modalités prises : on a regroupé les modalités qui n'étaient pas significativement différentes, du point de vue de la variable à "expliquer", lors des analyses à un critère. Ce faisant, on ne modifie que très peu la part de la variance "expliquée" comme en témoigne la première ligne du tableau. Dans une analyse de la variance à plusieurs critères, si l'on pense que l'effet du critère A peut être différent selon la modalité du critère B, on introduit dans le modèle un terme qui exprime l'interaction¹.

Tableau 11.6 : Pourcentage de variance expliquée dans les analyses de variance à un, deux et trois critères, après regroupement de certaines modalités des variables explicatives et prise en compte des interactions d'ordre 1.

	Localisation		Morphologie du bâtiment		Morphologie de l'îlot	
	% de var. expliquée	Gain	% de var. expliquée	Gain	% de var. expliquée	Gain
Un seul critère	27 %	0 %	22 %	0 %	22 %	0 %
Localisation (sans interaction)	30,5 %	38,6 %	30,3 %	37,7 %
Localisation (avec interaction)	32 %	45,5 %	31,9 %	45 %
Morphologie du bâtiment (sans interaction)	30,5 %	13 %	27,7 %	25,9 %
Morphologie du bâtiment (avec interaction)	32 %	18,5 %	28 %	27,3 %
Morphologie de la manzana (sans interaction)	30,3 %	12,2 %	27,7 %	25,9 %
Morphologie de la manzana (avec interaction)	31,9 %	18,1 %	28 %	27,3 %
Trois critères (sans interaction)	32,2 %	19,3 %	32,2 %	46,4 %	32,2 %	46,4 %

¹ : Voir L. LEBART , A. MORINEAU et J.P. FENELON : Traitement des données statistiques, méthodes et programmes, 2^e édition, DUNOD, 1982, Paris. p 231

L'effet discriminant de la morphologie des manzanas est-il par exemple notablement différent selon que l'on se trouve dans telle ou telle zone de Quito?

On peut répondre à ce type de question en examinant, à partir des résultats rapportés dans le tableau 11.6 si l'introduction des termes d'interaction est responsable d'une amélioration sensible des performances du modèle. En fait, à l'examen du tableau, il apparaît qu'aucune des interactions possibles n'a d'effet réellement important. Pour l'interaction entre morphologie de la manzana et morphologie du bâtiment par exemple, le test effectué ne permet pas de rejeter l'hypothèse selon laquelle l'effet du type de bâtiment est le même quelque soit le type de manzana. Quant au gain obtenu sur la part de variance expliquée par le modèle, il n'exède pas 5% de la variance expliquée par le modèle sans inter-action. *On pourra donc dans la suite considérer que, même si elles existent, on ne peut pas attendre de la prise en compte de ce type d'interactions (là encore par stratification croisée ou post-stratification) une amélioration sensible de la précision du sondage.*

11.2.3 Conclusions

Avant de résumer les principaux résultats acquis au cours de l'analyse du fichier "ménages", rappelons-en l'objectif principal. Il s'agissait, après évaluation et comparaison du pouvoir discriminant des facteurs explicatifs, de statuer sur la possibilité et l'intérêt d'une stratification (ou post-stratification) du second degré du sondage.

Concernant les possibilités de discrimination des caractéristiques démographiques des ménages par la morphologie du bâtiment ou du pâté de maisons dans lequel ils résident, ou leur localisation dans l'agglomération, les analyses de variances à un critère fournissent les indications suivantes.

Premièrement aucun des trois facteurs indépendants ne possède de pouvoir discriminant notable sur la durée moyenne de résidence des migrants du ménage. Deuxièmement, en ce qui concerne l'effectif des résidents du ménage et l'âge du chef de ménage, certaines possibilités de discrimination existent à l'aide de la localisation et du type morphologique de l'îlot, par contre le type morphologique du bâtiment a peu d'intérêt.

Enfin les facteurs explicatifs retenus permettent tous les trois une discrimination intéressante (autour de 25% de la variance totale) du revenu moyen des personnes du ménage. Le pouvoir "explicatif" des trois facteurs apparaît donc très inégal suivant qu'on considère l'un ou l'autre des caractères démographiques. C'est en particulier le cas pour la morphologie du bâtiment occupé par le ménage qui ne permet de discriminer que des classes de revenus. *On peut donc d'ores et déjà affirmer qu'une stratification du second degré du sondage sur ce critère n'améliorerait significativement les estimations que pour une minorité des indicateurs démographiques qui nous intéressent.*

En conséquence, la suite de l'expérience n'a porté que sur l'indicateur du revenu moyen des ménages ; elle s'attache, au moyen des analyses de la variance multi-critères, à l'évaluation du degré de redondance existant entre les informations apportées par les trois facteurs "explicatifs" ainsi qu'à l'examen des interactions entre ces facteurs. Les résultats obtenus montrent que la redondance d'information entre les trois facteurs "explicatifs" est importante sans être totale. Mais, on constate que le gain lors du passage de deux à trois critères est très faible lorsque le troisième facteur introduit est la morphologie du bâtiment. *En termes de procédures de sondage, cela veut dire qu'une fois prise en compte au niveau des unités primaires, la morphologie des pâtés de maisons (par stratification sur critère morphologique) et la localisation (par tirage systématique assurant la bonne répartition géographique de l'échantillon), il n'y a plus d'intérêt à stratifier le second degré du sondage sur un critère de morphologie du bâtiment. De même, s'agissant des interactions entre ces*

trois facteurs, on pourra considérer que même si elles existent, on ne peut pas attendre de leur prise en compte - là encore par stratification croisée ou post-stratification - une amélioration sensible de la précision du sondage.

Au terme de ces analyses effectuées sur les données concernant les unités secondaires du sondage, les ménages, la réponse à la question principale est donc négative. On n'a vraisemblablement pas intérêt à stratifier ou à post-stratifier le second degré du sondage en opérant une division morphologique à l'intérieur des pâtés de maisons. C'est un résultat qui, même s'il était prévu, est important pour la définition du plan de sondage final. Mais en revanche, nous allons voir maintenant qu'il existe de réelles possibilités d'amélioration de la précision des estimations si, en n'enquêtant qu'une fraction des ménages de chaque pâté de maisons, on parvient à augmenter la taille de l'échantillon d'îlots.

11.3 Détermination du second degré du sondage : échantillonnage de ménages dans un pâté de maisons.

11.3.1 Position du problème

Jusqu'à présent, dans toute l'expérience réalisée à Marseille et dans les premiers traitements des données recueillies à Quito, on a toujours considéré le pâté de maisons comme l'unité observée et cette unité était connue exhaustivement, soit que les données aient déjà été agglomérées à ce niveau comme c'est le cas pour le recensement de la population à Marseille, soit qu'on ait enquêté tous les ménages des îlots comme cela a été fait à Quito lors de l'enquête bâti/population. En termes techniques, *on n'a donc pratiqué jusqu'à maintenant que des sondage en grappe* : l'unité d'échantillonnage (le pâté de maisons) regroupe un certain nombre d'unités statistiques (les ménages), lesquelles sont toutes enquêtées. Connaissant à Quito, sur l'échantillon de pâtés de maisons enquêté, les caractéristiques démographiques moyennes de la population de l'îlot et celles de chacun des ménages résidents, on peut maintenant s'intéresser à des sondages à plusieurs degrés.

Comme nous l'avons vu, les arguments les plus souvent avancés en faveur de ce type de sondages concernent d'une part la difficulté qu'il y a à sélectionner directement les unités statistiques (c'est-à-dire pour nous les ménages) lorsqu'on ne dispose pas d'une liste à jour de ces dernières, et d'autre part la nécessité de concentrer l'échantillon dans des secteurs géographiques déterminés afin de réduire les temps et les coûts et de transport¹. Le premier argument est déterminant lorsqu'on réalise un sondage démographique; en ce qui nous concerne le second argument l'est moins puisque les coûts et temps de transport restent raisonnables en milieu urbain. *Pour nous, c'est au contraire parce qu'il permet, par rapport au sondage en grappe, d'augmenter le nombre et la dispersion des unités primaires et cela en conservant le même nombre de ménages dans l'échantillon, qu'un sondage à deux degrés peut être profitable.*

L'expression de la variance des estimations sous un tel plan comprend deux termes. Le premier représente l'aléa survenant lors de la sélection des unités primaires et il est fonction de la variabilité du caractère démographique estimé d'un pâté de maisons à l'autre. Le deuxième, dû au second degré du sondage, rend compte de la variabilité du caractère dans l'ensemble des ménages d'une même unité primaire. Dans chacun des deux termes interviennent évidemment les taux aux premier et second degré du

¹ : Voir par exemple L. M. ASSELIN 1984 : Techniques de sondage avec application à l'Afrique pp 105 & 106. GAETAN MORIN, Quebec. L'auteur ne s'inscrit pas exactement dans la même problématique que nous. L'ouvrage est principalement axé sur la production de statistiques agricoles et les sondages à plusieurs degrés y sont considérés comme une alternative aux sondages aléatoires à un degré sur listes et non pas, comme c'est le cas pour nous, à un sondage en grappe.

sondage. Les sondages en grappes considérés jusqu'à maintenant sont donc en fait des "cas limites" de sondages à deux degrés où, le second degré étant un recensement, le deuxième terme de la variance est nul. On conçoit bien intuitivement que, *si le premier terme est déterminant dans l'expression de la variance totale, on a intérêt pour réduire cette dernière tout en conservant un même nombre de ménages enquêtés, à augmenter le taux du premier degré quitte à n'enquêter qu'une fraction des ménages dans chaque ilot de l'échantillon.* C'est cette idée que nous allons maintenant formaliser pour quantifier ensuite, à partir des données de l'enquête préliminaire, les gains de précision qu'on peut en attendre et déterminer les taux qu'il convient d'appliquer aux premier et au second degré du sondage.

11.3.2 Formulaire, programme informatique

11.3.2.1 : Formulaire

1. Préliminaire : conventions, notations

Dans ce qui va suivre, pour réduire le volume du formulaire, on se limitera aux expressions des variances concernant l'estimation des totaux, l'étude de celles ayant trait aux ratios, en tout point semblable, n'apporterait rien de plus. L'expérience va porter sur deux plans de sondages, lesquels ne diffèrent en fait qu'au premier degré. Le premier plan, qu'on a toujours pris comme référence, est équiprobable au premier comme au second degré (c'est le sondage équiprobable sur liste, classique). Le second, que l'ensemble des résultats obtenus jusqu'ici nous porte à adopter, est le plan où, au premier degré, la probabilité d'appartenance d'une unité à l'échantillon est proportionnelle à sa surface totale, et où la sélection des unités secondaires se fait de manière équiprobable. Pour simplifier les calculs dans le cas du plan à probabilités inégales, on considère que le tirage des *unités primaires* se fait avec remise (pour les deux plans de sondage). Au second degré en revanche, il s'effectue sans remise. Enfin il ne s'agit pas ici de démontrer l'ensemble des résultats qui conduisent à l'expressions de la variances dans le cas des sondages à deux degrés. Ces résultats classiques sont, pour la plus part, repris des ouvrages de référence sur le sujet¹. l'objet des paragraphes suivants est seulement d'en faire *une synthèse qui nous conduira à l'expression de la différence qui nous intéresse : celle entre les variances associées respectivement au sondage en grappe et au sondage où l'on n'enquête qu'une fraction des ménages de chaque ilot. C'est sur l'expression de cette différence qu'est construit le programme qui permettra les tests.*

- Notations

Soient M le nombre total d'unités primaires (pâtés de maisons) contenues dans la base, m_i le nombre de ces unités incluses dans l'échantillon, N_i le nombre d'unités secondaires de l'unité primaire i , et n_i le nombre des ménages enquêtés dans cette unité (on note f_2 la fraction de sondage au second degré).

On considère donc un caractère y dont on cherche à estimer le total général Y à l'aide de l'estimateur \hat{Y} , y_{ij} étant la valeur de ce caractère pour le ménage j du pâté de maisons i . Le total de l'unité primaire i , noté Y_i , sera estimé par \hat{Y}_i et on note \bar{Y}_i la moyenne de y dans l'unité.

$V_1()$ sera la variance de l'estimateur employé au premier degré et $V_2()$ celle de l'estimateur employé au second degré.

¹ : On peut trouver des exposés complets dans tous les ouvrages classiques cités en bibliographie ; quant à nous, pour cette partie, nous nous sommes servi principalement de l'ouvrage de L. M. ASSELIN, déjà cité, auquel sont empruntées les notations adoptées.

Enfin on notera z_i la surface de l'unité primaire i et Z la surface totale de la base de sondage. Alors, la probabilité π_i que l'unité primaire i soit sélectionnée vaut :

$$\pi_i = m_1 z_i / Z.$$

2. Tirage au premier degré équiprobable

Les sélections étant équiprobables au premier comme au second degré, la variance de l'estimateur sans biais du total s'écrit : $V(\hat{Y}) = V_1(\hat{Y}_1) + \frac{M}{m_1} \sum_{i=1}^M V_2(\hat{Y}_i)$, avec

$$V_1(\hat{Y}_1) = \frac{M^2 \sigma^2}{m_1} \quad \text{et} \quad V_2(\hat{Y}_i) = N_i \left(\frac{N_i - n_i}{n_i} \right) S_i^2.$$

Les variances du caractère y , sur l'ensemble de la base de sondage (σ^2) et au sein de chaque unité primaires (S_i^2) ont respectivement les expressions suivantes :

$$\sigma^2 = \frac{\sum_{i=1}^M (Y_i - \bar{Y})^2}{M} \qquad S_i^2 = \frac{\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2}{N_i - 1}.$$

On obtient donc :
$$V(\hat{Y}) = \frac{M^2 \sigma^2}{m_1} + \frac{M}{m_1} \sum_{i=1}^M N_i \left(\frac{N_i - n_i}{n_i} \right) S_i^2.$$

Or, si l'on suppose f_2 constante, on a quelque soit i , $f_2 = n_i / N_i$; donc :

$$\forall i, \quad \frac{N_i - n_i}{m_1 n_i} = \frac{N_i - f_2 N_i}{m_1 f_2 N_i} = \frac{1 - f_2}{m_1 f_2}.$$

Donc, en posant :

$$A_1 = M^2 \sigma^2 \quad \text{et} \quad B_1 = M \sum_{i=1}^M N_i S_i^2.$$

on obtient :

$$V(\hat{Y}) = \frac{A_1}{m_1} + \left(\frac{1 - f_2}{m_1 f_2} \right) B_1.$$

3. Tirage au premier degré à probabilités proportionnelles aux surfaces

Le tirage au premier degré s'effectuant maintenant avec des probabilités inégales, la variance de l'estimateur sans biais du total s'écrit :

$$V(\hat{Y}) = V_1(\hat{Y}_1) + \sum_{i=1}^M \frac{V_2(\hat{Y}_i)}{\pi_i} = V_1(\hat{Y}_1) + \frac{Z}{m_1} \sum_{i=1}^M \frac{V_2(\hat{Y}_i)}{z_i}$$

avec $V_1(\hat{Y}_1) = \frac{1}{m_1} \times \left(Z \times \left(\sum_{i=1}^M \frac{y_i^2}{z_i} \right) - Y^2 \right)$ et toujours,

$$V_2(\hat{Y}_i) = N_i \left(\frac{N_i - n_i}{n_i} \right) S_i^2 \quad \text{et} \quad S_i^2 = \frac{\sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2}{N_i - 1}$$

On a donc : $V(\hat{Y}) = \frac{1}{m_1} \times \left(Z \times \left(\sum_{i=1}^M \frac{y_i^2}{z_i} \right) - Y^2 \right) + Z \times \sum_{i=1}^M \frac{N_i}{z_i m_1} \left(\frac{N_i - n_i}{n_i} \right) S_i^2$

A nouveau grâce au fait que f_2 est constante, on obtient :

$$V(\hat{Y}) = \frac{1}{m_1} \times \left(Z \times \left(\sum_{i=1}^M \frac{y_i^2}{z_i} \right) - Y^2 \right) + Z \times \sum_{i=1}^M \frac{N_i}{z_i m_1} \left(\frac{N_i - n_i}{n_i} \right) S_i^2$$

Donc, en posant :

$$A_2 = Z \times \left(\sum_{i=1}^M \frac{y_i^2}{z_i} \right) - Y^2 \quad \text{et} \quad B_2 = Z \sum_{i=1}^M \frac{N_i S_i^2}{z_i}$$

on obtient :

$$V(\hat{Y}) = \frac{A_2}{m_1} + \left(\frac{1-f_2}{m_1 f_2} \right) B_2$$

4. Discussion

Dans les deux cas envisagés pour le plan de sondage au premier degré, la variance de l'estimateur sans biais est donc de la forme : $V(\hat{Y}) = \frac{A}{m_1} + \left(\frac{1-f_2}{m_1 f_2} \right) B$, A et B variant suivant la probabilité de sélection retenue. A représente la variance entre unités primaires, due au premier degré de sondage. B représente la variance au sein des unités primaires, due au second degré.

Dans les deux cas également, lorsque $f_2 = 1$, le second terme s'annule et on obtient la variance du sondage en grappe de référence : $V_0(\hat{Y}) = \frac{A}{m_0}$, où m_0 est le nombre de pâtés de maisons enquêtés. On a donc, pour la différence entre les variance des deux sondages à comparer, l'expression suivante :

$$V_0(\hat{Y}) - V(\hat{Y}) = \left(\frac{1}{m_0} - \frac{1}{m_1} \right) A + \left(\frac{1-f_2}{m_1 f_2} \right) B.$$

Pour conserver, lorsque la fraction de sondage au second degré n'est plus égale à 1, le même nombre total de ménages enquêtés, il faut que f_2 vérifie : $m_1 f_2 = m_0$ ¹. On a alors : $V_0(\hat{Y}) - V(\hat{Y}) = (1 - f_2) \frac{A}{m_0} - (1 - f_2) \frac{B}{m_0}$, d'où l'expression du gain sur la variance qu'il faudra maximiser :

$$V_0(\hat{Y}) - V(\hat{Y}) = \frac{1 - f_2}{m_0} (A - B)$$

Tout dépend donc du signe et de la valeur de la quantité $(A - B)$. Si B est plus grand que A, alors $V(\hat{Y})$ est supérieur à $V_0(\hat{Y})$ et le sondage en grappe où chaque unité primaire est enquêtée exhaustivement est préférable. En revanche si B est plus petit que A, alors $V(\hat{Y})$ est inférieur à $V_0(\hat{Y})$ et cela d'autant plus que f_2 est petit : c'est, dans ce cas, le sondage à deux degrés qui est le plus efficace et il faut choisir la fraction de sondage au second degré la plus faible possible.

Le principe du test à partir des données collectées lors de l'enquête "Bâti / population" va donc être de *calculer, pour un ensemble de caractères démographiques, les quantités A et B* afin tout d'abord de vérifier que la répartition de la variance entre ses deux composantes est bien semblable pour tous les caractères considérés. Si c'est le cas, *on fixera la fraction de sondage à appliquer au second degré (f_2) et on évaluera, à partir des valeurs de A et B, le gain de précision obtenu par rapport au sondage en grappes.*

11.3.2.2 : Programme informatique

Le programme VAR2DEG (VARIance 2 DEGrés) effectue ces calculs à partir des fichiers de données constitués sous dbase III. Il est écrit dans le langage propre à ce logiciel² et réalise la stricte application du formulaire décrit plus haut : calcul des deux termes (A et B) de la variance d'estimation d'un total sous un plan de sondage à deux degrés. Le programme utilise deux fichiers de données correspondant aux deux degrés du sondage : le fichier des unités primaires (pâtés de maisons) et celui des unités secondaires (ménages). Le plan au premier degré, toujours considéré avec remise, peut être soit équiprobable soit à probabilité proportionnelle à une variable du choix de l'utilisateur dont la donnée doit figurer dans le fichier des unités primaires. Le plan au second degré est toujours équiprobable sans remise.

Pour lancer les calculs, on doit entrer successivement les noms des deux fichiers de données (unités primaires puis secondaires), le nom de la variable à estimer puis celui de la variable définissant la probabilité d'appartenance des unités primaires à l'échantillon, enfin le nom d'une variable identifiant les unités primaires dans les

¹ : En réalité, même une fois cette condition vérifiée, on n'enquêtera qu'approximativement le même nombre de ménage dans les deux cas, l'égalité n'étant rigoureusement vérifiée qu'en espérance. De ce fait, lorsque m_0 est grand on ne risque pas de trop mauvaises surprises.

² : On trouvera le listage complet du programme dans l'annexe A.3.3.

deux fichiers (le nom de cette variable devant bien entendu être identique dans les deux fichiers).

Comme pour VARECH, la version actuelle de VAR2DEG est à l'état de maquette. *Les deux programmes peuvent évidemment être appliqués aussi bien à des données exhaustives issues de recensement qu'à des données provenant d'enquêtes sur échantillon ; dans ce cas il faut bien sur tenir compte de la représentativité de ces données lorsqu'on formule les conclusions.* Ils sont en fait complémentaires, VARECH permettant d'effectuer des choix concernant la sélection des unités primaires et l'inférence au premier degré tandis que VAR2DEG aide à déterminer le taux au second degré du sondage. Dans leurs versions opérationnelles, ces deux programmes devraient donc constituer un outil informatique efficace pour évaluer les possibilités d'utilisation des techniques de sondage sur des sites où existent des données permettant une approche rigoureuse des choix à opérer.

11.3.3 Résultats

Le tableau 11.7 résume les résultats obtenus sur un échantillon de 1329 ménages répartis dans 44 pâtés de maisons enquêtés. Pour chacun des deux plans de sondage au premier degré, il donne la valeur des deux termes de la variance et la part de la variance totale due au premier degré.

Tableau 11.7 : Comparaison des variances dues aux premier et second degrés du sondage (Programme VAR2DEG)

Caractères	Equiprobable			Probabilités inégales(p=s)		
	A ₁ (x 10 ⁻⁶)	B ₁ (x 10 ⁻⁶)	$\frac{A_1}{A_1+B_1}$	A ₂ (x 10 ⁻⁶)	B ₂ (x 10 ⁻⁶)	$\frac{A_2}{A_2+B_2}$
EFFECTRES	29.02	0.13	0.99	40.75	0.25	0.99
EFFM5A	0,26	0,02	0,93	0,37	0,04	0,90
EFFPL5A	19.86	0.12	0.99	31.47	0.23	0.99
EFFMIG	4,97	0,11	0,98	7,59	0,24	0,97
EFFNMIG	8.02	0.16	0.98	13.16	0.32	0.98
EFFMIGM5	0,19	0,04	0,83	0,32	0,08	0,80
CEFFMIGM5	21.24	0.15	0.99	33.28	0.31	0.99

On constate que pour l'ensemble des caractères démographiques considérés, *la variabilité des caractères entre unités primaires est responsable de plus de 90 % de la variance totale et donc que le sondage à deux degrés permettant d'enquêter un maximum de pâtés de maisons est de beaucoup préférable à un sondage en grappe.* De plus, en ce qui concerne les variables que nous estimerons réellement (en caractères gras dans le tableau), le rapport A/(A+B) est si proche de 1 qu'on peut considérer que la variance totale est entièrement déterminée par le premier terme. Donc, si l'on

reprend les notations du formulaire, on a, entre la variance du sondage en grappe de référence et la variance du sondage à deux degrés, la relation linéaire suivante :

$$V(\hat{Y}) = f_2 V_0(\hat{Y})$$

Il est donc acquis qu'il faut pratiquer le taux de sondage le plus faible possible au sein des pâtés de maisons, afin de multiplier le nombre d'unités primaires enquêtées. Mais ici intervient une contrainte supplémentaire : il faut pouvoir estimer la variance des estimateurs de façon fiable d'après les données recueillies sur l'échantillon. On sait que dans l'estimateur de cette variance, pour chaque unité primaire, les quantités Y_i

et \bar{Y}_i (total et moyenne du caractère y sur l'unité primaire i) sont remplacées par leurs estimations obtenues à partir de l'échantillon de ménages enquêtés dans ce pâté de maisons. Pour que l'estimation de la variance soit correcte il faut donc que le nombre de ménages enquêtés dans chaque unité primaire soit suffisant pour permettre une estimation fiable du total et de la moyenne de y dans l'unité. De ce point de vue, une fraction de sondage au second degré constante, que l'on veut la plus faible possible, risque de poser des problèmes dans les pâtés de maisons peu peuplés. *Donc on enquêtera plutôt un nombre constant de ménages dans toutes les unités primaires*, ce qui correspond à une fraction f_2 variable suivant la taille du pâté de maisons (par taille du pâté de maisons, on entend ici le nombre de ménages qui l'occupent). Ce principe étant admis, reste donc à déterminer le nombre de ménages qu'il faut enquêter dans chaque unité pour obtenir le gain maximum sur la précision du sondage tout en conservant une estimation correcte de la variance des estimateurs. Dans ce but, on va constituer dans l'échantillon des groupes de pâtés de maisons ayant des tailles comparables. Ainsi, pour chaque groupe on pourra évaluer la part de la variance totale qui lui revient¹ et donc, quelle conséquence aura le choix des fractions de sondage au second degré sur la variance des estimateurs.

La distribution des pâtés de maisons et des ménages de l'échantillon de l'enquête bâti / population selon la taille de l'ilot (en nombre de ménages) est donnée dans le tableau suivant :

Taille (nb de ménages dans l'ilot)	$1 \leq n \leq 10$	$10 < n \leq 20$	$20 < n \leq 30$	$30 < n \leq 40$	$40 < n \leq 50$	$50 < n \leq 60$	$60 < n \leq 80$	$n > 80$
Nb d'ilots	10	16	9	1	3	1	2	2
Nb de ménages	59	241	227	35	137	52	139	439

Pour conserver suffisamment de pâtés de maisons dans chaque groupe on considérera les quatre groupes suivants :

$1 \leq n \leq 10$:	10 Manzanas	59 Ménages
$10 < n \leq 20$:	16 Manzanas	241 Ménages
$20 < n \leq 40$:	10 Manzanas	262 Ménages
$n > 40$:	8 Manzanas	767 Ménages
Total :	44 Manzanas	1329 Ménages

¹ : voir ci dessous comment on applique pour cela aux différents groupes le principe de sommation des variances dans le cas stratifié.

Dans le tableau 11.8 figurent les résultats de l'application successive du programme au fichier correspondant à chaque groupe. On n'a retenu que les quatres caractères démographiques que nous devons réellement estimer, pour lesquels on indique le rapport A/(A+B) dans les deux cas de plans de sondages possibles.

Comme en témoigne le tableau, le terme de la variance dû au premier degré du sondage est bien déterminant quelque soit la taille du pâté de maisons considéré (particulièrement dans le cas du sondage à probabilités inégales).

Tableau 11.8 : Comparaison des variances dues aux premier et second degrés du sondage selon la taille des pâtés de maisons (Programme VAR2DEG)

Taille des îlots	EFFCTRES		EFFPLUS1		EFFNMIG		CEFFMIGM5	
	$\frac{A_1}{A_1+B_1}$ Equip.	$\frac{A_2}{A_2+B_2}$ (P-s)	$\frac{A_1}{A_1+B_1}$ Equip.	$\frac{A_2}{A_2+B_2}$ (P-s)	$\frac{A_1}{A_1+B_1}$ Equip.	$\frac{A_2}{A_2+B_2}$ (P-s)	$\frac{A_1}{A_1+B_1}$ Equip.	$\frac{A_2}{A_2+B_2}$ (P-s)
1 ≤ n ≤ 10	0,85	0,96	0,83	0,95	0,83	0,92	0,88	0,96
10 < n ≤ 20	0,88	0,98	0,87	0,98	0,82	0,94	0,84	0,97
20 < n ≤ 40	0,86	0,99	0,88	0,98	0,75	0,96	0,85	0,98
n > 40	0,995	0,997	0,994	0,996	0,993	0,995	0,981	0,986

Il reste donc à quantifier la part de chaque groupe de pâtés de maisons dans la variance totale. Pour cela on va considérer la division opérée dans l'échantillon suivant la taille des pâtés de maisons comme une stratification. Dans ce cas la variance totale apparaît comme la variance d'une somme de variables aléatoires indépendantes et on a :

$$V(\hat{Y}) = \sum_{k=1}^L V(\hat{Y}_k).$$

Si l'on considère, d'après les résultats précédents, que seul compte le premier terme, tant pour la variance totale qu' à l'intérieur de chaque "strate", on peut écrire :

$$A/n \approx \sum_{k=1}^L A_k/n_k.$$

Comme $n=1/fN$ et, pour tout k , $n_k=1/fN_k$ on obtient : $A/fN \approx \sum_{k=1}^L A_k/fN_k.$

C'est-à-dire :

$A/N \approx \sum_{k=1}^L A_k/N_k .$

Le tableau 11.9, basé sur cette formule appliquée aux données recueillies sur les 44 pâtés de maison de l'échantillon, fournit une évaluation des variances obtenues avec les deux stratégies en concurrence :

1. le sondage en grappe où f_2 vaut 1 pour tous les pâtés de maisons sélectionnés
2. le sondage à deux degrés où l'on enquête exhaustivement les pâtés de maisons contenant jusqu'à 10 ménages et où l'on n'enquête qu'un échantillon de dix ménages dans tous les autres pâtés de maisons.

Tableau 11.9 : Variances des estimateurs, décomposées selon la taille des pâtés de maisons, pour deux stratégies de sondage au second degré (Programme VAR2DEG)

Taille (nb de ménages)	$1 \leq n \leq 10$	$10 < n \leq 20$	$20 < n \leq 40$	$n > 80$	Total
1) Caractéristiques de l'échantillon					
Nb d'îlots	10	16	10	8	44
Nb de ménages	59	241	262	767	1329
Contrib. à la variance : A_i/N_i	18232,3	66086,4	74786,1	757434,49	916539,4
Contrib. à la variance (%)	2,0	7,2	8,2	82,6	100
2) Sondage en grappes					
Fraction de sondage	1	1	1	1	1
Nb de ménages à enquêter	178	725	788	2309	4000
Nb d'îlots prévus	30	48	30	24	132
Contrib. à la variance	$0,02 V_0$	$0,07 V_0$	$0,08 V_0$	$0,83 V_0$	V_0
3) Sondage à deux degrés (f_2 variable)					
Fraction de sondage moyenne	1	10/15,1	10/26,2	10/95,9	
Nb de ménages à enquêter	177	730	790	2310	4007
Nb d'îlots prévus	30	73	79	231	413
Gain sur la variance (%)	0	34	62	90	80
Contrib. à la variance	$0,02 V_0$	$0,05 V_0$	$0,03 V_0$	$0,09 V_0$	$0,19 V_0$

Les conclusions à tirer de l'expérience apparaissent clairement à l'examen du tableau.

- 1) *Les pâtés de maisons les plus peuplés (contenant plus de 40 ménages) ont une contribution déterminante à la variance : à eux seuls ils fournissent plus de 80% du total. C'est donc sur cette fraction de l'échantillon qu'il importe d'avoir le taux au second degré le plus faible.*

2) *A l'inverse la contribution du groupe des pâtés de maisons peuplés de moins de dix ménages est presque négligeable : 2% de la variance totale. On ne peut donc pas espérer de gain appréciable de précision en n'enquêtant que partiellement ces îlots. D'autre part on a vu que pour obtenir une bonne estimation de la variance il convenait d'avoir un nombre suffisant de ménages enquêtés dans chaque unité primaire. Nous enquêterons donc tous les ménages des pâtés de maisons de ce groupe.*

3) Pour les pâtés de maisons contenant plus de dix ménages, un effectif enquêté constant de dix ménages est un bon choix. En premier lieu, il est suffisant pour éviter toute mauvaise surprise lors de l'estimation de la variance des estimateurs. D'autre part il permet de faire décroître la fraction de sondage au fur et à mesure qu'augmente le nombre de ménages vivant dans le pâté de maisons pour arriver à une fraction d'un dixième en moyenne dans le groupe des pâtés de maisons les plus peuplés. *Ainsi l'application de cette stratégie au second degré produit, pour un même nombre total de ménages enquêtés, une division par cinq de la variance par rapport au sondage en grappe (le coefficient de variation lui, sera divisé par 2,25).*

11.3.4 Conclusions

Au début de cette expérience visant à la détermination du second degré du sondage, nous avons posé le problème suivant :

Dans un sondage à deux degrés, l'expression de la variance comporte deux termes correspondant aux aléas dus à chacun des degrés ; si le terme correspondant à la sélection des unités primaires est déterminant dans l'expression de la variance totale, on a intérêt, pour réduire cette dernière tout en conservant un même nombre de ménages enquêtés, à augmenter le taux du premier degré quitte à n'enquêter qu'une fraction des ménages dans chaque îlot de l'échantillon.

Le programme VAR2DEG a permis le calcul de ces deux termes de la variance. On a alors constaté que la variabilité des caractères entre unités primaires est responsable de plus de 90 % de la variance totale et donc que le sondage à deux degrés permettant d'enquêter plus de pâtés de maisons est bien préférable à un sondage en grappe où les pâtés de maisons sont enquêtés exhaustivement. On doit donc d'une part choisir un taux au second degré le plus faible possible pour optimiser la précision du sondage, et d'autre part conserver un nombre suffisant de ménages dans chaque unité primaire pour pouvoir estimer précisément la variance des estimateurs. Ce double objectif est satisfait lorsqu'on enquête un nombre constant de ménages dans chaque pâté de maisons. En effet, lorsqu'on décompose le calcul de la variance pour évaluer la contribution de groupes de pâtés de maisons également peuplés, on constate que les pâtés de maisons les plus peuplés (contenant plus de 40 ménages) fournissent à eux seuls plus de 80% de la variance totale. En revanche, la contribution du groupe des pâtés de maisons peuplés de moins de dix ménages est négligeable. C'est donc sur la fraction de l'échantillon comprenant les pâtés de maisons les plus peuplés qu'il importe d'avoir le taux au second degré le plus faible, tandis qu'on peut enquêter exhaustivement les pâtés de maisons les moins peuplés.

Le résultat de l'expérience est la définition de la stratégie de sélection des ménages au sein des pâtés de maisons choisis au premier degré : *on enquêtera tous les ménages des pâtés de maisons contenant jusqu'à 10 ménages et seulement un échantillon de dix ménages dans tous les autres pâtés de maisons.* Enfin le programme mis au point, appliqué aux données recueillies sur l'échantillon de l'enquête préliminaire "Bâti-population", permet de comparer la précision du sondage ainsi défini avec celle du sondage en grappe de référence : *l'application de cette stratégie au second degré produit, pour un même nombre total de ménages enquêtés, une division par cinq de la variance.*

A l'issue des expériences menées d'une part sur les données du recensement de 1982 à Marseille et d'autre part sur les données provenant des enquêtes préliminaires réalisées à Quito, nous possédons maintenant tous les éléments pour élaborer un plan de sondage complet. Le chapitre suivant décrit d'abord ce plan dans son ensemble : étape de sélection des unités primaires et secondaires et étape d'inférence avec les formules d'estimation correspondant au premier et deuxième degré du sondage. Puis il indique comment ce plan a été mis en œuvre pour réaliser l'enquête démographique finale.

Chapitre 12

LE PLAN DE SONDAGE ET SA MISE EN ŒUVRE POUR L'ENQUÊTE

DEMOGRAPHIE / MIGRATIONS

12.1 Introduction : le double objectif méthodologique et thématique de l'enquête, présentation du questionnaire

L'enquête finale "Démographie Migration" que nous avons réalisée à Quito possède un double objectif. Tout d'abord, sur un plan méthodologique, *elle doit permettre de vérifier que la méthode mise au point possède bien les caractéristiques de rapidité, faible coût et précision que nous lui attribuons au terme de la phase d'expérience.* Toujours sur ce même plan méthodologique il importe également de savoir si la technique de sondage employée demeure efficace en dehors du domaine thématique restreint des indicateurs démographiques les plus classiques. C'est une des raisons pour lesquelles nous avons choisi d'aborder, dans le questionnaire de l'enquête finale, plusieurs thèmes.

Ensuite *l'enquête vise à produire une information fiable sur chacun des deux grands thèmes que permet de traiter le questionnaire (présenté dans les pages suivantes) : caractéristiques démographiques générales de la population quiténienne et systèmes de résidence*¹.

"La première fiche du questionnaire a pour objet de saisir les *caractéristiques démographiques et socio-économiques* de chacune des personnes vivant dans le logement enquêté ; en dehors de modifications mineures, elle correspond à la fiche d'enquête utilisée lors d'une première enquête démographique auprès de 1100 ménages quiténiens (enquête Bâti/population, premier passage en avril 1987, deuxième passage auprès du même échantillon d'ilots en novembre 1987). La fiche 2 permet d'analyser les statuts d'occupation du logement et les modes d'accès à la propriété, éléments essentiels pour appréhender les *modes d'insertion des migrants en milieu urbain.*

Les fiches 3 et 4 sont centrées sur la question des *systèmes résidentiels*. Si la fiche 4 correspond à un recueil classique de *biographies migratoires et professionnelles* avec un seuil de durée de résidence fixé à 6 mois, la fiche 3 vise à cerner la pratique des chefs de ménage en matière de systèmes résidentiels et d'activités, et à préciser l'efficacité des concepts de "*densité de résidence*" et "*densité d'activité*" : cette fiche 3 correspond au recueil *intégral*, sur une période de 2 ans, de tous les lieux de séjour et activités, sans critère de durée minimum. Doivent donc y figurer tous les lieux de séjour, y compris les voyages, et toutes les activités, même exercées durant un laps de temps très court

Enfin, par la fiche 5 relative aux caractéristiques d'activité et de résidence des

¹ : La présentation du questionnaire qui termine ce paragraphe d'introduction est reprise de la communication faite par F. DUREAU aux journées démographiques de l'ORSTOM 1988 : "A propos de l'analyse des systèmes résidentiels, présentation de l'enquête réalisée à Quito en décembre 1987", Paris, 17 pages.

RUBRIQUES		N° de personne <input type="text"/>	N° de personne <input type="text"/>	N° de personne <input type="text"/>	N° de personne <input type="text"/>	N° de personne <input type="text"/>	N° de personne <input type="text"/>
Où viviez-vous avant de résider à Quito ? Pays Province Paroisse Nom de la localité	
Quel est le dernier cours que vous avez suivi dans l'enseignement ?		Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....	Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....	Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....	Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....	Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....	Aucun <input type="checkbox"/> Primaire <input type="checkbox"/> année..... Secondaire <input type="checkbox"/> année..... Supérieur <input type="checkbox"/> année.....
Activité principale	Profession
	Branche d'activité
	Statut professionnel
	Revenus (sucres)	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>
Activité secondaire	Profession
	Branche d'activité
	Statut professionnel
	Revenus (sucres)	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>	<input type="text"/> par <input type="text"/> <i>Montant durée</i>
Avez-vous d'autres sources de revenus ? <i>(Pour chaque type, noter : - "non", si la personne n'a pas de revenus de ce type - la source précise et le montant mensuel, si la personne a des revenus de ce type)</i>		Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>	Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>	Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>	Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>	Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>	Transports <input type="text"/> Commerce <input type="text"/> Immobilier <input type="text"/> Agriculture <input type="text"/> Pension bourse <input type="text"/> Autre <input type="text"/> Refus de réponse <input type="checkbox"/>

2. VARIABLES SUR LES CONDITIONS D'OCCUPATION DU LOGEMENT

À quel titre occupez-vous ce logement?

- | | | |
|---|---|--|
| <input type="checkbox"/> Le chef de ménage est propriétaire | | |
| <input type="checkbox"/> Une autre personne vivant dans le logement est propriétaire | → Numéro de la personne | |
| <input type="checkbox"/> Locataire | → Montant du loyer mensuel (sucres) | |
| <input type="checkbox"/> Anticrèsis | → Montant mensuel (sucres) | |
| | → Montant de l'anticrèsis-même (sucres) | |
| <input type="checkbox"/> Occupation de faveur (gratuit ; prêt d'un ami, parent ou collègue) | | |
| <input type="checkbox"/> Pour service | → Type de service : | |
| | → Montant du loyer mensuel (sucres) | |
| <input type="checkbox"/> Autre type d'occupation | → Préciser : | |
| <input type="checkbox"/> Refus de réponse | | |

Date et coût de la construction ou de l'achat du logement :

- | | | |
|--|-----------------|--|
| <input type="checkbox"/> DON ou HERITAGE | | |
| | → Date | |
| <input type="checkbox"/> ACHAT | → Date | |
| | → Coût (sucres) | |
| <input type="checkbox"/> CONSTRUCTION : | | |
| → achat du terrain | → Date | |
| | → Coût (sucres) | |
| → construc. du bâtiment | → Date | |
| | → Coût (sucres) | |

**Si le propriétaire ne vit pas dans le logement :
de quel type de propriétaire s'agit-il ?**

- | | |
|---|--|
| <input type="checkbox"/> Personne physique ne vivant pas dans le logement | → Lien de parenté du chef de ménage avec le propriétaire : |
| <input type="checkbox"/> Société privée | <input type="checkbox"/> Aucun |
| <input type="checkbox"/> Organisme semi-public | <input type="checkbox"/> Lien de parenté : |
| <input type="checkbox"/> Organisme public | <input type="checkbox"/> Refus de réponse |
| <input type="checkbox"/> Organisme municipal | |
| <input type="checkbox"/> Organisme diplomatique | |
| <input type="checkbox"/> Refus de réponse | |

3. LIEUX DE RESIDENCE ET ACTIVITES DU CHEF DE MENAGE DURANT LES DEUX DERNIERES ANNEES

Calendrier	Date du changem. de résid.	LIEU DE RESIDENCE (Quelque soit la durée) <i>Pays, province, paroisse, localité, adresse</i>	Date du changem. d'activité	ACTIVITE PRINCIPALE <i>Profession, branche d'activité, statut professionnel, revenus mensuels</i>	RAISON DU CHANGEMENT DE RESIDENCE OU D'ACTIVITE
Décem. 87					
Novem. 87					
Octobre 87					
Septem. 87					
Août 87					
Juillet 87					
Juin 87					
Mai 87					
Avril 87					
Mars 87					
Février 87					
Janvier 87					
Décern. 86					
Novem. 86					
Octobre 86					
Septem. 86					
Août 86					
Juillet 86					
Juin 86					
Mai 86					
Avril 86					
Mars 86					
Février 86					

107

4. BIOGRAPHIE MIGRATOIRE ET PROFESSIONNELLE DU CHEF DE MENAGE DEPUIS SA NAISSANCE

Age (ans)	Date (jo.mo.an.)	LIEU DE RESIDENCE (> 6 mois) Pays, province, paroisse, localit� ou localit�, quartier et adresse, s'il s'agit de Quito	ACTIVITE Profession, branche d'activit�, statut professionnel, revenus mensuels (sucres)	RAISON DU CHANGEMENT DE RESIDENCE OU D'ACTIVITE
Nais. d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	
d	d		Prof Eran Statut Revenus []	

5 - VARIABLES SUR LES PARENTS DU CHEF DE MENAGE ET DE SON CONJOINT DE VIVANT PAS DANS LE LOGEMENT ENQUETE

Nom de parente	Nom et prénom	Sexe H=Homme M= Femme	Age ou année de naissance	Si décédé(e) année du décès	Lieu de résidence actuel Pays, province, paroisse, localité ou kazaïte, quartier et adresse, s'il s'agit de Outo	Activité principale actuelle	Relation économique avec ce parent (voir 2.1.1)
PERE du chef de ménage		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
MERE du chef de ménage		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
CONJOINT du chef de ménage		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
PERE du conjoint du chef de ménage		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
MERE du conjoint du chef de ménage		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
UNE FILLE, ou un fils du ménage (voir 2.1.1)		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
①		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
②		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
③		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
④		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑤		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑥		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑦		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑧		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑨		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑩		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑪		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑫		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>
⑬		<input type="checkbox"/>	age ou année nai.	<input type="checkbox"/>		Prof..... Bran..... Statut.....	<input type="checkbox"/>

1 = Le chef de ménage ne reçoit rien et n'envoie rien à ce parent
 2 = Le chef de ménage reçoit une aide financière de ce parent
 3 = Le chef de ménage reçoit une aide en nature (ex: produits agricoles) de ce parent

4 = Le chef de ménage héberge un enfant de ce parent
 5 = Ce parent reçoit une aide financière du chef de ménage
 6 = Ce parent reçoit une aide en nature du chef de ménage

7 = Ce parent héberge un fils du chef de ménage
 9 = Refus de réponse

membres de la famille ne vivant pas dans le ménage enquêté, nous tentons de *replacer les pratiques résidentielles et professionnelles des chefs de ménage enquêtés dans le groupe social, éventuellement spatialement éclaté, dans lequel elles s'inscrivent*

On observe facilement une certaine dichotomie dans le questionnaire. Tandis que les fiches 1, 2 et 4 font appel à de définitions et de méthodes d'observation tout à fait classiques en démographie, les fiches 3 et 5 visent justement à en renouveler certaines, en faisant avancer la connaissance sur les pratiques effectives en matière de résidence, d'activité et de reproduction sociale et économique au sein d'unités collectives spatialement segmentées. La présence, dans un même formulaire d'enquête, des deux types d'approches permet de resituer les observations "nouvelles" par rapport aux informations plus classiques, et de faire apparaître divergences, convergences et complémentarité des modes d'observation."

12.2 : Description du plan de sondage, étapes de sélection et d'inférence

Au cours des expériences menées durant toute cette recherche à Marseille et à Quito, on a formulé un ensemble de conclusions qui vont maintenant trouver leur application pratique dans la définition du plan de sondage retenue pour l'enquête "Démographie / Migrations". Ces conclusions ont été exposées, au fur et à mesure de leur obtention, tout au long des chapitres précédents. Pourtant, il nous semble opportun, avant de décrire dans le détail le plan de sondage, de faire une synthèse rapide des principaux résultats, en les distinguant suivant qu'ils s'appliquent à l'une ou l'autre des étapes de la définition du plan.

12.2.1 : Résumé des résultats obtenus lors des phases d'expériences

- **Base de sondage et stratification** : parmi les stratifications actuellement disponibles sur l'image satellite, la *plus efficace est une stratification en six niveaux de densité du bâti dont le premier niveau permet d'isoler les espaces non construits, exclus de la base. Dans la base de sondage, on aura donc cinq strates.*
- **Probabilité d'appartenance des unités primaires à l'échantillon** : la surface bâtie ou, plus exactement, l'approximation que l'on peut en obtenir à partir de l'information satellitaire, est une mesure de la taille des unités primaires bien meilleure que leur surface totale au sens de la précision du sondage à but démographique; mais, *une fois la base de sondage stratifiée sur le critère de densité du bâti, une probabilité de sélection proportionnelle à la surface totale suffit à obtenir des précisions équivalentes.*
- **Inférence au premier degré** : les deux mesures de taille évoquées ci-dessus, et qui sont pour le moment les seules disponibles, ne sont ni l'une ni l'autre suffisamment corrélées aux variables démographiques pour qu'il y ait intérêt à les intégrer comme information exogène dans des estimateurs par le ratio. *On retient donc des estimateurs sans biais pour les totaux et l'estimation par le ratio pour les quotients.*
- **Règle d'allocation de l'échantillon aux strates** : Grâce à une bonne stabilité des règles d'allocation optimale, quels que soient les caractères estimés, *on peut définir une allocation aux strates qui améliore l'ensemble des estimations*(voir, pour les taux dans chaque strate, la description de l'étape de sélection des unités primaires)

- Second degré du sondage : la solution retenue est un sondage à deux degrés avec comme unités primaires les pâtés de maisons et comme unités secondaires les ménages. Le test effectué sur les données provenant de l'enquête préliminaire "bâti/population" *conclue à l'intérêt de pratiquer le taux le plus fort possible au premier degré, dans la limite du budget disponible et sous la contrainte d'avoir un nombre suffisant de ménages dans chaque pâté de maisons.*

- Spécificités du plan envisagé : Par rapport aux sondages les plus couramment pratiqués, par exemple les sondages aléatoires équiprobables sur listes, le sondage que nous proposons présente deux particularités principales :

1 : En tant que sondage à probabilités inégales, il nécessite, pour mener correctement l'inférence à partir de l'échantillon, la connaissance d' *informations exogènes* :

- *la taille (surface) de chaque unité primaire (pâté de maisons) enquêtée*, il s'agit de la surface totale et il faut donc disposer d'un moyen fiable de la mesurer.

- *La taille (surface) de chaque strate*, sa mesure sera faite sur l'image satellite.

2 : *Le caractère systématique du sondage, en assurant une bonne répartition géographique de l'échantillon, améliore la précision des estimations*. On évalue ce gain à 20 ou 30 % de la variance du sondage strictement aléatoire au même taux. Si l'on considère l'effet cumulé du tirage systématique et de la stratification, l'allocation aux strates étant optimale, le gain total se situe, suivant les caractères démographiques estimés, entre 45 et 95 % de la variance de départ (tirage aléatoire non stratifié).

L'ensemble de ces conclusions conduisent donc au plan de sondage décrit dans le paragraphe suivant.

12.2.2 : Description du plan de sondage : étapes de sélection aux premier et second degrés

1: Base de sondage

La base de sondage est l'ensemble de l'aire métropolitaine matérialisée sur l'image satellite par la limite de la zone considérée comme urbaine; à l'intérieur de cet ensemble, les zones non bâties sont exclues.

2: Stratification

Cet espace est stratifié en cinq niveaux de densité du bâti (stratification obtenue par classification de l'image satellite).

3: Tirage au premier degré

Les unités primaires sont les pâtés de maisons. Au premier degré nous pratiquons un sondage aréolaire systématique, la sélection de l'échantillon se faisant à l'aide d'une grille de points. Lorsque le placement initial de la grille sur l'image est aléatoire, on peut considérer que la probabilité d'appartenance à l'échantillon d'une unité spatiale donnée est proportionnelle à sa surface. La maille de la grille sera de taille variable pour respecter les taux dans chaque strate fixés par la règle d'allocation; ces taux sont

donnés dans le tableau suivant.

strates	taux approximatifs au premier degré (en % du nombre de manzanas de la strate)	Coefficient appliqué à la strate
1 : densité très faible	2,5	0,4
2 : densité faible	2,2	0,3
3 : densité moyenne	6,7	1
4 : densité forte	8,9	1,5
5 : densité très forte	4,4	0,9
Ensemble de la base	5,8	1

4. Tirage au deuxième degré

Les unités secondaires sont les ménages, la sélection est là encore systématique; elle s'effectue sur liste après recensement de tous les ménages vivant dans l'îlot. La fraction de sondage f_2 au second degré est variable suivant le nombre de ménages vivant dans l'îlot :

- moins de dix ménages dans le pâté de maisons : enquête exhaustive ($f_2 = 1$)
- plus de dix ménages dans le pâté de maisons : dix ménages sont enquêtés (f_2 décroît lorsque la taille du pâté de maisons augmente)

Le taux final obtenu sera, nous le verrons, inférieur à 1,5 % des ménages.

12.2.3 : Formulaire d'estimation pour l'étape d'inférence

13.2.3.1 : Préliminaire, les problèmes posés par l'inférence dans le cas du sondage systématique

Comme on l'a déjà signalé, et c'est pourquoi nous avons utilisé une méthode d'approximation au chapitre 8¹, les sondages systématiques posent un double problème théorique s'agissant de l'inférence statistique.

Premièrement, les probabilités d'appartenance à l'échantillon sont-elles suffisamment contrôlées pour que le sondage soit justiciable du calcul des probabilités?

Deuxièmement, il n'existe pas, dans le stricte cadre de la théorie, d'expression rigoureuse de la variance des estimations. Alors quelle solution retenir pour l'estimation de la variance?

En ce qui concerne le premier problème, J. DESABIE le pose en ces termes : " Le sondage systématique est effectivement un sondage aléatoire, à condition que la base de la progression arithmétique soit effectivement déterminée par tirage au sort"². L'auteur formule cette règle à propos du sondage systématique dans une liste et, lorsque la condition est satisfaite, le sondage est assimilé à un sondage équiprobable sans remise. Il n'y a aucun obstacle à la transposition de cette règle au cas du sondage systématique aréolaire à l'aide d'une grille de points. Si, et c'est bien le cas ici, le

¹ : voir chapitre 8 : Apport du tirage aréolaire systématique.

² : J. DESABIE, 1966 : Théorie et pratique des sondages, DONOD, Paris, p 98.

premier point de la grille a été placé aléatoirement, *le sondage est équivalent, en ce qui concerne les formules d'estimation des totaux, moyennes, quotients etc., à un sondage aléatoire à probabilités inégales, proportionnelles à la surface des unités de la base.*

Le second problème, plus compliqué, est généralement formulé ainsi¹ : le sondage systématique est un cas particulier de sondage en grappe, sa variance dépend donc de la variance inter-grappes; comme on ne sélectionne qu'une seule grappe dans l'échantillon, il est impossible de calculer un estimateur non biaisé de la précision. Pour lever cet obstacle on trouve, dans la littérature, trois types de solutions "classiques"² :

1. Négliger l'influence sur la précision du caractère systématique du sondage en l'assimilant au sondage aléatoire qui lui correspond.

J.M. GROBRAS³ par exemple dit, à propos du sondage systématique sur liste : "Si le rangement des unités dans le fichier peut être considéré comme aléatoire, ou tout au moins sans corrélation avec la variable étudiée, le tirage systématique est rigoureusement assimilable à un sondage à probabilités égales sans remise et donc de même efficacité (même espérance, même variance)". Cette assimilation peut donc être plus ou moins justifiée.

Elle l'est parfaitement dans le cas du second degré du sondage que nous utiliserons : la sélection systématique des ménages à enquêter dans le pâté de maisons peut être considérée comme aléatoire sans que l'on risque de sur ou sous estimer la variance intra-unité primaire. C'est donc la solution que nous retiendrons pour l'étape d'estimation de totaux (moyennes ou quotients) dans les unités primaires et pour estimer la variance de ces estimations : le sondage sera alors considéré strictement équivalent à un sondage équiprobable sans remise.

Le problème est sensiblement différent pour le premier degré. Le tirage de l'échantillon de pâtés de maisons devra être assimilé à un tirage à probabilités inégales, proportionnelles aux surfaces des unités. Les expériences faites à Marseille et à Quito ont démontré que la bonne répartition géographique de l'échantillon permise par le tirage systématique à l'aide d'une grille de points améliore nettement l'efficacité du sondage. *Il est donc certain que si l'on néglige cette caractéristique du sondage dans le formulaire d'estimation des variances, on surestimera les variances des estimations.* Les deux autres solutions visent à réduire cette surestimation.

2. Envisager le sondage systématique du premier degré comme un sondage stratifié où l'on sélectionne deux unités par strate.

Cette méthode, proposée par J. DESABIE⁴, et que nous avons déjà utilisée pour évaluer l'apport du tirage systématique, pourra fournir une bonne approximation des variances des estimations mais risque de conduire à des calculs assez lourds. On peut néanmoins, comme nous le verrons, en rendre l'application plus légère en sélectionnant plus de deux unités par strates (c'est à dire en réduisant le nombre de strates).

¹ : notamment par L.M. ASSELIN (Techniques de sondages avec application à l'Afrique, p 49), J. DESABIE (Théorie et pratique des sondages, p 98 et p 171) ou J.M. GROBRAS (Méthodes statistiques des sondages p29 et 30).

² : On ne retient pas dans cet inventaire les solutions reposant sur l'application de la théorie des processus stochastiques ou fonction aléatoire développées récemment (voir par exemple l'article de X.GUIDICELLI, J.P. LANLY, J.B. OUAKAM et M. PIETRI, 1972 dans Annales des Sciences Forestières, INRA). Pour une justification de ce choix, voir dans l'introduction générale, le paragraphe consacré à la présentation de la méthode appliquée et à sa justification .

³ : J.M. GROBRAS : Méthodes statistiques des sondages, 1987, ECONOMICA, PARIS; p 29.

⁴ : J. DESABIE , 1966 : Théorie et pratique des sondages, DONOD, Paris. Chap. 7.7, p 171 : Le tirage systématique envisagé comme un sondage stratifié.

3. Utiliser des méthodes robustes d'estimation de variance.

Lorsque, comme c'est le cas ici, les méthodes classiques d'estimation de variance sont mises en défaut, on peut utiliser les méthodes dites "robustes" qui permettent de calculer des intervalles de confiance même si l'on ignore les lois de probabilité des estimateurs. On peut donc, en utilisant de telles méthodes, faire abstraction de la structure probabiliste du sondage et en particulier la technique de sélection des unités d'échantillonnage importe peu. Parmi ces méthodes, les plus couramment employées sont les méthodes du Jackknife et du Bootstrap, dont on trouvera un exposé pratique dans le livre de J.M. GROSBRAS¹. Ces méthodes, a priori plus performantes que la méthode d'approximation par un sondage stratifié, nécessitent malheureusement des calculs complexes que nous ne pouvions programmer nous même. Nous ne disposons pas non plus à Quito des programmes correspondants et n'avons pas eu le temps, dans les délais prévus pour la rédaction de cette thèse, de chercher à les obtenir en France. Nous nous sommes donc contenté d'appliquer aux résultats de l'enquête finale la méthode mise au point à Marseille qui s'inspire de l'idée de DESABIE.

Cette méthode, puisqu'elle repose sur un principe de stratification, ne modifie en rien les expressions des variances au premier degré qui seront donc celles, classiques, associées au plan à probabilités inégales avec remise. Ce sont ces expressions qui figurent dans le formulaire suivant. Le lecteur intéressé par les autres méthodes pourra se reporter aux ouvrages cités en référence.

12.2.3.2 : Formulaire

Conformément aux conclusions du paragraphe précédent et aux résultats obtenus durant toute cette recherche, le formulaire d'estimation correspond à la structure de sondage suivante :

Premier degré : sondage aléatoire stratifié avec probabilités de sélection proportionnelles aux surfaces des unités primaires.

Second degré : sondage aléatoire équiprobable; recensement dans les petites unités primaires (jusqu'à dix ménages), dix ménages enquêtés dans toutes les autres unités.

Estimateurs employés : Aux deux degrés du sondage on emploie des estimateurs sans biais pour l'estimation des totaux ou des moyennes et des estimateurs par le ratio pour l'estimation directe des quotients.

Bien sûr, les résultats que fournira ce formulaire sont des approximations, du moins en ce qui concerne les variances et leurs estimations. Comme on l'a déjà dit, l'assimilation de la procédure de sélection systématique sans remise à un plan de sondage aléatoire avec remise au premier degré doit logiquement avoir pour conséquence la surestimation des erreurs d'échantillonnage et donc une évaluation pessimiste de la précision du sondage. C'est néanmoins la seule solution compatible avec l'équipement logiciel dont nous disposons parcequ'elle ne réclame pas un travail de programmation trop important; c'est donc celle qui sera retenue, au moins dans la première étape d'exploitation des résultats de l'enquête. Par la suite il serait intéressant, si les moyens en temps et programmes informatiques le permettent, d'appliquer aux résultats des méthodes de calcul de variance mieux adaptées au sondage mis au point, en particulier pour tenter d'évaluer l'apport de la sélection systématique.

De plus, pour simplifier les expressions des variances et de leurs estimateurs, nous avons choisis ici de négliger les termes correspondant au second degré du sondage. L'évaluation, faite au chapitre 11, des deux composantes de la variance des caractères démographiques - variance inter-unité primaire et intra-unité primaire -, justifie

¹ : J.M. GROSBRAS : Méthodes statistiques des sondages, 1987, ECONOMICA, PARIS; P 329.

pleinement cette approximation et les résultats obtenus grâce à l'enquête finale confirment, nous le verrons, sa légitimité. Cependant on peut imaginer que dans certains contextes d'application de la méthode, la composante intra-unité primaire de la variance ne soit pas négligeable. Nous exposerons au chapitre suivant la méthode générale permettant d'intégrer dans l'expression de l'estimateur de la variance le terme correspondant à la variance intra-unité primaire et construirons cet estimateur dans le cas des estimations de totaux.

Plutôt qu'une présentation "littéraire", qui alourdirait beaucoup un exposé qui se veut avant tout pratique, nous avons choisi de présenter le formulaire sous forme synthétique dans les trois tableaux suivants. Cette présentation vise également à permettre par la suite une référence rapide à la bonne formule, pour chaque cas de figure. Le tableau 12.1 fixe les notations retenues ici, qui sont celles classiquement employées en sondage ; pour notre part nous les avons empruntées à l'ouvrage de L.M. ASSELIN¹. Les tableaux 12.2 et 12.3 donnent ensuite les formules à employer pour l'inférence au premier (tableau 12.3) et second (tableau 12.2) degrés du sondage, pour les trois types de caractères envisagés : totaux, moyennes et quotients.

¹ : L.M. ASSELIN 1984 : Techniques de sondages avec application à l'Afrique, Gaëtan Morin éditeur, QUEBEC. Les notations sont exposées dans les chapitres 2 et 3.

Tableau 12.1 : Notations adoptées dans le formulaire

Objets	Notations	Significations
1) Base de sondage et échantillon	N_i n_i M_k m_k N_k	Nombre d'unités secondaires dans l'unité primaire i Nombre d'US de l'UP i contenues dans l'échantillon Nombre d'UP dans la strate k (inconnu) Nombre d'UP de la strate k contenues dans l'échantillon Nombre d'US dans la strate k (inconnu)
2) Variables dont on doit estimer le total ou la moyenne (y) ou intervenant au dénominateur du quotient à estimer (x)	y_j, x_j y_i, x_i \bar{y}_i, \bar{x}_i \hat{y}_i, \hat{x}_i $\hat{\bar{y}}_i, \hat{\bar{x}}_i$ Y_k, X_k \bar{Y}_k, \bar{X}_k \hat{Y}_k, \hat{X}_k \hat{N}_k $\hat{\bar{Y}}_k, \hat{\bar{X}}_k$	Valeurs des caractères pour l'US j Totaux des caractères dans l'UP i Moyennes des caractères dans l'UP i Estimateurs des totaux dans l'UP i Estimateurs des moyennes dans l'UP i Totaux des caractères dans la strate k Moyennes des caractères dans la strate k Estimateurs des totaux dans la strate k Estimateur du nombre d'US dans la strate k Estimateurs des moyennes dans la strate k

Tableau 12.1 (suite) : Notations adoptées dans le formulaire

Objets	Notations	Significations
3) Variable définissant la probabilité d'inclusion dans l'échantillon des unités primaires (z)	z_i Z_k	Surface totale de l'UP i Surface totale de la strate k
4) Quotients à estimer	r_i \hat{r}_i R_k \hat{R}_k	Valeurs du quotient pour l'UP i ($r_i = y_i/x_i$) Estimateurs du quotient dans l'UP i ($\hat{r}_i = \hat{y}_i/\hat{x}_i$) Valeurs du quotient pour la strate k ($R_k = Y_k/X_k$) Estimateurs du quotient dans la strate k ($\hat{R}_k = \hat{Y}_k/\hat{X}_k$)
5) Sommations	$\sum_{j=1}^N$ $\sum_{j=1}^n$ $\sum_{i=1}^M$ $\sum_{i=1}^m$	Somme sur toutes les US de l'UP i Somme sur les US de l'UP i appartenant à l'échantillon Somme sur toutes les UP de la strate k Somme sur les UP de la strate k appartenant à l'échantillon

Tableau 12.1 (fin) : Notations adoptées dans le formulaire

Objets	Notations	Significations
6) Erreurs d'échantillonnage et estimation des erreurs	$V(\hat{y}_i)$ et $V(\hat{\bar{y}}_i)$	Variances des estimateurs de totaux et de moyennes dans l'UP i (second degré du sondage)
	$\hat{V}(\hat{y}_i)$ et $\hat{V}(\hat{\bar{y}}_i)$	Estimations des variances des estimateurs de totaux et de moyennes dans l'UP i (second degré du sondage)
	$E(\hat{r}_i - r_i)^2$	Ecart quadratique moyen de l'estimateur de quotient dans l'UP i (second degré du sondage)
	$\hat{E}(\hat{r}_i - r_i)^2$	Estimation de l'écart quadratique moyen de l'estimateur de quotient dans l'UP i (second degré du sondage)
	$V(\hat{Y}_k)$ et $V(\hat{\bar{Y}}_k)$	Variances des estimateurs de totaux et de moyennes dans la strate k (premier degré du sondage)
	$\hat{V}(\hat{Y}_k)$ et $\hat{V}(\hat{\bar{Y}}_k)$	Estimations des variances des estimateurs de totaux et de moyennes dans la strate k (premier degré du sondage)
	$E(\hat{R}_k - R_k)^2$	Ecart quadratique moyen de l'estimateur de quotient dans la strate k (premier degré du sondage)
	$\hat{E}(\hat{R}_k - R_k)^2$	Estimation de l'écart quadratique moyen de l'estimateur de quotient dans la strate k (premier degré du sondage)

Tableau 12.2 : Formulaire d'estimation de totaux, moyennes et quotients, pour une unité primaire i (deuxième degré).

	Estimateur	Variance	Estimateur de la variance
Second degré : dans l'unité primaire i, le sondage est systématique, équiprobable, assimilé à un sondage aléatoire équiprobable sans remise (N_i et n_i sont notés N et n).			
Total	$\hat{y}_i = N \sum_{j=1}^n \frac{y_j}{n}$	$V(\hat{y}_i) = \frac{N(N-n)}{n} \sum_{j=1}^N \frac{(y_j - \bar{y}_i)^2}{(N-1)}$	$\hat{V}(\hat{y}_i) = \frac{N(N-n)}{n} \sum_{j=1}^n \frac{(y_j - \hat{y}_i)^2}{(n-1)}$
Moyenne	$\hat{\bar{y}}_i = \sum_{j=1}^n \frac{y_j}{n}$	$V(\hat{\bar{y}}_i) = \frac{N-n}{nN} \sum_{j=1}^N \frac{(y_j - \bar{y}_i)^2}{(N-1)}$	$\hat{V}(\hat{\bar{y}}_i) = \frac{N-n}{nN} \sum_{j=1}^n \frac{(y_j - \hat{\bar{y}}_i)^2}{(n-1)}$
Ratio : $r_i = y_i/x_i$	<p>Estimateur</p> $\hat{r}_i = \frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n x_j}$ <p>Ecart quadratique</p> $E(\hat{r}_i - r_i)^2 = \frac{N(N-n)}{n \bar{x}^2} \sum_{j=1}^N \frac{(y_j - r_i x_j)^2}{(N-1)}$ <p>Estimateur de l'écart quadratique</p> $\hat{E}(\hat{r}_i - r_i)^2 = \frac{N(N-n)}{n \hat{x}^2} \sum_{j=1}^n \frac{(y_j - \hat{r}_i x_j)^2}{(n-1)}$		

Tableau 12.3 : Formulaire d'estimation de totaux, moyennes et quotients, pour une strate donnée k, négligeant la variance intra-unité primaire. (premier degré)

	Estimateur	Variance	Estimateur de la variance
Premier degré : sélection des unités primaires systématique, avec probabilité proportionnelle à leur surface, assimilée à un sondage aléatoire à probabilité inégale avec remise			
Total	$\hat{Y}_k = \frac{z_k}{m} \sum_{i=1}^m \frac{\hat{y}_i}{z_i}$	$v(\hat{Y}_k) = \frac{z_k}{m} \left(\left(\sum_{i=1}^m \frac{y_i^2}{z_i} \right) - \frac{Y_k^2}{z_k} \right)$	$\hat{v}(\hat{Y}_k) = \frac{1}{m(m-1)} \sum_{i=1}^m \left(\frac{z_k \hat{y}_i}{z_i} - \hat{Y}_k \right)^2$
Moyenne : $\bar{Y}_k = Y_k / N_k$	$\hat{\bar{Y}}_k = \hat{Y}_k / \hat{N}_k$ où, $\hat{N}_k = \frac{z_k}{m} \sum_{i=1}^m \frac{N_i}{z_i}$	$v(\hat{\bar{Y}}_k) = \frac{z_k}{N_k^2 m} \left(\left(\sum_{i=1}^m \frac{y_i^2}{z_i} \right) - \frac{Y_k^2}{z_k} \right)$	$\hat{v}(\hat{\bar{Y}}_k) = \frac{1}{m^2 - m} \sum_{i=1}^m \frac{z_k^2}{\hat{N}_k^2} \left(\frac{\hat{y}_i - N_i \hat{\bar{Y}}_k}{z_i} \right)^2$
Ratio : $R_k = \frac{Y_k}{X_k}$	Estimateur $\hat{R}_k = \frac{\frac{z_k}{m} \sum_{i=1}^m \frac{\hat{y}_i}{z_i}}{\frac{z_k}{m} \sum_{i=1}^m \frac{\hat{x}_i}{z_i}}$	Ecart quadratique $E(\hat{R}_k - R_k)^2 = \frac{z_k}{m X_k^2} \sum_{i=1}^m \left(\frac{y_i^2}{z_i} + R_k^2 \frac{x_i^2}{z_i} - 2R_k \frac{y_i x_i}{z_i} \right)$	Estimateur de l'écart quadratique $\hat{E}(\hat{R}_k - R_k)^2 = \frac{1}{m(m-1)} \sum_{i=1}^m \frac{z_k^2}{\hat{X}_k^2} \left(\frac{\hat{y}_i - \hat{R}_k \hat{x}_i}{z_i} \right)^2$

12.3 : Mise en œuvre du plan de sondage : sélection de l'échantillon et étapes du travail de collecte

Nous abordons maintenant, dans l'ordre chronologique de leur réalisation, les différentes étapes de la mise en œuvre pratique de la méthode : définition de la base de sondage, sélection de l'échantillon au premier et second degré et réalisation des opérations de collecte sur le terrain. Les techniques employées à chaque étape respectent le plus rigoureusement possible le plan de sondage défini ci dessus.

On fera très souvent appel, par la suite, aux résultats obtenus par les autres membres de l'équipe de recherche qui ont entièrement réalisé les deux premières étapes : définition et stratification de la base de sondage. Pour avoir des informations supplémentaires sur les méthodes et techniques employées le lecteur se reportera aux rapports et articles rédigés par l'équipe sur ces questions¹.

12.3.1 : Définition de la base de sondage sur l'image: la limite urbaine

12.3.1.1 : Les documents utilisés

Plusieurs documents photographiques réalisés à l'Atelier de télédétection de Bondy ont servi pour parvenir à la délimitation finale de la zone urbaine :

- deux images panchromatiques SPOT :
 - Juin échelle 1/30 000 (2 tirages noir et blanc différemment contrastés afin de tenir compte de la dynamique entre les différents espaces végétaux et urbanisés) ;
 - Août échelle 1/50 000 (2 tirages noir et blanc différemment contrastés) ;
- un indice de végétation échelle 1/50 000 (2 tirages noir et blanc différemment contrastés) ;
- un écart-type local réalisé sur l'image panchromatique, copie couleur d'écran, échelle approximative 1/60 000 ;
- 1 indice de brillance , copie couleur d'écran, échelle approximative 1/60 000
- 1 coefficient de densité du bâti en 6 modalités, copie couleur d'écran, échelle approximative 1/60 000.

12.3.1.2 : Mise en œuvre et règles de décision

Quelques règles simples ont été discutées, puis retenues afin de définir de la façon la plus précise possible, le concept de "limite urbaine" applicable pour une enquête à objectif démographique. Nous avons décidé que la limite serait tracée là où se terminait *une zone continue d'espace bâti*, en suivant si possible une limite (naturelle ou artificielle) visible sur l'ensemble des documents et sur le terrain. Une telle définition, volontairement très large, doit permettre d'inclure suffisamment d'espace pour pouvoir prendre en considération l'extension de la ville entre la date d'enregistrement de l'image (juin 1986) et la date de réalisation de l'enquête démographique (novembre 1987). Les lotissements visiblement prévus, repérables grâce à l'ébauche de leur réseau de voirie, ont logiquement été inclus à l'intérieur de notre limite. *Un espace au bâti très peu dense et éparse a été délimité en se basant sur les frontières visibles (routes, rangées d'arbres, ravines, etc.) et non pas le long de la dernière maison afin de ne pas induire une densité artificiellement élevée dans les*

¹ : Par exemple, pour une première approche et une bibliographie sur le thème de la stratification d'après l'image satellite, on peut se reporter à la communication faite au colloque PIXIM par A. MICHEL, F.DUREAU, B. LORTIC, O. BARBARY, M.SOURIS : Segmentation et classification d'une image satellite en milieu urbain, 17 pages, Paris 1988. La bibliographie complète de l'équipe est donnée à la fin de l'ouvrage.

ilots périphériques Pour la même raison, les incursions profondes du contour à l'intérieur de la ville devaient être limitées et n'être prise en considération que lorsque cela permettait de détourner une zone importante d'espace non bâti.

Les différentes limites obtenues indépendamment par deux personnes (F. DUREAU et A. MICHEL) ont été confrontées. Leurs tracés respectifs n'offrent pas de différences fondamentales. Une limite de "synthèse" a donc été retenue. La discrimination de la limite externe s'est faite pour l'essentiel sur l'image panchromatique brute, mais aussi à partir de l'image de l'indice de végétation, de l'écart-type local et d'un coefficient de densité du bâti. Enfin, nous sommes allés vérifier la limite ainsi obtenue sur le terrain, cette vérification a également permis de lever certaines ambiguïtés concernant un espace recouvert de nuages sur l'image satellite.

12.3.1.3 : Validation

Nous avons eu à notre disposition la couverture aérienne photographique noir et blanc de novembre 1987 à l'échelle approximative 1/40 000 de la ville. Cette opportunité a été mise à profit pour estimer la fiabilité de la limite déterminée sur les documents satellitaires. La même démarche que précédemment, respectant les mêmes règles de décision a donc été appliquée, et une "limite urbaine" a été dessinée sur ces photographies par deux interpréteurs. *La comparaison de la limite issue de l'image SPOT avec celle issue de la couverture aérienne met en évidence la fiabilité du travail effectué sur les images satellite.* La seule véritable erreur d'interprétation est imputable aux carrières qui présentent sur le panchromatique et dans une moindre mesure sur la composition colorée une apparence d'espace bâti.

12.3.2 : Stratification de la base de sondage sur l'image : les critères retenus et la réalisation pratique, mesures de surface de chaque strate

12.3.2.1 : Utilisation de la classification des pixels de l'image SPOT selon la densité du bâti

Le zonage à l'intérieur de l'espace urbain a été réalisé à partir d'une copie d'écran représentant une classification de la densité du bâti. Cette classification de type barycentrique en 6 postes a été supervisée en fonction des statistiques descriptives calculées sur l'échantillon de 164 ilots précisément décrites et mesurées lors de l'enquête Morphologie. Les canaux (XS1) et néo-canaux (Ecart-type local sur panchromatique, Indice de végétation, et Indice de brillance) utilisés pour réaliser cette classification ont été choisis en fonction de leur aptitude à rendre compte de la densité du bâti. Les délais d'élaboration de cette classification ont été très courts, l'enquête devant impérativement commencer avant la fin de l'année 1987 : on s'est donc contenté de reproduire le type de classification expérimenté sur le site de Marseille.

12.3.2.2 : Evaluation de la classification

En raison des délais très courts, la classification a été initialisée et évaluée sur un même échantillon. Cela est contraire à la rigueur scientifique, mais l'absence d'étude précise sur la détermination des canaux et néo-canaux les plus aptes à discriminer la densité de l'espace bâti nous a légitimement conduit à initialiser la classification sur le plus grand nombre possible d'individus afin de réduire autant que possible les classements de pixels ambigus. Ce choix s'est donc, dès le départ effectué au détriment de la validation de la classification. Cependant, il est nécessaire d'évaluer cette

classification, en particulier pour connaître les confusions entre les classes les plus importantes. La matrice de confusion qui suit a été établie (en effectif), en croisant la modalité de référence avec la classe d'affectation.

Matrice de confusion						
Confusion	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Référence 1	2	2	2	0	0	0
Référence 2	8	29	16	5	2	0
Référence 3	1	3	10	20	15	0
Référence 4	1	1	2	5	14	0
Référence 5	1	1	2	3	17	2
Référence 6	0	1	0	0	1	3

La fiabilité de cette classification est :

39% en 6 classes

60% en 3 classes

On remarque qu'un décalage important subsiste entre classe et modalité de référence. La classification a tendance à surestimer la densité. Etant donné l'objectif de la classification, stratifier l'espace quiténien selon la densité du bâti, ce décalage n'est pas gênant : cette classification respecte bien une hiérarchie de densité, même si celle-ci est décalée par rapport à une référence.

12.3.2.3 : *Elaboration de la stratification*

L'interprétation de cette classification de pixels, a nécessité un important travail de zonage, non seulement en fonction de la valeur des pixels, mais surtout en fonction des *associations de valeurs* (de la texture) de l'image classée. En effet, il est aberrant de concevoir un indicateur de densité par pixel, et l'on doit tenir compte du voisinage, à fortiori s'il s'agit d'effectuer un zonage dans l'optique d'une stratification de l'espace. Les plages homogènes de petite surface n'ont donc pas été détournées. Ce zonage, pour lequel nous avons conservé les 6 postes initiaux a été effectué indépendamment par F. DUREAU et A. MICHEL, puis reporté, après confrontation, sur le tirage panchromatique au 1/30.000 qui comporte déjà la limite urbaine.

12.3.2.4 : *Mesures de surface de chaque strate*

Le zonage établi sur un tirage du canal panchromatique SPOT au 1/30.000 a été reporté sur un agrandissement du même canal au 1/15.000. C'est à cette échelle qu'a été mesurée chacune des 232 zones composant la base de sondage à l'aide d'un planimètre électronique.

Le tableau ci dessous présente les surfaces occupées par chacune des 5 strates construites, ainsi que la superficie occupée par les zones non construites incluses à l'intérieur du périmètre urbain.

Strates	non bâti	1 densité très faible	2 densité faible	3 densité moyenne	4 densité forte	5 densité très forte	Total de l'aire métropol- -itaine
Surfaces S _k (ha)	2065,8	3485,4	3258,2	3287,2	3207,1	1036,9	16340,6

12.3.3 : Sélection de l'échantillon de pâtés de maisons

Une fois définie et stratifiée la base de sondage, le tirage de l'échantillon de pâtés de maison s'effectue indépendamment dans chaque strate, comme toujours pour un sondage stratifié. Pour respecter la règle d'allocation de l'échantillon aux strates, on doit construire une grille de sélection pour chaque strate, dont la taille de la maille sera déterminée pour respecter le taux prévu. Les étapes du processus de sélection, que nous allons maintenant décrire, sont illustrées par la figure 12.6 pour une partie de la base de sondage.

12.3.3.1 : Détermination du nombre total de pâtés de maisons de l'échantillon et de la répartition entre les strates

Le budget alloué à l'enquête et la période de temps dont nous disposons pour sa réalisation permettent d'enquêter entre 3000 et 3500 ménages. Combien de pâtés de maisons doit on sélectionner pour obtenir cette taille d'échantillon? La réponse à la question est immédiate si le nombre de ménages enquêtés dans chaque pâté de maisons est constant. Dans notre cas, tel qu'a été défini le plan de sondage au second degré, le nombre total de ménages enquêtés dépend du nombre de ménages que nous devons enquêter dans les pâtés de maisons de moins de dix ménages inclus dans l'échantillon. Examinons plus précisément ce point.

Appelons m le nombre de ménages total à enquêter, n le nombre de pâtés de maisons à sélectionner dans l'échantillon, p_1 la proportion dans l'échantillon des pâtés de maisons contenant moins de dix ménages et m_1 le nombre moyen de ménages que contiennent ces pâtés de maisons. Puisqu'on enquête exhaustivement les pâtés de maisons de moins de dix ménages et que l'on choisit dix ménages dans les autres, on a :

$$m = m_1 n p_1 + 10n(1-p_1), \text{ donc}$$

$$n = \frac{m}{p_1 \cdot m_1 + (1-p_1) \cdot 10} \quad (12.3.3.1)$$

Pour calculer n il faut donc prévoir les valeurs qu'auront p_1 et m_1 dans l'échantillon. Nous baserons cette prévision sur les données recueillies lors de l'enquête préliminaire "bâti population", mais il faut faire une remarque à ce sujet.

On se souvient que l'échantillon correspondant à cette enquête (53 pâtés de maisons) a été choisi de manière raisonnée au sein des 199 pâtés de maisons de l'enquête préliminaire "morphologie", eux même sélectionnés à l'aide d'une grille de points plaquée sur les plans réalisés par l'INEC¹ pour le recensement de 1982. Ce mode de sélection a deux conséquences :

1. Les probabilités d'appartenance des unités à l'échantillon ne sont pas contrôlées et en particulier, rien ne permet d'affirmer qu'elles aient été égales entre elles ou proportionnelles à la surface des pâtés de maisons. On considèrera la proportion des pâtés de maisons de moins de dix ménages dans cet échantillon et la "taille" moyenne de ces pâtés de maisons comme des estimateurs de p_1 et de m_1 mais il est clair qu'ils peuvent être biaisés et avoir des variances élevées; donc les prévisions seront très approximatives.

2. La base de sondage correspondant à cet échantillon, définie par la limite urbaine adoptée par l'INEC en 1982, est beaucoup plus restreinte que celle que nous retenons pour cette enquête. La plupart des zones supplémentaires que nous considérons comme appartenant maintenant à la ville sont des zones périphériques relativement peu denses. Il est certain que la proportion d'îlots de moins de dix ménages observée sur l'échantillon des 53 îlots de l'enquête préliminaire est nettement inférieure à ce qu'elle est dans la base de sondage définie. La formule 12.3.3.1, lorsqu'on remplace p_1 et m_1 par les valeurs observées sur l'échantillon de l'enquête préliminaire fournira donc un nombre de pâtés de maisons trop faible. C'est la raison pour laquelle pour calculer n , nous avons fixé à 4000 le nombre total de ménages m .

Dans l'échantillon des 53 pâtés de maisons, on constate la répartition suivante :

- Pâtés de maisons contenant jusqu'à dix ménages : 22,7 %, en moyenne 6,9 ménages par pâté de maisons

- Pâtés de maisons de plus de dix ménages : 77,3 % .

$$\text{On a donc } n = \frac{4000}{0.227 \times 6.9 + 0.773 \times 10} \approx 430 .$$

Reste à répartir ces 430 pâtés de maisons dans les cinq strates. Le principe est simple : disposant pour chaque strate de la surface totale mesurée sur l'image satellite (S_k) et de la surface moyenne des pâtés de maisons de la strate estimée d'après l'échantillon de 199 pâtés de maisons de l'enquête préliminaire "morphologie" (\bar{s}_k), on calcule tout d'abord une approximation du nombre de pâtés de maisons que contient chaque strate (N_k).

Le quotient $\frac{430}{\sum_{k=1}^5 N_k}$ fournit ensuite le taux global sur l'ensemble de la base qui, appliqué

à chacune des strates donne la répartition dans le cas de l'allocation proportionnelle. Enfin, par application des coefficients qui définissent l'allocation aux strates (C_k : voir tableau du paragraphe 12.2.2), on obtient le nombre de pâtés de maisons à sélectionner

¹ : INEC : Instituto Nacional de Estadística y Censo, institut national de statistique équatorien.

dans chaque strate(n_k). Cependant, la répartition entre les strates de l'échantillon n'est pas exactement la même selon que l'on considère les 53 pâtés de maisons de l'enquête "bâti/population" ou les 199 pâtés de maison de l'enquête "morphologie". Comme l'allocation du paragraphe 12.2.2 a été définie à partir de l'échantillon de 53 îlots (de plus cette allocation est le résultat d'une moyenne d'allocations optimales), si l'on applique tels quels les coefficients définis, l'échantillon obtenu compte trop peu de pâtés de maisons. Il est donc nécessaire d'ajuster les coefficients à la répartition entre strates observée sur l'échantillon des 199 pâtés de maisons qui sert de base au calcul. Le tableau 12.4 ci dessous résume les calculs effectués et la dernière ligne donne la répartition de l'échantillon qui, finalement, a été adoptée.

Tableau 12.4 : résumé des calculs effectués pour la détermination du nombre de pâtés de maisons à enquêter dans chaque strate.

Strates	1 densité très faible	2 densité faible	3 densité moyenne	4 densité forte	5 densité très forte	Total de la base
S_k (m ²)	34.853.537	32.582.087	32.872.461	32.070.924	10.369.293	142.748.300
\bar{s}_k (m ²)	31.473	25.875	20.374	12.443	12.106	19.256
N_k	1.107	1.259	1.613	2.577	857	7.413
n_k (Alloc. prop.)	64	73	94	149	50	430
C_k (Originaux)	0,4	0,3	1	1,5	0,9	0,95
C_k (Ajustés)	0,44	0,38	1,15	1,54	0,76	1
n_k	28	28	108	230	38	43

12.3.3.2 : Définition des grilles

Pour obtenir, pour chaque strate, la surface de la maille de la grille de tirage, il suffit de diviser la surface totale de la strate mesurée sur l'image par le nombre d'îlots à sélectionner. On calcule ensuite la longueur du côté de la maille. Le tableau 12.5 ci-après fournit les dimensions des grilles obtenues pour chaque strate.

Tableau 12.5 : calcul des mailles de grilles dans chaque strate.

Strates	1 densité tres faible	2 densité faible	3 densité moyenne	4 densité forte	5 densité tres forte
S_k (en mm ² sur l'image ¹)	161.963	151.408	152.757	149.033	48.186
N_k	28	28	108	230	38
Surface de la maille (en mm ² sur l'image)	5784,4	5407,4	1414,4	648,0	1268,1
Côté de la maille(en mm sur l'image)	76,1	73,5	37,6	25,5	35,6

12.3.3.3: Sélection

Une fois construites les cinq grilles sur papier calque millimétré, on a procédé au tirage en cinq étapes en commençant par la strate la plus dense (strate 5) pour finir par la strate la moins dense (strate 1). Une précaution importante doit être prise lors des placements initiaux des grilles sur l'image : il faut éviter d'orienter les axes principaux de la grille (horizontal, vertical ou diagonal) parallèlement à un axe principal du réseau de voirie. En effet, une telle orientation pourrait aboutir à une sur-représentation dans l'échantillon des îlots situés en bordure (ou à distance fixe) d'un grand axe, lesquels ont des caractéristiques morphologiques et démographiques spécifiques. On risquerait alors par exemple de surestimer les activités de commerce et service ou même de sous-estimer le total de population résidente. Remarquons que ce type d'erreur d'échantillonnage est en fait le pendant, dans le cas du tirage systématique aréolaire, des erreurs qui surviennent dans un sondage systématique sur liste lorsque la variable à estimer présente des variations périodiques dans la liste et que le pas de tirage est approximativement un multiple de la période².

Les pas des grilles sont calculés, comme on l'a vu plus haut, pour fournir dans chaque strate le nombre n_k d'îlots souhaité. Il se peut cependant que le placement initial de la grille sur l'image conduise en définitive à un nombre d'îlots légèrement différent. Cette situation ne s'est pas produite lors du processus de sélection et, par chance, le premier placement de la grille a toujours donné le nombre exact de pâtés de maisons. Dans un cas moins favorable il aurait fallu répéter l'opération en défessant aléatoirement une nouvelle position de la grille jusqu'à obtenir le bon nombre de points inclus dans la strate.

¹ : rappelons que ces mesures sont effectuées sur une image à une échelle approximative de 1/15 000.

² : Voir sur ce point par exemple J. DESABIE, "Théorie et pratique des sondages" P. 99 ou J.M. GROBRAS, "Méthodes statistiques des sondages" P.29.

Une fois que l'on s'est assuré d'avoir dans la strate le nombre de points voulu, l'étape suivante consiste à définir sur l'image, partout où c'est possible (c'est-à-dire partout où le réseau de voirie apparaît sans ambiguïtés), l'îlot à enquêter correspondant à chaque point de la grille. Lorsque le point est situé clairement à l'intérieur de l'îlot c'est évidemment l'îlot en question qui est sélectionné. Si le point tombe dans la voirie séparant deux (ou plusieurs) pâtés de maisons, on applique une règle systématique : nous avons pour notre part retenu les îlots situés immédiatement au nord-ouest des points. Pour faciliter l'identification des pâtés de maisons par les enquêteurs et leur repérage sur le terrain, l'ensemble des îlots ainsi repérés est reporté sur les plans utilisés par l'INEC pour le recensement (voir figure 12.6). Enfin, lorsque l'identification précise de l'îlot est impossible sur l'image ou que l'îlot est situé en dehors de la zone cartographiée par l'INEC, on effectue alors le repérage et/ou l'identification des limites directement sur le terrain avec le superviseur chargé de la zone (parfois même avec l'enquêteur en charge de l'îlot). La localisation du point sur le terrain au vu de sa situation sur l'image n'a jamais posé de problème ; en revanche l'accès au point en question a parfois été difficile.

Signalons pour finir que le respect du tirage sans remise a été total et qu'il n'a posé aucun problème de remplacement ou de répétition de la sélection puisque, lors du tirage, aucun pâté de maisons n'a été sélectionné deux fois. Ce fait n'est d'ailleurs pas surprenant : au taux pratiqué, la maille de la grille dans chaque strate est suffisamment grande pour que le risque de sélection répétée d'une même unité, qui correspond à l'existence d'unités assez grandes pour contenir plusieurs points de la grille, soit quasiment nul.

Figure 12.6 : illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage (I)

L'image satellite ayant servi de base de sondage (fragment au centre de quito)



Figure 12.6 : illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage (II)

Première étape : La limite urbaine, telle qu'elle a été desssinée sur l'image support de la sélection. Ici, une partie du nord moderne de la ville, comprise entre le parc "la Carolina" et l'aéroport.

 : zones non bâties exclues de la base de sondage

 : zones incluses dans la base de sondage

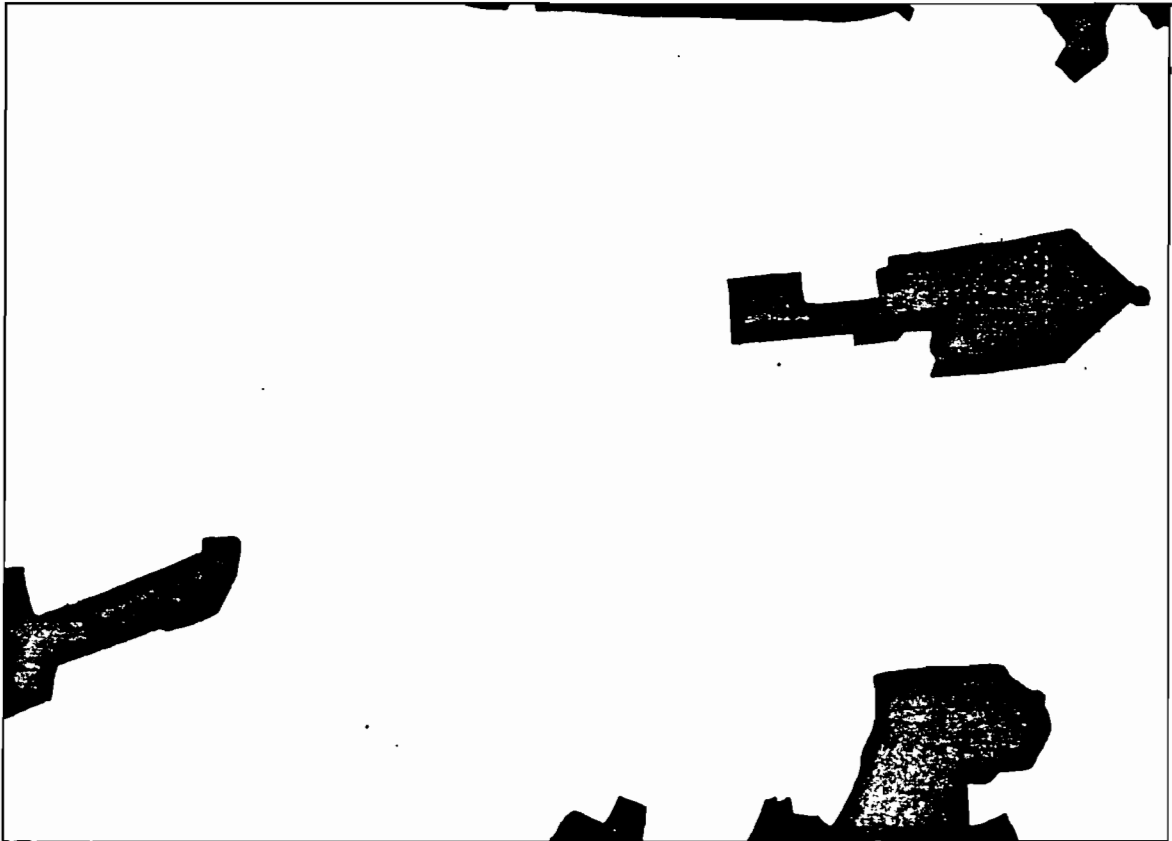



Figure 12.6 : illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage (III)

Deuxième étape : Stratification de la base de sondage sur le critère de la densité du bâti :

-  : zones non bâties exclues de la base de sondage
1. Densité très peu dense
 2. Densité peu dense
 3. Densité moyenne
 4. Densité forte
 5. Densité très forte.

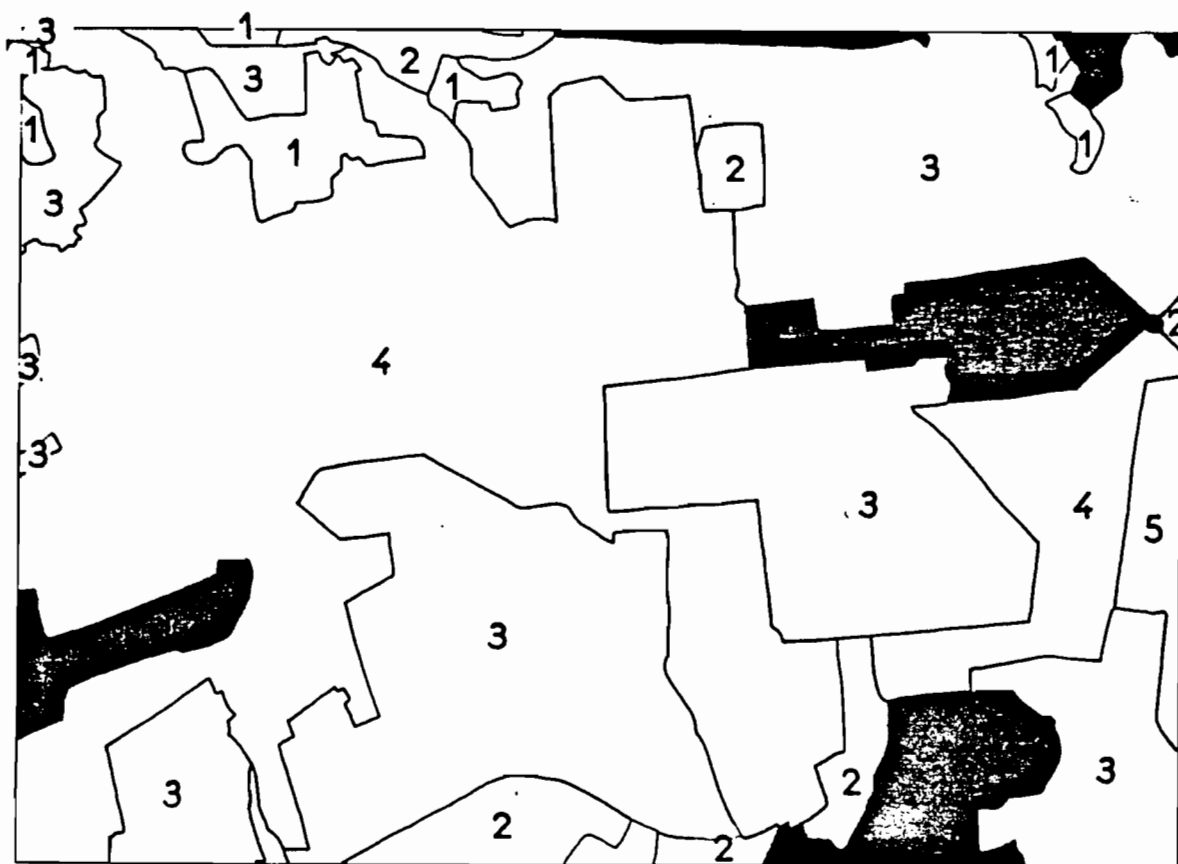


Figure 12.6 : illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage (IV)

Troisième étape : Dans chaque strate, la grille superposée à l'image permet la sélection des unités primaires de l'échantillon. L'allocation aux différentes strates est gérée en faisant varier la taille de la maille de la grille pour sélectionner, dans chaque strate, le nombre d'îlots voulu.
Ici, les grilles plaquées sur les strates 3 et 4.

 : zones non bâties exclues de la base de sondage

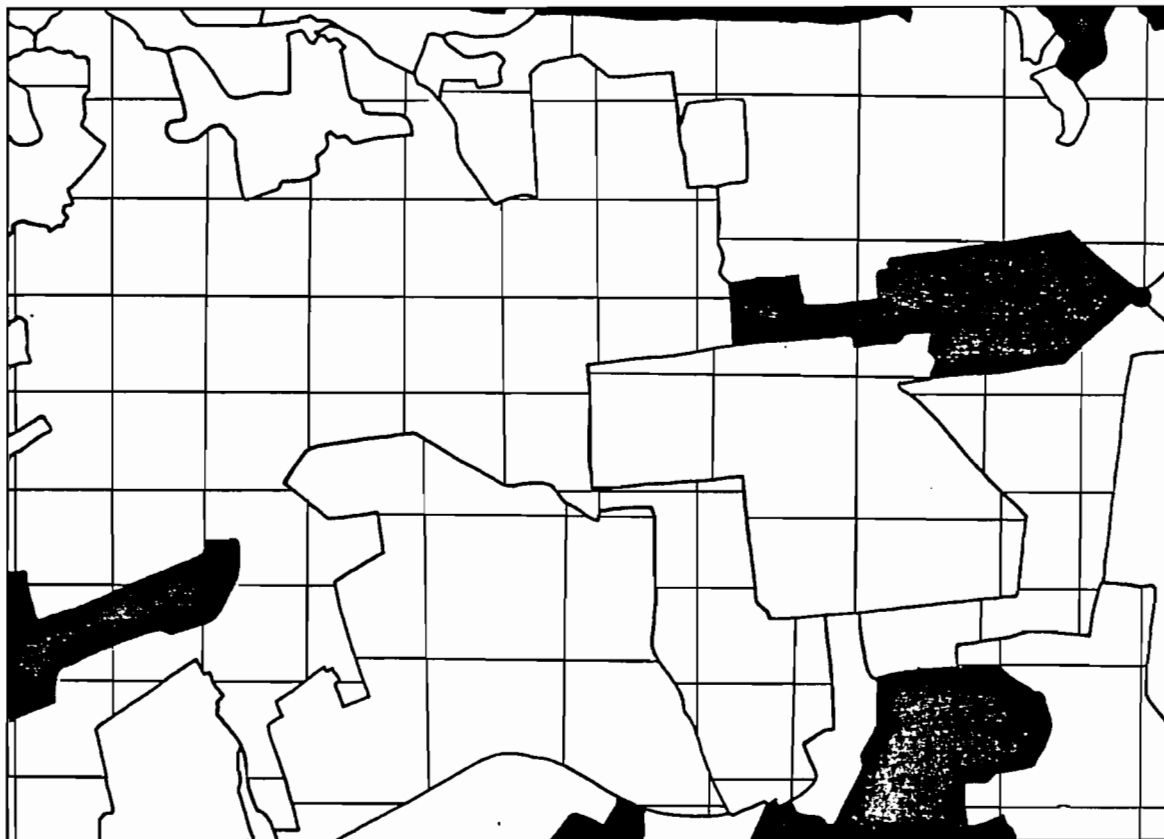


Figure 12.6 : illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage (V)

Quatrième étape : Les ilots à enquêter sont identifiés sous chaque point de la grille. Ici, les ilots sélectionnés dans les strates 3 et 4.

 : zones non bâties exclues de la base de sondage



12.3.4 : Recensement des ménages dans les pâtés de maisons et sélection des ménages à enquêter

Une fois réalisée l'opération de sélection de l'échantillon d'unités primaires, le travail de collecte de l'information relative aux pâtés de maisons peut commencer ; il a un double objectif :

1. effectuer le recensement des ménages qui occupent le pâté de maisons pour pouvoir sélectionner l'échantillon de ménages à enquêter et, par la suite, pratiquer l'inférence au second degré (estimation des totaux, moyennes et quotients au sein des unités primaires)
2. mesurer la surface de chaque unité primaire pour pouvoir pratiquer l'inférence au premier degré (estimation des totaux, moyennes et quotient de la base entière, des strates ou des éventuels domaines d'étude).

Nous indiquons ici rapidement comment ont été atteints ces objectifs.

12.3.4.1 : Recensement des ménages

Le travail de recensement des ménages dans chaque unité primaire est confié aux superviseurs qui réalisent sur le terrain un schéma du pâté de maisons permettant son identification sans ambiguïté par l'enquêteur et sur lequel figure, outre les indications que le superviseur juge utiles pour le travail de l'enquêteur, l'ensemble des bâtiments constituant l'îlot, ainsi que le nombre de ménages qui occupent chacun des bâtiments (figure 12.7). C'est ce document qui sert de base au comptage des ménages, puis à la numérotation des différents bâtiments occupés. Il est également utilisé par l'enquêteur chargé des mesures de surfaces lorsque ces dernières doivent être effectuées sur le terrain.

Signalons enfin que dès cette première étape du travail de collecte et durant toute la période d'enquête, un quartier populaire situé à l'extrême nord-ouest de la ville a été le théâtre d'affrontements armés qui ont fait plusieurs morts parmi la population¹. Il était donc exclu d'y mener l'enquête et l'échantillon réellement enquêté ne comprend de ce fait que 426 îlots au lieu des 432 initialement sélectionnés.

12.3.4.2 : Sélection des ménages à enquêter

Sur le schéma de l'îlot réalisé, le superviseur indique à l'enquêteur la liste des ménages constituant l'échantillon. Hormis pour les pâtés de maisons comptant jusqu'à dix ménages, pour lesquels l'enquête est exhaustive, la règle de sélection systématique des ménages est la suivante : en divisant le nombre total de bâtiments par dix, le superviseur obtient le pas k du tirage ; il sélectionne ensuite "au hasard" un bâtiment dont le numéro est compris entre 1 et k puis, en partant de ce numéro il retient un bâtiment tous les k . L'enquêteur devra choisir un ménage par bâtiment figurant dans la liste. En cas de refus de réponse ou d'absence prolongée pour un ménage donné, l'enquêteur le remplace par un ménage du même bâtiment ou, si c'est impossible, par un ménage choisi dans un bâtiment voisin. Ces règles s'appliquent sans problèmes à l'immense majorité des cas à Quito, ville qui compte très peu de grands ensembles. *Néanmoins, pour assurer une représentation équitable, au sein de l'échantillon, des ménages vivant dans de l'habitat collectif, les bâtiments contenant un grand nombre de ménages sont traités à part lors du processus de sélection des ménages. La règle est alors de sélectionner d'office un ménage dans chaque bâtiment contenant plus d'un dixième des ménages du pâté de maisons, deux ménages dans les bâtiments en contenant plus de deux dixième, etc...*

¹ : Il s'agit du quartier "d'invasions" regroupant les coopérative "Pisuli" et "Jaime Roldos" dont il a été question dans l'introduction générale.

En toute rigueur, la stratégie de sélection des ménages au sein des unités primaires correspond donc elle même à un sondage à deux degrés dont le premier degré consiste à choisir un échantillon de bâtiments avec probabilité d'inclusion approximativement proportionnelle au nombre de ménages occupant le bâtiment. Mais, comme cette probabilité n'est que très grossièrement respectée, il nous semble illusoire voir dangereux, d'intégrer ce degré intermédiaire de sondage dans le formulaire d'inférence. On considèrera en fait que l'échantillon de ménages dans chaque îlot est prélevé selon un plan équiprobable sans remise.

À l'issu du processus de sélection et de l'opération de collecte, l'échantillon des ménages comprend 3161 ménages répartis dans les 426 îlots enquêtés, ce qui correspond à un taux final d'environ 1,3 % des ménages.

12.3.5 : Mesure de surface des unités primaires¹

12.3.5.1 : Les méthodes de mesure employées

Pour la mesure des surfaces des unités primaires, deux techniques ont été employées et validées :

- mesure sur le tirage photographique du canal panchromatique SPOT au 1/15 000 employé pour le tirage de l'échantillon d'îlots,
- mesure sur le terrain.

256 des 426 îlots enquêtés ont pu être repérés sans ambiguïté sur le tirage photographique au 1/15 000 ; nous avons alors mesuré la surface par superposition d'un plastique transparent millimétré.

Pour les 170 îlots non repérables précisément sur image satellite, un enquêteur a été envoyé sur le terrain. Il a établi avec le maximum de précision possible un croquis mentionnant les angles et les longueurs de chaque segment droit délimitant l'îlot ainsi que leur pente (Figure 12.8).

Les longueurs sont mesurées au décamètre sur les terrains en pente et au pas sur les terrains plats, la longueur du pas ayant été préalablement étalonnée par comparaison avec les mesures effectuées au décamètre ; les angles sont mesurés à la boussole à chaque extrémité d'un segment, ce qui autorise une vérification. La pente de chaque segment est mesurée au clinomètre.

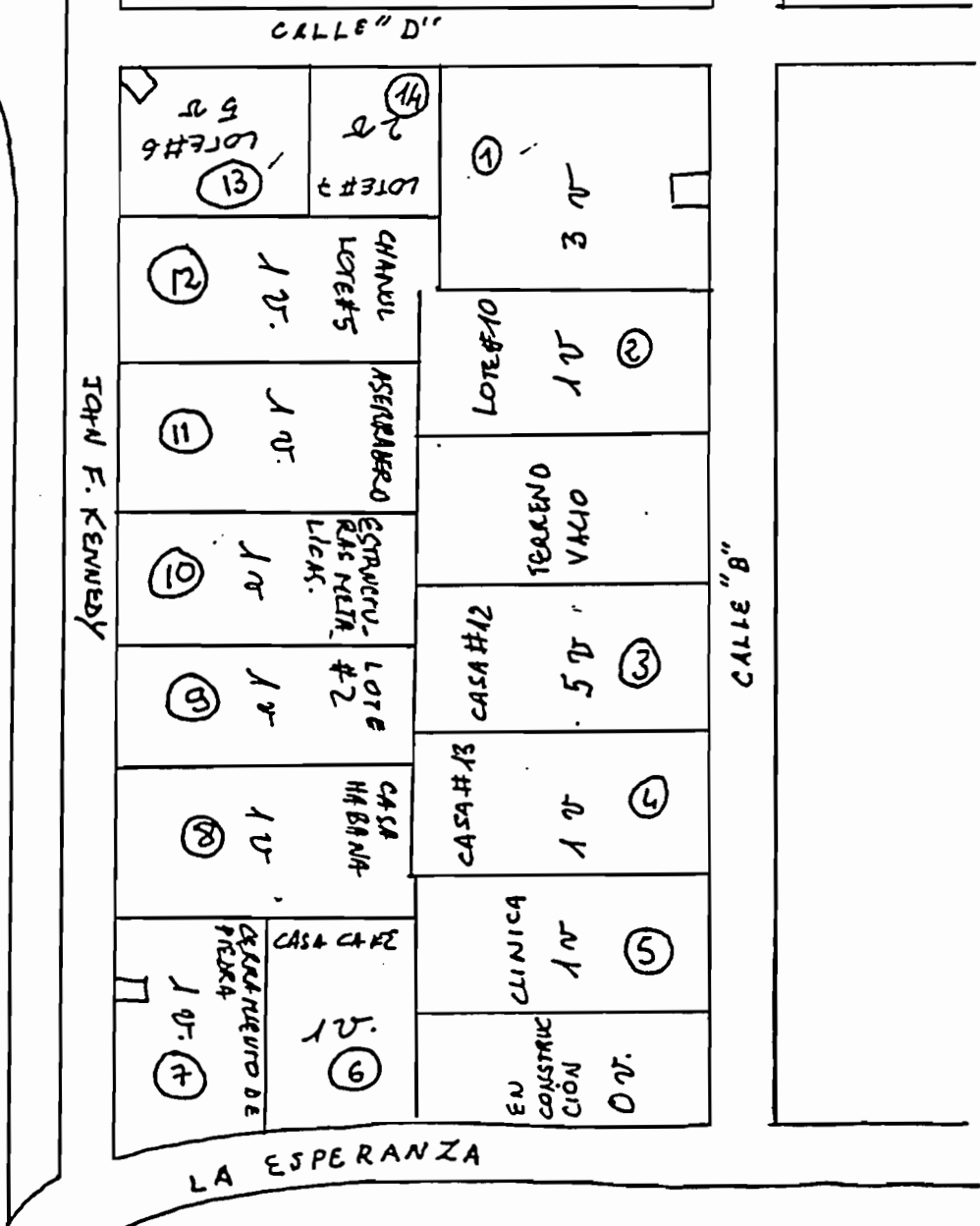
La surface est ensuite calculée automatiquement à l'aide du logiciel MacDraft™ sur Apple MacIntosh. Il suffit de dessiner à l'échelle choisie les contours de l'îlot en respectant les longueurs préalablement corrigées en fonction de la pente, et les angles prescrits (Figure 12.9).

¹ : Ce paragraphe est repris dans ses grande lignes de la note de travail qu'a rédigé A. MICHEL en avril 88 : "Evaluation de la précision des mesures de surfaces des îlots, étude sur le site de Quito. Quito, 13 p.

Figure 12.7 : Schéma du pâté de maisons établi à l'issue du recensement des ménages, servant pour la sélection des ménages à enquêter

Nombre SUPERVISOR <u>HENRY ALBAN</u> Nombre ENCUESTADOR <u>JUAN PUENTE</u>	N° Manzana <u>253</u> Número de viviendas : TOTAL <u>125</u>
---	---

ESQUEMA DE LA MANZANA



MUESTRA DE VIVIENDAS Lista de las viviendas o método de selección <u>1 vivienda en cada de los edificios.</u> Siguiendo <u>1-4-6-8-10-12-14-13-</u> <u>viviendas en el edificio 3.</u> Método de selección de la viviendas de sustitución.....	Recolección /Codificación Encuesta <input checked="" type="checkbox"/> Verif. encuesta (superv) <input checked="" type="checkbox"/> Verif. encuesta (Orstom) <input checked="" type="checkbox"/> Correcciones question. <input checked="" type="checkbox"/> Codificación <input type="checkbox"/> Verif. codificación <input type="checkbox"/>
---	---

Figure 12.8 : Croquis d'ilot pour les mesures de surface sur le terrain

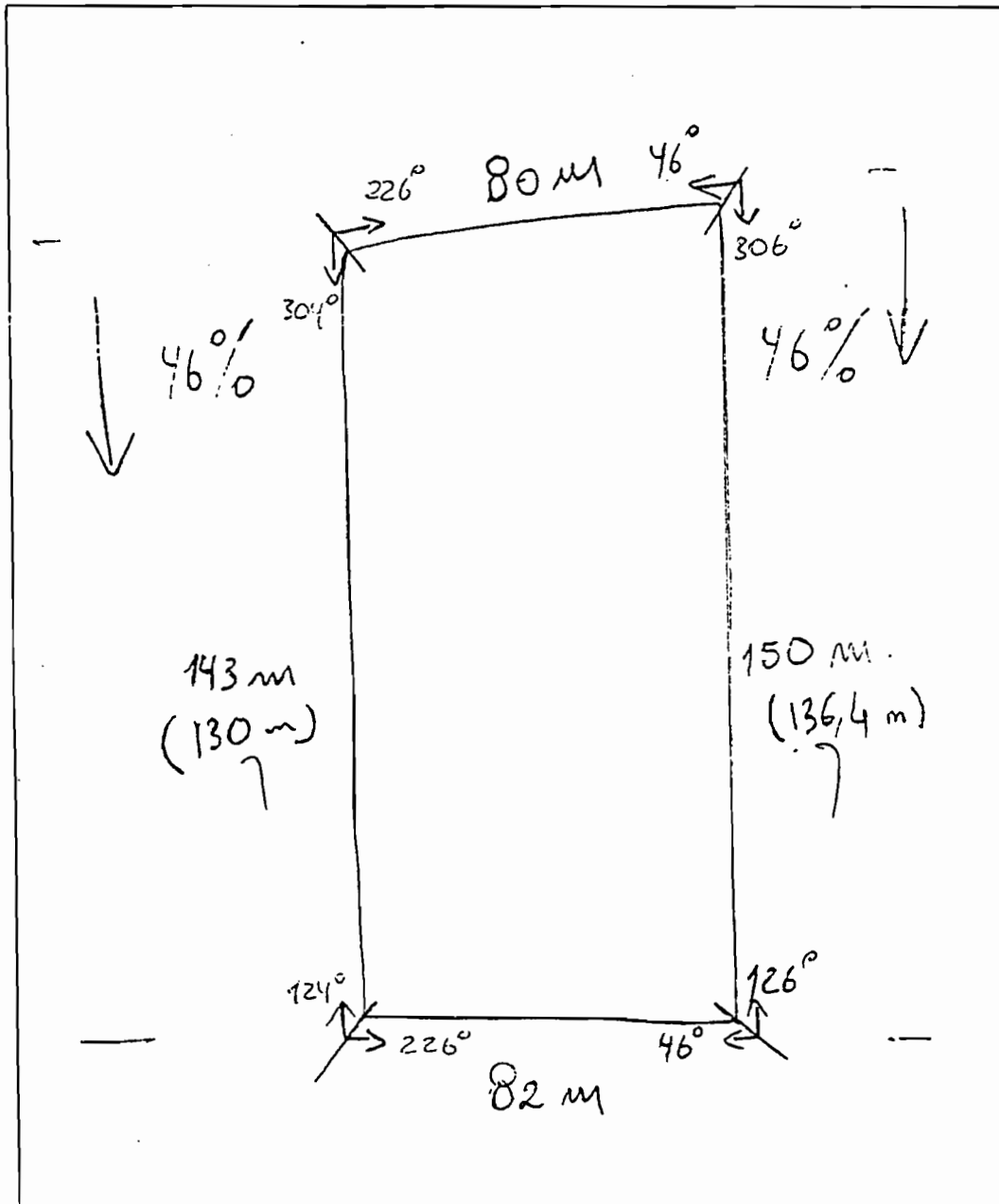
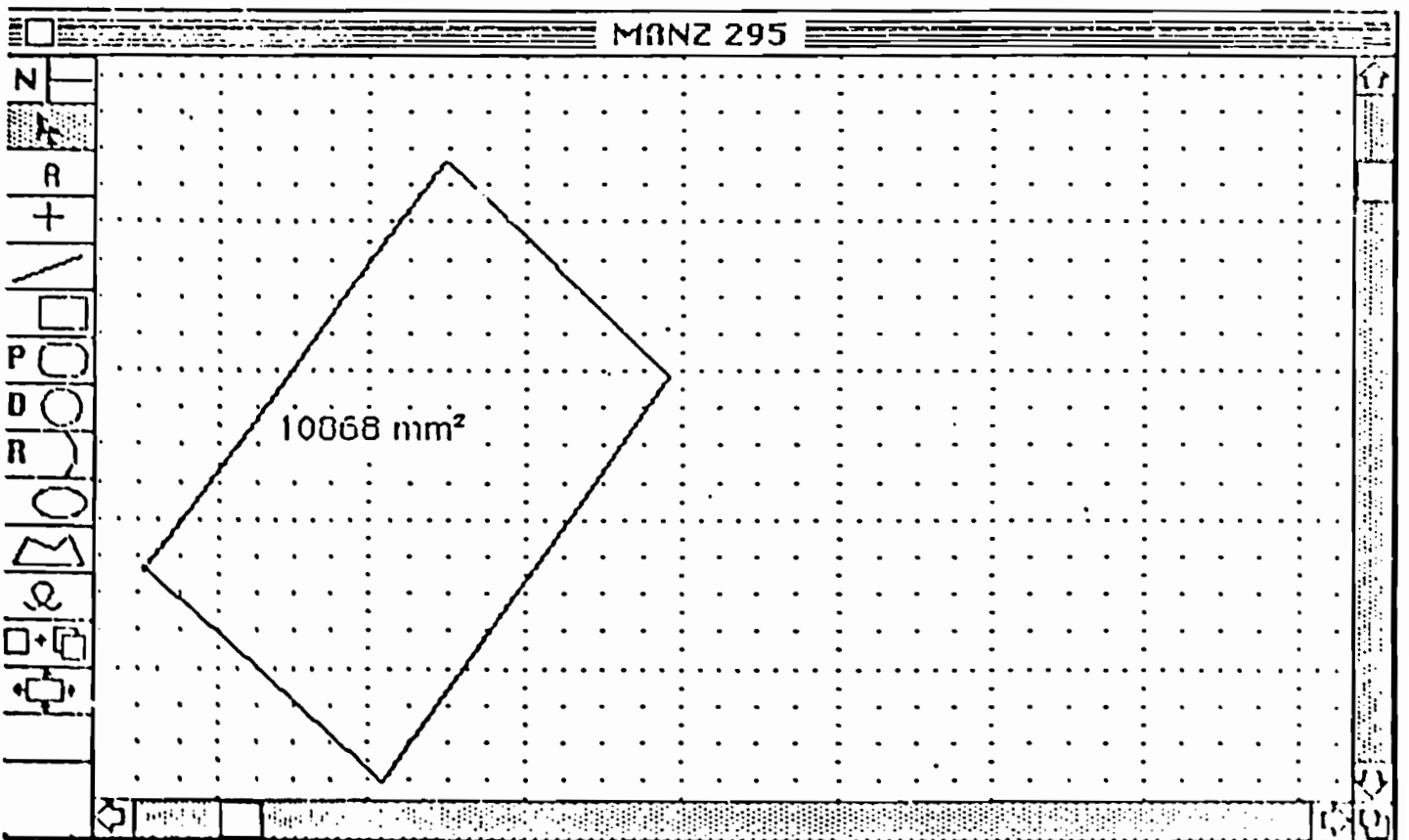
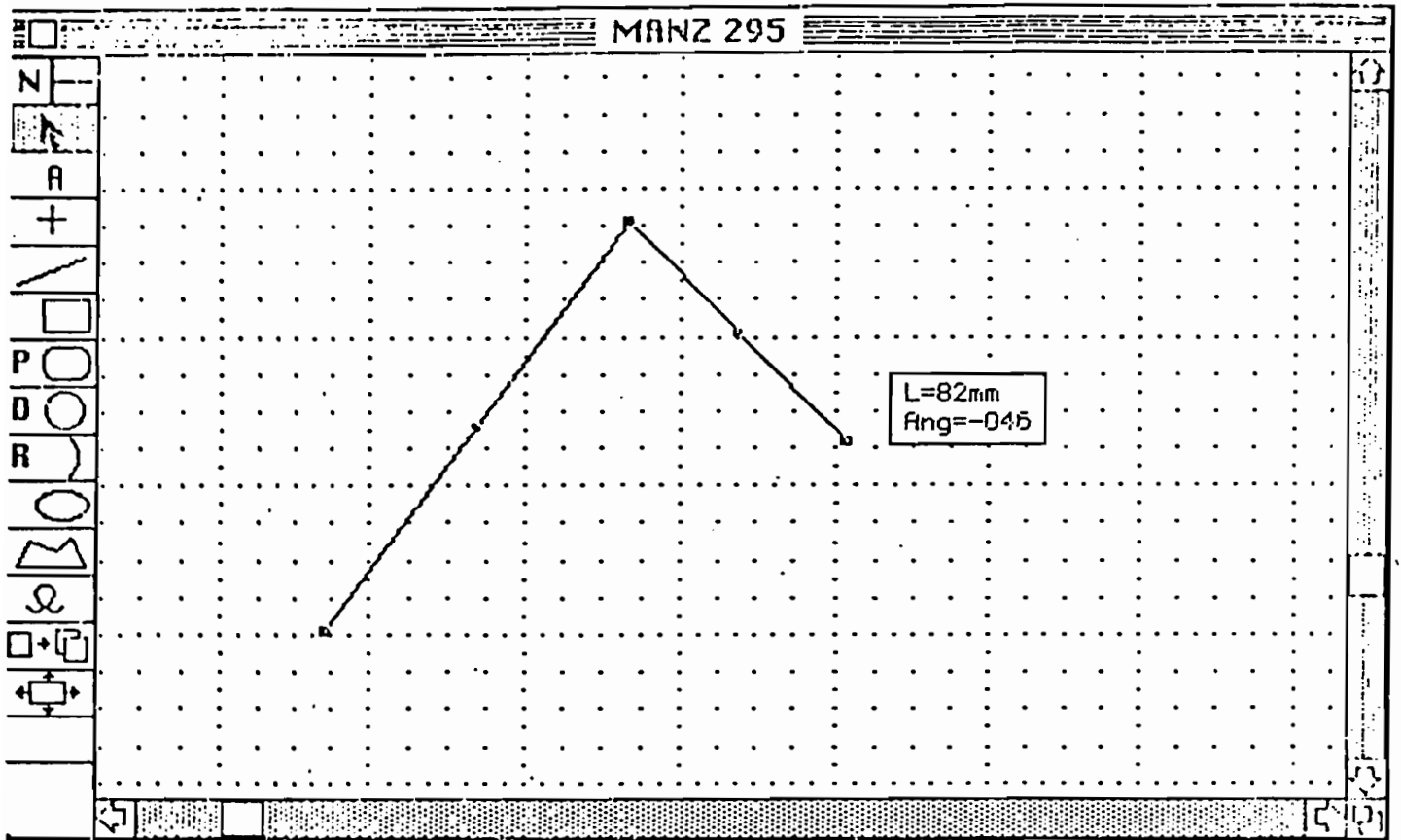


Figure 12.9 : Exemple de calcul de surface effectué à l'aide du logiciel MacDraft™ sur Macintosh



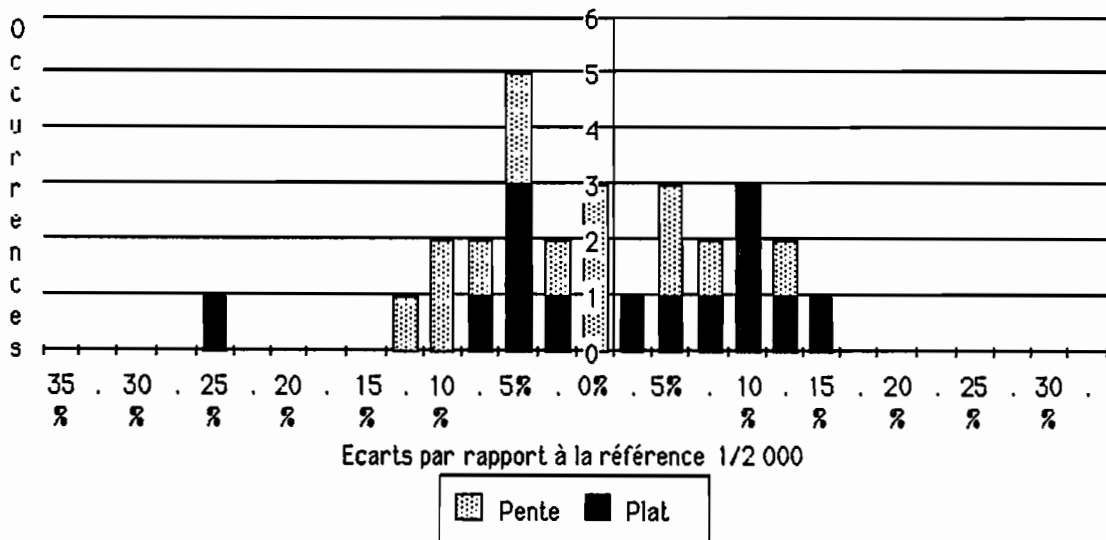
12.3.5.2 : Evaluation de la précision des techniques de mesures

Pour deux échantillons de 30 ilots correspondant chacun à une technique de mesure, nous avons comparé les mesures ainsi obtenues avec celles, considérées comme référence, issues de la cartographie au 1/2 000 de l'IGM.

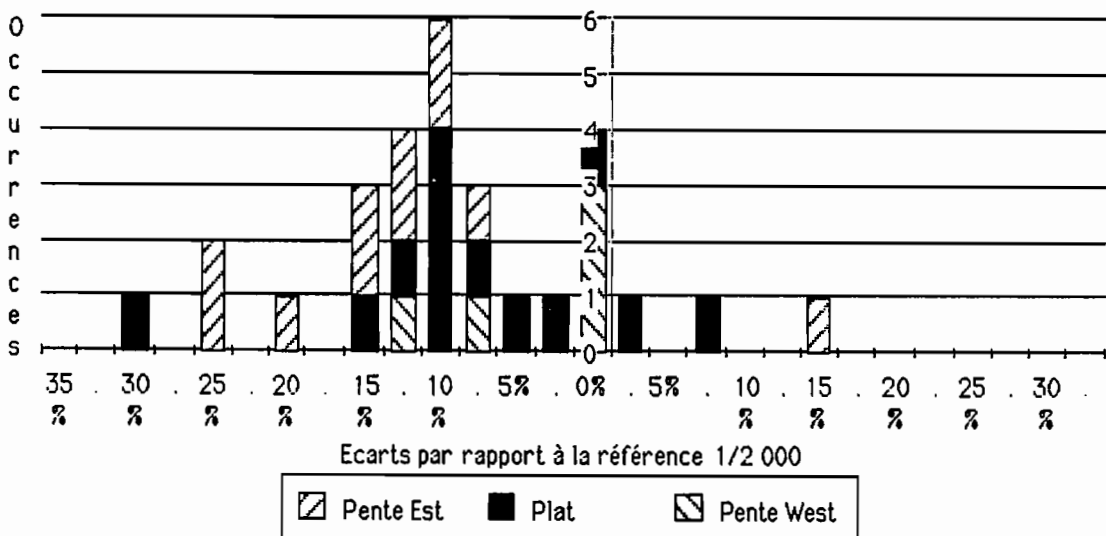
Les graphiques qui suivent correspondent aux fréquences des différences relatives entre la mesure sur image ou sur le terrain et la surface de référence. Les fréquences correspondant aux différentes modalités de pente sont superposées les unes sur les autres. Ces écarts (en pourcentage) sont classiquement négatifs à gauche de l'axe des ordonnées (*effectif*) et positifs à droite. L'axe des abscisses (*différences relatives*) est gradué de 2,5 en 2,5%.

Graphique 12.10 : Mesures de surfaces

Fréquences des écarts relatifs
Mesures sur le terrain



Fréquences des écarts relatifs
Mesures sur tirage photo



Les surfaces mesurées sur le terrain apparaissent comme remarquablement centrées autour de 0%, variant de -15% à 15%. Les mesures sur pente sont légèrement plus nombreuses à avoir été sous-estimées.

Le centrage du graphique autour de -10% montre très nettement que les surfaces, toutes modalités de pente confondues ont été sous-estimées par cette méthode de mesure. Il s'agit ici d'un biais *systématique* ; il est donc légitime de multiplier la surface ainsi obtenue par un coefficient de valeur 1.09 (1/0,92) pour tenir compte de cette sous-estimation.

L'opposition entre îlots exposés à l'Est (situés en majorité à gauche de l'abscisse -10%) et îlots exposés à l'Ouest (en majorité à droite de l'abscisse -10%) apparaît clairement sur ce graphique.

En résumé, l'analyse des écarts relatifs entre les surfaces mesurées sur image ou sur le terrain et les surfaces de référence conduit aux conclusions suivantes :

- Fiabilité des différentes mesures :

On note avec intérêt, que *c'est la mesure la plus facilement praticable dans les pays en développement (mesure sur le terrain) qui apparaît comme la plus fiable.*

- Influence de la taille de l'îlot:

Les surfaces sont globalement sous-estimées pour les îlots de grande taille quelque soit la technique de mesure employée. Sur image satellite, les mesures de surface sont plus imprécises sur des îlots de petite taille ; pour les mesures terrain, ce sont pour les îlots de taille moyenne qu'elles sont les moins précises.

- Influence de la pente:

La pente affecte peu les mesures de surface effectuées sur le terrain ; par contre sur le document satellite, très logiquement, les surfaces situées sur un flanc Ouest sont *sous-estimées* (conjugaison de la pente et de l'angle de visée du satellite SPOT), les surfaces situées sur un flanc Est *"redressées"* (influence opposée de la pente et de l'angle de visée) et les surfaces situées sur un espace plan *sur-estimées*

12.4 : Conclusion

En conclusion, rappelons rapidement les principales caractéristiques du plan de sondage appliqué à l'enquête finale "Démographie/migrations".

1. La base de sondage est l'image satellite SPOT stratifiée en cinq niveaux de densité du bâti, les zones non bâties sont exclues du tirage.

2. Tirage au premier degré : les unités primaires sont les pâtés de maisons, la sélection est aréolaire, systématique à l'aide d'une grille de points à maille variable respectant la règle d'allocation fixée. Le placement initial de la grille sur l'image est aléatoire, sous la contrainte de respect des taux. Le tirage fournit un échantillon de 432 îlots dont 426 pourront être réellement enquêtés.

3. Tirage au deuxième degré : les unités secondaires sont les ménages, la sélection est là encore systématique; elle s'effectue sur liste après recensement de tous les ménages vivants dans l'îlot. Le taux au second degré est variable suivant le nombre de ménages vivant dans l'îlot :

- moins de dix ménages dans l'îlot : enquête exhaustive
- plus de dix ménages dans l'îlot : dix ménages sont enquêtés

L'échantillon de ménages enquêtés comprend finalement 3161 ménages, ce qui correspond à un taux d'environ 1,3 % des ménages.

4. Les estimateurs employés :

- estimateur sans biais sous plan à probabilités proportionnelles aux surfaces des unités primaires pour l'estimation des totaux.
- estimateur du ratio sous plan à probabilités proportionnelles aux surfaces des unités primaires pour l'estimation des quotients.

Les opérations de collecte, commencées le lundi 23 novembre 87 se sont poursuivies jusqu'à la fin décembre 87. La phase de codification et de saisie des résultats s'est achevée à la fin du mois d'avril 88. Dans la quatrième et dernière partie de cet exposé, nous allons présenter les résultats de la première exploitation des données recueillies lors de cette enquête. Cette exploitation, qui fait l'objet du premier chapitre, est bien sûr très partielle ; son seul objectif est en fait d'évaluer "définitivement" *la précision des estimations que permet d'obtenir la méthode mise au point.*

QUATRIEME PARTIE
EVALUATION DES PERFORMANCES ET DES CONDITIONS D'APPLICATION
DE LA METHODE

INTRODUCTION

Dans cette dernière partie, nous nous proposons, après la première application réalisée à Quito, de donner les principaux éléments qui permettent d'évaluer "a posteriori" les performances de la méthode qui a été mise au point durant cette recherche à l'ORSTOM. Le terme un peu vague de "performances" doit être précisé. Si on résume très rapidement l'objectif poursuivi tout au long de ce travail, on peut dire qu'il s'est agi de mettre au point une méthode de collecte d'informations démographiques par sondage précise, peu chère et adaptée au milieu urbain des pays en développement. Les performances de la méthode doivent donc être évaluées de ces trois points de vue : *précision des estimations, domaine et conditions d'application de la méthode et enfin coûts et délais nécessaires à sa mise en œuvre*. Nous aborderons successivement ces trois points.

Le chapitre 13 présente les premiers résultats de l'exploitation de l'enquête Démographie / migrations. Après avoir décrit les principes et les outils du calcul des estimations, les résultats sont donnés en même temps que la mesure de l'erreur d'échantillonnage qui leur est associée. C'est cette mesure qui permet d'aboutir aux conclusions sur l'efficacité réelle de la méthode en terme de précision. On aborde enfin un deuxième type d'évaluation de la précision, avec un rapide examen des cohérences existant entre les résultats de l'enquête et les chiffres du recensement de 1982.

Dans le chapitre 14, on tente de définir les domaines, géographiques et thématiques, auxquels peut s'appliquer la méthode, et les conditions nécessaires à cette application : contraintes liées à l'emploi de l'image satellite et aux étapes de définition et de mise en œuvre du plan de sondage. On s'efforcera également dans ce chapitre, de faire apparaître les différentes possibilités d'adaptation de la méthode à des domaines ou à des conditions différents de ceux sur lesquels elle a été développée.

Enfin, le dernier chapitre aborde les délais et les coûts que nécessitent la définition du plan de sondage, puis les phases de collecte, de codification et de saisie des données, bref la mise en œuvre de la méthode. A partir de ces éléments, on essaiera de définir les fonctions de coût et de précision qui caractérisent cette méthode.

Au terme de ces exposés, le lecteur disposera des éléments lui permettant d'apprécier les qualités de précision, de souplesse et d'économie que possède la technique que nous proposons, et il sera en mesure, pensons-nous, de l'utiliser sur d'autres sites.

Chapitre 13

PRECISION DE LA METHODE : LES PREMIERS RESULTATS DE L'ENQUETE REALISEE A QUITO

Introduction

Comme nous l'avons vu au chapitre 12, l'enquête Démographie / Migration réalisée à Quito porte sur un échantillon relativement léger : 426 pâtés de maisons où 3157 ménages¹ ont été enquêtés soit 14 553 personnes, ce qui représente d'après les estimations issues de l'enquête 1,3% de la population totale. Cependant, la quantité d'information recueillie grâce aux cinq fiches du questionnaire est, elle, très importante : par exemple le fichier concernant tous les individus couverts par l'enquête, qui correspond au seul dépouillement de la fiche 1, contient près de 50 variables. Il était donc exclu, dans le cadre de cette thèse, d'effectuer et de présenter une exploitation exhaustive de ces résultats. De plus, évaluer à partir des résultats de l'enquête la précision de la méthode de sondage mise en œuvre ne nécessite pas de disposer de chiffres de précision concernant toutes les variables du questionnaire. A la date où nous rédigeons ces lignes, la totalité de l'information correspondant au questionnaire est codifiée et saisie ; cependant l'important travail de contrôle de cohérence et de correction des données nécessaire pour que toutes les variables soient exploitables est, lui, loin d'être terminé². Actuellement, 15 variables sont disponibles pour effectuer des estimations. Elles correspondent à un choix d'indicateurs socio-démographiques classiques qui couvre trois mesures d'effectifs de populations, deux moyennes par unité secondaire (ménages) et dix quotients illustrants les caractéristiques des ménages et de la population (voir tableau 13.3).

Pour évaluer, à partir de ces 15 indicateurs, la précision des estimations que permet la méthode que nous proposons, nous résumons tout d'abord dans une première partie les principes de la démarche d'inférence que nous avons appliquée aux résultats de l'enquête et qui conduisent aux programmes informatiques fournissant les estimations. Dans une seconde partie, nous exposons et commentons les résultats des calculs effectués : estimations et précision de ces estimations. La troisième partie est consacrée à un parallèle entre les résultats de l'enquête que nous avons réalisée en 1987 et de ceux du recensement de la population effectué par l'INEC³ en 1982. L'ensemble de ces indications chiffrées nous permettront, en conclusion, de statuer sur les performances de précision de la méthode.

¹ : Ce chiffre correspond aux ménages "ordinaires" c'est à dire exclusion faite des ménages collectifs présents dans l'échantillon (12 ménages). Dans la suite de ce chapitre, toutes les estimations effectuées se rapporte à la population hors ménages collectifs.

² : Ces étapes de codification, saisie et contrôle de l'information sont résumées par F. DUREAU dans l'annexe A.3.2

³ : Instituto Nacional de Estadísticas y Censos : institut national de statistiques équatorien.

13.1 : Démarche d'inférence : les programmes d'estimation

Les principes de la démarche d'inférence attachée à la structure du sondage appliqué à Quito ont été détaillés, tout au long des deuxième et troisième parties de cet exposé. Le formulaire général utilisé ici est celui résumé dans les tableaux 12.1 à 12.3 du chapitre précédent, sauf pour les estimations de totaux, où nous l'avons complété en intégrant dans l'estimateur de la variance le terme correspondant au second degré du sondage. On trouvera ce complément de formulaire dans le paragraphe 13.1.1.

Avant de présenter les programmes de calcul des estimations et de leurs précisions (§ 13.1.3), nous allons examiner, dans les paragraphes 13.1.1 et 13.1.2, les deux points particuliers suivants, liés à la spécificité du plan de sondage employé :

1. Le sondage est à deux degrés, mais nous avons vu dans le chapitre 12 que l'on peut utiliser les estimateurs de la variance donnés dans le tableau 12.3 (ceux que l'on obtient en négligeant la variance intra-unités primaires). L'emploi de ces estimateurs a déjà été justifié, au chapitre 11, par les résultats de l'exploitation de l'enquête préliminaire Bâti / Population. Cependant il peut être intéressant, pour vérifier à partir d'un échantillon moins restreint que la deuxième composante de la variance est bien négligeable, *d'estimer séparément la variance due au premier et au second degrés du sondage*. C'est ce que nous avons choisi de faire, au moins dans le programme d'estimation des totaux que nous présenterons ci-après.

2. La sélection de l'échantillon de pâtés de maisons est systématique, effectuée à l'aide d'une grille de points. Nous avons abordé plusieurs fois dans les chapitres précédents¹ la question de l'estimation de la variance pour un tel plan de sondage, et nous nous sommes finalement arrêté au choix d'une méthode d'approximation par assimilation à un sondage stratifié sur un critère de localisation. C'est donc à partir du croisement d'une telle stratification et de la stratification basée sur la morphologie des quartiers, que s'effectuera l'estimation de l'erreur d'échantillonnage associée aux résultats produits.

13.1.1 : Estimation séparée des deux termes de la variance (premier et second degrés)

1. Expression de l'estimateur de la variance totale

L.M. ASSELIN² énonce ainsi le principe de construction de l'estimateur de la variance dans le cas d'une structure de sondage à deux degrés, en notant $\hat{V}_1(\hat{Y}_1)$ l'estimateur de la variance associé au seul premier degré du sondage³ et $\hat{V}_2(\hat{y}_i)$ l'estimateur de la variance du total \hat{y}_i estimé dans l'U.P. i ⁴ :

- " 1. substituer dans $\hat{V}_1(\hat{Y}_1)$ les estimateurs \hat{y}_i aux vraies valeurs y_i ;
2. substituer, dans l'estimateur \hat{Y}_1 relatif au premier degré de sondage, les variances estimées $\hat{V}_2(\hat{y}_i)$ aux valeurs y_i ;
3. additionner les deux termes ainsi obtenus."

¹ : Voir les chapitres 3 (§ 3.3) , 8 et 12 (§ 12.3.1).

² : L.M. ASSELIN : Techniques de sondages avec applications à l'Afrique, p. 109. Pour une démonstration voir A. DES RAJ : Sampling theory (1968) pp. 215, 216.

³ : Ces estimateurs sont donnés dans le tableau 12.3.

⁴ : Ces estimateurs sont donnés dans le tableau 12.2.

Pour l'estimation du total de Y dans la strate k, avec les notations adoptées au chapitre 12, on a :

- au deuxième degré (sondage équiprobable sans remise, tableau 12.2)

$$\hat{y}_i = N \sum_{j=1}^n \frac{y_j}{n} \quad \text{et} \quad \hat{V}(\hat{y}_i) = \frac{N(N-n)}{n} \sum_{j=1}^n \frac{(y_j - \hat{y}_i)^2}{(n-1)};$$

- au premier degré (approximation avec remise, tableau 12.3)

$$\hat{Y}_k = \frac{z_k}{m} \sum_{i=1}^m \frac{\hat{y}_i}{z_i} \quad \text{et} \quad \hat{V}(\hat{Y}_k) = \frac{1}{m(m-1)} \sum_{i=1}^m \left(\frac{z_k \hat{y}_i}{z_i} - \hat{Y}_k \right)^2.$$

En appliquant les règles 2 et 3 formulées ci-dessus (la première substitution étant déjà faite), on obtient l'expression de l'estimateur de la variance de \hat{Y}_k :

$$\hat{V}(\hat{Y}_k) = \frac{1}{m(m-1)} \sum_{i=1}^m \left(\frac{z_k \hat{y}_i}{z_i} - \hat{Y}_k \right)^2 + \frac{z_k}{m} \sum_{i=1}^m \frac{\hat{V}(\hat{y}_i)}{z_i} \quad (13.1)$$

2. Estimations séparées des deux termes de la variance

On peut montrer que le second terme de l'estimateur ci dessus n'est pas un estimateur sans biais du second terme de la variance, et il en est évidemment de même pour les premiers termes¹. Le second terme, qui représente la variance due au second degré du sondage est estimé de manière non biaisée par² :

$$\hat{V}_k \text{ Intra} = \frac{z_k^2}{m^2} \sum_{i=1}^m \frac{\hat{V}(\hat{y}_i)}{z_i^2} \quad (13.2).$$

On obtient ensuite l'estimateur sans biais du premier terme par différence avec l'estimateur de la variance totale :

$$\hat{V}_k \text{ Inter} = \hat{V}(\hat{Y}_k) - \hat{V}_k \text{ Intra} \quad (13.3).$$

Ce sont les estimateurs donnés par les formules 13.1 à 13.3 que nous utiliserons dans le programme d'estimation de totaux. Pour les moyennes et quotients nous emploierons le formulaire plus simple des tableaux 12.2 et 12.3. Nous verrons dans la deuxième partie de ce chapitre, grâce à l'estimation séparée des premier et second termes de la variance, que lors du sondage réalisé à Quito l'effet de la variance intra-unité primaire sur l'erreur d'échantillonnage totale était bien négligeable.

¹ : Voir par exemple DES RAJ, op. cit., pp. 118, 119.

² : Voir par exemple DESABIE, Théorie et pratique des sondages (1966), pp. 300.

13.1.2 : Approximation de la variance du sondage systématique

La méthode (rappel)

La méthode d'approximation de l'erreur d'échantillonnage que nous allons appliquer au sondage systématique pratiqué à Quito est celle que nous avons mise au point au chapitre 8 lors de l'expérience réalisée sur les données de Marseille. Son principe ayant été exposé en détail au paragraphe 8.1 auquel on peut se reporter, nous n'y reviendrons pas ici. Cependant il faut faire une remarque à propos de l'application de cette méthode au sondage réalisé à Quito.

On se souvient que lors de la mise au point de cette méthode, une difficulté était apparue concernant le calcul des taux spécifiques à chaque strate résultant du croisement des deux stratifications, calcul que le programme VARECH ne pouvait pas effectuer. Du fait de cette difficulté, nous avons dû opter pour un calcul ne correspondant pas exactement au problème posé : celui de l'erreur d'échantillonnage qui aurait caractérisé un sondage aréolaire à taux constant sur l'ensemble de la base¹. Ce problème ne se pose plus dans le cas de l'application de la méthode à des données recueillies par sondage. En effet dans ce cas, on connaît à posteriori le nombre d'unités primaires (ilots) enquêtées dans chaque strate de la stratification "croisée", ainsi que la surface de chaque îlot et les surfaces totales des strates. Ces éléments étant les seuls nécessaires au calcul des divers estimateurs, il est inutile de chercher à connaître le taux, effectivement variable, pratiqué dans chaque strate.

Le seul travail nécessaire à l'application de la méthode est donc l'affectation de chaque unité primaire de l'échantillon à l'une des strates définies par le croisement des stratifications basées respectivement sur la morphologie des quartiers et sur leur localisation au sein de l'agglomération.

La stratification "croisée"

Nous avons décrit la stratification morphologique employée pour l'enquête finale au chapitre 12 ; rappelons qu'elle est constituée de cinq strates de densité du bâti et d'une strate non-bâtie, exclue de la base de sondage. Par ailleurs au chapitre 11, pour analyser la liaison existant entre les caractéristiques démographiques des ménages et leur localisation au sein de la ville, nous avons effectué un zonage de l'agglomération en 12 sous-ensembles géographiques, issus du croisement d'une division Nord/Sud en 4 zones avec une division Est/Ouest en 3 zones². C'est ce même zonage que nous avons utilisé pour prendre en compte, dans l'exploitation des résultats de l'enquête finale, la localisation des îlots de l'échantillon dans l'agglomération.

Dans la base de sondage retenue, le croisement des deux critères (morphologie en cinq modalités et localisation en 12 modalités) devrait, en théorie, produire une stratification en 60 classes. En réalité, les strates morphologiques les plus densément bâties (strates 4 et 5) étant concentrées dans la partie centrale de Quito, 7 modalités du critère résultant, parmi les 60 possibles, n'existent pas. D'autre part, pour conserver dans chaque strate un nombre d'unités primaires qui permette d'estimer la variance des estimations avec une précision acceptable, nous avons effectué certains regroupements à l'issue desquels nous retenons 29 strates "croisées" ; les tableaux 13.1 et 13.2 présentent cette stratification résultante.

13.1.3 : Les programmes d'estimation

Pour appliquer la démarche d'inférence, que nous avons décrite dans les paragraphes précédents, aux données collectées lors de l'enquête Démographie / Migration, il

¹ : Voir à ce propos le paragraphe 8.1.

² : A propos de cette classification des quartiers en fonction de leur localisation dans l'agglomération, voir le paragraphe 11.2.1.2

Tableau 13.1 : Croisement des stratifications sur les critères de densité du bâti et de localisation

(la case courante du tableau indique le numéro de la strate de la stratification "croisée")

Localisation	-1- Sud Ouest	-2- Sud	-3- Sud Est	-4- Centre Sud Ouest	-5- Centre Sud	-6- Centre Sud Est	-7- Centre Nord Ouest	-8- Centre Nord	-9- Centre Nord Est	-10- Nord Ouest	-11- Nord	-12- Nord Est
Densité												
-1- Densité très faible	1	2	3	1	2	3	4	4	4	4	4	4
-2- Densité faible	5	6	7	5	6	7	8	8	8	8	8	8
-3- Densité moyenne	9	10	11	12	10	13	14	15	16	14	17	18
-4- Densité forte	19	20	21	19	22	21	...	23	24	25	26	27
-5- Densité très forte	28	28	28	29	29	29

... : Modalités inexistantes

Tableau 13.2 : la stratification résultante

Strate	1	2	3	4	5	6	7	8
Densité du bâti	Très faible	Très faible	Très faible	Très faible	Faible	Faible	Faible	Faible
Localisation	Sud ouest	Sud	Sud est	Nord	Sud ouest	Sud	Sud est	Nord
Nb. d'U.P.	7	7	5	9	6	4	6	10
Surface (Ha)	1 094	534	616	1 100	591	571	636	1 310

Tableau 13.2 : la stratification résultante (suite)

Strate	9	10	11	12	13	14	15	16
Densité du bâti	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne
Localisation	Sud sudouest	Sud	Sud sud est	Sud ouest	Sud est	Nord ouest	Nord	Nord est
Nb. d'U.P.	5	14	7	8	16	4	18	11
Surface (Ha)	218	421	212	257	483	170	533	276

Strate	17	18	19	20	21	22	23	24
Densité du bâti	Moyenne	Moyenne	Forte	Forte	Forte	Forte	Forte	Forte
Localisation	Nord nord	Nord nord est	Sud ouest	Sud sud	Sud est	Sud	Nord	Nord est
Nb. d'U.P.	14	7	14	12	19	41	36	9
Surface (Ha)	394	203	157	153	326	561	477	128

Strate	25	26	27	28	29	Total
Densité du bâti	Forte	Forte	Forte	Très Forte	Très Forte	...
Localisation	Nord nord ouest	Nord nord	Nord est	Sud	Nord	...
Nb. d'U.P.	4	82	13	25	13	426
Surface (Ha)	53	1 153	200	651	386	13 864 ^[1]

¹ : La différence de 411 hectares existant entre ce chiffre et celui donné au chapitre 12 correspond aux surfaces occupées par les coopératives "Pisuli" et "Jaime Roldos" que nous n'avons pas pu enquêter.

fallait construire les programmes informatiques correspondants. Le logiciel SPSS/PC + contient un ensemble de fonctions¹ bien adaptées aux calculs qu'il est nécessaire d'effectuer sur les différents fichiers : fichier d'unités secondaires (ménages), fichier d'unités primaires (ilots) et fichiers de strates. C'est pourquoi nous avons écrit les programmes d'estimation sous ce logiciel ; le prix à payer pour les facilités qu'il offre est l'absence d'interactivité, donc de souplesse, qui caractérise son mode de fonctionnement "en batch".

Les trois programmes écrits (estimations de totaux, de moyennes et de quotients) se présentent donc comme des ensembles de commandes SPSS, organisées en procédures qui correspondent aux différentes étapes des calculs à réaliser. Le chaînage des procédures permet d'obtenir les estimations par strate et sur le total de la base, et de calculer les estimateurs des variances associées à ces estimations. La mesure de l'erreur d'échantillonnage que fournissent les programmes est le coefficient de variation de l'estimateur, exprimé en pourcentage de la quantité à estimer. Pour être soumises aux programmes, les données issues de l'enquête doivent être organisée, nous l'avons dit, en trois fichiers². Le tableau ci-dessous récapitule les informations que doivent contenir chacun de ces trois fichiers.

FICHIERS		
Unités secondaires	Unités primaires	Strates
<p>1. Identification de l'U.P. à laquelle appartient l'U.S</p> <p>2. Variables à estimer</p>	<p>1. Identification de l'U.P. (identique à celle utilisée dans le fichier U.S.</p> <p>2. Nombre d'U.S. recensées dans l'U.P.</p> <p>3. Surface de l'U.P.</p> <p>4. Identification de la strate à laquelle appartient l'U.P.</p>	<p>1. Identification de la strate (identique à celle utilisée dans le fichier strates)</p> <p>2. Surface totale de la strate</p>

Les trois programmes fournissent deux types de sortie : un premier tableau donne les estimations obtenues pour chaque unité primaire enquêtée, un second donne l'estimation, la variance et le coefficient de variation estimés pour chaque strate de la stratification croisée, puis pour chaque strate morphologique et enfin pour le total de la base. Dans le programme d'estimation des totaux on estime séparément, comme nous l'avons dit, les variances dues aux deux degrés du sondage; au contraire dans les programmes d'estimation de moyennes et de quotients, seule la variance totale est estimée. Rappelons enfin que les estimations de totaux sont sans biais alors que pour

¹ : Les deux fonctions les plus utiles sont AGGREGATE (agglomération de cas au travers d'un fichier partitionné par une variable de rupture) et JOIN (merge et concaténation de deux ou plusieurs fichiers). En outre on dispose dans ce logiciel d'un éventail d'opérateurs mathématiques déjà programmés, ce qui facilite d'autant le travail de programmation.

² : Ces fichiers devront être construits à partir de la base de données contenant les résultats de l'enquête, base de données qui possède son organisation propre. Au sujet des différents fichiers constituant cette base et de leur structures, on se reportera à l'annexe A.3.2.

celles des quotients et des moyennes¹, on emploie l'estimateur par le ratio légèrement biaisé.

Le listage complet des trois programmes est donné dans l'annexe A.4.1. Nous allons maintenant examiner les résultats que ces programmes ont permis d'obtenir.

13.2 : Estimation des caractères démographiques

Comme nous le signalions dans l'introduction, cette première exploitation des résultats de l'enquête Démographie / Migration, porte sur 15 caractères démographiques. Le tableau 13.3 résume leurs définitions. Nous ne ferons pas ici l'analyse démographique de ces premiers résultats bien que ce soit également un moyen d'apprécier leur cohérence et donc indirectement les performances de cette méthode d'enquête ; la dernière partie de ce chapitre fournira à ce sujet quelques éléments. Pour le moment, à partir des seules estimations des erreurs d'échantillonnage fournies par les programmes, nous voudrions examiner quatre questions :

- à partir des indications sur la précision des estimations "globales" (pour l'ensemble de la base), peut-on conclure que la méthode que nous proposons est concurrentielle, face aux méthodes de sondage plus traditionnelles?
- la précision des estimations "locales" (estimations par strate) est-elle suffisante pour permettre de rendre compte des variations des caractères démographiques à l'intérieur de l'agglomération?
- le choix du sondage à deux degrés, effectué après l'analyse des résultats de l'enquête démographique préliminaire, est-il judicieux? comment la variance totale des estimations est-elle répartie entre les deux degrés du sondage?
- enfin, quelle est la précision des estimations au sein des unités primaires de l'échantillon, c'est à dire pour chaque îlot enquêté? l'information qu'elles fournissent est-elle exploitable?

13.2.1 : Précision des estimations globales

Le tableau 13.4 donne, pour les 15 caractères démographiques considérés ici, dans chaque strate morphologique et pour l'ensemble de la base, les estimations obtenues, ainsi que les valeurs estimées du coefficient de variation et de l'intervalle de confiance à 95% attachés à chaque estimation. D'une certaine façon, ces chiffres parlent d'eux-mêmes ; on peut cependant faire deux remarques générales à leur sujet. Tout d'abord, rappelons que la méthode d'estimation des variances adoptée dans les programmes, du fait de son caractère approximatif, surestime nécessairement les variances propres au sondage systématique employé. De plus, les coefficients de variation et les intervalles de confiance du tableau 13.4, doivent bien sûr être considérés relativement à deux facteurs : la taille de l'échantillon enquêté² et la fréquence de l'événement observé sous-jacent à chacun des indicateurs démographiques retenus ; il est clair que l'on n'estimera pas avec la même précision relative les effectifs (ou les pourcentages de population) de catégories d'importances numériques très différentes, telles que le nombre de femmes et la population de 60 ans

¹ : Par "moyenne" on entend ici la moyenne du caractère par unité statistique (par ménage), c'est à dire le quotient de l'estimation du total du caractère sur la base de sondage, par l'estimation du nombre de ménages composant cette base. Ces quantités sont aléatoires tant au numérateur qu'au dénominateur; il s'agit donc en fait d'un cas particulier d'estimation par le ratio.

² : L'influence de ce facteur sur la précision des estimations et le coût de mise en œuvre de la méthode sera examiné en détail au chapitre 15.

Tableau 13.3 : Liste et définitions des caractères démographiques estimés

Caractères	Définitions
1) Effectifs	
1. POPTOT	Effectif de population hors ménages collectifs
2. POPFEM	Effectif de population féminine (hors mén. coll.)
3. POPHOM	Effectif de population masculine (hors mén. coll.)
2) Moyennes	
4. POPMOY	Taille moyenne des ménages (hors mén. coll.)
5. AGEMOY	Age moyen des chefs de ménage (hors mén. coll.)
3) Quotients	
6. PCMEN1	Pourcentage de ménages d'une personne (hors mén. coll.)
7. PCMEN7	Pourcentage de ménages de sept personnes et plus (hors mén. coll.)
8. PCMENP	Pourcentage de ménages propriétaires de leur logement (hors mén. coll.)
9. SEXRAT	Sexe ratio : Nombre d'hommes / Nombre de femmes (hors mén. coll.)
10. PCPM5A	Pourcentage de population de moins de cinq ans (hors mén. coll.)
11. PCPP60	Pourcentage de population de soixante ans ou plus (hors mén. coll.)
12. PCPNEQ	Pourcentage de population née à Quito (hors mén. coll.)
13. PCPNEE	Pourcentage de population née hors d'Equateur (hors mén. coll.)
14. PCP6NS	Pourcentage de population de six ans ou plus non scolarisée (hors mén. coll.)
15. PCP6ES	Pourcentage de population de six ans ou plus ayant fait des études supérieures (hors mén. coll.)

Tableau 13.4 : Estimations des caractères démographiques à Quito et leur précision

Caractères	-1- <u>POPTOT</u>	-2- <u>POPPEM</u>	-3- <u>POPHOM</u>	-4- <u>POPMOY</u>	-5- <u>AGEMOY</u>
Estimations					
<u>1. Densité très faible</u>					
Estimation :	101 009	49 544	51 465	4,47	42,36
Coeff. de variation :	57,38 %	60,73 %	54,39 %	1,03 %	2,57 %
Interv. de conf. à 95% :	+/- 115 918	+/- 60 176	+/- 55 984	+/- 0,092	+/- 2,18
<u>2. Densité faible</u>					
Estimation :	78 065	38 980	39 085	4,84	40,08
Coeff. de variation :	27,09 %	27,30 %	27,49 %	4,36 %	2,76 %
Interv. de conf. à 95% :	+/- 42 296	+/- 21 283	+/- 21 489	+/- 0,422	+/- 2,21
<u>3. Densité moyenne</u>					
Estimation :	252 233	127 881	124 351	4,55	44,64
Coeff. de variation :	9,62 %	9,96 %	9,56 %	2,33 %	3,37 %
Interv. de conf. à 95% :	+/- 48 530	+/- 25 474	+/- 23 776	+/- 0,212	+/- 3,01
<u>4. Densité forte</u>					
Estimation :	475 951	245 798	230 153	4,71	44,83
Coeff. de variation :	5,88 %	5,98 %	5,95 %	1,19 %	0,95 %
Interv. de conf. à 95% :	+/- 55 972	+/- 29 397	+/- 27 388	+/- 0,112	+/- 0,85
<u>5. Densité très forte</u>					
Estimation :	213 733	110 013	103 720	4,12	43,47
Coeff. de variation :	10,63 %	10,80 %	10,82 %	4,01 %	2,23 %
Interv. de conf. à 95% :	+/- 45 440	+/- 23 763	+/- 22 445	+/- 0,330	+/- 1,94
<u>Ensemble de la base</u>					
Estimation :	1 120 991	572 216	548 775	4,535	43,96
Coeff. de variation :	6,73 %	6,85 %	6,70 %	1,10 %	1,02 %
Interv. de conf. à 95% :	+/- 150 885	+/- 78 394	+/- 73 536	+/- 0,10	+/- 0,90

Tableau 13.4 : Estimations des caractères démographiques à Quito et leur précision (suite)

Caractères	-6- <u>PCMEN1</u>	-7- <u>PCMEN7</u>	-8- <u>PCMENP</u>	-9- <u>SEX RAT</u>	-10- <u>PCPM5A</u>
Estimations					
<u>1. Densité très faible</u>					
Estimation :	6,18	10,72	18,76	1,039	8,65
Coeff. de variation :	5,98 %	2,56 %	25,94 %	5,06 %	6,69 %
Interv. de conf. à 95% :	+/- 0,74	+/- 0,55	+/- 9,75	+/- 0,105	+/- 1,16
<u>2. Densité faible</u>					
Estimation :	4,12	11,60	53,00	1,003	13,05
Coeff. de variation :	31,56 %	16,87 %	9,43 %	7,31 %	10,26 %
Interv. de conf. à 95% :	+/- 2,60	+/- 3,91	+/- 10,00	+/- 0,147	+/- 2,68
<u>3. Densité moyenne</u>					
Estimation :	4,92	11,11	48,77	0,972	10,89
Coeff. de variation :	20,46 %	13,62 %	4,51 %	3,18 %	5,83 %
Interv. de conf. à 95% :	+/- 2,01	+/- 3,03	+/- 4,40	+/- 0,062	+/- 1,27
<u>4. Densité forte</u>					
Estimation :	2,33	10,09	55,98	0,936	8,25
Coeff. de variation :	17,66 %	9,77 %	3,06 %	1,95 %	4,41 %
Interv. de conf. à 95% :	+/- 0,82	+/- 1,97	+/- 3,43	+/- 0,037	+/- 0,73
<u>5. Densité très forte</u>					
Estimation :	8,61	7,04	33,36	0,943	9,02
Coeff. de variation :	26,42 %	26,40 %	7,53 %	3,96 %	7,31 %
Interv. de conf. à 95% :	+/- 4,55	+/- 3,72	+/- 5,02	+/- 0,077	+/- 1,32
<u>Ensemble de la base</u>					
Estimation :	4,70	9,84	46,01	0,959	9,36
Coeff. de variation :	11,96 %	6,80 %	2,49 %	1,50 %	2,86 %
Interv. de conf. à 95% :	+/- 1,12	+/- 1,34	+/- 2,29	+/- 0,029	+/- 0,535

Tableau 13.4 : Estimations des caractères démographiques à Quito et leur précision (fin)

Caractères	-11- <u>PCPP60</u>	-12- <u>PCPNEO</u>	-13- <u>PCPNEE</u>	-14- <u>PCP6NS</u>	-15- <u>PCP6ES</u>
Estimations					
<u>1. Densité très faible</u>					
Estimation :	3,02	68,53	0,53	2,79	8,89
Coeff. de variation :	10,38 %	4,34 %	106,87 %	38,80 %	7,28 %
Interv. de conf. à 95% :	+/- 0,63	+/- 5,95	+/- 1,13	+/- 2,16	+/- 1,29
<u>2. Densité faible</u>					
Estimation :	3,59	63,99	1,52	8,76	3,02
Coeff. de variation :	24,67 %	6,34 %	56,59 %	18,13 %	50,14 %
Interv. de conf. à 95% :	+/- 1,77	+/- 8,11	+/- 1,72	+/- 3,18	+/- 3,03
<u>3. Densité moyenne</u>					
Estimation :	5,28	61,57	2,78	3,19	15,46
Coeff. de variation :	15,60 %	2,69 %	36,25 %	13,41 %	8,01 %
Interv. de conf. à 95% :	+/- 1,65	+/- 3,31	+/- 2,02	+/- 0,86	+/- 2,48
<u>4. Densité forte</u>					
Estimation :	6,31	62,38	2,27	1,59	22,91
Coeff. de variation :	6,04 %	1,75 %	13,93 %	12,28 %	4,30 %
Interv. de conf. à 95% :	+/- 0,76	+/- 1,09	+/- 0,63	+/- 0,39	+/- 1,97
<u>5. Densité très forte</u>					
Estimation :	7,27	64,26	1,04	2,13	20,89
Coeff. de variation :	12,92 %	4,09 %	37,75 %	20,03 %	10,77 %
Interv. de conf. à 95% :	+/- 1,88	+/- 5,26	+/- 0,78	+/- 0,85	+/- 4,50
<u>Ensemble de la base</u>					
Estimation :	5,77	63,22	1,94	2,64	18,26
Coeff. de variation :	5,40 %	1,38 %	14,69 %	7,94 %	3,68 %
Interv. de conf. à 95% :	+/- 0,62	+/- 1,74	+/- 0,57	+/- 0,42	+/- 1,34

et plus. Taille de l'échantillon et fréquence de l'événement observé sont donc les deux principaux paramètres qu'il faut prendre en compte si l'on veut comparer les précisions du tableau 13.4 à celles que fournissent d'autres méthodes de sondages.

Comparaison au sondage équiprobable sur liste

Probablement à cause de son caractère très répandu et de la simplicité de sa théorie, la méthode du sondage aléatoire équiprobable sur liste est la seule pour laquelle nous disposons, dans la littérature sur les sondages, d'informations relatives à la précision des résultats fournis¹. Comme tout au long de ce travail nous avons évoqué ce type de sondage pour lui comparer la méthode que nous proposons, c'est encore lui qui nous servira ici de référence. La donnée dont nous disposons concernant le sondage équiprobable sur liste est l'intervalle de confiance à 95% qui caractérise cette méthode² lorsqu'elle est appliquée, avec diverses tailles d'échantillons, à l'estimation de taux démographiques (natalité, mortalité, accroissement naturel)³. N'ayant pas pratiqué, à partir de l'enquête démographique effectuée à Quito, l'estimation de ces mêmes taux, nous baserons la comparaison sur deux indicateurs qui présentent des ordres de grandeur comparables :

- le pourcentage de population âgée de 60 ans et plus (PCPP60=5,4%), approximativement comparable au taux de natalité (de l'ordre de 4,5%, dans l'exemple que nous utilisons)
- le pourcentage de population âgée de six ans ou plus non scolarisée (PCP6NS=2,6%), comparable au taux d'accroissement naturel (de l'ordre de 2,5%, dans l'exemple que nous utilisons).

Le tableau ci-après permet de comparer les précisions obtenue par les deux méthodes de sondage sur un échantillon de la taille de celui que nous avons enquêté à Quito (environ 14 500 personnes). Mais avant d'examiner ce tableau, une remarque importante s'impose. Si les indicateurs démographiques comparés ici on bien, deux à deux, des valeurs du même ordre de grandeur, ils n'ont en revanche certainement pas les mêmes variances au sein de la base de sondage. Or ces variances pésent d'un poids déterminant sur la précision des estimations et cela, quelque soit la structure du sondage. Malheureusement nous manquons totalement d'éléments, dans la documentation dont nous disposons à Quito, pour évaluer l'ordre de grandeur des variances des taux de natalité et d'accroissement naturel. Nous ne pouvons donc pas chiffrer l'écart de variance existant entre les caractères retenus, cependant on peut sans risque d'erreur affirmer que la variance du pourcentage de population âgée de soixante ans et plus est probablement légèrement supérieure à celle du taux de natalité et que la variance du pourcentage de population âgée de six ans ou plus non scolarisée est

¹ : On peut déplorer que dans la plupart des expériences où sont mis en œuvre des plans de sondage, quelque'en soit d'ailleurs la structure, les comptes rendus publiés ne comportent que fort peu d'indications (coefficients de variation ou intervalles de confiance) relatives à l'erreur d'échantillonnage. Ces absences rendent impossible toute évaluation rigoureuse de la précision des méthodes employées et par conséquent, interdisent qu'on utilise ces expériences comme références pour comparer entre elles des méthodes de sondage différentes.

² : Il s'agit d'intervalles de confiances "théoriques", c'est à dire qu'ils ne sont pas le résultat d'une enquête réellement effectuée mais de calculs faits à partir de la formule donnant la variance de ce type d'estimations lorsqu'on connaît la taille de l'échantillon et la variance du caractère à estimer ; ces calculs sont du même type que ceux que nous avons effectués dans la deuxième partie à partir du recensement de Marseille. Par ailleurs, la mesure de la précision de la méthode qui est donnée suppose l'existence d'une base de sondage (liste d'individus ou au moins de ménages) complète et à jour. Enfin la source dont sont tirées ces indication ne précise pas si le sondage considéré est stratifié ou non.

³ : Ces données sont tirées du "Manuel des enquêtes démographiques par sondage en Afrique" publié par les Nations-Unies et la CEA en 1974, pp136-137, cité par D. TABUTIN (1984).

certainement très nettement supérieure à celle du taux d'accroissement naturel. La comparaison que nous allons effectuer ici est donc largement biaisée et il nous semble légitime de tenir compte de ce biais dans les conclusions que nous en tirerons.

Tableau de comparaison au sondage équiprobable sur liste

Méthodes	Précision des estimations (intervalles de confiances à 95%)	
Sondage aléatoire équiprobable sur liste :	Taux de natalité (4,5%) I \approx +/- 7.6%	Taux d'accroissement naturel (2,5%) I \approx +/- 16.7%
Sondage aréolaire systématique sur image satellite :	Pourcentage de population de soixante ans et plus (5,4%) I \approx +/- 10.8%	Pourcentage de population de six ans et plus non scolarisée (2,6%) I \approx +/- 15.9%

Compte tenu de la remarque précédente, on peut tirer deux enseignements de ce tableau.

Premièrement, même abstraction faite de toute considération sur les variances intrinsèques des caractères estimés, l'ordre de grandeur des intervalles de confiance est comparable pour les deux méthodes. C'est un résultat important compte tenu des nombreux obstacles qui, comme nous l'avons dit, s'opposent souvent dans les villes des pays en développement, à l'application des méthodes de sondage classiques. Les critères de choix entre les deux méthodes pourrait donc être le coût de mise en œuvre caractérisant chacune d'elles et leur praticabilité respective. A cet égard, pour la méthode que nous proposons, nous verrons dans les chapitres suivants qu'elle peut être adaptée à un grand nombre de contextes différents et que son coût dépend largement des informations disponibles concernant la ville où l'on compte réaliser l'enquête. De façon parallèle, pour les méthodes de sondage sur listes, la praticabilité et le coût de la mise en œuvre dépendront de l'existence d'une base de sondage utilisable, c'est à dire en fait, des facilités dont on disposera sur le site d'application pour actualiser cette base. Dans les cas, comme nous l'avons dit fréquents, où une telle actualisation est impossible ou trop onéreuse, la méthode proposée ici possède donc un avantage décisif.

Deuxièmement, par rapport à la référence choisie ici, les performances de précision de la méthode que nous proposons s'avèrent différentes suivant le caractère démographique estimé. En tenant compte des écarts de variance existant entre les indicateurs démographiques que nous comparons ici, on peut dire que lorsque le caractère démographique à estimer est peu corrélé à l'information morphologique prise en compte par la stratification (on peut penser que c'est le cas du pourcentage de population de 60 ans et plus), le sondage aréolaire sur image satellite serait à peu près équivalent, du point de vue de la précision, au sondage aléatoire équiprobable sur liste. En revanche, et c'est le cas pour le pourcentage de population de six ans ou plus

non scolarisée. *si le caractère estimé est bien corrélé à la morphologie, la méthode que nous proposons est très probablement nettement plus précise qu'un sondage aléatoire sur liste, en tous cas si ce dernier n'est pas stratifié (nous supposons ici que les indications fournies par les Nations Unies concernent un sondage aléatoire sur liste non stratifié).*

Ainsi, à partir de cette première exploitation des résultats de l'enquête réalisée à Quito, nous pensons pouvoir répondre affirmativement à la première question que nous nous posons au début de ce paragraphe : la méthode que nous proposons semble concurrentielle par rapport aux méthodes de sondage plus traditionnelles.

13.2.2 : Estimations par strate et variation des caractères au sein de la base de sondage

L'appréciation, à partir du tableau 13.4, des variations suivant les strates des caractères démographiques estimés et de la précision de ces estimations est malaisée. La figure 13.5 donne une représentation graphique des intervalles de confiance associés à cinq de ces estimations choisies pour représenter les différents cas de figures : "mauvaise" précision des estimations par strate (PCMEN7), précision "moyenne" (POPTOT) et "bonne" précision (PCMENP, PCP6NS et PCP6ES). A l'examen de ces graphiques apparaît une des limites de la méthode que nous avons déjà soulignée lors de l'expérience réalisée à Marseille : à l'échelle de variation des indicateurs démographiques estimés, les amplitudes des intervalles de confiance à 95% sont, du moins pour certaines d'entre elles, trop importantes pour obtenir une image nette des fluctuations de ces indicateurs au sein de l'agglomération. Deux remarques doivent cependant être faites pour nuancer cette affirmation.

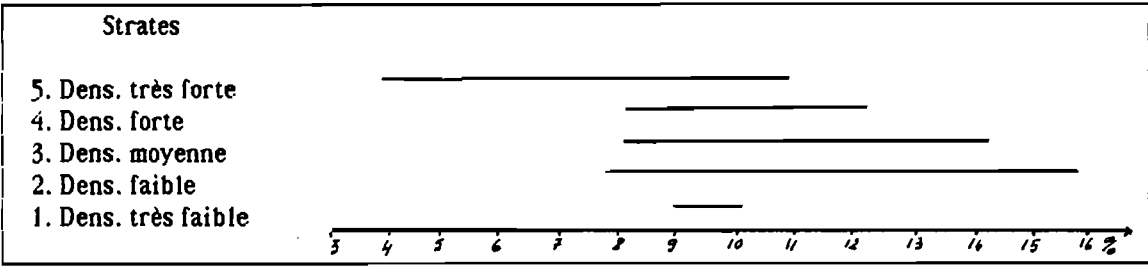
- La stratification de la base sur un critère morphologique (ici la densité du bâti) ne peut être également pertinente pour toutes les variables du questionnaire. Certains indicateurs démographiques construits à partir des résultats de l'enquête peuvent donc ne varier que faiblement d'une strate à l'autre (c'est par exemple le cas du pourcentage de ménages de sept personnes et plus). Dans ce cas, outre que les variations à observer sont quasiment inexistantes, la stratification profite alors peu à la précision des estimations. Il serait donc vain de chercher, en augmentant la taille de l'échantillon ou en perfectionnant la méthode, à améliorer, pour ces caractères, les estimations par strate. En revanche pour d'autres caractères les variations d'une strate morphologique à l'autre sont importantes et ceci nous amène à la seconde remarque.

- Pour les quatre autres indicateurs pris en compte dans la figure 13.5, il existe des variations inter-strates significatives, mais les intervalles de confiance présentent parfois des plages de recouvrement trop grandes ou, plus généralement, ont des amplitudes trop importantes (c'est fréquemment le cas pour les strates de faible densité de bâti, 1 et 2). D'une part ces intervalles de confiance élevés sont en eux-mêmes une information thématiquement intéressante puisqu'ils témoignent d'une grande variabilité des caractères observés dans les quartiers constituant les strates en question. D'autre part, rappelons que lors de l'enquête réalisée à Quito, le budget disponible nous a imposé une taille limitée de l'échantillon. La figure 13.5 montre que, lors d'autres applications de la méthode, un accroissement raisonnable de la taille de l'échantillon, par exemple son doublement (ce qui n'entraînerait, comme nous le verrons dans le chapitre 15, qu'une augmentation limitée du coût), ou même une simple diminution du taux au second degré permettant d'enquêter un plus grand nombre d'ilots, *devrait permettre à ce type d'enquête de rendre compte de façon satisfaisante des variations des caractères démographiques au sein de l'agglomération.*

Figure 13.5 : Intervalles de confiance des estimations par strate

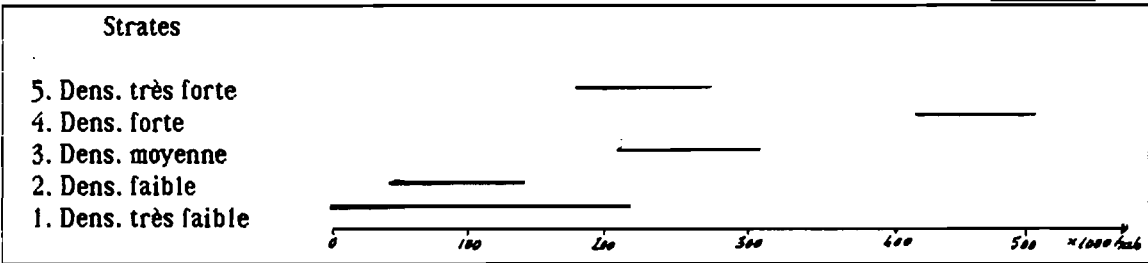
1. Pourcentage de ménages de sept personnes et plus :

PCMEN7



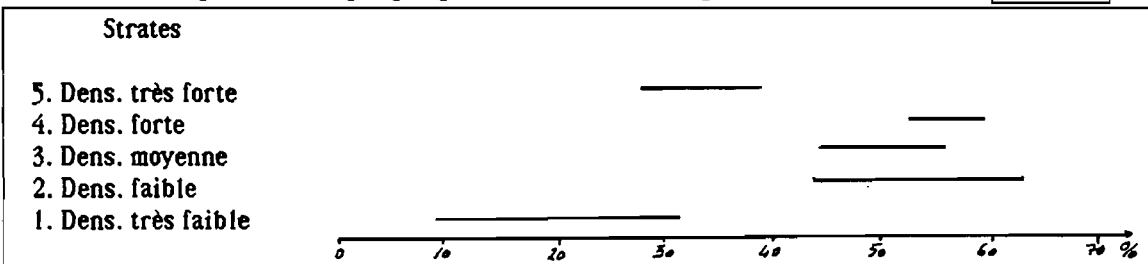
2. Population totale :

POPTOT



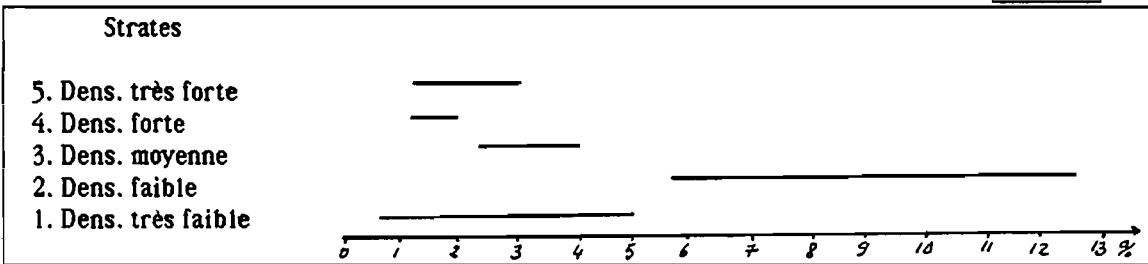
3. Pourcentage de ménages propriétaires de leur logement :

PCMENP



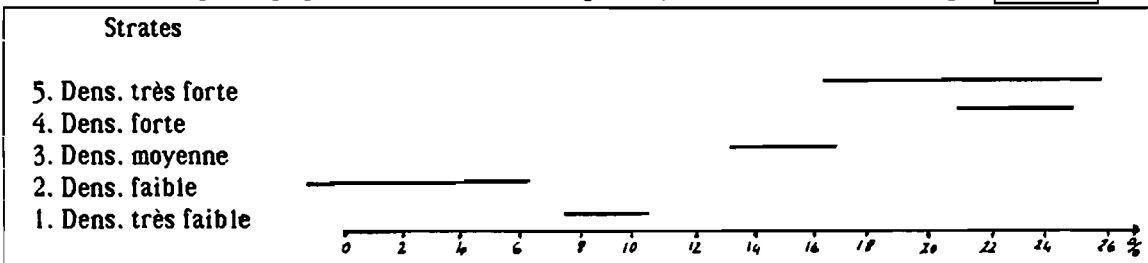
4. Pourcentage de population de six ans et plus non scolarisée :

PCP6NS



5. Pourcentage de population de six ans et plus ayant fait des études sup. :

PCP6ES



13.2.3 : Estimation séparée des deux termes de la variance

Comme nous l'avons dit, c'est seulement pour les totaux que le programme estime séparément les termes de la variance correspondant aux premier et deuxième degrés du sondage. Le tableau ci dessous présente les résultats de cette estimation pour les trois totaux retenus ici :

	Variance inter-unités primaires	Variance intra-unités primaires	Pourcentage <u>variance intra UP</u> / variance totale
POPTOT	5 546,4 x 10 ⁶	148,4 x 10 ⁶	2,61%
POPFEM	1 478,3 x 10 ⁶	59,8 x 10 ⁶	3,89%
POPHOM	1 289,9 x 10 ⁶	64,0 x 10 ⁶	4,73%

La dernière colonne du tableau est éloquente : comme le montraient déjà les résultats de l'enquête préliminaire, la composante intra-unités primaires de la variance est négligeable par rapport à la variance totale. Le choix d'un taux minime au second degré est donc parfaitement justifié, *on peut même préconiser de le réduire encore si l'on veut, avec un même nombre de ménages enquêtés, améliorer les précisions des estimations par strates et sur l'ensemble de la base.*

13.2.4 : Précision des estimations dans les unités primaires de l'échantillon

Produire des estimations fiables pour chacune des unités primaires de l'échantillon n'est évidemment pas le but premier de la méthode de sondage proposée ici, ni même un objectif retenu lors de sa mise au point. Néanmoins les informations que l'enquête permet d'obtenir sur chacun des îlots de l'échantillon peuvent être intéressantes en soi si les estimations au niveau des unités primaires bénéficient d'une précision suffisante. Ainsi, dans un quartier comme le centre historique de Quito, des informations détaillées concernant la population des quinze îlots de l'échantillon situés dans cette zone intéressent vivement le service de planification de la municipalité qui s'attache en ce moment à définir un plan directeur d'aménagement du quartier. A un autre niveau, celui de la recherche, la base de données localisées, qui se constitue actuellement au sein du projet d'atlas informatisé de la ville de Quito, intégrera les résultats de l'enquête que nous avons menée. Si l'exploitation des résultats concernant les îlots de l'échantillon s'avère légitime, la base de données constituée permettra aux chercheurs du projet d'effectuer des croisements d'informations au niveau de l'îlot, etc.

Au paragraphe précédent, nous avons souligné combien la variance à l'intérieur des îlots est petite en comparaison de la variance entre ces mêmes îlots. En bonne logique, les estimations au sein des unités primaires (îlots) ne doivent donc pas être entachées d'erreurs d'échantillonnage trop importantes. Le tableau 13.6 permet de vérifier le bien fondé de cette hypothèse : pour 50 îlots (10 îlots sélectionnés dans chaque strate de manière systématique), il donne le nombre de ménages recensés dans l'îlot, le nombre de ménages enquêtés, l'estimation du total de population de l'îlot (hors ménages collectifs) et le coefficient de variation de cette estimation ; pour chaque strate nous rappelons, comme base de comparaison, le coefficient de variation de l'estimation sur l'ensemble de la strate.

Tableau 13.6 : Estimation du total de population dans 50 ilots de l'échantillon

Strate 1 : densité très faible (ensemble de la strate : cdv = 57,38 %)

N° de l'îlot	405	408	411	414	417	418	422	425	428	431
Nb de ménages	0	5	0	4	0	32	70	3	5	0
Nb de ménages enquêtés	0	4	0	4	0	10	10	3	4	0
Estimation de la population totale	0	22,5	0	13	0	179,2	280	16	25	0
Coeff. de var. de l'estimation (%)	0	8,6	0	0	0	9,4	15,0	0	3,6	0

Strate 2 : densité faible (ensemble de la strate : cdv = 27,09 %)

N° de l'îlot	377	378	380	383	386	389	392	395	398	402
Nb de ménages	19	3	13	10	30	0	25	0	9	0
Nb de ménages enquêtés	10	3	7	8	10	0	10	0	9	0
Estimation de la population totale	95	18	70,6	48,7	153	0	105	0	48	0
Coeff. de var. de l'estimation (%)	10,3	0	11,5	4,4	11,1	0	16,0	0	0	0

Strate 3 : densité moyenne (ensemble de la strate : cdv = 9,62 %)

N° de l'îlot	271	281	292	302	313	323	334	344	355	370
Nb de ménages	21	42	20	16	57	64	5	64	19	18
Nb de ménages enquêtés	10	10	10	10	10	10	4	10	10	10
Estimation de la population totale	128,1	252	88	78,4	330,6	204,8	23,7	390,4	85,5	79,2
Coeff. de var. de l'estimation (%)	8,7	7,8	7,3	5,4	19,9	10,3	7,1	14,1	10,3	8,5

Tableau 13.6 : Estimation du total de population dans 50 ilots de l'échantillon (suite)

Strate 4 : densité forte (ensemble de la strate : cdv = 5,88 %)

N° de l'îlot	45	68	91	114	137	160	183	206	229	252
Nb de ménages	4	51	24	0	2	20	25	45	23	2
Nb de ménages enquêtés	3	10	10	0	2	10	10	10	10	2
Estimation de la population totale	16	255	112,8	0	8	94	105	238,5	82,8	8
Coeff. de var. de l'estimation (%)	14,4	8,5	6,4	0	0	8,7	8,2	9	9,9	0

Strate 5 : densité très forte (ensemble de la strate : cdv = 10,63 %)

N° de l'îlot	2	6	10	14	18	22	26	30	34	38
Nb de ménages	50	13	148	90	56	191	66	142	69	92
Nb de ménages enquêtés	10	8	10	10	10	10	10	10	10	10
Estimation de la population totale	210	56,9	532,8	297	229	687,6	277,2	568	262,2	386,4
Coeff. de var. de l'estimation (%)	7,6	9,6	5,9	16,0	10,1	13,5	11,2	14,4	11,9	8,1

Mis à part dans les flots vides ou enquêtés exhaustivement, pour lesquels il n'y a évidemment pas d'erreur d'échantillonnage, *les coefficients de variation des estimations par flots varient entre 5 et 20%, la majorité se situant entre 7 et 12%*. Il appartient aux urbanistes, géographes et démographes de décider si de telles précisions peuvent satisfaire aux besoins de certaines analyses.

13.3 : Cohérence des estimations avec les résultats du recensement de 1982 : évolution entre 1982 et 1987¹

L'analyse conjointe des estimations issues de l'enquête Démographie/ migrations et de quelques résultats du recensement de 1982 peut également servir à l'évaluation de cette enquête : des biais éventuels au niveau de la sélection de l'échantillon (définition même ou application du plan de sondage) ou de la collecte des données sur le terrain se traduiraient par des résultats incohérents ou en contradiction avec les évolutions constatées par les personnes ayant une solide connaissance de la démographie quiténienne.

Si l'on considère l'effectif total de population de Quito, l'estimation fin 1987 par l'enquête (1 120 991) donne un taux de croissance moyen annuel de 5,26 % sur la période 1982 - 1987 ; d'après les recensements antérieurs, ce taux était de 4,47 % pour les périodes 1950-1962 et 1962-1974, et de 4,70 % pour la période 1974-1982. Le niveau légèrement plus élevé du taux calculé sur la période 1982-1987 peut être mis au compte, mais seulement en partie, des définitions spatiales de la ville adoptées lors de chacune des deux opérations de collecte : le recensement de 1982 a exclu certaines périphéries, habitées, que nous avons incluses lors de l'enquête, par application d'une définition physique de la ville ne tenant aucun compte des limites administratives de paroisse. Néanmoins, ce facteur ne peut à lui seul expliquer le taux de croissance observé sur la période 1982-1987 : on doit conclure à une poursuite de l'accélération du phénomène migratoire en direction de la capitale équatorienne, déjà observée sur les deux périodes intercensitaires précédentes, 1950-1962 et 1962-1974.

Cette observation rejoint celles faites par le CONADE², qui table sur un taux de croissance plus élevé qu'avant 1982 et estime à l'heure actuelle la population quiténienne à 1 135 000 habitants.

Le phénomène d'accroissement de la migration vers Quito, observé à partir des effectifs de population, émerge aussi de la comparaison des sex ratio de 1982 et 1987 (0,929 en 1982, 0,959 en 1987) : les flux migratoires à destination urbaine ayant une dominante masculine, la légère masculinisation de la population quiténienne traduirait ce phénomène.

Si l'on passe maintenant au domaine du statut d'occupation du logement, autre thème important traité par l'enquête, on observe une augmentation sensible du pourcentage de ménages propriétaires de leur logement : en 1982, 36,3 % des ménages étaient propriétaires tandis qu'en 1987 ce pourcentage atteint 46,01 %. Cette évolution correspond à une mutation importante, souvent décrite par les urbanistes et sociologues équatoriens : le transfert d'une partie de la population quiténienne soit en direction des ensembles de logements construits par l'Etat, les banques ou le Municipio, soit vers les terrains périphériques pour construire eux-mêmes leur logement.

Initiés dès les années soixante dix, ces deux phénomènes ont connu une ampleur particulière au cours de la période qui nous intéresse, 1982-1987 : une part non négligeable de la population, aux revenus modestes, est ainsi passée du statut de

¹ Ce paragraphe est entièrement rédigé par F.DUREAU.

² CONADE, 1987. Poblacion y cambios sociales. Diagnóstico Sociodemográfico del Ecuador, 1950-1982. Voir p. 296, Tableau 5: "ECUADOR : PROYECCION DE LA POBLACION TOTAL SEGUN REGIONES Y AREAS URBANAS Y RURALES. PERIODO 1985-2000".

locataire dans la partie centrale de l'agglomération au statut de propriétaire, en accédant aux logements des programmes (principalement localisés dans le sud) ou en construisant leur propre logement sur les marges de la ville.

Cette brève analyse comparée de quelques indicateurs n'a pas prétention à statuer précisément sur la validité du plan de sondage, les paragraphes précédents l'ayant déjà fait ; mais elle montre, à partir de quelques exemples, la *cohérence, interne et par rapport au recensement, des données collectées*. Au vu de la quinzaine d'estimateurs disponibles actuellement, il n'apparaît pas de biais sensible, tant au niveau de l'échantillonnage que de la collecte. Bien évidemment, seule l'exploitation complète de l'information recueillie au cours de l'enquête permettra de conclure plus précisément sur ce point.

Conclusion

Les principaux éléments à retenir de cette première exploitation des résultats de l'enquête Démographie / migrations réalisée à Quito, sont de deux ordres. Méthodologiques d'une part : ils ont trait aux types de calculs qui doivent être effectués pour obtenir les estimations et leur précision. D'autre part ils concernent bien entendu l'évaluation de la précision de la méthode de sondage que cette exploitation a permis de réaliser.

1) S'agissant de la méthode d'inférence qu'il convient d'appliquer lors de l'exploitation des résultats de ce type de sondage, on peut formuler deux conclusions :

- Puisque le terme intra-unités primaires des variances est négligeable, on peut utiliser pour le calcul des estimateurs le formulaire limité au premier degré du sondage donné dans le tableau 12.3.

- La méthode d'évaluation de l'erreur d'échantillonnage du sondage systématique par grille de points utilisée ici (approximation à un sondage stratifié sur un critère de localisation) conduit à une surestimation peut être importante des variances. Il serait sûrement plus efficace d'utiliser un programme type JACKKNIFE ou BOOTSTRAP.

2) Concernant la précision de la méthode qu'il s'agissait ici d'évaluer, les points importants sont les suivants.

- Par rapport au sondage équiprobable sur liste, que nous avons considéré tout au long de ce travail comme une sorte de référence, le sondage que nous proposons possède plusieurs avantages. Sa précision est au moins équivalente et elle est meilleure dès que le caractère estimé est bien corrélé à la morphologie. L'absence de liste qui la caractérise en fait, comme nous le verrons au chapitre suivant, une méthode facilement adaptable à d'autres bases de sondage spatiales que l'image satellite ; elle est donc beaucoup plus souvent praticable qu'un sondage sur liste. Enfin, si le coût d'actualisation de la liste nécessaire à la mise en œuvre d'un sondage équiprobable "traditionnel" est variable et peut être prohibitif dans bien des villes de pays en développement, le coût lié à l'application du sondage aréolaire sur image satellite possède en revanche une borne supérieure fixe quelque soit la ville à laquelle on s'intéresse.

- Pour permettre à ce type d'enquête de rendre compte de façon satisfaisante des variations des caractères démographiques au sein de l'agglomération, deux possibilités s'offrent qui n'entraînent pas une augmentation excessive du budget : l'accroissement raisonnable de la taille de l'échantillon ou la diminution du taux au second degré permettant d'enquêter un plus grand nombre d'îlots. Ces deux possibilités ne s'excluent évidemment pas l'une l'autre.

- Enfin la précision relativement bonne des estimations par unité primaire fournit une information très complète sur l'échantillon d'ilots enquêtés qui peut être intéressante pour elle même.

Pour conclure sur la précision des résultats obtenus lors de cette première application de la méthode développée par l'équipe, nous voudrions rappeler que pour stratifier la base de sondage spatiale que constitue l'image satellite, nous nous sommes volontairement limité à exploiter l'information relative à la densité du bâti issue de cette même image. C'était là la règle du jeu pour évaluer en temps que telle l'efficacité d'une stratification d'après l'image. Il est cependant certain que l'on peut améliorer de beaucoup cette stratification, ne serait ce qu'en faisant appel à la connaissance empirique de la ville qui existe toujours dans les milieux de la gestion et de la recherche urbaines. Il n'y a bien entendu aucune raison de se priver de ce recours si l'on cherche à appliquer la méthode avec une efficacité maximum.

A examiner l'ensemble des indications chiffrées concernant la précision des estimations que cette première exploitation de l'enquête a permis de réunir, ainsi que les perspectives d'amélioration existantes, il nous semble qu'on peut conclure que la méthode que nous proposons peut, dans bien des cas, être avantageuse par rapport aux méthodes de sondage plus traditionnelles.

Chapitre 14

DOMAINES ET CONDITIONS D'APPLICATION DE LA METHODE

Introduction

A quels domaines, géographiques et thématiques, et dans quelles conditions peut s'appliquer la méthode de production de données démographiques en milieu urbain qui a été mise au point au cours de ce projet de recherche? Pour répondre à cette question, nous aborderons successivement dans ce chapitre chacun de ses termes. Tout d'abord dans une première partie, nous tenterons de définir les caractéristiques générales des sites urbains d'application, puis les champs thématiques auxquels la méthode paraît le mieux adaptée. Nous examinerons ensuite les conditions nécessaires à son application : contraintes liées à l'acquisition de l'image satellite et à son traitement et enfin celles propres à la définition du plan de sondage et à sa mise en œuvre. Tout au long de ce chapitre, nous identifierons les exigences propre à la méthode telle qu'elle a été appliquée à Quito, *mais nous tenterons également, pour chacune de ces exigences, de proposer des solutions de remplacement lorsque le site, les données disponibles ou le budget ne permettent pas d'appliquer la méthode "originale"*.

14. 1 : Sites et thématiques possibles pour l'application de la méthode

La méthode que nous pouvons proposer au terme des recherches pluridisciplinaires effectuées par l'équipe ORSTOM a été mise au point et testée sur des sites, et à partir d'une thématique qui, comme nous l'avons vu, ont certaines particularités.

Marseille et Quito sont des villes de taille comparable (autour du million d'habitant) et, même si elle sont situées a des latitudes fort différentes, elles n'en présentent pas moins certaines similitudes quant à l'organisation de leur espace : double fonction du centre ville -quartier d'affaires et quartier d'habitation parfois densément peuplé-, vaste extension spatiale et densité de population très variable, ségrégation spatiale et sociale marquée, c'est à dire forte "spécialisation socio-économique" des quartiers. Ces caractéristiques générales sont d'ailleurs communes à un grand nombre de villes, sous toutes les latitudes.

Le thème de l'enquête finale réalisée à Quito couvre, outre les caractéristiques démographiques générales de la population, des aspects variés ayant trait aux conditions d'occupation du logement, au niveau socio-économique, aux migrations et systèmes de résidence et aux réseaux familiaux d'entraide.

Mais, comme nous l'avons dit, dès le départ de cette recherche nous visions un domaine d'application le plus large possible du point de vue géographique comme du point de vue thématique. Les sites, la thématique et les conditions sur lesquels fut basé le développement de cette méthode n'en limitent pas, comme nous allons le voir, la portabilité.

14.1.1 : Les sites urbains d'application

Les sites d'application de la méthode sont bien sûr des villes et plus spécialement les villes des pays en développement. Deux questions se posent alors :

- De quelles données doit-on disposer concernant la ville étudiée pour envisager de réaliser une telle enquête?
- Sur quels types de villes la méthode fournira-t-elle de bons résultats à un coût raisonnable?

Comme pour toute méthode de sondage, la condition indispensable à la mise en œuvre est l'existence d'une base de sondage. Si cette base existe, l'ensemble des autres informations utiles pour appliquer la méthode que nous proposons peuvent être collectées grâce à des enquêtes préliminaires légères (toutes les informations utilisées lors de la mise au point de la méthode n'étant d'ailleurs pas indispensables, comme nous le verrons plus loin). Le résultat principal obtenu par l'équipe étant que l'image satellite peut fournir, en même temps, une bonne base spatiale de sondage et l'information nécessaire à sa stratification, *la seule condition indispensable à l'application de la méthode est donc de disposer d'une image satellite à haute résolution (SPOT ou Thematic Mapper) de la ville.* Comme nous le verrons au chapitre suivant, *le coût de l'image se révèle très concurrentiel face aux coûts de constitution de bases de sondage traditionnelles (listes d'individus ou de ménages) ou d'autres bases de sondages spatiales (cartes, cadastres, photographies aériennes).* Il peut même être intéressant d'acquérir l'image plutôt que de se lancer dans l'actualisation d'une base de sondage trop ancienne et inexacte (recensement ou cartographie non actualisés). Cependant, s'il existe une base de sondage spatiale actualisée et fiable couvrant toute la ville à laquelle on s'intéresse (cartes, cadastre ou photographies aériennes) et des informations suffisantes pour stratifier cette base (recensement ou enquête déjà effectués, bonne connaissance empirique de la ville chez des personnes d'expérience), on peut bien entendu envisager d'appliquer la même méthode du sondage aréolaire à deux degrés sans acquérir l'image. *A la condition que l'identification des unités primaires (ilots) soit possible sans ambiguïté sur la base de sondage spatiale retenue, l'adaptation de la méthode à cette base ne posera, à notre sens, aucun problème particulier.*

La seconde question, ayant trait aux types de villes auxquelles cette méthode peut être appliquée avec profit, doit être abordée du double point de vue du coût de l'enquête et de la précision des résultats que l'on souhaite obtenir. Comme il a été dit, la taille de l'échantillon que l'on doit enquêter pour obtenir une précision relative fixée, est à peu près indépendante de la taille de la population étudiée. *C'est dire que le coût de l'enquête ne diminuera que faiblement si on désire conserver les mêmes précisions pour estimer les caractéristiques de petites villes.* Que faut-il entendre, en l'occurrence, par "petite ville"?

À Quito, un échantillon d'environ 14 500 personnes nous a permis d'obtenir un coefficient de variation de l'ordre de 6% (peut être moins...) sur l'estimation du total de population de la ville. La précision atteinte peut donc être considérée comme bonne compte tenu du taux très faible pratiqué (rappelons le, moins de 1,3 % des ménages). L'exigence de précision peut bien sûr être variable suivant la taille des villes auxquelles on s'intéresse mais *si l'on veut conserver ce même ordre de grandeur, le taux de sondage à appliquer par exemple à des villes de moins de 100 000 habitants devient vite, comme on le voit, assez important.* Sur ce type de ville, l'économie réalisée par rapport à un recensement diminuant, les frais à engager ainsi que l'ampleur de l'opération à mettre en œuvre peuvent paraître disproportionnés par rapport aux moyens dont elles disposent. En revanche il est évident que plus la taille de la ville augmente plus la méthode présente d'intérêt, cela même si l'exigence de précision augmente. Nous détaillerons d'ailleurs ce point à la fin du chapitre 15.

14.1.2 : Les thématiques que l'on peut aborder

Tout d'abord rappelons que l'image satellite, en dehors de l'usage qui en est fait pour réaliser l'enquête, est toujours porteuse d'une information riche relative à la morphologie de la ville étudiée, à son organisation spatiale, à son environnement, etc... Cette information intéresse bien sûr, pour elle-même, les gestionnaires et les urbanistes. Mais nous avons vu qu'elle joue également un rôle central dans la définition du plan de sondage et par conséquent elle contribue à définir le domaine thématique auquel peut s'appliquer la méthode. En effet, d'une part la thématique de l'enquête doit être "attachée" à des individus statistiques compatibles avec la structure du sondage proposé, et d'autre part le gain de précision que permet l'application de cette méthode est fonction comme nous l'avons vu des corrélations existant entre le thème de l'enquête et la morphologie urbaine. Examinons un peu plus en détail ces deux points.

- La thématique de l'enquête et la structure du sondage

L'utilisation de l'image satellite comme base de sondage nous semble, au terme de cette recherche, *un principe applicable à toute enquête portant sur des unités statistiques que l'on puisse rattacher à une unité primaire repérable sur l'image sans équivoque*. Lors de l'application à l'étude de la population de Quito, cette unité primaire était l'îlot physique et, plus généralement, l'îlot apparaît comme l'unité primaire adéquate dès que l'unité enquêtée est l'individu, le ménage, l'immeuble, l'établissement (atelier, commerce, etc.) ... ou toute unité pouvant être facilement rattachée à des unités de ce type. Partant, *le domaine thématique concerné par ce type d'unité, et par conséquent pour lequel la méthode est applicable, est très vaste ; il couvre en fait pratiquement tout le champ des sciences sociales en milieu urbain : géographie humaine, démographie, sociologie, épidémiologie, économie etc...*

Cependant, même en l'état actuel des possibilités d'interprétation de l'image satellite en milieu urbain, l'îlot physique n'est pas la seule unité primaire que l'on peut imaginer de repérer sur l'image. La placette, c'est à dire la surface définie par une figure géométrique (généralement un carré) que l'on centre sur chaque point de l'image sélectionné dans l'échantillon, est une unité primaire très utilisée, nous l'avons vu au chapitre 2, pour obtenir des statistiques d'occupation du sol. Son repérage sur l'image ne pose aucun problème et on peut donc envisager d'appliquer au milieu urbain les méthodes développées en statistique agricole *pour estimer les surfaces couvertes dans l'agglomération par tel ou tel type d'utilisation du sol*. La méthode de sondage est alors, comme nous l'avons vu, nettement différente et beaucoup plus simple du moins en ce qui concerne l'étape d'inférence. Elle convient sans doute beaucoup mieux, du moins en théorie, que la méthode que nous proposons pour étudier l'occupation du sol, mais le repérage sur le terrain des unités primaires et la collecte de l'information peuvent poser, comme on imagine, des problèmes pratiques beaucoup plus difficiles qu'en milieu rural. *Malgré ces difficultés on peut penser que ce type d'application de la télédétection devraient intéresser les architectes et les urbanistes*.

- La thématique de l'enquête, liaisons existant avec l'information morphologique

Mais revenons à la méthode que nous proposons. Comme nous venons de le constater, son domaine potentiel d'application est vaste ; mais peut-on espérer que ses performances soient équivalentes lorsque le thème de l'enquête est différent de celui à partir duquel nous l'avons mise au point? *Une réponse définitive à cette question ne peut venir que de l'accumulation d'expériences dans les autres domaines possibles d'application de cette méthode*. Contentons nous donc prudemment d'avancer une hypothèse qui nous semble intuitivement raisonnable. L'amélioration de la précision des estimations, que ce soit par le biais de la stratification ou par celui du tirage systématique, repose, nous l'avons souvent répété, sur l'existence de corrélations entre les variables à estimer et les caractéristiques morphologiques des quartiers. Or

on peut penser, et nous ne faisons là qu'énoncer une évidence, que la plupart des phénomènes sociaux auxquels s'intéressent les sciences humaines ont une traduction morphologique et spatiale. Déchiffrer cette traduction, tenter d'en repérer les manifestations sur les images satellites et les intégrer dans des stratégies de sondage, c'est la démarche qu'a suivie l'équipe à laquelle nous appartenons. C'est cette même démarche qui devrait être poursuivie pour adapter la méthode à d'autres thématiques.

14. 2 : Les conditions d'application liées à l'utilisation de l'image satellite, à la définition du plan de sondage et à sa mise en œuvre

Nous l'avons dit dans la première partie de ce chapitre : la seule condition indispensable à l'application de la méthode, en l'absence d'autre base de sondage, est de disposer d'une image satellite à haute résolution couvrant l'agglomération étudiée. Nous avons également dit que *l'ensemble des informations nécessaires pour stratifier l'image et définir complètement le plan de sondage peuvent être obtenue grâce à des enquêtes préliminaires légères*. Ainsi définies, les conditions d'application de la méthode semblent faciles à satisfaire et, effectivement, elles ouvrent selon nous à ce type d'enquête un vaste domaine géographique et thématique. Cependant, si minimales soient-elles, ces conditions doivent être détaillées si l'on veut d'une part, définir précisément les exigences de la méthode en termes de qualité des images, de matériels informatiques et/ou de budget nécessaires aux traitements des différentes informations etc..., et d'autre part, proposer des alternatives lorsque les coûts de certaines étapes sont élevés relativement au budget disponible.

14.2.1 : Conditions liées à l'utilisation de l'image

Ces conditions sont de deux types.

- D'une part celles concernant l'image elle même : l'image satellite doit être exploitable c'est à dire couvrir la totalité de la surface étudiée et avoir été prise dans des conditions météorologiques acceptables. La nébulosité est évidemment le facteur météorologique déterminant mais, comme nous allons le voir selon le type d'exploitation que l'on envisage, d'autres facteurs peuvent intervenir.
- D'autre part celle liées aux traitements de l'image que l'on doit réaliser pour en extraire toute l'information nécessaire. Ces traitements peuvent être photographiques, numériques ou relever de l'interprétation visuelle. Les exigences concernant le matériel et l'information exogène nécessaire à ces traitements sont bien sûr différentes suivant les cas.

1) Conditions d'obtention d'une image exploitable comme base de sondage

En Septembre 1987, A. MICHEL a réalisé, durant une mission à Quito, une étude des facteurs météorologiques qui déterminent la probabilité d'obtention d'une image satellite SPOT ou Thematic Mapper (cette étude a fait l'objet d'une note de travail ¹). Les conclusions auxquelles il aboutit ne sont évidemment applicables qu'au site de Quito ou éventuellement à des ville connaissant des conditions météorologiques comparables. Il ne nous semble donc pas utile de les reproduire intégralement ici. Cependant, quelque soit le site considéré, *la probabilité d'obtention d'une image exploitable est un facteur déterminant lors de l'évaluation des possibilités d'application de la méthode*. La

¹ : A. MICHEL : Influence des facteurs météorologiques sur la probabilité d'obtention d'une image satellite SPOT ou THEMATIC MAPPER. Etude sur le site de Quito, Equateur. Quito, septembre 1987, 17 pages.

démarche à suivre pour évaluer ces probabilités est, dans ses grandes lignes, identique quelque soit le site, nous résumerons donc ici la problématique et les principales étapes du travail d'A. MICHEL.

En télédétection, la détermination de la probabilité d'obtention d'une image exploitable résulte toujours de la prise en compte de deux facteurs principaux : d'une part la nébulosité ou, plus généralement, l'ensemble des facteurs climatiques qui peuvent bruite le signal enregistré par les capteurs et d'autre part l'intérêt thématique de chaque mois ou période de l'année. A. MICHEL expose ainsi cette problématique générale :

- Quelle est la probabilité d'obtenir une image pour un mois déterminé, en admettant une nébulosité maximum? En ce qui concerne SPOT, on prendra en compte les différents angles de visée possible.
- Quelle est l'importance d'autres facteurs météorologiques : vitesse du vent, précipitations? Ces facteurs interviennent-ils de façon significative pendant les périodes où la nébulosité est minimum?
- Existe-t-il des mois qui thématiquement sont prépondérants ou indispensables?"¹

Dans la méthode que nous proposons, l'image a, comme nous l'avons vu, deux fonctions : elle constitue tout d'abord la base de sondage, en outre elle est la source d'information morphologique qui permet la stratification de cette base. C'est à partir de ces considérations qu'il faut définir les caractéristiques que devront avoir les images :

"De nombreux traitements d'image sont indispensables pour parvenir à discriminer les différents types morphologiques d'espaces qui constituent une cité comme Quito. Ces traitements requièrent des images panchromatiques ou multispectrales, enregistrées à des dates déterminées par le type de traitement désiré (...)"²

Parvenu à ce point de l'exposé, A. MICHEL dresse la liste détaillée des images utiles (parfois indispensables) à la mise au point des méthodes d'interprétation de l'information satellitaire en milieu urbain. L'objet de la recherche est alors beaucoup plus vaste que la seule application aux besoins de la stratification de l'image et de son utilisation comme base de sondage. C'est pourquoi nous ne retiendrons ici de cette liste qu'une seule image réellement indispensable :

"- un indicateur de la densité de l'espace bâti, lequel est sensible à la densité de végétation ; une période à forte activité photo-synthétique est donc privilégiée."³

Une fois définies les dates et les caractéristiques des images que l'on souhaite obtenir, "la connaissance et le traitement des données météorologiques permettent de calculer les probabilités" d'obtention de ces images. A Quito, A. MICHEL se base sur les relevés quotidiens de la station météorologique de l'aéroport. Les données concernent la nébulosité aux heures de passage des deux satellites SPOT et Thematic Mapper, les précipitations durant la période précédant leur passage et la vitesse maximale des vents pour chaque mois. *Les probabilités, qu'il calcule à partir de la première donnée (fréquence relative de la nébulosité par mois), sont celles d'obtenir une image "pour un mois donné, pour une nébulosité maximum donnée et pour un angle de visée maximum fixé"*. Par exemple, à Quito en juillet, la probabilité d'obtenir une image comportant moins de 2/8 de ciel couvert, avec un angle de visée inférieur à 10°, est de 79/100 pour une image SPOT et de 67/100 pour une image Thematic Mapper. A partir

¹ : A. MICHEL, Op. cit., p. 1.

² : A. MICHEL, Op. cit., p. 14.

³ : A. MICHEL, Op. cit., p. 14.

des données sur les précipitations et les vitesses des vents, il examine ensuite si ces facteurs risquent d'introduire des perturbations dans les images prises durant les périodes intéressantes thématiquement et où la probabilité d'obtention des images est bonne.

Les conclusions prennent en compte simultanément les probabilités d'obtention et les impératifs thématiques :

"Les différentes probabilités d'obtention d'image calculées indiquent que la conjonction de la thématique et de la météorologie est favorable pendant l'époque sèche. En période humide, seul le mois de novembre paraît praticable ; la nébulosité des autres mois paraît trop forte pour espérer obtenir une image correcte dans un délai raisonnable."

Comme nous l'avons dit, ces conclusions et les probabilités obtenues, sont spécifiques au site étudié. Soulignons cependant l'ordre de grandeur acceptable de ces probabilités sur un site tout de même assez souvent nuageux. Mais ce qu'il faut surtout retenir du travail effectué par A. MICHEL, c'est la démarche qui conduit à l'obtention de ces probabilités.

2) Conditions nécessaires au traitement de l'image en vue de la stratification de la base

Constituer le support du sondage à partir de l'image satellite ne nécessite pas, comme nous l'avons vu au chapitre 12, de traitement complexe de l'image : on utilise un tirage sur papier du canal panchromatique de l'image SPOT qui offre une résolution spatiale de 10 mètres. Une échelle approximative de 1/15 000 nous semble un bon choix pour réaliser toute l'opération de sélection de l'échantillon. En revanche, pour définir la limite urbaine, c'est à dire délimiter précisément la base de sondage, et pour réaliser la stratification de cette base, nous avons soumis l'image à de nombreux traitements numériques, photographiques et visuels. Il n'y a pas lieu ici d'exposer dans le détail les techniques et méthodes qui ont été utilisées, nous en serions d'ailleurs incapable. Le lecteur intéressé peut consulter la thèse de A. MICHEL (1988) et le rapport final rédigé par l'équipe en avril 1988 pour la convention CNES-ORSTOM¹. Les quelques éléments qui suivent, concernant les techniques et les matériels employés, sont repris de ce rapport. Ils permettent simplement de *situer dans quelles conditions méthodologiques et matérielles les traitements de l'image conduisant à une stratification peuvent être réalisés*.

Dans la phase de mise au point de la méthode, les recherches concernant l'exploitation possible des images SPOT de Quito, ont porté sur six thèmes :

1. détermination de la limite urbaine
2. évaluation de la précision des mesures de surfaces
3. mise en évidence des réseaux de voirie - segmentation de l'image
4. classification supervisée de l'image en fonction de la densité de l'espace bâti
5. analyse numérique des textures et structures des quartiers
6. détection des changements intervenus entre deux dates"²

Les résultats obtenus sur les deux premiers thèmes ont été directement utilisés dans la phase d'application à Quito, nous les avons d'ailleurs évoqués de façon plus détaillée au chapitre 12. Principalement pour des raisons de calendrier, les résultats obtenus concernant le thème 3 n'ont pas pu être utilisés à Quito. Pour stratifier l'image ayant servi au tirage de l'échantillon, on s'est basé sur *une classification supervisée des pixels* rendant compte de la densité de bâti (thème 4, voir chapitre 12). Cependant les tests effectués à partir d'une stratification obtenue par classification d'îlots sur

¹ : Rapport final pour la convention CNES - ORSTOM : L'utilisation de l'imagerie SPOT pour l'observation démographique en milieu urbain, avril 1988, 195 pages (disponible au CNES et à l'ORSTOM).

² : Equipe ORSTOM, Op. cit. p.13, 14 et 15.

lesquels on considère la moyenne radiométrique des pixels montrent qu'une telle stratification est meilleure que celle que nous avons utilisée. Quant aux thèmes 5 et 6, bien que de grand intérêt, ils ne sont pas directement liés au problème de la définition et de la stratification de la base de sondage (soit en raison du thème en lui-même - thème 6 - , soit en raison du caractère exploratoire des recherches). Après ce rapide tour d'horizon, il apparaît que *les seules opérations indispensables à la mise en œuvre du sondage sont celles liées aux thèmes 1 et 4* (les opérations d'évaluation de la précision des mesures de surfaces, indispensables lors d'une première application de la méthode, n'ont pas à être répétées pour les applications futures).

Durant cette même phase de mise au point, les traitements numériques d'image satellite ont été réalisés en France, (logiciel PLANETE de l'ORSTOM), sur le matériel suivant. A partir du dernier trimestre 1987, l'atelier de télédétection du L.I.A.¹ de l'ORSTOM à Bondy a été équipé de stations de travail graphiques SUN 3/110C et 3/160C, d'un disque dur de 500 Mo, d'un dérouleur de bande 1600/6250 BPI, d'une table traçante BENSON 1332 et d'une recopie d'écran. Les traitements photographiques ont été réalisés à l'Atelier de Traitement Photochimique de l'ORSTOM à Bondy.

Evidemment, *un tel environnement matériel n'est pas nécessaire pour toute application de la méthode* et la configuration citée ci-dessus n'est donnée qu'à titre d'information. Concernant le matériel informatique et photographique réellement nécessaire, deux remarques doivent être faites :

- d'une part, comme il est dit dans le rapport pour la convention CNES-ORSTOM : "La plupart des auteurs s'accordent à affirmer que *les données SPOT sont utilisables, visuellement, au même titre que les photographies aériennes à condition toutefois de ne pas descendre à l'échelle du bâtiment* ; aussi, dans un premier temps, l'interprétation visuelle des images, méthode bien connue et fiable, peut apporter plus d'information que ne le ferait un traitement sophistiqué. Il est donc raisonnable de dresser une typologie de l'espace urbain à l'intérieur de la commune (...)"² ; il n'est donc pas exclu d'effectuer la stratification de l'image à l'aide d'une interprétation visuelle sur tirage papier ;

- d'autre part, si l'on choisit de baser cette stratification sur des traitements numériques et si les traitements nécessaires sont connus, on peut bien sûr envisager de les *sous-traiter à un organisme équipé*.

14.2.2 : Conditions liées à la définition du plan de sondage et à sa mise en œuvre

De même que les recherches en télédétection, la mise au point de la méthode de sondage, a nécessité le recours à de nombreux outils informatiques. Durant la phase marseillaise en particulier, nous avons utilisé les bibliothèques de programmes comme celle du CIRCE ou du CNUSC, qui regroupent la plupart des logiciels statistiques conçus pour de gros systèmes. Nous avons vu, dans la première partie de cette thèse, quel usage a été fait du logiciel SAS et des programmes de l'ADDAD. Lorsque l'on envisage simplement d'appliquer la méthode, le recours au traitement informatique des données a simplement pour but d'adapter le plan de sondage aux spécificités du site considéré ; *les besoins matériels et logiciels sont donc heureusement bien moindres*. Ils existent néanmoins et nous les examinerons tout d'abord. Par ailleurs, cette adaptation du plan de sondage, nécessite également de disposer d'informations concernant les caractéristiques générales morphologiques et démographiques de la

¹ : Laboratoire d'Informatique Appliquée.

² : Equipe ORSTOM, Op. cit. p. 7.

ville. Nous nous intéresserons donc ensuite aux *différentes possibilités existant pour réunir ces informations*. Enfin nous signalerons les particularités de l'organisation de la collecte induite par l'utilisation de l'image satellite et par la structure du sondage.

1) Environnement informatique nécessaire

Il y a plus de deux ans, lorsque nous envisagions la phase d'application de la méthode sur la ville de Quito, nous posions ainsi le problème de l'équipement informatique¹ :

"Les problèmes matériels et logiciels se posent bien sur de manière assez différente dans la phase d'application à Quito.

En supposant résolu par la phase de mise au point le problème de la sélection des méthodes d'analyse de données adaptées au traitement de l'information morphologique, on devrait donc disposer au moment du départ à Quito d'une "trousse à outils statistiques" qui permette d'adapter la méthode de stratification(...) et le plan de sondage à la spécificité de la ville.

Parmi les partenaires équatoriens du projet, seul l'Institut Géographique Militaire semble en mesure d'offrir un jour, un environnement informatique comparable à celui dont nous disposons en France. L'I.G.M. envisage en effet de s'équiper d'une configuration informatique centrée sur un ordinateur "Digital Equipment" de type "Micro Vax", un mini ordinateur puissant, autour duquel sera implanté un réseau de stations de travail "Appolo" (...).

Une telle configuration permettrait de développer à Quito, sur un même système, les applications statistiques, infographiques et de télédétection liées à la recherche et plus généralement au projet d'atlas urbain informatisé auquel elle sera intégrée. Malheureusement, aux dernières nouvelles, le calendrier prévu actuellement par l'IGM, ne laisse pas augurer que ce système puisse être opérationnel avant un délai minimum d'un an. Ce délai est incompatible avec le calendrier propre du projet ainsi qu'avec celui de la thèse.

D'autre part, développer l'ensemble de l'application à Quito dans un environnement informatique aussi favorable que le serait la configuration prévue par l'IGM, ne risquerait-il pas de limiter dangereusement la portabilité de la méthode ?

Pour la partie statistique du travail, il apparaît donc un objectif supplémentaire : la chaîne de traitements mise au point à Marseille devra utiliser les logiciels les plus portables possible et ne dépendant pas trop d'un environnement informatique coûteux et très rare dans la majorité des sites potentiels d'application dans les pays en développement."

Et nous avons conclu :

"(...) une tendance s'affirme actuellement à la réalisation de versions "micro" de beaucoup de logiciels statistiques (...). Il n'est donc pas interdit de penser que la chaîne de programmes nécessaire à l'application de cette nouvelle méthode de collecte de données démographiques puisse se satisfaire d'une configuration de type micro-informatique."

Nous ne pensons pas alors si bien dire puisque, comme nous l'avons vu, mise à part l'étape d'analyse typologique des données morphologiques de Marseille que nous avons réalisée en France en utilisant des gros systèmes, *tout le développement de la méthode et son application à Quito a été réalisé sur le seul équipement dont nous avons*

¹ : O. BARBARY : Projet de thèse de mathématiques appliquées aux sciences de l'homme, plan de travail, Paris 1985, 20 pages.

disposé dès le départ : un micro ordinateur IBM AT, en utilisant trois logiciels écrits pour cet ordinateur : Dbase III, SPSS/PC et STAT ITCF. La preuve est donc faite que dans une phase d'application de la méthode, la définition complète du plan de sondage ne réclame pas un équipement informatique lourd. Quant aux logiciels, nous avons souligné, au cours des deuxième et troisième parties, l'usage constant qui a été fait du système de gestion de bases de données (Dbase III) ; un système équivalent sera également indispensable pour toute autre application. Le logiciel statistique SPSS/PC +, malgré les imperfections que contient encore la dernière version¹, possède à peu près toutes les fonctions statistiques nécessaires à la définition du plan de sondage et offre les possibilités de manipulation et de contrôle des données indispensables à l'exploitation de l'enquête finale. Cependant il faut noter ici qu'avec la taille d'échantillon que nous avons adoptée (environ 14 500 individus) et la quantité d'information collectée, on atteint, pour l'exploitation de l'enquête, des temps de traitements fastidieux et on sature facilement les capacités de mémoire morte (40 Mo). Cela dit on peut penser qu'au rythme actuel d'évolution des matériels et des logiciels, ces inconvénients n'existeront bientôt plus.

2) Etapes de la mise au point du plan : les enquêtes préliminaires

Lorsque nous parlons d'*adaptation de la méthode aux spécificités du site*, qu'entendons nous par là? Quelles sont les caractéristiques du plan de sondage susceptibles d'être modifiées lors de cette adaptation?

La structure du plan : nombre de degrés, taux et probabilités de sélection aux différents degrés, ainsi que les estimateurs employés, ne sont pas modifiés ; seule la stratification change. Il faut donc *redéfinir les strates et la répartition de l'échantillon entre ces strates*. De quelles informations a-t-on besoin pour réaliser ces deux opérations?

Définition des strates

Si l'on choisit de définir la stratification en s'appuyant sur des traitements numériques de l'image, quelle que soit la méthode retenue (classification de pixels ou de zones), il faudra superviser la classification ; c'est à dire définir des pixels ou des zones d'entraînement. Ceci implique de réunir, sur un échantillon d'îlots représentatif du milieu urbain étudié, les informations relatives aux critères morphologiques selon lesquels on veut classer l'image : densité du bâti, densité et état de la végétation, matériaux de couverture des bâtiments etc. , la liste de ces informations dépendra des progrès réalisés dans les années à venir en télédétection urbaine. Comme nous l'avons vu, pour l'instant le critère déterminant est la densité du bâti. Pour une stratification portant sur ce seul critère, l'enquête préliminaire "*Morphologie*" que l'équipe ORSTOM a réalisée à Quito est certainement trop détaillée, elle avait, comme nous l'avons vu, un objectif plus vaste. On pourra dans une phase d'application limiter les informations recueillies au(x) seul(s) thème(s) de la stratification ; néanmoins la taille de l'échantillon doit être suffisante pour permettre l'initialisation de la classification de l'image satellite.

Evidemment, cette enquête préliminaire peut être extrêmement légère si l'on dispose d'autres source d'informations plus ou moins actualisées (cadastre, plan détaillé ou photographies aériennes) ; on pourra alors n'enquêter qu'un échantillon dans les zones non couvertes par ces autres sources ou ayant connu d'importantes modifications depuis la date d'établissement de ces sources.

¹ : Etant donné son prix d'achat élevé pour un logiciel sur micro-ordinateur, on nous permettra de trouver abusif que soient facturées des versions actualisées qui contiennent toujours des imperfections notées dans les versions antérieures !

Répartition de l'échantillon entre les strates

Tout d'abord rappelons *qu'aucune information supplémentaire concernant les variables que l'on cherche à estimer par l'enquête n'est nécessaire pour effectuer la répartition de l'échantillon entre les strates si l'on adopte une allocation proportionnelle*, c'est à dire un même taux dans toute les strates. Dans ce cas, la seule donnée indispensable est une estimation, la plus précise possible, du nombre d'ilots contenu dans chaque strate (ou, ce qui revient au même, de la surface moyenne des ilots par strate¹. On peut l'obtenir, comme nous l'avons fait à Quito, à partir des mesures de surface des ilots réalisées lors de l'enquête préliminaire morphologie, mais aussi à partir de toute autre source permettant des mesures de surface fiables sur un échantillon d'ilots représentatif de l'ensemble des strates ; en particulier l'obtention d'une bonne estimation du nombre d'ilots par strate à partir de l'image satellite elle même ne devrait pas poser de problème. *Une telle allocation, si elle ne permet pas d'optimiser le gain de précision que peut apporter la stratification, améliorera cependant toujours les résultats (cette amélioration est évidemment fonction de la pertinence de la stratification elle même).* Les expériences faite à Marseille et à Quito permettent de chiffrer ce gain à environ 50% du gain fournit par une allocation optimale.

Pour améliorer la répartition de l'échantillon entre les strates, nous avons vu qu'il faut adopter une allocation favorisant les strates où la variance des caractères à estimer est forte. Comment obtenir une information concernant ces variances? Il ne s'agit pas de réunir des données permettant, pour l'ensemble des variables du questionnaire, de calculer ces variances : d'une part, l'allocation que l'on cherche à définir ne peut être qu'un compromis entre les allocations optimales correspondant à chacune des variables, et d'autre part le gain de précision obtenu sur les estimations est, nous l'avons vu, peu sensible à des variations de détails dans la répartition de l'échantillon entre les strates. L'ordre de grandeur des variances aux sein des différentes strates de deux ou trois variables clefs du questionnaire suffit en général pour déterminer une allocation proche de l'optimal.

Si l'on choisit de réaliser, pour obtenir cette information, une enquête préliminaire, le questionnaire peut donc être très court et l'échantillon très restreint. De plus il faut se souvenir ici que la variance d'un caractère dans une strate donnée est presque entièrement déterminée par la variance inter unités-primaires, c'est à dire par les variations d'un ilot à l'autre². *Il est donc inutile, dans ce type d'enquête préliminaire, de collecter l'information pour tous les ménages de l'échantillon.* Par exemple, pour estimer l'ordre de grandeur de la variance du total de population au sein d'une strate, on pourra se contenter, pour une quinzaine d'ilots représentatifs de cette strate, d'effectuer le comptage du nombre de ménages vivant dans ces ilots et de noter la taille de dix ménages dans chaque ilot.

Cependant, dans certains cas on disposera, au moins sur une grande partie de la ville, d'autre sources d'informations (recensement, même ancien, ou autre enquête). Comme pour l'information morphologique, l'enquête préliminaire ne portera alors que sur les zones non couvertes par ces sources ou pour lesquelles l'information qu'elles fournissent est trop périmée.

Enfin, notons que les caractéristique de souplesse, rapidité et faiblesse des coût que nous avons chercher à donner à cette méthode devraient permettre son application répétée à une même ville et à un même thème pour en suivre l'évolution. Dans un tel cas de figure l'enquête préliminaire démographique est bien sur inutile, le plan de sondage (stratification, allocation aux strates) devrait s'améliorer à chaque nouvelle application en profitant des résultats de l'enquête précédente.

¹ : voir au chapitre 12, § 12.3.3, A. Détermination du nombre total de pâtés de maisons et de leur répartition dans les strates.

² : Voir à ce sujet le chapitre 11, § 11.3.3

3) Conditions de la collecte

Lors de la réalisation des opérations de collecte à Quito, pour des raisons pratiques et de respect du calendrier, nous avons effectué le repérage de la majorité des îlots de l'échantillon sur les cartes au 1/10 000 établie par l'INEC pour le recensement de 1982 ; le recours à cette cartographie a facilité le travail sur le terrain des superviseurs et des enquêteurs. Lorsque sur le site d'application, des plans fiables à une échelle convenable existent, il n'y évidemment aucune raison de ne pas les utiliser, mais ce recours n'est pas indispensable. Nous avons vu que l'on pouvait, sans difficulté majeure, identifier avec l'image sur le terrain l'îlot auquel appartient chaque point de la grille. Cependant, l'orientation sur le terrain l'image en main demande un minimum de pratique et peut dérouter des superviseurs et des enquêteurs, même s'ils ont été bien sélectionnés et connaissent la ville. Si l'on ne dispose d'aucun autre document que l'image satellite, *il faut donc veiller à ce que le premier repérage des îlots sur le terrain à l'aide de l'image soit fait par une (des) personnes(s) ayant cette pratique.*

L'autre particularité, quant à l'organisation de la collecte, de la méthode que nous proposons est liée à la structure du sondage sur lequel elle repose : les deux degrés du sondage font chacun l'objet d'une collecte d'information. *Une organisation rationnelle de ces deux étapes de la collecte doit permettre de minimiser les coûts, les délais et les risques d'erreur.*

- après avoir identifié sans équivoque les îlots de l'échantillon, on doit s'attacher dans un premier temps au recensement des ménages et à la mesure des surfaces sur le terrain lorsqu'elles ne peuvent pas être effectuées directement sur l'image ;
- dans un second temps, une fois sélectionnés les ménages à enquêter dans chaque îlot, les enquêteurs visiteront chaque ménage et rempliront le questionnaire.

Conclusion

Pour conclure, résumons tout d'abord les étapes que doit suivre l'application de la méthode, en ne supposant l'existence dans la ville étudiée d'aucune source d'information complémentaire :

1. Acquisition de l'image.
2. Enquêtes préliminaires légères.
3. Exploitation des enquêtes préliminaires sur micro-ordinateur
4. Stratification : traitement numérique de l'image (éventuellement sous-traité à un organisme équipé) ou interprétation visuelle ; et définition de l'allocation aux strates
5. Tirage de l'échantillon
6. Collecte en deux étapes. 1 : informations sur les unités primaires, 2 : informations sur les unités statistiques.
7. Codification, saisie et exploitation de l'enquête sur micro-ordinateur.
8. Publication des résultats.

Comme nous l'avons vu tout au long de ce chapitre, ce schéma admet de nombreuses variantes (dans les étapes 1 à 4) si l'on dispose sur le site d'une autre base de sondage ou d'informations pouvant aider à stratifier l'image. Nous n'avons pas examiné dans le détail les modalités de l'adaptation de la méthode à de telles situations ; cette adaptation aura évidemment des répercussions sur les coûts et les délais de mise en œuvre de l'enquête ainsi que sur la précision de ces résultats. Si la méthode est utilisée, l'évaluation de ces répercussions se fera au fur et à mesure des applications dans des contextes variés, et on peut penser qu'apparaîtra alors toute une gamme de solutions

aux problèmes du traitement de l'image, de la stratification, de la répartition de l'échantillon entre les strates, etc. ; la méthode évoluera ainsi avec chaque nouvelle application. Nous pensons et nous espérons l'avoir montré dans ce chapitre, que la méthode que nous proposons est "souple" et qu'elle peut donc s'adapter à un large domaine géographique et thématique.

Chapitre 15

COÛTS ET DELAIS DE LA MISE EN ŒUVRE

Introduction

L'évaluation de la méthode proposée ne peut être complète sans examiner une dernière question : quelle doit être la taille de l'échantillon? c'est à dire, plus précisément, quelle taille d'échantillon, autrement dit quel budget, permettra d'atteindre une précision donnée?

En effet, comme le remarque ASSELIN¹,

"Le problème de la taille (...) de l'échantillon se pose toujours dans un contexte dominé par deux contraintes inconciliables : la contrainte *coût* et la contrainte *précision* du sondage. Si l'on met l'accent sur la précision, le coût augmentera d'autant. Par contre si l'on veut diminuer le coût, il faut sacrifier en partie la précision. La solution du problème exige donc qu'on effectue une opération préalable : construire une fonction coût et une fonction précision qui dépendent des différentes tailles d'échantillon associées à la structure de sondage."

C'est précisément la démarche que nous adoptons dans ce dernier chapitre. Cependant, dans sa généralité, le problème du calcul de ces deux fonctions dans le cas d'une structure de sondage à plusieurs degrés est assez complexe². Leurs expressions doivent, en toute rigueur, faire intervenir le nombre d'unités sélectionnées à chacun des degrés du sondage puisque chaque degré a une répercussion spécifique sur le coût et la précision globale. En ce qui concerne la fonction de précision, c'est à dire la fonction qui donne la variance des estimations pour une structure du sondage et une taille d'échantillon données, nous avons vu dans les chapitre 11 et 12 que l'effet du second degré du sondage peut être négligé sans que les valeurs des variances soient significativement affectées. Quant au coût de la mise en œuvre du sondage, nous allons voir qu'on peut l'exprimer comme une fonction linéaire du nombre d'ilots enquêtés plus une constante représentant les frais fixes.

Pour parvenir à ces résultats, nous donnerons dans une première partie les éléments permettant de calculer les coûts afférant aux différentes étapes de la mise en œuvre de la méthode. Ces éléments nous ont été fournis par F. DUREAU qui a tenu la comptabilité des frais engagés durant toute la phase d'application à Quito. Dans les tableaux présentés figurent également les durées nécessaires à l'exécution de chaque opération. Dans une seconde partie, nous donnerons les expressions des fonctions de coût et de précision qui caractérisent la méthode *telle que nous l'avons appliquée à Quito*, puis nous exprimerons le budget nécessaire à l'application de la méthode comme une fonction de la précision que l'on souhaite atteindre.

¹ : L.M. ASSELIN : Techniques de sondage avec application à l'Afrique, 1984, p. 225.

² : Ce problème général relève en fait de la recherche opérationnelle.

15. 1 : **Eléments pour le calcul des coûts et des délais de mise en œuvre de la méthode**¹

L'objet de ce paragraphe est de fournir tous les éléments nécessaires au calcul du coût de la mise en œuvre d'une enquête par sondage stratifié sur image satellite. Outre les tarifs relatifs aux équipements et achats de données satellitaires, sont récapitulés les différents éléments composant les budgets des trois enquêtes que nous avons réalisées à Quito² : enquête Morphologie urbaine (octobre-novembre 1986), enquête Bâti/population (avril 1987), enquête Démographie/migrations (décembre 1987).

A l'issue d'une série de tableaux relatifs à chaque type de dépense, un tableau final de synthèse fait le bilan financier général de ces trois opérations de collecte.

15.1.1 : Equipement

1. Equipement informatique

. Matériel

Station SUN 3/110C, unité de disque dur 140 Mo :	200 000 FF
IBM AT (ou compatible), disque dur 20 Mo, imprimante :	15 000 FF

. Logiciels

Saisie et traitements statistiques des données d'enquête sur IBM AT :	
- SPSS/ PC+ (module de base et module Data Entry) :	12 900 FF
Gestion de base de données relationnelles sur IBM AT :	
- DBASE III + :	6700 FF
Traitements images satellite :	
- Logiciel PLANETE, développé par l'ORSTOM :	Gratuit

2. Equipement de mesure

Pour la mesure sur le terrain des flots de l'échantillon :

Boussole avec viseur, Clinomètre (mesure des pentes), Décamètre :	2 500 FF
---	----------

Pour la mesure des surfaces des strates sur image satellite ou plan :

Planimètre électronique :	6 000 FF
ou planimètre manuel :	1 500 FF

¹ : Les tableaux et commentaires qui constituent le paragraphe 15.1 sont repris d'une note de travail rédigée par F. DUREAU : Eléments de calcul des coûts d'enquête à Quito, Août 1988, 13 p. Quito.

² : La conversion des sucres en francs français a été systématiquement faite selon le taux de change de l'ambassade en vigueur au moment de la dépense.

15.1.2 Images satellites (prix au 1er semestre 1987)

1. Bandes magnétiques 6250 BPI

Image THEMATIC MAPPER, bande 6250 BPI, 7 canaux :

- scène entière (180 x 180 Km) existant au catalogue :	3 300 \$ 19 800 FF
- quart de scène (90 x 90 Km) existant au catalogue :	1 650 \$ 9 900 FF
- supplément pour acquisition spéciale (par scène) :	1 600 \$ 9 600 FF
- supplément pour couvert nuageux maximum (par scène) :	275 \$ 1 650 FF

Image SPOT (60 x 60 Km), bande 6250 BPI, canal panchromatique :

- scène existant au catalogue :	9 500 FF
- scène avec choix localisation le long de la trace du satellite (à partir de deux scènes existantes) :	15 000 FF
- supplément pour acquisition spéciale (par scène) :	2 000 FF

Image SPOT (60 x 60 Km), bande 6250 BPI, canaux XS1, XS2, XS3 :

- scène existant au catalogue :	9 500 FF
- scène avec choix localisation le long de la trace du satellite (à partir de deux scènes existantes) :	15 000 FF
- supplément pour acquisition spéciale (par scène) :	2 000 FF

2. Films

THEMATIC MAPPER, échelle 1/ 1 000 000 (tarif pour 7 canaux) :	1 120 \$ 6 720 FF
SPOT, échelle 1/ 400 000, canal panchromatique :	5 000 FF
SPOT, échelle 1/ 400 000, canaux XS (tarif pour les 3 canaux) :	4 600 FF

15.1.3 : Sélection de l'échantillon

	MORPHOLOGIE	BATI/POPULATION	DEMO/MIGRATIONS
Technique de sélection	Tirage systématique par grille superposée aux plans INEC 1/ 10 000	Choix raisonné dans liste de 199 îlots classes par type morphologique	Tirage aréolaire systématique stratifié, par grille superposée à l'image SPOT panchro. 1/15 000
Nombre îlots sélectionnés	199	53	426
Etapes de la sélection et durée	<p>1/2 journée, 2 pers. : Dessin de la grille sur calque millimétré Tirage des points Dessin des îlots de l'échantillon sur les plans au 1/ 10 000</p> <p>9 jours, 1 pers. : Agrandissement au pantographe des plans d'îlot (1/ 500)</p>	<p>1/2 journée, 2 pers. : Sélection raisonnée, au vu de la liste des îlots classés par type, du tableau décrivant la morphologie des îlots, et des diapo. de ces 199 îlots</p>	<p>2 jours, 2 pers. : Tracé de la limite urbaine sur image SPOT 1/ 30 000</p> <p>4 jours, 2 pers. : Tracé de la stratification en 6 niveaux de densité du bâti sur image SPOT 1/ 30 000</p> <p>2 jours, 2 pers. : Mesure de la superficie des strates sur image SPOT 1/ 30 000</p> <p>9 jours, 2 pers. : Report de la stratif. sur image SPOT 1/ 15 000 Dessin des grilles de tirage (1 par strate bâtie) Tirage des points d'enquête Dessin des îlots de l'échantillon sur plans 1/ 10 000</p>
DUREE TOTALE	9 jours et demi	1/2 journée	17 jours

15.1.4 : Collecte des données sur le terrain**1. Morphologie des flots**

	ENQUETE MORPHOLOGIE
Nombre flots	199
ENQUETEURS	
Nombre	2
Nb jours de travail	50 jours à temps complet
Salaire total des enquêteurs	200 000 sucres (base : 40 000/mois/enquêteur)
SUPERVISEURS (photo. des flots)	
Nombre	1 superviseur + 1 chercheur ORSTOM
Nb jours de travail	20 jours à mi-temps
Salaire total du superviseur	40 000 sucres
REPRODUCTION DOCUMENTS ENQUETE	
Plans	10 000 sucres
Questionnaires	15 000 sucres
Manuels d'enquête	700 sucres
Dévelop. diapos	15 000 sucres
COUTTOTAL	280 700 sucres (11 200 FF)
COUT PAR ILOT	1 410 sucres (56,4 FF)

2. Données démographiques

	ENQUETE BATI/POPULATION	ENQUETE DEMO/MIGRATIONS
Nombre ilots	53	426
Nombre logements	1700 (dont 1108 occupés)	3157 (tous occupés)
ENQUETEURS		
Nombre	5	16
Nb jours de travail	21 jours pour 4 enquêteurs 14 jours pour 1 enquêteur	1 mois pour 13 enquêteurs 15 jours pour 1 enquêteur 1 mois 1/2 pour 2 enquêteurs
Salaire total des enquêteurs	145 000 sucres (base 40 000 s/mois/enquêteur)	900 000 sucres (base 50 000 s/mois/enquêteur)
SUPERVISEURS		(contrôle enq. + recensement et sélection des logements dans les ilots de l'échantillon)
Nombre	1	3
Nb jours de travail	21 jours	1 mois
Salaire total des superviseurs	37 000 sucres (base 50 000 s/mois/superviseur)	180 000 sucres (base 60 000 s/mois/superviseur)
REPRODUCTION DOCUMENTS ENQUETE		
Plans	2 000 sucres	8 000 sucres
Questionnaires	35 000 sucres	95 000 sucres
Manuels d'enquête	1 000 sucres	10 000 sucres
COUT TOTAL	220 000 sucres (8 800 FF)	1 193 000 sucres (26 250 FF)
COUT PAR LOGEMENT OCCUPE	199 sucres (7,96 FF)	378 sucres (8,31 FF)

3. Mesures des ilots sur le terrain

Nombre ilots	170 (parmi les 426 de l'enquête Demo./mig.)
MESURE TERRAIN	
Nombre enquêteurs	1
Nb jours de travail	45 jours (5 heures par jour)
Salaire total par enquêteur	100 000 sucres (base 60 000 s/mois/enq. + frais de déplacement)
CALCUL SURFACES	
Nombre enquêteurs	1
Nb jours de travail	15 jours (2 heures par jour)
Salaire total par enquêteur	30 000 sucres
REPRODUCTION DOCUMENTS ENQUETE	
Plans de repérage	1000 sucres
Questionnaires	1700 sucres
Manuels d'enquête	-
COUT TOTAL	132 700 sucres (1 600 FF)
COUT PAR ILOT	780 sucres (9,36 FF)

15.1.5. Codification des données

	MORPHOLOGIE	BATI/POPULATION (Codif. partiel. faite sur questionnaire)	DEMO/MIGRATIONS
Nombre îlots	199	53	426
Nombre logements	-	1700 (1108 occupés)	3157 (tous occupés)
Nombre personnes	-	4803	14553
Nombre de fichiers	2	2	5
FICHER 1	ILOTS	LOGEMENTS	ILOTS
Nb enregistrements	199	1700	426
Nb var./enregist.	64	16	5
Nb car./enregist.	158	29	14
Nb total caractères	31442	49300	5964
FICHER 2	GENRES BATIMENTS	INDIVIDUS	LOGEMENTS
Nb enregistrements	2133	4803	3157
Nb var./enregist.	14	22	19
Nb car./enregist.	41	45	50
Nb total caractères	87453	216135	157850
FICHER 3			PARENTS
Nb enregistrements	-	-	13462
Nb var./enregist.	-	-	18
Nb car./enregist.	-	-	37
Nb total caractères	-	-	498094
FICHER 4			INDIVIDUS
Nb enregistrements	-	-	14553
Nb var./enregist.	-	-	43
Nb car./enregist.	-	-	99
Nb total caractères	-	-	1440747
FICHER 5			MIGRATIONS
Nb enregistrements	-	-	20601
Nb var./enregist.	-	-	26
Nb car./enregist.	-	-	66
Nb total caractères	-	-	1359666
LIEU CODIFICATION	Quito - ORSTOM	Quito	Quito
TEMPS CODIFICATION	1 chercheur 1 mois travail mi-temps	1 codificateur 22 jours travail 8 heures/jour	5 codificateurs 65 jours travail 5 heures/jour
NB TOTAL CARACT.	118895	265 435	3 462 321
COUT TOTAL CODIF.	-	40 000 sucres	502 500 sucres
PRIX DU CARACTERE	-	0, 15 sucres	0, 15 sucres

(suite Tableau Codification)

	MORPHOLOGIE	BATI/POPULATION	DEMO/MIGRATIONS
COUTCODIFICATION / LOGEMENT OCCUP. Question. Bâti/pop	-	34 sucres	76 sucres
Questi. démo complet	-	-	159 sucres
COUT REPRODUCTION			
FICHER 1			
Manuel	-	-	-
Fiches	3 000 sucres	5 100 sucres	270 sucres
FICHER 2			
Manuel	-	1 800 sucres	450 sucres
Fiches	3 000 sucres	-	3 200 sucres
FICHER 3			
Manuel	-	-	900 sucres
Fiches	-	-	10 000 sucres
FICHER 4			
Manuel	-	-	630 sucres
Fiches	-	-	24 000 sucres
FICHER 5			
Manuel	-	-	450 sucres
Fiches	-	-	28 000 sucres
COUT TOTAL REPRODUCTION	6 000 sucres	6 900 sucres	67 900 sucres
COUT REPRODUCTION / LOGEMENT OCCUP. Question. Bâti/pop	-	6 sucres	9 sucres
Questi. démo complet	-	-	21 sucres
COUTTOTAL	6 000 s (249 FF)	46 900 s (1 800 FF)	570 400 s (8 500 FF)
COUTTOTAL / LOGEMENT OCCUP.			
Question. Bâti/pop	-	40 sucres (1,52 FF)	85 sucres (1,27 FF)
Questi. démo complet	-	-	180 sucres (2,70 FF)

15.1.6 : Saisie des données

	MORPHOLOGIE	BATI/POPULATION	DEMO/MIGRATIONS
Nombre îlots	199	53	426
Nombre logements	-	1700 (1108 occupés)	3157 (tous occupés)
Nombre personnes	-	4803	14553
Nombre de fichiers	2	2	3
FICHER 1	ILOTS	LOGEMENTS	ILOTS
Nb enregistrements	199	1700	426
Nb var./enregist.	64	16	5
Nb car./enregist.	158	29	14
Nb total caractères	31442	49300	5964
FICHER 2	GENRES BATIMENTS	INDIVIDUS	LOGEMENTS
Nb enregistrements	2133	4803	3157
Nb var./enregist.	14	32	19
Nb car./enregist.	41	89	50
Nb total caractères	87453	427467	157850
FICHER 3			PARENTS
Nb enregistrements			13462
Nb var./enregist.			18
Nb car./enregist.			37
Nb total caractères			498094
FICHER 4			INDIVIDUS
Nb enregistrements			14553
Nb var./enregist.			43
Nb car./enregist.			99
Nb total caractères			1440747
FICHER 5			MIGRATIONS
Nb enregistrements			20601
Nb var./enregist.			26
Nb car./enregist.			66
Nb total caractères			1359666
LIEU SAISIE	Quito - ORSTOM	France - Société serv.	Quito - Société serv.
TEMPS SAISIE		1 semaine	3 mois
NB TOTAL CARACT.	118895	476767	3462321
COUTTOTAL	-	8345 Francs	1187531 s (17 800 FF)
PRIX DU CARACTERE	-	1,75 centimes	0,34 sucres (0,51 cent.)
PRIX / LOG. OCCUP.	-		
Question. Bâti/pop	-	7,25 Francs	172 sucres (2,58 FF)
Questi. démo complet	-	-	375 sucres (5,64 FF)

15.1.7 : Synthèse des coûts

1. Coûts de collecte/ codification/ saisie de données démographiques

A partir des expériences des enquêtes Bâti/population et Démographie/migrations, on obtient les résultats suivants :

ENQUETE BATI/POPULATION	Coût total par logement Si saisie en France : 16, 73 FF Si saisie en Equateur : 12, 06 FF	Collecte : 7, 96 FF Codification : 1, 52 FF Saisie (FRANCE) : 7, 25 FF Saisie (EQUATEUR) : 2, 58 FF
ENQUETE DEMOGRAPHIE/MIG.	Coût total par logement 16, 65 FF	Collecte : 8, 31 FF Codification : 2, 70 FF Saisie : 5, 64 FF

Comme le montre le tableau ci-dessus, les coûts de codification et saisie de l'information collectée sont sensiblement équivalents à ceux de la collecte sur le terrain elle-même.

Hormis le salaire du personnel qualifié pour l'organisation et l'encadrement de la collecte, de la codification et de la saisie de l'information (4 mois), si les opérations de codification et de saisie sont effectuées en Equateur, le coût par logement enquêté est de :

- 12 FF pour un questionnaire démographique tel que celui pratiqué lors de l'enquête Bâti/ population : âge, sexe, lieu de naissance, durée d'occupation du logement, durée de résidence à Quito, lieu de résidence antérieur, conditions d'occupation du logement, activité, autres sources de revenus (pour chacune des personnes vivant dans le logement enquêté) ;

- 16 FF pour un questionnaire démographique tel que celui pratiqué lors de l'enquête Démographie/ migrations : outre les caractéristiques démographiques générales de l'enquête Bâti/population, biographie migratoire et professionnelle du chef de ménage, recueil intégral de tous les changements de résidence et d'activité du chef de ménage durant les deux dernières années, et caractéristiques d'activité et de résidence des membres de la famille ne vivant pas dans le logement enquêté.

2. Coût de la sélection de l'échantillon de l'enquête démographique par tirage aréolaire systématique stratifié sur image SPOT

La mise en oeuvre d'un plan de sondage aréolaire systématique stratifié sur image satellite

entraîne les dépenses suivantes :

	EQUIPEMENT	FONCTIONNEMENT
IBM AT	15 000 FF	
Logiciels	19 600 FF	
Planimètre		
Boussole, clinomètre décamètre	4 000 FF	
Achat et traitements d'une image SPOT panchro et XS (scène existante)		25 000 FF
Relevés morphologie urbaine		Collecte : 56,4 FF par ilot Codif.: 1 mois mi-temps chercheur Saisie : 3 FF par ilot
Sélection échantillon		Morpho. : 9,5 jours, 2 chercheurs Dém. : 17 j, 2 chercheurs
Mesure des ilots		9,36 FF par ilot

Pour une enquête démographique auprès de 3000 ménages répartis dans 400 ilots, les coûts afférant à la sélection de l'échantillon se répartiraient donc comme suit :

- Equipement : 38 600 FF
- Fonctionnement :
 - achat et traitements image SPOT : 9 500 FF
 - relevés de morphologie urbaine (200 ilots) : 11 880 FF
 - mesures de 400 ilots sur le terrain : 3 750 FF
 - 3 mois de salaire du personnel qualifié pour le tirage de l'échantillon (exploitation de l'image, mise au point et réalisation du plan de sondage)

Cette estimation ne tient pas compte de la réalisation d'une enquête sur les relations bâti/population, telle que celle que nous avons réalisé à Quito ; au cas où on ne disposerait d'aucune information, même ancienne, sur la démographie de la ville à enquêter, cette enquête préliminaire serait nécessaire à la définition du plan de sondage, pour déterminer l'allocation aux strates. Son coût serait alors très variable suivant l'objectif de l'enquête finale. Lorsque l'objectif est d'estimer les effectifs et les caractéristiques générales de la population, on peut réaliser une enquête très légère dont le coût, si l'on se base sur les coûts de collecte, codification et saisie donnés dans ce paragraphe, ne devrait pas excéder 6.050 FF¹. En revanche si l'on s'intéresse à des catégories de population très minoritaires

¹ : Ce coût se décompose de la façon suivante :

- Collecte, 50 ilots à 10 logements enquêtés par ilots : 4000 FF
(7,96 FF par logement)
- Codification et saisie, fichier de 500 enregistrements : 2050 FF
(4,10 FF par logement, c'est à dire par enregistrement)

(migrants ou population d'une tranche d'âge réduite etc...), ou à des événements "rares" (natalité, mortalité, fécondité etc...), l'enquête devrait alors porter sur un grand nombre de ménages ; dans ce cas, il paraît plus raisonnable de s'en dispenser et de choisir simplement le même taux pour toutes les strates.

3. Coût total de l'enquête selon la taille de l'échantillon

Nous considérons deux tailles d'échantillon qui pourraient être retenues pour réaliser une enquête démographique par sondage dans une ville de 1 million d'habitants :

	Ech. 1 : 400 ilots, 3000 ménages	Ech. 2 : 800 ilots, 6000 ménages
EQUIPEMENT	38 600 FF	38 600 FF
FONCTIONNEMENT Achat et traitements d'une image SPOT panchro et XS (scène existante)	25 000 FF	25 000 FF
Relevés morphologie Sélection échantillon Mesure ilots	14 950 FF	18 700 FF
Collecte démo. (collecte, codif., saisie questio. complet)	49 950 FF	99 900 FF
COUTTOTAL	128 500 FF + 7 mois personnel qualifié	182 200 FF + 7 mois personnel qualifié
COUT PAR MENAGE (équipement inclus, person. qualif. exclus)	42, 8 FF	30, 4 FF

Du fait de l'existence de postes fixes de dépenses importants (équipement, achat et exploitation de l'image satellite avec initialisation sur 200 îlots, premières étapes nécessaires à la sélection de l'échantillon sur l'image), le coût d'enquête par ménage décroît sensiblement lorsque la taille de l'échantillon augmente.

15.2 : Coûts nécessaires à l'obtention d'une précision fixée à l'avance

15.2.1 : Les fonctions de coût attachées aux types d'enquêtes réalisées à Quito

Les données du paragraphe précédent permettent d'évaluer les coûts afférent à deux types d'enquêtes qui ont été différenciées :

- des enquêtes visant le recueil des caractéristiques démographiques générales de la population telles que l'enquête que nous appelons à Quito "Bâti / Population" et que nous appellerons ici pour simplifier enquête à "questionnaire court".

- des enquêtes où le questionnaire a une taille comparable à celle du questionnaire de l'enquête "Démographie / Migrations" ; nous dirons ici enquête à "questionnaire long".

Nous tenterons maintenant de construire une fonction de coût pour ces deux types d'enquête¹.

D'après le paragraphe précédent, il est clair que ces fonctions comprendront un terme constant (d'ailleurs égal dans les deux cas), qui représente les postes de dépense fixes (c'est à dire n'étant pas affectés par la taille de l'échantillons) : matériel informatique, achat et traitement de l'image satellite et réalisation de l'enquête préliminaire sur la morphologie urbaine, et un terme variant en fonction de la taille de l'échantillon et de l'importance du questionnaire. Nous appellerons C_c le terme constant, C_{1v} et C_{2v} respectivement, les termes variables dans le cas du questionnaire court (1) et du questionnaire long (2). Ainsi, pour chaque type d'enquête, on aura les coûts globaux :

$$C_{1T} = C_{1v} + C_c$$

$$C_{2T} = C_{2v} + C_c$$

Evaluation du terme constant

Nous ne prendrons pas en compte dans ce qui va suivre les dépenses liées à la réalisation d'une enquête préliminaire ayant pour but d'estimer les ordres de grandeur des variances des caractères à estimer dans chaque strate, ni celles liées aux salaires des personnels qualifiés encadrant l'enquête. Ces deux postes sont, nous l'avons vu, très variables et il nous paraît préférable qu'ils soient chiffrés pour chaque application spécifique. Notons qu'il s'agit bien de postes fixes ; il suffira donc d'ajouter ces dépenses au terme constant, que nous allons maintenant chiffrer grâce aux données du paragraphe 15.7.

matériel informatique :	38 600 F
achat et traitement de l'image satellite :	25 000 F
réalisation de l'enquête préliminaire sur la morphologie urbaine :	11 880 F
Total (C_c)	75 480 F

¹ : Tous les coûts donnés ici sont exprimés en Francs Français.

Evaluation des termes variables

Plusieurs opérations doivent être intégrées dans ces termes : la mesure des surface des îlots, le recensement des ménages et la collecte démographique sur l'échantillon de ménages (deux opérations que nous ne séparerons pas ici), la codification de l'information et enfin, sa saisie informatique. Dans les données du paragraphe précédent, les coûts relatifs à chacune de ces opérations sont exprimé par îlot mesuré pour la première et par logement occupé pour les trois autres :

- Mesure des surface des îlots :	9,36 F par îlot
- Recensement des ménages et collecte démographique sur l'échantillon de ménages :	
Questionnaire court :	7,96 F par logement
Questionnaire long :	8,31 F " "
- Codification de l'information :	
Questionnaire court :	1,27 F " "
Questionnaire long :	2,70 F " "
- Saisie informatique :	
Questionnaire court :	2,58 F " "
Questionnaire long :	5,64 F " "

Dans l'examen général qu'il fait des structures de sondages à deux degrés, DESABIE¹ propose d'exprimer le coût de la collecte en fonction des paramètres m , nombre d'unités primaires, et \bar{n} , nombre moyen d'unités secondaires par unité primaire :

$$C = c_1 m + c_2 m \bar{n}, \text{ où :}$$

c_1 représente approximativement le coût des déplacements pour atteindre chaque unité primaire, y dresser éventuellement la liste des unités secondaires, et c_2 représente le coût de collecte des renseignements pour une unité secondaire."

En ce qui nous concerne, nous ne séparerons pas, comme nous l'avons dit, le coût d'établissement des listes d'unités secondaire du coût global de la collecte ; en revanche la mesure de surface des îlots est opération attachée aux seules unités primaire. Nous aurons donc comme valeur de c_1 le coût moyen de cette mesure : 9,36 F. Dans c_2 , nous ajouterons aux coût de collecte les coûts de codification et de saisie pour obtenir l'ensemble des coûts attachés aux unités secondaires (plus exactement aux logements occupés, c'est à dire enquêtés) :

$$\text{- pour un questionnaire court : } c_2 = 7,96 + 1,27 + 2,58 = 11,81$$

$$\text{- pour un questionnaire long : } c_2 = 8,31 + 2,70 + 5,64 = 16,65$$

Enfin la valeur de \bar{n} correspond pour nous au nombre moyen de logements enquêtés par îlot lors de l'enquête "Démographie / Migration" soit : $\bar{n} = 3157/426 = 7,41$. Nous obtenons donc les expressions suivantes des termes variables :

$$\text{- pour un questionnaire court : } C_{1V} = 9,36 m + 11,81 \times 7,41 m = 96,87 m$$

$$\text{- pour un questionnaire long : } C_{2V} = 9,36 m + 16,65 \times 7,41 m = 132,74 m$$

¹ : J. DESABIE, 1966 : Théorie et pratique des sondages, pp. 264 et 265.

Les fonctions de coûts associées aux deux types d'enquêtes sont donc, m étant, rappelons le, le nombre d'unités primaires de l'échantillon, c'est à dire le nombre d'îlots sélectionnés :

- pour un questionnaire court : $C_{1T} = C_{1v} + C_c = 75\,480 + 96,87 m$
- pour un questionnaire long : $C_{2T} = C_{2v} + C_c = 75\,480 + 132,74 m$

15.2.2 : La fonction précision

Comme nous l'avons dit en introduction, nous négligerons ici l'effet du second degré du sondage sur la variance des estimations. De plus, comme au chapitre 12, nous approximerons le tirage effectué par un tirage avec remise pour simplifier l'expression de la variance. repartons donc de l'expression de la variance de l'estimation d'un total donnée dans le tableau 12.3 :

$$V(\hat{Y}_k) = \frac{z_k}{m} \left(\left(\sum_{i=1}^M \frac{y_i^2}{z_i} \right) - \frac{Y_k^2}{z_k} \right), \text{ que l'on peut écrire : } V(\hat{Y}_k) = \frac{1}{m} S_y^2, \text{ où la quantité}$$

$$S_y^2 = \left(\sum_{i=1}^M \frac{z_k y_i^2}{z_i} \right) - Y_k^2, \text{ est constante quelque soit } m.$$

Exprimons, en fonction de m et de S_y^2 , l'amplitude de l'intervalle de confiance à 95 %, en pourcentage du total à estimer Y_k :

$$A = \frac{2 \times 100}{m^{1/2}} \times \frac{(S_y^2)^{1/2}}{Y_k}$$

A Quito, pour le total de population hors ménages collectifs, les estimations sont les suivantes (avec $m = 426$):

$$\hat{Y}_k = 1\,120\,991$$

$$\hat{V}(\hat{Y}_k) = 5\,694\,772\,444, \text{ soit } (\hat{S}_y^2)^{1/2} = (5\,694\,772\,444 \times 426)^{1/2} = 1\,557\,553,5$$

On obtient donc la fonction de précision suivante (représentée sur la figure 15.1) :

$$A \approx \frac{278}{\sqrt{m}}$$

15.2.3 : Budget nécessaire pour atteindre une précision fixée

A partir des expressions donnant le coût et la précision des enquêtes en fonction de la taille de l'échantillon, on exprime facilement le coût d'une précision donnée :

- pour un questionnaire court : $C_{1T} = 75\,480 + \frac{7\,475\,733}{A^2}$

- pour un questionnaire long : $C_{2T} = 75\,480 + \frac{10\,243\,923}{A^2}$;

où la précision est mesurée par A : l'amplitude de l'intervalle de confiance à 95 %, en pourcentage du total à estimer. Sur la figure 15.2 on a représenté ces deux paraboles pour des précisions s'étallant entre +/- 30% et +/- 5%. Le tableau 15.3, quant à lui, résume les résultats obtenus (coûts et précision) avec une série d'échantillons dont les tailles varient entre 85 et 3 000 ilots.

Nous ne nous risquons pas ici à faire un commentaire "économique" de ces résultats, contentons nous de quelques remarques.

1. Du fait de l'existence des frais fixes liés au matériel, à l'achat et au traitement de l'image, et compte tenu également des précisions médiocres que l'on obtiendrait alors, il paraît peu "rentable", si l'on est obligé d'effectuer l'ensemble de ces dépenses fixes, d'appliquer cette méthode à des enquêtes trop légères. Pour des échantillons inférieurs à 250 ilots, l'investissement de départ représente plus de 75% du coût total de l'enquête et l'amplitude des intervalles de confiance reste importante (supérieure à 20% du total de population). En revanche, si l'on se trouve dans un contexte où l'achat du matériel et de l'image ne sont pas obligatoires, la méthode peut être appliquée à des villes beaucoup plus petites que celles où nous l'avons testée.

2. Pour des tailles d'échantillons inférieures à 1200 ilots, c'est à dire pour des intervalles de confiance supérieurs à +/- 8%, la pente de la parabole donnant le coût en fonction de la précision reste inférieure à 2. Le coût d'une augmentation de la précision est alors raisonnable : pour passer de +/- 30 % à +/- 10 %, il faut doubler le budget c'est à dire en fait, consacrer autant d'argent à l'enquête proprement dite qu'aux frais fixes. En revanche, dans la partie supérieure de la courbe (tailles d'échantillons supérieures à 1200 ilots), une petite augmentation de précision coutera très cher : il faut à nouveau doubler le budget pour passer de +/- 8% à +/- 5%.

3. Toujours à cause des frais fixes, le coût du passage d'un questionnaire "court" à un questionnaire "long" est faible pour les échantillons de tailles "petites" ou "moyennes" : j'usqu'à 1000 ilots enquêtés il représente moins de 20% (30 000 F) du budget de l'enquête "courte". La différence s'accroît ensuite : 25 % (72 000 F) du budget pour 2 000 ilots et 30 % (108 000 F) pour 3 000 ilots.

Figure 15.1 : Précision en fonction de la taille de l'échantillon

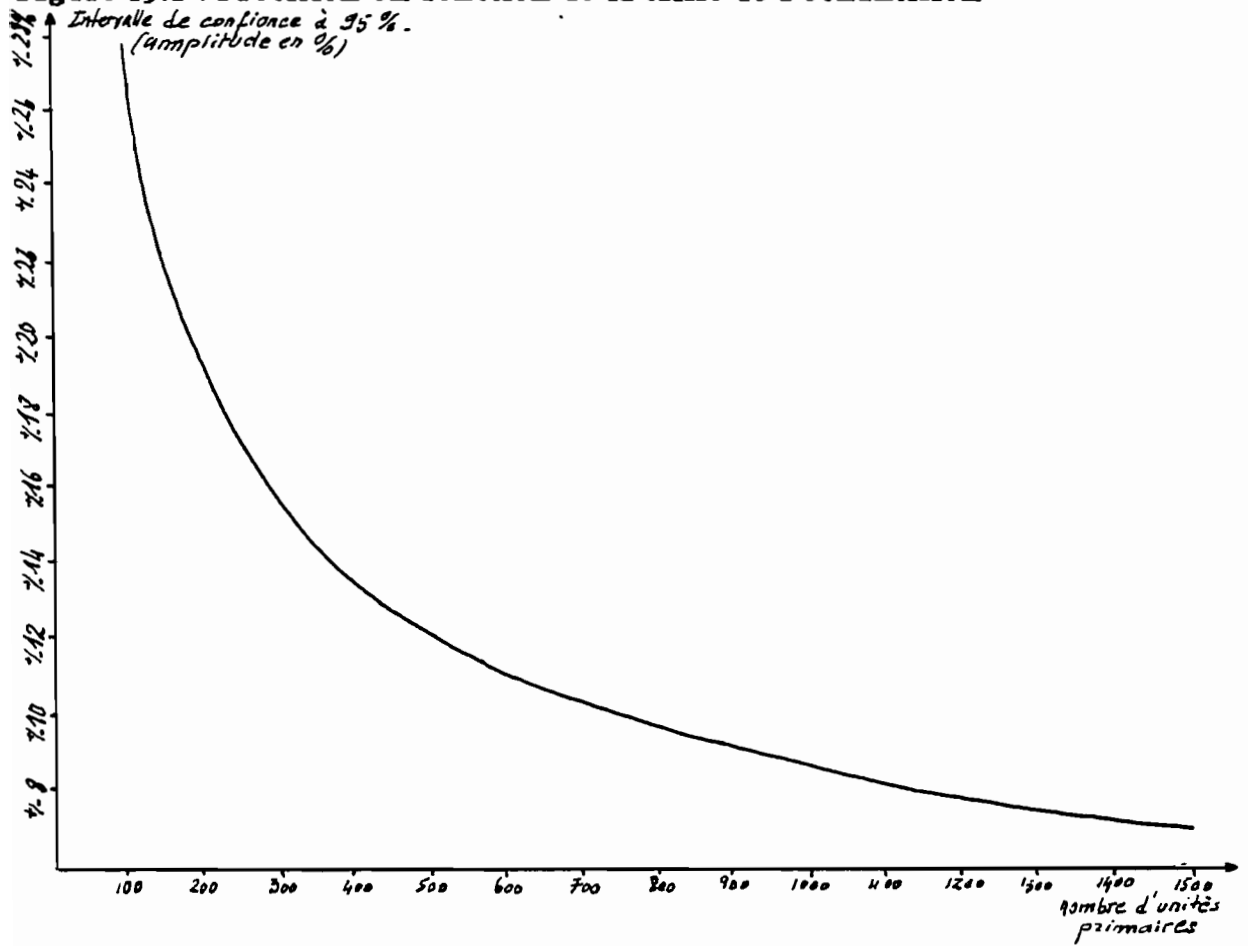


Figure 15.2 : Coûts nécessaires à une précision donnée

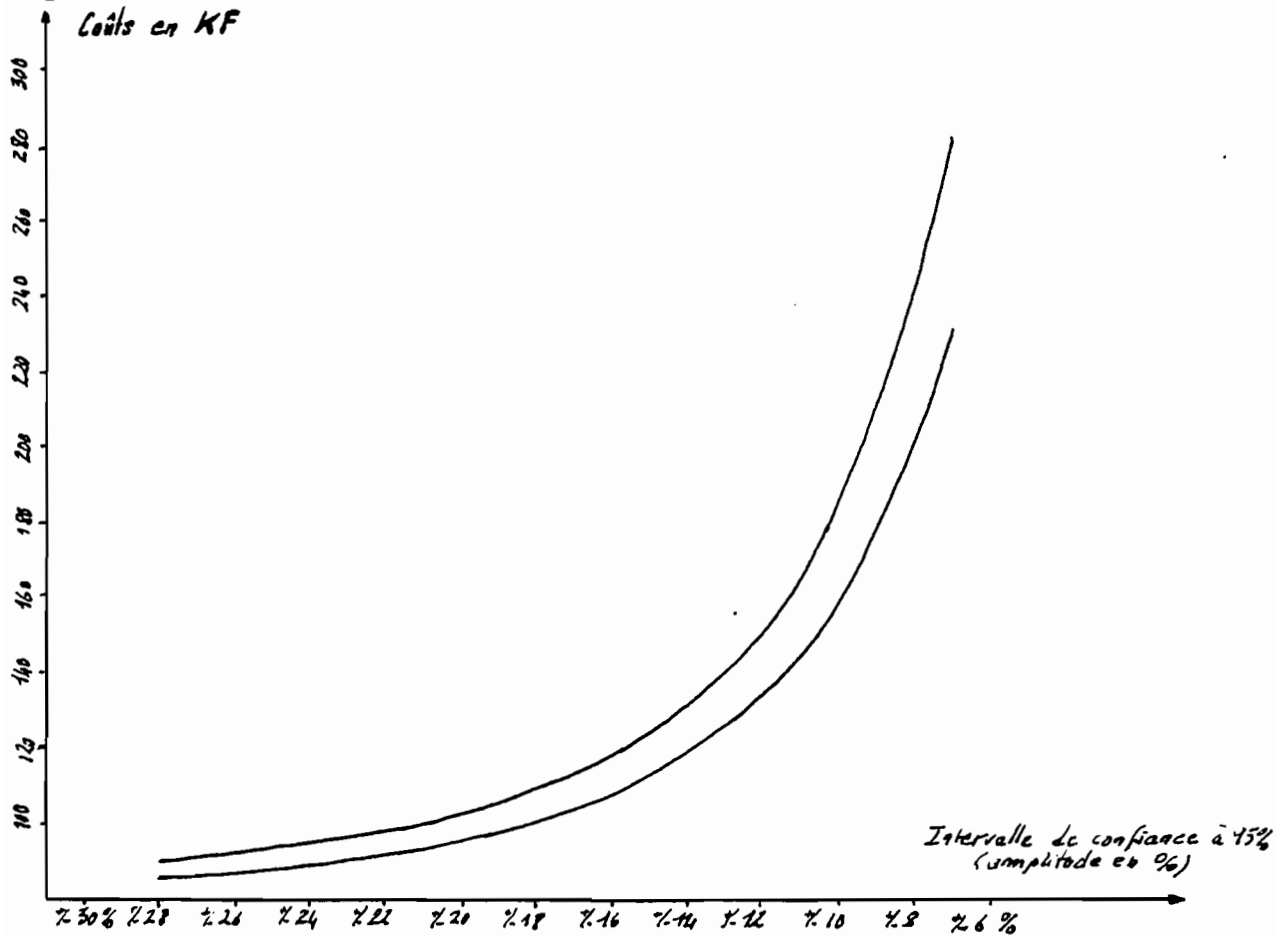


Tableau 15.3 : Taille des échantillons, coûts et précisions pour l'enquête par sondage pratiquée à Quito.

m Taille des échantillons (Nombre d'îlots enquêtés)	M Nombre de logements enquêtés	P Nombre de personnes enquêtées	C _{1T} Coût question- naire court (FF)	C _{2T} Coût question- naire long (FF)	A Précision : intervalle de confiance à 95% (en % de la pop. totale)
85	640	2 900	84 000	87 000	+/- 30,0
100	740	3 400	85 000	89 000	+/- 27,8
150	1 110	5 000	90 000	95 000	+/- 22,7
200	1 480	6 700	95 000	102 000	+/- 19,6
250	1 850	8 400	100 000	109 000	+/- 17,6
300	2 220	10 100	105 000	115 000	+/- 16,0
400	2 960	13 400	114 000	129 000	+/- 13,9
500	3 700	16 800	124 000	142 000	+/- 12,4
600	4 450	20 200	134 000	155 000	+/- 11,3
800	5 930	26 900	153 000	182 000	+/- 9,8
1 000	7 410	33 600	172 000	208 000	+/- 8,8
1 200	8 890	40 300	192 000	235 000	+/- 8,0
1 500	11 120	50 400	221 000	275 000	+/- 7,2
2 000	14 820	67 200	269 000	341 000	+/- 6,2
2 500	18 520	84 000	318 000	407 000	+/- 5,6
3 000	22 230	100 000	366 000	474 000	+/- 5,0

Avec :

$$M = 7,41 \times m \quad C_{1T} = 96,87 \times m + 75\,470 \quad A \approx \frac{278}{\sqrt{m}}$$

$$P = 33,6 \times m \quad C_{2T} = 132,74 \times m + 75\,470$$

Conclusion

Pour conclure, soulignons que, pour une taille d'échantillon fixée, les coûts comme les précisions varieront en fonction du site d'application : le contexte économique du pays influera sur les coûts de mise en œuvre et les spécificités morphologiques et démographiques de la ville étudiée se répercuteront sur la précision des estimations. De plus, comme nous l'avons vu au chapitre précédent, la méthode peut faire l'objet de nombreuses adaptations aux conditions particulières dont on peut bénéficier sur le site. Nous ne pouvions pas chiffrer ici les conséquences de l'ensemble de ces facteurs. L'évaluation que nous venons de présenter, basée sur l'expérience que nous avons réalisée à Quito, n'a donc qu'une valeur d'exemple. Néanmoins, à partir d'une part, des éléments de coûts donnés dans la première partie, que l'on peut modifier en fonction du contexte économique de l'application, et d'autre part du formulaire du chapitre 12, qu'il faut adapter aux modifications éventuelles du plan de sondage, on peut obtenir des approximations des fonctions de coût et de précision qui caractérisent cette méthode sur un autre site.

CONCLUSION

**SYNTHESE DES PRINCIPAUX RESULTATS, PERSPECTIVES POUR LES
RECHERCHES FUTURES**

CONCLUSION

SYNTHESE DES PRINCIPAUX RESULTATS, PERSPECTIVES POUR LES RECHERCHES FUTURES

Au début de ce travail, nous nous posions la question suivante : face à la demande croissante d'information sur les villes des pays en développement et les modalités de leur croissance, que proposer pour tenter d'améliorer la collecte démographique en milieu urbain? Les coûts et les délais d'exploitation des recensements sont tels que cette source classique de données n'est pas adaptée à l'observation des villes du tiers monde et à la croissance rapide qu'elles connaissent en majorité. C'est donc dans l'amélioration des techniques de sondage qu'il convient de rechercher la solution au problème. Bien que régulièrement utilisées, particulièrement dans les pays en développement, ces techniques d'enquêtes se heurtent souvent au problème de la constitution ou de l'actualisation de la base de sondage indispensable à leur mise en œuvre. En démographie, le recours à des listes de ménages (les plus exhaustives possibles), le plus souvent issue des recensements, est la solution généralement pratiquée. Dans les villes qui nous intéressent, l'absence de ces listes, ou leur caractère obsolète, sont très fréquemment des obstacles insurmontables à la réalisation d'enquêtes par sondage avec une périodicité intéressante. Ce sont ces considérations qui ont donné naissance au projet de recherche dans lequel s'est inscrit ce travail : *utiliser les images satellite à haute résolution comme base de sondage spatiale et comme source d'informations morphologiques pour stratifier cette base doit permettre de s'affranchir de la plupart des contraintes qui limitent encore l'emploi des techniques de sondages classiques en milieu urbain et donc de mettre au point une méthode de collecte par sondage de données démographiques rapide, peu chère et précise, bref une méthode adaptée à l'observation des villes du tiers monde.*

Ayant rappelé quel fut l'objectif principal de ce travail, nous voudrions, au terme de cet exposé, faire une rapide synthèse des résultats obtenus, c'est à dire caractériser dans ses grandes lignes la méthode mise au point, dresser un premier bilan de ses performances (précision, praticabilité et coût) et enfin évoquer les possibilités d'amélioration de cette méthode en suggérant quelques directions pour les recherches futures.

1. La méthode mise au point.

Si la méthode développée se veut souple dans ses modalités d'application et donc adaptable aux conditions particulières des sites urbains ou les thèmes d'enquêtes, en revanche les contraintes liées à l'utilisation d'une base de sondage spatiale comme l'image satellite nous impose un type de sondage bien particulier. Cette spécificité se répercute à trois niveaux : sur la structure générale du sondage, sur la définition précise du plan de sondage à appliquer et enfin sur les étapes de la mise en œuvre de l'enquête.

Structure générale : un sondage aréolaire stratifié à deux degrés

L'utilisation de l'image satellite pour définir et mettre en œuvre un sondage démographique entraîne deux contraintes principales liées, pour la première, à l'emploi de l'image comme base spatiale de sondage, et pour la seconde, à l'utilisation de l'information morphologique issue de l'image pour améliorer la précision des estimations.

- Pour substituer l'image satellite à une base de sondage classique d'enquête démographique en milieu urbain, il faut prendre en compte une double contrainte : d'une part les unités d'échantillonnage doivent être identifiables sur l'image qui servira de base, d'autre part, dans le cadre qui est le nôtre d'observation des caractéristiques de la population, l'unité statistique enquêtée reste le ménage. L'îlot physique étant la seule unité spatiale repérable sans ambiguïté sur l'image, le sondage sera donc un sondage aréolaire à deux degrés ; au premier degré les aires sélectionnées sont des îlots (unités primaires) et au second degré les unités enquêtées sont des ménages (unités secondaires).

- L'hypothèse fondamentale sur laquelle repose l'idée qu'on peut améliorer la précision du sondage en utilisant une information exogène sur la morphologie des quartiers est qu'il existe, entre cette information et les variables démographiques que nous cherchons à estimer, des liaisons statistiques significatives. Pour exploiter ces liaisons dans la définition du plan de sondage, plusieurs techniques, que nous ne rappellerons pas ici, ont été envisagées au cours de cette recherche. Un des principaux résultats obtenus est que la technique la plus efficace (on réduit d'environ 40% l'erreur d'échantillonnage) et la mieux adaptée aux possibilités actuelles d'interprétation des images satellite en milieu urbain (les classifications d'images sur ces thèmes ont fait leur preuves), est la stratification de la base de sondage sur des critères morphologiques. La stratification retenue lors de l'expérience finale à Quito prend en compte le seul critère de la densité du bâti, mais d'ores et déjà les spécialistes en télédétection de l'équipe ont pu produire des stratifications plus complexes, certainement plus efficaces. Quoiqu'il en soit, pour tirer tout le parti de l'information morphologique que recèlent les images satellites, un sondage basé sur ces documents se doit d'être stratifié en utilisant cette information.

Plan de sondage

La définition complète d'un plan de sondage praticable dans le contexte que nous venons de décrire est le résultat principal de cette recherche. Les étapes et les résultats de ce travail ont été décrits en détail dans les deuxième et troisième parties de l'exposé, nous ne ferons ici que reprendre les principales conclusions exposées au chapitre 12.

1. Base de sondage

L'ensemble de l'aire métropolitaine est matérialisée sur l'image satellite par la limite de la zone considérée comme urbaine ; à l'intérieur de cet ensemble, les espaces non bâtis sont exclus de la base de sondage. C'est aux urbanistes et aux géographes, plus généralement aux thématiciens qui conçoivent l'enquête et le domaine d'étude, de définir, en collaboration avec le statisticien chargé du plan de sondage, les critères qui présideront à l'élaboration des limites de la base.

2. Stratification

Lors de l'enquête réalisée à Quito, l'agglomération était stratifiée en cinq niveaux de densité du bâti, d'après la classification de l'image satellite sur ce thème. Cette stratification, rappelons le "expérimentale", ne doit évidemment pas être considérée

comme une référence et le champ ouvert pour son amélioration apparaît dès maintenant très vaste. A Quito, grâce à une bonne stabilité des règles d'allocations optimales aux strates pour l'ensemble des caractères que nous voulions estimer, nous avons pu définir, à partir des données fournies par l'enquête démographique préliminaire, une allocation qui améliore l'ensemble des estimations. Nous pensons que la démarche adoptée peut être appliquée à d'autres sites, en utilisant les résultats d'une enquête préliminaire ou d'autres sources disponibles.

3. Sélections des échantillons aux premier et second degrés

Un autre résultat important de la recherche est qu'un sondage spatial systématique, parce qu'il assure une bonne répartition géographique de l'échantillon, améliore la précision des estimations (diminution de 20 à 30% de la variance par rapport au sondage strictement aléatoire au même taux). Au premier degré du sondage, on peut donc recommander de pratiquer un sondage aréolaire systématique : la sélection des îlots de l'échantillon se fait à l'aide d'une grille de points, le placement initial de la grille sur l'image étant aléatoire. S'agissant du second degré du sondage, la comparaison des variances inter et intra-unités primaires, d'où ressort le caractère marginal de la seconde, permet de conclure à l'intérêt de pratiquer le taux le plus fort possible au premier degré, dans la limite du budget disponible et sous la contrainte de conserver un nombre suffisant de ménages enquêtés dans chaque pâté de maison si l'on tient à estimer la variance au second degré. Une solution conciliant ces différents impératifs est l'adoption d'un taux variable au second degré suivant le nombre de ménages vivant dans l'îlot. Nous avons pour notre part enquêté dix ménages dans tous les îlots comptant plus de cet effectif, et la totalité des ménages partout ailleurs. Là encore, il ne s'agit nullement d'une règle stricte et la diminution du taux au second degré peut être une solution pour améliorer la précision des résultats d'ensemble. La sélection au second degré se fait de façon systématique sur une liste de ménages, elle peut être considérée comme équiprobable, ce qui n'est pas le cas, comme nous allons le voir, du tirage au premier degré.

4. Inférence

En ce qui concerne les estimateurs employés, du moins au premier degré, la spécificité du sondage que nous proposons découle de l'emploi d'une base de sondage spatiale (plus exactement de la définition des unités primaires qu'elle entraîne) et du mode de sélection de l'échantillon. Un tirage par points appliqué à des unités spatiales de tailles variables induit des probabilités d'appartenance à l'échantillon inégales. Les estimateurs qu'il convient d'utiliser, tant pour l'estimation sans biais des totaux que pour celle légèrement biaisée des quotients, intègrent donc, pour chaque unité primaires des sommes calculées, une pondération par l'inverse de la probabilité d'inclusion. Le formulaire développé dans la seconde partie et résumé dans le tableau 12.3, correspond à la théorie du sondage à probabilités inégales avec remise ; il a une portée générale, c'est à dire qu'il peut être appliqué quelle que soit la probabilité d'inclusion retenue. Dans l'expérience réalisée à Quito, la technique de sélection des unités primaires utilisée fournit, comme nous l'avons vu, des probabilités d'appartenances proportionnelles aux surfaces des pâtés de maisons. Les tests effectués concernant d'autres mesures de la taille des unités primaires à partir desquelles on pourrait définir d'autres probabilités d'inclusion n'ont pas débouché pour l'instant sur des conclusions pratiquement exploitables, mais il s'agit là nous le verrons d'une des voies de recherche prometteuse.

Mise en œuvre de l'enquête

Pour mettre en œuvre le plan de sondage défini et réaliser les opérations de collecte de l'information, deux étapes sont nécessaires. La première a pour but le recensement des ménages dans les îlots de l'échantillon ainsi que la mesure sur le terrain de la

surface de ces îlots si cette dernière ne peut être effectuée directement sur l'image. La seconde correspond au recueil de l'information visée par l'enquête, c'est à dire au remplissage des questionnaires pour chaque ménage sélectionné au second degré. Les techniques de collecte appliquées à chacune de ces étapes lors de l'enquête réalisée à Quito, qu'il serait trop long de rappeler ici, ont démontré leur efficacité. Le plan de sondage développé est praticable et permet de recueillir, rapidement et avec une bonne fiabilité (c'est à dire avec peu d'erreurs de mesure), une grande quantité d'information. Quant à l'erreur d'échantillonnage, nous pensons à l'issue des tests menés à Marseille et à Quito que l'effet cumulé de la stratification avec allocation optimale et du tirage systématique spatial permet de réduire l'amplitude des intervalles de confiance de 30 à 80 % selon les caractères estimés par rapport à un sondage aléatoire non stratifié. La première exploitation faite des résultats de cette enquête permet de chiffrer plus précisément les performances de la méthode.

2. Les performances

Comme nous l'avons dit en présentant dans le chapitre 13 les premiers résultats de l'enquête Démographie/migration, l'exploitation de cette première application "en vraie grandeur" de la méthode n'en est qu'au tout début. L'évaluation des performances de cette méthode d'après ses résultats ne peut donc être que très incomplète. Néanmoins, les éléments rassemblés dans la quatrième partie de l'exposé ont permis d'obtenir des indications intéressantes sur la précision des estimations, les domaines et conditions d'application de la méthode et enfin les coûts et délais nécessaires à sa mise en œuvre.

Précision des estimations

Les chiffres de précisions que l'on obtient sur les premières estimations effectuées doivent être appréciés en tenant compte de deux éléments. D'une part la taille relativement réduite de l'échantillon imposée par le budget attribué à l'expérience, d'autre part la surestimation de l'erreur d'échantillonnage due à l'emploi du formulaire avec remise et surtout à la méthode retenue pour évaluer la variance du sondage systématique pratiqué (les chiffres donnés ne tiennent compte d'aucune de ces deux surestimations). Avec un échantillon de 14 500 personnes réparties dans 426 îlots, soit environ 1,3 % de la population totale, le total de population de Quito a été estimé à 1 120 991 habitants avec un coefficient de variation de 6,73%. Ces coefficients sont évidemment très variable suivant les indicateurs démographiques estimés ; par exemple la taille moyenne du ménage (4,53 personnes) est estimée avec un coefficient de 1,10% et le pourcentage de population âgée de six ans ou plus ayant fait des études supérieures (18,6 %) avec un coefficient de 3,68 %.

La comparaison de ces précisions à celles que fournissent d'autre méthodes de sondage est rendue très difficile par la rareté des indications que fournit la littérature spécialisée concernant la précisions des différentes méthodes de sondage appliquées à des enquêtes démographiques en milieu urbain. La tentative que nous avons pu effectuer, d'ailleurs très imparfaite, devra être répétée à partir d'informations plus complètes pour statuer définitivement sur les qualités de précision du plan de sondage et de la méthode dans son ensemble. Pour l'instant on peut simplement affirmer, à partir des seules indications sommaires dont nous disposons concernant la précision des sondages démographiques équiprobables sur liste, que lorsque les caractères démographiques à estimer sont bien corrélés à la morphologie des quartiers, la méthode que nous proposons est vraisemblablement nettement plus précise qu'un sondage aléatoire simple.

Enfin, concernant la précision du sondage mis au point, il faut rappeler que des possibilités d'améliorations existent d'ores et déjà, qui ne nécessitent aucune autre

recherche méthodologique (elles ne sont d'ailleurs pas spécifiques à la méthode particulière que nous avons développée au cours de ce projet de recherche), et dont les répercussions sur le coût de la mise en œuvre de l'enquête sont nulles ou raisonnables :

- la stratification morphologique simple, constituée sur le seul critère de la densité du bâti, peut être améliorée, sans aucune dépense supplémentaire, par la prise en compte d'autres critères en faisant simplement appel à la connaissance empirique de la ville qui existe toujours dans les milieux de la gestion et de la recherche urbaines ;

- l'augmentation du nombre d'ilots sélectionnés au premier degré, compensée par une diminution équivalente du nombre de ménages enquêtés dans chaque ilot permet, pour un même nombre total de ménages enquêtés d'améliorer significativement la précision des estimations d'ensemble, à nouveau sans dépenses supplémentaires notables ;

- enfin au chapitre 15, nous avons montré que, du fait de l'existence de frais indépendants de la taille de l'échantillon, une augmentation raisonnable de la taille de l'échantillon n'entraîne qu'un surcoût limité ; par exemple pour réduire de 30% l'amplitude des intervalles de confiance, il faudrait doubler la taille de l'échantillon et il en coûterait une augmentation d'environ 40% du budget.

Si nous n'avons retenu aucune de ces améliorations lors de l'application de la méthode à Quito, c'est uniquement, rappelons le, pour respecter les contraintes qui étaient fixées dans le plan de l'expérience : respect d'une stratification à tester concernant la première possibilité, nécessité d'estimer la variance au second degré pour la seconde et contrainte de budget pour la troisième.

Domaines et conditions d'application de la méthode

Nous pensons que les qualités de souplesse et d'économie de la méthode en font un outil applicable à un large domaine thématique et géographique.

Le domaine thématique d'application de ce type d'enquête par sondage est principalement déterminé par deux contraintes : d'une part l'enquête doit porter sur des individus statistiques compatibles avec la structure du sondage proposé et d'autre part, si l'on veut que la méthode soit efficace, il faut qu'existent certaines corrélations entre les thèmes abordés dans le questionnaire et la morphologie urbaine. Par rapport à la première contrainte, l'utilisation de l'image satellite comme base de sondage est un principe applicable à tout domaine thématique où l'on observe des unités statistiques rattachées à des unités d'échantillonnage repérables sans ambiguïté sur l'image ; en l'état actuel des possibilités d'interprétation des images satellite en milieu urbain, l'ilot physique est l'unité d'échantillonnage praticable. Le domaine thématique concerné par ce type d'unité couvre donc théoriquement tout le champ des sciences sociales en milieu urbain : géographie humaine, démographie, sociologie, économie, épidémiologie et médecine sociale etc... Reste à savoir si les liaisons statistiques existant entre ces différents domaines potentiels d'application et l'information morphologique permettront que les performances de la méthode soient équivalentes à celles que nous avons obtenues lors de la mise au point de la méthode dans le domaine démographique. La réponse définitive à cette question ne peut venir que de l'accumulation des expériences d'application de la méthode à ces différents thèmes.

Les sites potentiels d'application de la méthode sont eux aussi très nombreux. L'équipement informatique nécessaire à la définition du plan de sondage et à l'exploitation de l'enquête n'est pas un obstacle à sa portabilité puisque l'ensemble des tâches peuvent être réalisées sur une configuration micro-informatique standard. Le traitement de l'image peut être limité à la production des seules sorties sur papier de canaux, néo-canaux ou classifications nécessaires à la définition et à la stratification de la base de sondage ; il sera éventuellement sous-traité à une institution nationale ou

étrangère, possédant l'équipement nécessaire. En outre on peut penser à utiliser la même méthode de sondage à partir d'une autre base de sondage spatiale existant sur le site : cartes détaillées, cadastre, photographies aériennes. Les limites de l'application de la méthode tiennent plutôt à la taille de la ville à laquelle on pense l'appliquer et à l'exigence de précision que l'on se fixe : pour conserver la précision obtenue à Quito sur une ville de 100 000 habitants, le taux de sondage devrait être d'environ 15% et le coût de l'enquête peut alors être lourd relativement au budget dont on dispose dans ce type de villes. Enfin rappelons que l'utilisation de l'image satellite suppose l'acquisition d'images exploitables et par conséquent : une étude préliminaire des probabilités d'obtention de ces images sur le site ou l'on envisage d'appliquer la méthode est fortement recommandée.

Coûts et délais de la mise en œuvre de la méthode

Nous ne reviendrons pas ici sur les indications chiffrées fournies au chapitre 15 qui ont conduit aux expressions des fonctions de coût et de précision qui caractérisent la méthode telle qu'elle a été appliquée à Quito. Ces chiffres et les fonctions qui en découlent devront d'ailleurs être adaptés en fonction du contexte économique du pays où se déroule l'enquête et des conditions particulières dont on peut bénéficier sur le site. Là encore, c'est la démarche qui reste valide. On retiendra simplement des conclusions de ce chapitre deux faits marquants.

Premièrement l'analyse conjointe des coûts et des précisions qui leur correspondent a permis de définir un intervalle de variation de la taille de l'échantillon au sein duquel le coût d'une augmentation significative de la précision reste raisonnable, une sorte de zone de bon rapport qualité/prix. Cette zone couvre l'intervalle allant d'un échantillon de 250 ilots à un échantillon de 1200 ilots. Grossièrement, en deçà de cet intervalle la méthode est trop imprécise, au delà elle est trop chère.

Deuxièmement l'investissement que représente l'achat et le traitement de l'image satellite apparaît extrêmement concurrentiel par rapport au coût d'une actualisation importante d'une base de sondage classique (liste de ménages), surtout si l'on considère que l'image peut servir à de nombreuses autres applications sur un même site (nous songeons par exemple au développement rapide des observatoires urbains). Il n'est d'ailleurs pas interdit de penser que, dans un avenir assez proche, beaucoup de villes feront régulièrement l'acquisition de ces images pour de toutes autres raisons et que ces documents feront en quelque sorte partie du domaine public ou que du moins le coût d'achat en sera réparti entre plusieurs utilisateurs. Face à ces "perspectives radieuses" d'applications et même s'il faut les considérer avec sérénité et patience, il n'est donc pas vain de chercher à perfectionner la méthode. C'est à l'examen des possibilités ouvertes pour son amélioration que nous allons nous attacher pour finir.

3. Perspectives de recherche pour l'amélioration de la méthode

Au terme de ce travail, quatre directions principales de recherche concernant les plans de sondages aréolaires systématiques en milieu urbain nous semblent prometteuses, les trois premières impliquant d'ailleurs autant les spécialistes de télédétection que les statisticiens.

D'autres stratifications

Si la stratification sur la seule densité du bâti utilisée dans l'expérience quiténienne est, comme nous l'avons dit, perfectible par la simple prise en compte de connaissances empiriques sur le milieu urbain, il n'est reste pas moins que la mise au point de nouvelles stratifications morphologiques d'après l'information satellitaire est une voie de recherche importante pour l'amélioration de la méthode. Au chapitre 7 nous proposons de suivre, dans ces recherches futures concernant les stratifications,

une démarche analytique et faite "d'allers et retours" entre télédétection et statistique. L'étude des liaisons statistiques entre la morphologie urbaine et les différents thèmes possibles d'application de la méthode doit être généralisée pour permettre de sélectionner des critères pertinents de stratifications adaptés à chaque thème. L'identification de ces thèmes morphologiques de stratification sur l'image satellite est ensuite la tâche des spécialistes de télédétection. Enfin la validation des stratifications obtenues est complète lorsqu'on applique efficacement ces stratifications à des enquêtes par sondages réelles.

D'autres probabilités d'appartenance à l'échantillon

Nous avons souligné dès le départ de cette recherche l'intérêt qu'il y a à sélectionner les unités primaires d'échantillonnage (pâtés de maisons) avec une probabilité proportionnelle à une "bonne" mesure de leur taille, c'est à dire à une mesure bien corrélée aux variables que l'on cherche à estimer. C'est potentiellement une méthode d'échantillonnage très efficace mais elle se heurte à deux obstacles principaux que seules des recherches plus poussées pourraient lever. Le premier est l'appréciation sur l'image ou d'après l'information morphologique qu'elle fournit de la mesure de taille des unités primaires que l'on envisage d'utiliser pour définir les probabilités d'inclusion dans l'échantillon : approximations de la surface bâtie des îlots, de leur surface habitable etc... Pour être concevable, une telle opération exige des recherches approfondies en télédétection sur l'identification et la caractérisation d'après l'image des surfaces bâties en milieu urbain. Le second obstacle est celui de la gestion de ces probabilités inégales lors de la sélection de l'échantillon. On l'a vu, le recours à des "stratégies" de sélection des îlots associées à des sondage par points pose des problèmes pratiques difficilement solubles. De plus ce type de méthode atteint vite ses limites lorsque la définition de la mesure de taille devient complexe. Pour gérer correctement tout type de probabilité, il faut songer à constituer une liste des unités primaires et des tailles qui leurs sont attachées. Là encore ce sont les recherches en télédétection qui peuvent fournir la clef du problème. L'identification du réseau de voirie et la segmentation de l'image en unités spatiales utilisables comme unités d'échantillonnage sont les conditions indispensables à la poursuite des recherches statistiques sur ce thème.

D'autres estimateurs

Au fur et à mesure des progrès de l'interprétation des images satellite en milieu urbain, les recherches en sondages spatiaux devront s'intéresser aux possibilités d'intégration des informations exogènes fournies par les images dans les estimateurs. Le redressement des estimations par une information morphologique fine peut utiliser des estimateurs par le ratio, par la différence ou par la régression et conduire à une amélioration significative de la précision du sondage.

D'autres techniques d'évaluation de la variance du sondage systématique

Enfin, et cette fois indépendamment de l'utilisation qui sera faite des images satellite, l'estimation de l'erreur d'échantillonnage liée aux sondages systématiques est un domaine où les recherches sont loin d'être closes. Comme nous l'avons souvent dit au cours de cet exposé, la solution retenue ici est une approximation relativement grossière et elle a pu engendrer une surestimation importante des variances. Outre les méthodes robustes d'estimations de la variance, qui ont déjà fait leur preuves, on peut penser, suivant en cela la tendance des recherches actuelles, à appliquer à ce problème la théorie des processus stochastiques, plus précisément l'application qu'en a fait MATHERON à l'étude des variables régionalisées. L'analyse des corrélogrammes et covariogrammes spatiaux attachés aux variables que l'on cherche à estimer peut

fournir des éléments de solution au problème ; il ne faut cependant pas occulter le fait qu'elle réclame qu'on dispose de données démographiques localisées à un niveau très fin, données qui sont pour l'instant assez rarement disponibles.

Nous venons d'évoquer quatre voies de développement de futures recherches sur les sondages spatiaux en milieu urbain. Ce rapide inventaire n'a pas la prétention d'être exhaustif ; son objectif principal était de montrer que les recherches menées par l'équipe ORSTOM, concernant l'amélioration des méthodes de collectes démographiques dans les villes des pays en développement, débouchent sur un large champ d'investigations. Ainsi la nouvelle méthode de sondage que nous proposons, déjà opérationnelle sous sa forme actuelle, peut certainement être perfectionnée, du moins en ce qui concerne la précision des résultats qu'elle fournit.

ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES

*SONDAGES AREOLAIRES POUR L'ESTIMATION DE DONNEES
DEMOGRAPHIQUES EN MILIEU URBAIN*

*ESSAI DE DEFINITION D'UNE METHODE DE COLLECTE
INTEGRANT L'INFORMATION SATELLITAIRE, APPLICATION
AUX VILLES DES PAYS EN DEVELOPPEMENT*

Thèse de doctorat

Soutenue le

par

Olivier BARBARY

Direction de thèse :
Marc BARBUT, EHESS
Jean Paul GILG, EHESS

Membres du jury :

ANNEXES

Liste des annexes

	Pages
Annexes a la deuxième partie	
A.2.1 : Obtention, analyse statistique et codage de l'information morphologique à Marseille.....	385
A.2.2 : Les indicateurs démographiques intégrés dans la base de données sur Marseille ; définitions, variations selon les types morphologiques et étude statistique préliminaire.....	402
A.2.3 : L'outil informatique mis au point dans la phase d'expérience à Marseille.....	425
Annexes à la troisième partie	
A.3.1 : Les enquêtes préliminaires, méthodes de collecte, questionnaires, codification.....	471
1) Enquête Morphologie urbaine : instructions aux enquêteurs	472
2) Enquête Morphologie urbaine : codification.....	491
3) Enquête Bâti / population : instructions aux enquêteurs.....	507
4) Enquête Bâti / population : codification.....	535
A.3.2 : L'enquête démographie / Migrations :, collecte codification et saisie.....	555
1) Manuel d'instructions aux enquêteurs.....	556
2) Condition de collecte, codification et saisie des données.....	583
A.3.3 : L'outil informatiques pour la définition du second degré du sondage : le programme VAR2DEG (listage).....	592
Annexe à la quatrième partie	
A.4.1 : Programmes d'estimations pour l'exploitation de l'enquête finale : listage des procédures SPSS.....	600

Annexe A.2.1

OBTENTION, ANALYSE STATISTIQUE ET RECODAGE DE L'INFORMATION MORPHOLOGIQUE A MARSEILLE

O. BARBARY et F. DUREAU - Avril 1986

Introduction

Les caractéristiques morphologiques constituent une des trois informations de base intégrées dans la base de données sur Marseille. Les chapitres 5 et 6 ont permis d'analyser l'utilisation qui était faite de ces données et comment elles ont conduit à une typologie de la morphologie urbaine à Marseille. Mais pour passer des photographies aériennes au tableau disjonctif complet sur lequel est basée cette typologie, un important travail a été nécessaire.

Cet annexe présente tout d'abord les documents qui ont servi à l'élaboration de cette information. Ensuite nous exposons la démarche adoptée pour l'analyse visuelle de la mosaïque photographique et pour la constitution du tableau de description morphologique des zones. Enfin, on présente les analyses statistiques et les recodages qui ont conduit au tableau soumis à l'analyse¹.

1. Données utilisées

1.1 Les photographies aériennes

Pour satisfaire les objectifs assignés à la photo-analyse, les photographies aériennes devaient remplir trois conditions :

- apporter une information exhaustive et homogène sur l'ensemble de la zone d'étude : ce qui suppose une couverture aérienne, à une même échelle et à une date unique (ou des dates très voisines) sur l'ensemble de la zone d'étude
- avoir été enregistrées à la date la plus proche possible de celles des deux autres sources de données : recensement de mars 1982 et image Thematic Mapper de février 1983
- être à une échelle adéquate pour analyser le tissu urbain et les objets bâtis: 1/15 000 à 1/40 000 environ.

Les photos aériennes réalisées en août 1983 à Marseille dans le cadre de la couverture systématique du littoral français (mission IPLI - Lot 23 - échelle voisine de 1/23 000 - panchromatique - noir et blanc) satisfont à ces trois conditions : le décalage temporel est réduit au minimum puisque les photographies aériennes ont été prises à une date intermédiaire entre le recensement et l'image Thematic Mapper, et l'échelle est tout à fait satisfaisante pour l'utilisation prévue ici.

¹ : Cet annexe A.2.1 correspond à une partie de l'annexe 6 du rapport rédigé par l'équipe pour l'ATP "télé-détection" du CNRS : "Etablissement d'une typologie de la morphologie urbaine à partir d'une analyse de la mosaïque photographique de Marseille", par F. DUREAU et O. BARBARY, 66 p., Paris, avril 1986.

La mission de 1981 (1/15 000 - panchromatique - noir et blanc), que nous avons envisagé d'utiliser, aurait apporté des éléments d'information supplémentaires sur les objets urbains ; malheureusement, elle ne couvre que la moitié nord de notre zone d'étude. Il faut souligner à ce propos que l'agglomération urbaine n'est visiblement pas une notion retenue par l'IGN pour ses missions aériennes : la plupart des couvertures à grande échelle réalisées sur Marseille depuis une dizaine d'années ne couvre qu'une partie de la commune.

Une *mosaïque photographique non contrôlée* a été réalisée à l'ATOB ; c'est sur ce document qu'a été réalisée la photo-analyse.

Une certaine perte d'information existe entre les photographies originales (tirages papier 24 x 24 cm fournis par l'IGN) et la mosaïque. Du fait de la moindre netteté de la mosaïque, le zonage est plus aisé : l'oeil ne perçoit plus nettement les objets urbains bien individualisés, et une certaine intégration, plus proche du niveau d'information recherché pour satisfaire les objectifs de l'analyse visuelle (établissement d'une typologie, et référence de terrain pour la télédétection spatiale), devient ainsi immédiate.

1.2. Contrôles terrain

Les informations extraites de la photo-analyse ont été contrôlées, par sondage, lors d'une enquête sur le terrain début novembre 1985.

Le principe du tirage de l'échantillon de zones morphologiques à contrôler au sol a été le suivant : chacun des 17 types morphologiques définis par les premiers traitements d'analyse des données, AFC (Analyse Factorielle des Correspondances) et CAH (Classification Ascendante Hiérarchique), devaient être représentés par plusieurs zones dans l'échantillon. On a donc procédé à un tirage systématique sur la liste des zones classées par type morphologique, avec un pas variable selon les types :

Nombre de zones par type	Taux de sondage	Nombre de zones dans l'échantillon
226	1 / 15	16
14 à 35	1 / 3	5 à 11
6 à 11	1 / 2	3 à 5
2 à 5	1 / 1	2 à 5

Au total, l'échantillon tiré selon cette procédure comportait 100 zones morphologiques. La majorité de ces zones occupent une superficie trop importante pour être contrôlées exhaustivement : c'est pourquoi, un deuxième tirage a été effectué au sein de 2/3 des zones les plus grandes de l'échantillon. Il s'agissait de déterminer sur les photographies aériennes, pour chacune de ces grandes zones, un groupe de pâtés de maisons représentatifs.

La fiche de relevé (fig. A.2.1.1), remplie pour chacune des zones de l'échantillon, rassemble des informations sur :

- le plus grand bâtiment et le plus petit bâtiment de la zone (longueur, largeur, et nombre d'étages). La détermination de ces deux bâtiments a été faite sur photographie aérienne : même si ces bâtiments se trouvaient hors du groupe d'ilots sélectionnés, nous sommes allés les mesurer.

Figure A.2.1.1. Contrôles terrain (novembre 85) : fiche de relevé

FICHE LLOT: PHYSIQUE	N° Classe	N° Zone Morpho
Croquis lot (Noter photo)	Mesures de l'élément le plus grand dans la zone:	
	Longueur: _____	
	Largeur: _____	
	Hauteur: _____	
	Mesures de l'élément le plus petit dans la zone:	
	Longueur: _____	
	Largeur: _____	
	Hauteur: _____	

	Bâtiment 1	Bâtiment 2	Bâtiment 3	Bâtiment 4	Bâtiment 5
Longueur					
Largeur					
Nbre étages					
Nature du toit					
Nature du Non Bâti					

Changements depuis 1982: _____

- les caractéristiques des éléments bâtis majoritaires (longueur, largeur, nombre d'étages, nature du toit), et nature du non bâti. Dans le cas des grandes zones, ces renseignements sont issus de l'observation du groupe d'ilots sélectionnés. Nous nous sommes efforcés de décrire tous les types de bâtiments présents dans chaque zone morphologique.
 - un croquis et des photographies au sol viennent compléter l'information. Près de 150 photographies au sol ont été prises lors des relevés sur le terrain, illustrant tous les types de morphologie urbaine.
 - des observations sur les changements intervenus depuis 1982.
- Ce contrôle sur le terrain, comme nous le verrons, a été à l'origine de certaines modifications dans le codage issu de l'analyse visuelle.

2. Analyse visuelle de la mosaïque photographique au 1/23 000

La méthodologie retenue pour analyser la mosaïque s'est appuyée sur les enseignements issus des travaux présentés dans la littérature sur le sujet, et l'expérience en photographie de B. LORTIC ; un test sur un petit nombre de zones a permis de cerner la praticabilité des critères retenus a priori. Enfin, la réalisation même de l'analyse a permis de rectifier certaines options prises à l'issue du test.

2.1 Etablissement de la méthodologie d'analyse

Comme le rappellent J.P. GILG et son équipe¹, la démarche de l'interprète humain est :

- d'abord globale : délimitation de zones d'égale apparence ou isophènes,
- puis analytique : description des zones, qui valide et module la stratification initiale.

L'efficacité de cette démarche n'est plus à démontrer. Il faut toutefois souligner les problèmes que soulève le zonage manuel dans son application : B. LORTIC² évoque *trois difficultés* :

- les variations continues et progressives de l'image : les ruptures existant au sol étant la plupart du temps progressives, il est difficile de localiser précisément les limites des zones homogènes.
- la tendance à privilégier un paramètre visuel au détriment des autres : les variables couleur et valeur sont privilégiées spontanément au détriment des variables texture et structure, qui peuvent pourtant, dans certains domaines, apporter l'information la plus intéressante.
- l'existence de différents niveaux d'homogénéité : le seuil de "différence" au-dessous duquel on décrète qu'il y a homogénéité est difficile à déterminer et à maintenir constant sur l'ensemble de l'image.

¹ : J.P. GILG et collaborateurs (1976) : Aperçu d'une expérience de télé-analyse et photo-analyse. Informatique et sciences humaines n°30, pp. 21-38.

² : B. LORTIC (1984), "Une méthode d'analyse visuelle d'une image spatiale. Objectivation du zonage manuel", mémoire EHESS.

Pour trouver ce seuil, B. LORTIC précise qu'il est nécessaire de procéder par itération :

- 1) Réaliser un premier zonage sans critère bien défini, en détournant chaque zone semblant homogène et se différenciant de son environnement
- 2) A partir de ce zonage, déterminer précisément les critères discriminants et les seuils d'homogénéité les plus pertinents
- 3) Reprendre le premier zonage en décrivant chaque zone à partir de ces critères.

C'est cette démarche itérative que nous avons retenue pour l'analyse visuelle de la mosaïque photographique de Marseille, et que nous exposerons maintenant en détail:

- 1) Zonage en unités d'égale apparence
- 2) Caractérisation de chacune de ces zones à l'aide de descripteurs, avec modifications ponctuelles du zonage initial.

2.2 Zonage en unités d'égale apparence

Le zonage a été établi sur un calque posé sur la mosaïque photographique. La démarche suivie était de détourner les zones d'égale apparence sur le document photographique.

Le découpage spatial ne tient compte que de ce paramètre, et non du découpage en pâtés de maisons ou en districts de recensement : un certain nombre de pâtés de maisons sont donc divisés, s'ils ne sont pas homogènes. Pour cette étape, on ne prend pas en compte des paramètres bien définis : seule compte la différenciation entre zones, corrolairement l'homogénéité de la zone. 524 zones ont ainsi été délimitées sur la zone d'étude.

Il faut insister sur le fait que la perte d'information de la mosaïque par rapport aux photographies IGN a facilité le zonage, évitant l'écueil majeur en milieu urbain, qui est d'isoler le bâtiment, objet se démarquant généralement de son voisinage immédiat ; nous avons ainsi pu éviter, dans la majorité des cas, d'établir un zonage à un niveau trop fin pour notre application. Toutefois, il existe un certain nombre de zones très petites, dues à la présence d'un objet urbain relativement grand, et très contrasté par rapport à son environnement. Isolées de ce fait, ces zones ne peuvent pas être décrites de la même façon que les autres zones : le milieu urbain est caractérisé par l'importance de l'effet d'échelle. Lorsque l'on passe du groupe d'îlots à l'îlot lui-même, les variables descriptives valides ne sont plus les mêmes.

2.3 Test des descripteurs

Cette opération avait pour but de tester le degré de praticabilité d'une première série de descripteurs, et le nombre de modalités qui pouvait être retenues pour chacun d'entre eux.

Ce test a été réalisé sur un petit nombre de zones, dix seulement ; ces zones ont été choisies de manière à couvrir un large éventail des situations rencontrées sur la mosaïque de Marseille. Ont été testés les descripteurs suivants :

- taille des éléments bâtis,
- forme des éléments bâtis,
- agencement du bâti dans l'îlot,
- la valeur (niveau de gris) dominante du bâti,
- la variabilité de valeur du bâti,
- la valeur de la voirie.

Ce premier essai nous a montré que tous ces descripteurs étaient praticables sans problème majeur à l'exception de celui sur la variabilité de la valeur ; d'autre part, nous avons testé le critère de densité du bâti dans la zone, en se fixant des zones étalons.

Le test nous a conduit à étoffer la liste des descripteurs, la première liste choisie s'avérant insuffisante pour rendre compte des diverses situations rencontrées. Quatre axes nous ont guidés dans le choix de ces nouveaux descripteurs :

- distinguer la valeur dominante de la valeur sous-dominante sur la zone, plutôt que de chercher à caractériser directement la variabilité de cette valeur
- séparer la voirie et le non bâti-non voirie, pour la caractérisation de la valeur
- préciser les mesures des éléments bâtis (longueur, largeur, surface et hauteur)
- introduire des descripteurs de l'agencement du réseau de voirie, et de la densité de ce dernier.

De plus, nous avons choisi des descripteurs analytiques. S'ils sont apparemment plus difficiles à maîtriser sur la photographie que des indicateurs synthétiques, il est toujours aisé de construire ensuite un indice à partir de critères observés sur la photographie ; on peut ainsi tester différents indices sans avoir à revenir sur le document photographique. C'est ce qui a été fait, par exemple, pour la longueur et la largeur, qui ont permis de construire deux indices : la surface, et le degré d'allongement de l'élément bâti majoritaire.

D'après les enseignements tirés des tests, nous avons établi une liste de descripteurs, dont le nombre de modalités a été fixé en essayant de les appliquer aux zones les plus diverses possibles de Marseille. La description des zones a donc débuté avec un "guide d'analyse" comprenant:

- la liste des descripteurs
- le mode de collecte de chacun de ces descripteurs (mesure directe à la règle, ou estimation visuelle) et le nombre de modalités dans le cas de descripteurs estimés visuellement, définies par référence à des zones témoins de la mosaïque.

Le tableau A.2.1.2 présente la liste des descripteurs et leur mode de recueil sur la mosaïque.

2.4 Description des zones d'égale apparence

La description des zones a conduit à *modifier partiellement le premier zonage* en zones d'égale apparence : des limites de zones ont été modifiées lors de la description, de même, certaines zones ont été supprimées, et d'autres ajoutées. C'est d'ailleurs ce dernier cas qui a été le plus courant : nous sommes ainsi passé de 524 zones à 542 zones après caractérisation. Ce phénomène est dû au fait que certains voisinages très contrastés tendent à donner une homogénéité apparente à des zones qui, lors de la description, s'avèrent ne pas pouvoir être décrites de façon homogène ; d'où des cas de subdivision lors de la caractérisation.

Notons que nous avons choisi de procéder à la *description des zones critère par critère, ou par couple de critères similaires* (exemples : longueur et largeur de l'élément bâti majoritaire, valeur dominante et sous-dominante du bâti), et non zone par zone.

Etant donné la surface à analyser et le nombre important de zones à décrire, il nous a en effet semblé préférable de procéder ainsi pour garder une certaine constance dans l'appréhension des critères qualitatifs : il nous semble en effet que le risque est déjà grand de ne pas rester constant pour des critères qui réclament plusieurs jours de travail pour être établis sur l'ensemble de la zone d'étude. En procédant par zone, l'écart se chiffre en semaines entre le début et la fin du travail de caractérisation : il

est exclu que l'on puisse rester constant dans ses jugements sur une période de temps aussi longue. Procéder par critère ou type de critères permet, à notre avis, de limiter le degré de variabilité de la description visuelle, qui ne peut jamais être totalement éliminée pour des critères tels que la valeur, dont la perception par l'oeil est influencée par des paramètres tels que l'ensoleillement de la pièce où est réalisée l'analyse photographique.

Pour une partie des descripteurs, outre la hauteur des bâtiments qui nécessitait la stéréoscopie, nous avons dû retourner au document d'origine, c'est-à-dire les photographies aériennes de l'IGN, la mosaïque se révélant d'une précision insuffisante. Ce gain de précision ne remet pas en cause la méthode sur mosaïque "lissée" pour établir le zonage.

Sur les 17 variables retenues initialement, 6 ont été éliminées dès la phase de description des zones. L'examen de la mosaïque pour les 11 premiers critères ont en effet mis en évidence :

- soit que certains de ces 6 critères étaient redondants avec ceux déjà établis : ainsi, l'agencement du bâti dans l'îlot a été abandonné après application à la moitié des zones, car ce critère n'apportait aucune information nouvelle par rapport aux autres descripteurs.
- soit qu'ils n'étaient pas praticables pour un grand nombre de zones, trop petites pour qu'ils puissent être appliqués : c'est le cas des caractéristiques du réseau de voirie (agencement et densité), qui supposent que la zone contienne plusieurs pâtés de maisons pour qu'elles aient un sens.

2.5 Le tableau de description des zones

A l'issue de la photo-analyse, nous avons donc obtenu un tableau croisé, décrivant les 542 zones par 11 variables analytiques de la morphologie urbaine (voir Tableau A.2.1.2). Ce tableau a été saisi sur support informatique, afin de servir à l'établissement de la typologie de la morphologie urbaine, à l'étude statistique des relations bâti/population, et à l'interprétation des images satellites. Le zonage, reporté sur carte topographique IGN au 1/25 000 a été digitalisé par l'Unité d'Infographie.

Avant tout traitement, il importe de bien préciser les caractéristiques de ce tableau initial de données sur la morphologie urbaine :

- *les variables sont de nature différente, au sens mathématique* : on trouve des variables métriques (mesures de longueur et largeur) aussi bien que des variables qualitatives (valeurs, densité du bâti, par exemple)
- *étant donné leur mode de production, toutes les variables n'ont pas exactement le même degré de fiabilité*
- *les différentes modalités prises par une variable ordonnée n'ont pas un écart constant entre elles, mais font référence à des situations types* : il existe des phénomènes de "saut" qui font que l'on ne peut rencontrer tous les cas de figures pour une variable. On ne peut donc assimiler une échelle de modalités ordonnées à une variable quantitative continue qui aurait été seuillée.

Tableau A.2.1.2 - Liste des descripteurs

Descripteur	Mode de relevé	Tableau de description						
Densité du bâti dans la zone	Estimation visuelle avec zones étalon du rapport S. bâtie/S. totale. 6 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (6 modalités)						
Valeur dominante du bâti	Estimation visuelle des niveaux de gris du bâti majoritaire dans la zone. 4 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (4 modalités)						
Valeur sous-dominante du bâti	id	id						
Valeur dominante de l'espace intersticiel	Estimation visuelle du niveau de gris majoritaire du non-bâti non-voirie. 4 modalités	Conservée dans tableau de description morpho. (4 modalités)						
Valeur sous-dominante de l'espace intersticiel	id	id						
Hauteur du bâti	Estimation visuelle par examen stéréoscopique. 5 modalités dont 1 pour les zones non bâties. (2 niveaux de hauteur x hauteur mixte ou homogène)	Conservée dans tableau de description morpho. (5 modalités)						
Surface minimum des éléments bâtis	Mesure de la surface du plus petit élément bâti de la zone, en centièmes de mm ²	Conservée dans tableau de description morpho. (seuillée)						
Surface maximum des éléments bâtis	Mesure de la surface du plus grand élément bâti de la zone, en centièmes de mm ²	id						
Longueur de l'élément bâti majoritaire	Mesure de la longueur de l'élément bâti majoritaire de la zone, en dixièmes de mm	Conservée dans tableau de description morpho. (non seuillée), utilisée pour variables dérivées*						
Largeur de l'élément bâti majoritaire	Mesure de la largeur de l'élément bâti majoritaire de la zone, en dixièmes de mm	id						
Forme de l'élément bâti majoritaire	Observation visuelle. 4 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (4 modalités) Mais abandonnée par la suite dans l'analyse typologique						
<p>Descripteurs supprimés au cours de l'analyse visuelle :</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">Agencement du bâti dans l'îlot</td> <td style="width: 50%;">Valeur de la voirie</td> </tr> <tr> <td>Agencement du réseau de voirie</td> <td>Densité du réseau de voirie</td> </tr> <tr> <td>Variabilité de la valeur du bâti</td> <td>Type de non bâti</td> </tr> </table>			Agencement du bâti dans l'îlot	Valeur de la voirie	Agencement du réseau de voirie	Densité du réseau de voirie	Variabilité de la valeur du bâti	Type de non bâti
Agencement du bâti dans l'îlot	Valeur de la voirie							
Agencement du réseau de voirie	Densité du réseau de voirie							
Variabilité de la valeur du bâti	Type de non bâti							

* Création de 2 variables dérivées à partir des descripteurs de longueur et largeur : surface et rapport longueur/largeur

- la caractérisation visuelle doit être complétée par une enquête terrain, pour confirmer certaines opérations d'interprétation, volontairement limitées mais néanmoins effectives. Cette enquête est également nécessaire pour l'utilisation qui sera faite de cette analyse visuelle de photographies aériennes : il s'agit de passer de l'analyse à l'interprétation, c'est-à-dire faire correspondre aux objets présents sur la photographie les objets présents sur le terrain, afin de pouvoir utiliser la photographie aérienne comme élément de référence terrain lors de mise au point et de la validation des méthodes d'extraction de l'information sur image satellite.

3. Analyses statistiques univariées et recodage des descripteurs morphologiques

3.1 Les distributions des descripteurs morphologiques qualitatifs

Remarque préliminaire

La variable qualitative "FORM", qui décrit la forme de l'élément bâti majoritaire en deux modalités (parallélépipèdes et autres formes) apparaît dès l'examen de sa distribution, sans grand intérêt. 417 zones sur les 441 existantes ont été codées avec la première modalité. En fait, quelle que soit la forme des constructions, on a toujours pu donner un sens aux mesures de longueur et de largeur sur le bâtiment typique de la zone. On préférera donc remplacer la variable qualitative "FORM" par la mesure calculée du rapport (longueur/largeur) caractérisant l'élément bâti majoritaire. La distribution et le recodage de cette variable seront décrits au paragraphe concernant les variables quantitatives du tableau.

Description des distributions (fig. A.2.1.3)

1 : Densité du bâti (5 modalités fig. A.2.1.3.1)

La majorité des zones (380 sur 441) a une densité de bâti faible ou moyenne. Les zones denses et très denses sont relativement peu nombreuses.

2 : Valeur dominante dans l'espace bâti (3 modalités, fig. A.2.13.2)

Les modalités dominantes correspondent aux valeurs moyennes (220 zones) et claires (159 zones). Seules 62 zones ont une valeur majoritaire foncée de leur bâti. Ce sont, soit des zones au bâti ancien, à toits de tuiles canal sombres, soit des zones où dominent de vieux bâtiments industriels, hangars à toits de tuiles mécaniques .

3 : Valeur sous-dominante du bâti (3 modalités, fig. A.2.1.3.3)

On observe, grossièrement, la même répartition que pour la valeur dominante.

4 : Valeur dominante de l'environnement du bâti, voirie exclue (4 modalités, fig. A.2.13.4)

La distribution est concentrée sur les trois modalités de rangs 2, 3 et 4 et il y a approximativement equi-répartition entre elles. La modalité 4 correspond à une dominante boisée de l'environnement, la modalité 3 aux pelouses, prairies, terrains agricoles ou parterres et la modalité 2, déjà assez claire, est le plus souvent liée aux espaces bitumés : parking, entrée d'immeuble, etc., ou aux terrains vagues. La modalité 1, qui est marginale (11 zones), n'apparaît que dans les endroits où la réflectance est très forte : sols nus, roches et carrières, etc.

5 : Valeur sous-dominante de l'environnement (4 modalités, fig. A.2.13.5)

Les fréquences vont en croissant avec les modalités.

Figure A.2.1.3 : Histogrammes des descripteurs morphologiques qualitatifs

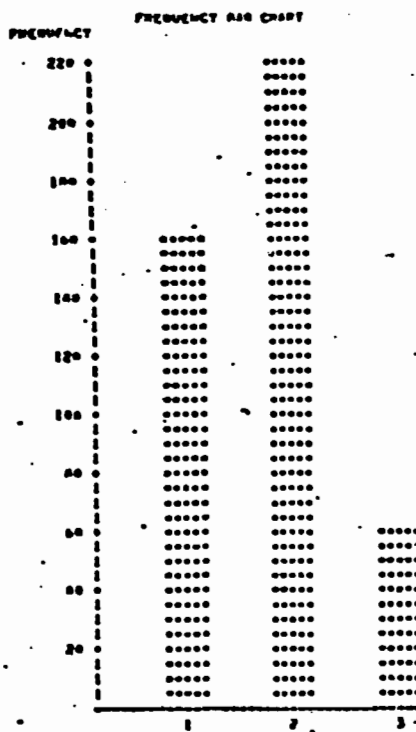
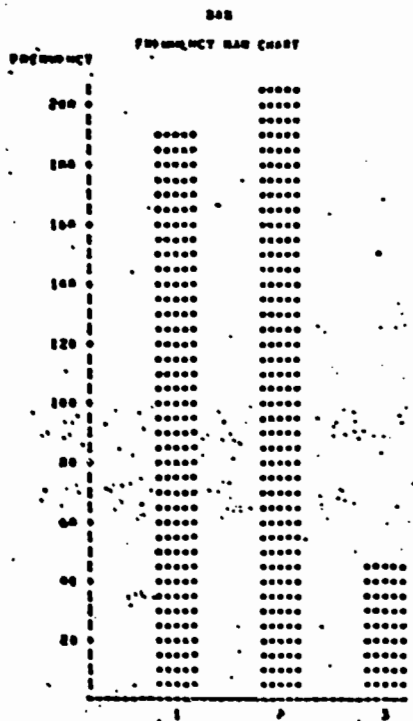
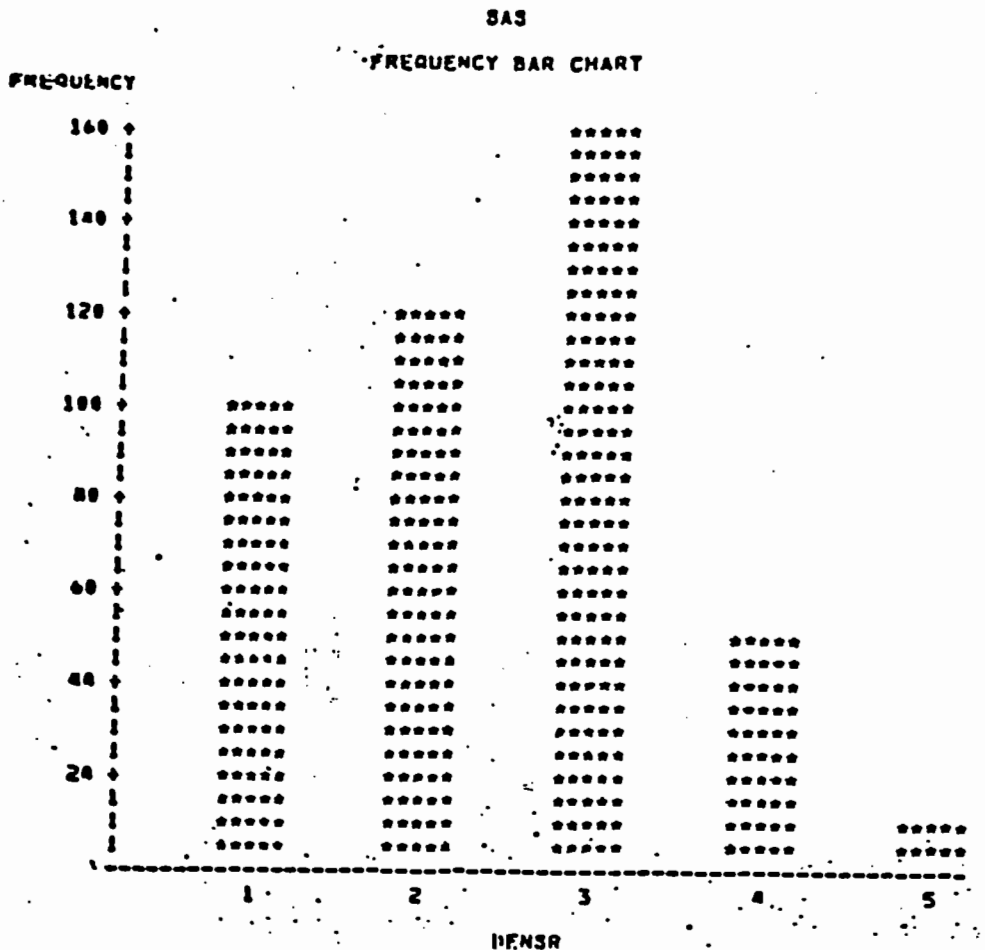
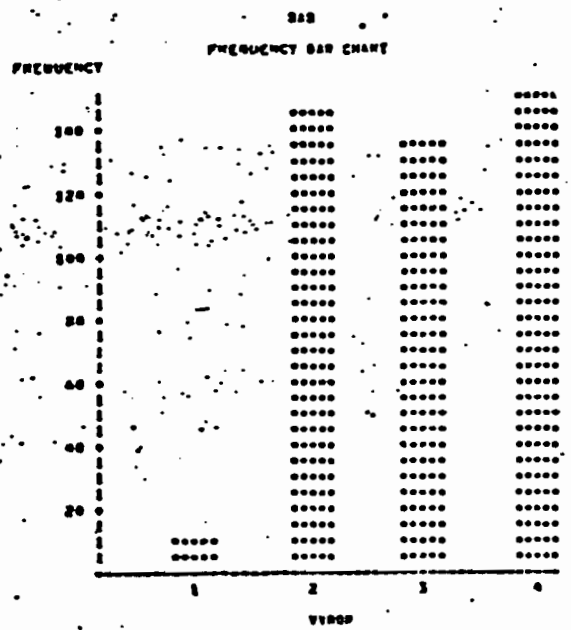
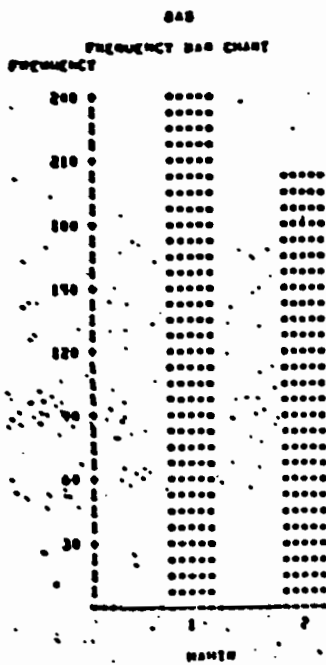
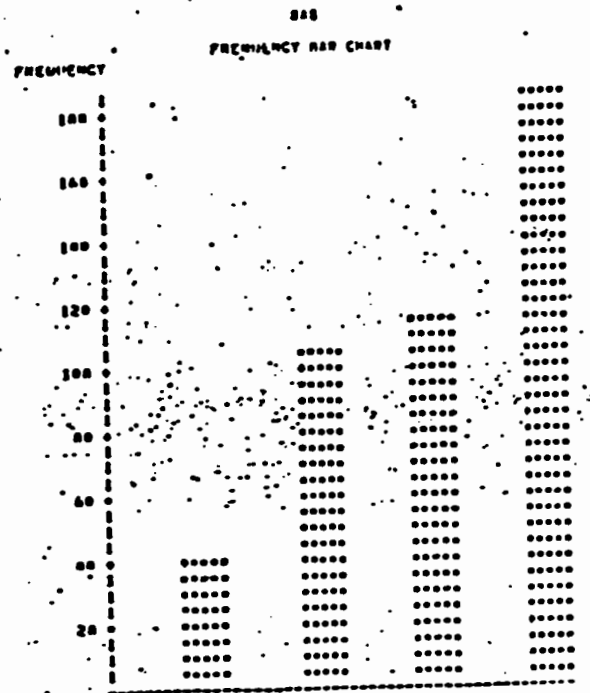
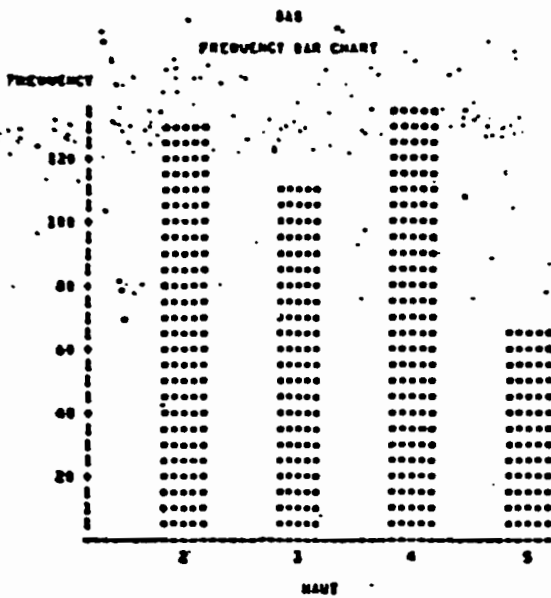


Figure A.2.1.3 : Histogrammes des descripteurs morphologiques qualitatifs (suite)



6 : Elévation du bâti (fig. A.2.1.3.6 et A.2.1.3.7)

A l'origine, la description de l'élévation du bâti a été faite en quatre modalités. Par la suite, considérant d'une part que la distinction entre les modalités 1, 2 et 3, 4 pose parfois problème lorsqu'on se contente d'une évaluation rapide (sans stéréoscope) et que, d'autre part, dans l'optique typologique qui nous intéresse pour l'instant, c'est le seuil entre dominante haute et dominante basse qui importe le plus, on a pris le parti de regrouper les modalités 1 et 2 d'une part, 3 et 4 d'autre part.

Il en résulte une description dichotomique de l'élévation du bâti, où les zones à dominante basse sont légèrement plus nombreuses que celles à dominante haute (241 pour 200).

On verra plus tard que, pour pallier à la description très sommaire de ce caractère morphologique important du point de vue thématique, on sera amené à augmenter son poids dans l'analyse par pondération de ses deux modalités.

3.2. Distribution des caractères quantitatifs, seuillage et recodage en qualitatif

Remarque préliminaire

Les quatre variables quantitatives s'attachent à la description des éléments bâtis de chaque zone. Leur surface maximum et leur surface minimum ont été mesurées directement sur l'image grâce à des rectangles étalons. On a également mesuré la longueur et la largeur de l'élément majoritaire dans la zone.

Plutôt que de retenir ces deux descripteurs tels quels, on a préféré, pour faciliter l'interprétation des résultats d'analyse, calculer deux indicateurs synthétiques : surface de l'élément bâti majoritaire et rapport de forme (longueur/largeur) du bâtiment en question.

Les distributions des descripteurs et leur recodage (Tableau A.2.1.4 et Fig. A.2.1.5)

Il s'agit ici, en étudiant les histogrammes de ces quatre variables, d'identifier les modes principaux de chaque distribution pour fixer ensuite les bornes des intervalles qui résumeront aux mieux ces mesures en autant de variables qualitatives; chaque modalité représentera alors un mode principal. Pour mieux évaluer ce qui est conservé et ce qui est perdu dans l'information qu'apporte chaque descripteur, on pourra comparer les histogrammes originaux aux histogrammes des variables qualitatives obtenues, qui sont juxtaposés sur les figures A.2.1.5.1 à A.2.1.5.4. Dans le tableau ci-dessous, on se limite à donner les intervalles de chaque modalité et à indiquer les fréquences et les principales valeurs "centrales".

Tableau A.2.1.4 : descriptions des modalités retenues pour les descripteurs quantitatifs

Descripteurs Modalités	SURF	RAPT	SMAX	SMIN
Intervalle de def.*	0-30	0-1	0-60	0-24
-1- Valeurs centrales	25	1	50	10 & 20
Effectif	138	177	76	97
Frequence en %	31	40	17	22
Intervalle de def.*	30-90	1-2	60-149	25-35
-2- Valeurs centrales	50	1,4 & 2	80 & 100	25
Effectif	89	113	120	252
Frequence en %	20	26	27	57
Intervalle de def.*	90-190	2-5	149-299	35-65
-3- Valeurs centrales	25	3 ; 4 & 5	50	10 & 20
Effectif	122	100	105	48
Frequence en %	28	21	24	11
Intervalle de def.*	190-360	5-40	299-1100	65-4000
-4- Valeurs centrales	200 à 250	6 ; 8 & 10	300 à 500	80 & 100
Effectif	57	51	107	44
Frequence en %	13	13	24	10
Intervalle de def.*	360-950	...	1100-10000	...
-5- Valeurs centrales
Effectif	22	...	33	...
Frequence en %	5	...	7	...
Intervalle de def.*	950-3000
-6- Valeurs centrales
Effectif	13
Frequence en %	3

* : L'intervalle définissant la classe est donné borne inférieure exclue et borne supérieure incluse.

... : Donnée sans objet.

Figure A.2.1.5 : Histogrammes des descripteurs morphologiques quantitatifs

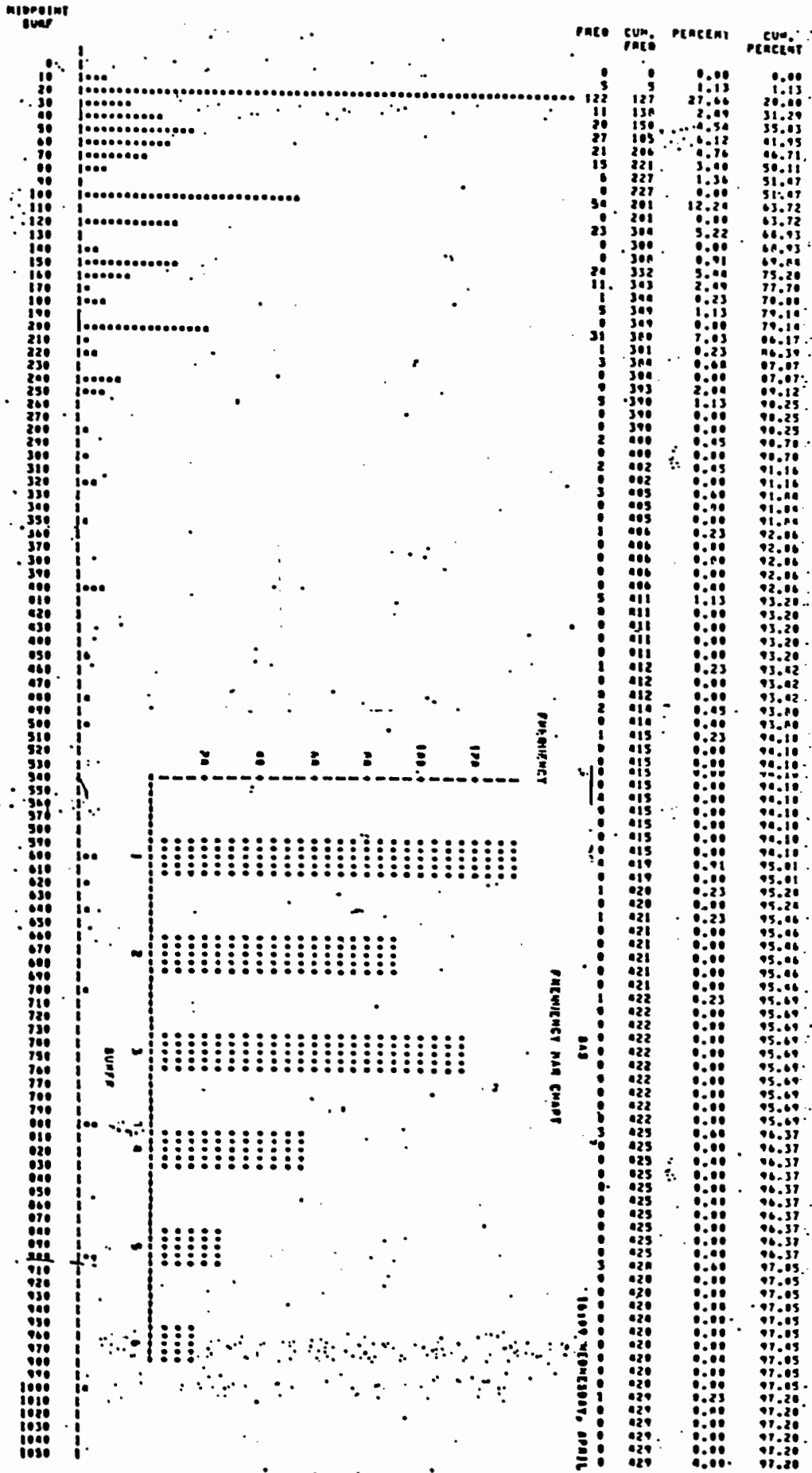
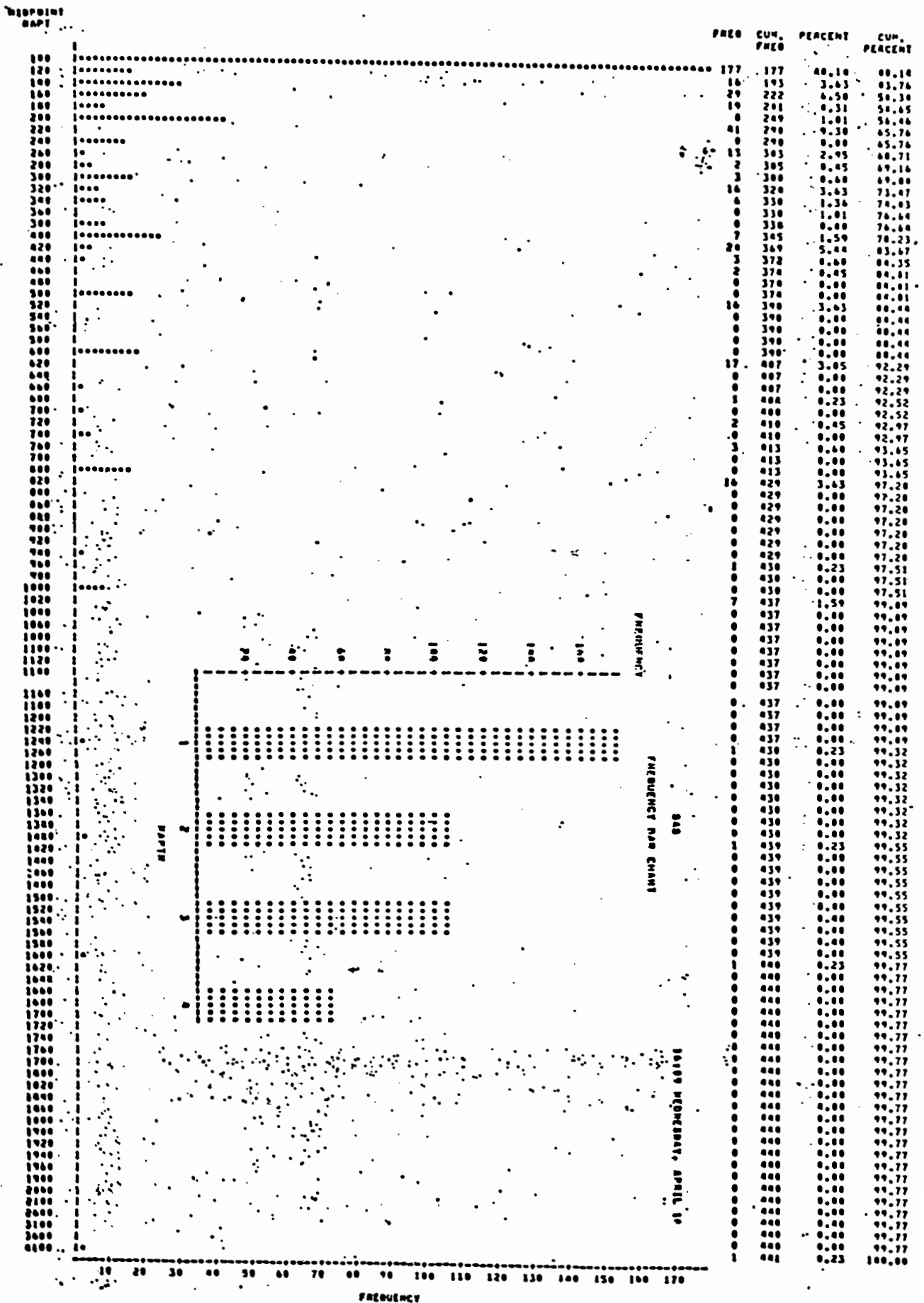


Figure A.2.1.5 : Histogrammes des descripteurs morphologiques quantitatifs (suite)



SAS
FREQUENCY BAR CHART

MIDPOINT
SMIN

FREQ CUM. PERCENT CUM.
FREQ PERCENT

MIDPOINT SMIN	FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
0	0	0	0.00	0.00
10	52	52	11.79	11.79
20	261	313	63.72	75.51
30	12	325	2.72	78.23
40	8	333	1.81	80.05
50	39	372	8.84	88.89
60	3	375	0.68	89.57
70	1	376	0.23	89.80
80	26	402	5.90	95.69
90	0	402	0.00	95.69
100	15	417	3.40	99.09
110	0	417	0.00	99.09
120	0	417	0.00	99.09
130	0	417	0.00	99.09
140	0	417	0.00	99.09
150	0	417	0.00	99.09
160	0	417	0.00	99.09
170	0	417	0.00	99.09
180	0	417	0.00	99.09
190	0	417	0.00	99.09
200	1	418	0.23	99.32
210	0	418	0.00	99.32
220	0	418	0.00	99.32
230	0	418	0.00	99.32
240	0	418	0.00	99.32
250	0	418	0.00	99.32
260	0	418	0.00	99.32
270	0	418	0.00	99.32
280	0	418	0.00	99.32
290	1	419	0.23	99.55
300	0	419	0.00	99.55
310	0	419	0.00	99.55
320	0	419	0.00	99.55
330	0	419	0.00	99.55
340	0	419	0.00	99.55
350	0	419	0.00	99.55
360	0	419	0.00	99.55
370	0	419	0.00	99.55
380	0	419	0.00	99.55
390	0	419	0.00	99.55
400	0	419	0.00	99.55
410	0	419	0.00	99.55
420	0	419	0.00	99.55
430	0	419	0.00	99.55
440	0	419	0.00	99.55
450	0	419	0.00	99.55
460	0	419	0.00	99.55
470	0	419	0.00	99.55
480	0	419	0.00	99.55
490	0	419	0.00	99.55
500	0	419	0.00	99.55
510	0	419	0.00	99.55
520	0	419	0.00	99.55

FREQUENCY

FREQUENCY BAR CHART

DATA REORGANIZED, APRIL 10, 1968

Annexe A.2.2

LES INDICATEURS DEMOGRAPHIQUES INTEGRES DANS LA BASE DE DONNEES SUR MARSEILLE : DEFINITIONS, VARIATIONS SELON LES TYPES MORPHOLOGIQUES ET ETUDE STATISTIQUE PRELIMINAIRE

O. BARBARY et F. DUREAU - avril & décembre 1986

Introduction

Les données démographiques utilisées sur Marseille dans le cadre du programme sur "L'utilisation de la télédétection pour l'observation des populations urbaines" sont issues du Recensement Général de la Population, réalisé par l'INSEE en mars 1982.

Comme on utilise constamment ces données tout au long de cette recherche sur les sondages spatiaux en démographie, il importe de rappeler les définitions utilisées par l'INSEE pour cette opération de collecte : ce sera l'objet de la première partie de cette annexe.

Dans un deuxième temps, nous nous intéressons aux quelques indicateurs socio-démographiques intégrés dans la base de données sur Marseille et à leurs variations selon les strates morphologiques¹.

Enfin une troisième partie de cette annexe est consacrée à l'étude statistique de trois indicateurs attachés aux districts de recensement : la surface, la population sans doubles comptes et la densité de cette population, ainsi qu'à une première approche, par le biais du calcul de la variance inter et intra-strate, du gain que la stratification morphologique peut apporter à l'estimation de ces caractères².

¹ : Les première et deuxième parties de cette annexe (§ 1 & 2) sont repris de l'annexe 12 du rapport ATP final : "Les indicateurs démographiques intégrés dans la base de données sur Marseille : définitions et variations selon les types morphologiques", F. DUREAU et O. BARBARY, 12 p., Paris, décembre 1986.

² : La troisième partie de cette annexe (§ 3) est repris de l'annexe 13 du rapport ATP final : "Caractères démographiques et stratification morphologique, analyse statistique", O. BARBARY, 13 p., Paris, octobre 1986.

1. Définitions utilisées par l'INSEE

1.1. Les différentes catégories de population

L'INSEE distingue quatre catégories de population, selon "le type de logement et le mode de vie, centré sur la famille ou organisé de façon communautaire" ¹ :

- **La population des ménages ordinaires** : "Un ménage ordinaire est constitué de l'ensemble des occupants d'une unité d'habitation privée (local séparé et indépendant) occupée comme résidence principale."

- **La population "Hors ménage ordinaire"**, qui comprend :

- **La population des ménages collectifs** : "Un ménage collectif est un groupement de personnes vivant en communauté, logeant dans certains établissements en chambres individuelles ou collectives et prenant souvent leur repas en commun." Le tableau A.2.2.1 précise les 9 catégories de personnes constituant les ménages collectifs.

- **La population comptée à part** : cette catégorie est composée de 6 types de population, répertoriés dans le tableau A.2.2.1.

- **La population des habitations mobiles** : il s'agit "des personnes itinérantes logeant en caravane ou roulotte".

A partir de ces catégories de population, l'INSEE calcule la population des communes selon trois définitions :

- population municipale
- population totale (ou légale)
- population totale sans doubles comptes.

Dans le tableau A.2.2.1, sont indiqués les modes de calcul de ces trois types de population.

1.2. Les caractéristiques individuelles

Nous ne rappellerons ici que les définitions des caractéristiques individuelles que nous avons retenues pour l'intégration dans le système d'information géographique SAVANE (voir liste des indicateurs démographiques intégrés dans la base Marseille dans le chapitre 4, Tableau 4.7).

Age

Dans les statistiques du recensement de 1982, "l'âge indiqué est l'âge atteint au cours de l'année 1982 (différence entre 1982 et l'année de naissance) qui est également l'âge en années révolues au 31 décembre 1982".²

C'est à dire que les personnes classées "10 ans" au recensement de 1982 avaient en fait entre 9 ans 2 mois et 10 ans 2 mois, à la date de référence du recensement, soit le 4 mars 1982.

¹ : Toutes les citations entre guillemets du paragraphe 1. 1 sont du guide publié par l'INSEE : (Recensement général de la population de 1982. Guide d'utilisation).

² : Toutes les citations entre guillemets des paragraphes 1.2 et 1.3 sont de l'INSEE (Recensement général de la population de 1982. France métropolitaine).

Nationalité

La population est répartie en "Français de naissance", "Français par acquisition" (classés selon leur nationalité antérieure), et "Etrangers" (classés selon leur nationalité).

Parmi les étrangers présents en France au moment du recensement de 1982, n'ont été recensés que "ceux qui ont leur résidence permanente en France et ceux qui y travaillent ou étudient (ainsi que leur famille, le cas échéant)", y compris "le personnel étranger (administratif, technique ou de service) des ambassades résidant de façon permanente en France".

N'ont pas été recensés les "travailleurs saisonniers et les travailleurs frontaliers" ainsi que "les étrangers membres du corps diplomatique".

Type d'activité

Toutes les personnes recensées en 1982 atteignant leur quinzième anniversaire l'année du recensement sont classées selon le type d'activité, c'est-à-dire dans une des rubriques suivantes :

- ACTIFS : "La population active comprend la population active ayant un emploi et les chômeurs".
- INACTIFS : cette catégorie regroupe :
 - les anciens actifs (retraités n'ayant pas d'emploi et n'en cherchant pas
 - les étudiants ou élèves (sans activité professionnelle)
 - les militaires du contingent
 - les autres inactifs

Dans le cadre du recensement, la catégorie " *chômeurs*", qui, comme nous venons de le voir font partie des "actifs", regroupe les personnes qui :

- ne sont pas étudiants ou stagiaires rémunérés
- et* - n'ont pas répondu aux questions 12 à 15 relatives à la profession exercée actuellement (voir bulletin individuel du recensement ci-joint)
- et* - ont répondu à la question 16a "être chômeur, ou mère de famille, ou retraité" (ou pas de réponse)
- et* - ont répondu à la question 16b "rechercher actuellement un travail".

Une personne sans travail n'en recherchant pas est classée "inactive". Un étudiant ou élève recherchant un travail est classé "étudiant ou élève". Par contre, un retraité recherchant un travail est classé "chômeur".

1. 3. Immeubles et logements

Est considéré comme *immeuble* :

- " toute construction habitée au moment du recensement, quelle que soit sa nature, les matériaux utilisés, y compris les immeubles vétustes, les habitations de fortune, et les immeubles en cours de construction partiellement habités ;
- toute construction à usage d'habitation, même sans occupant à l'époque du recensement, à l'exception des immeubles en démolition ou devenus totalement inhabitables par vétusteté ;
- tout ensemble de bâtiments administratifs, industriels ou commerciaux, même inhabités."

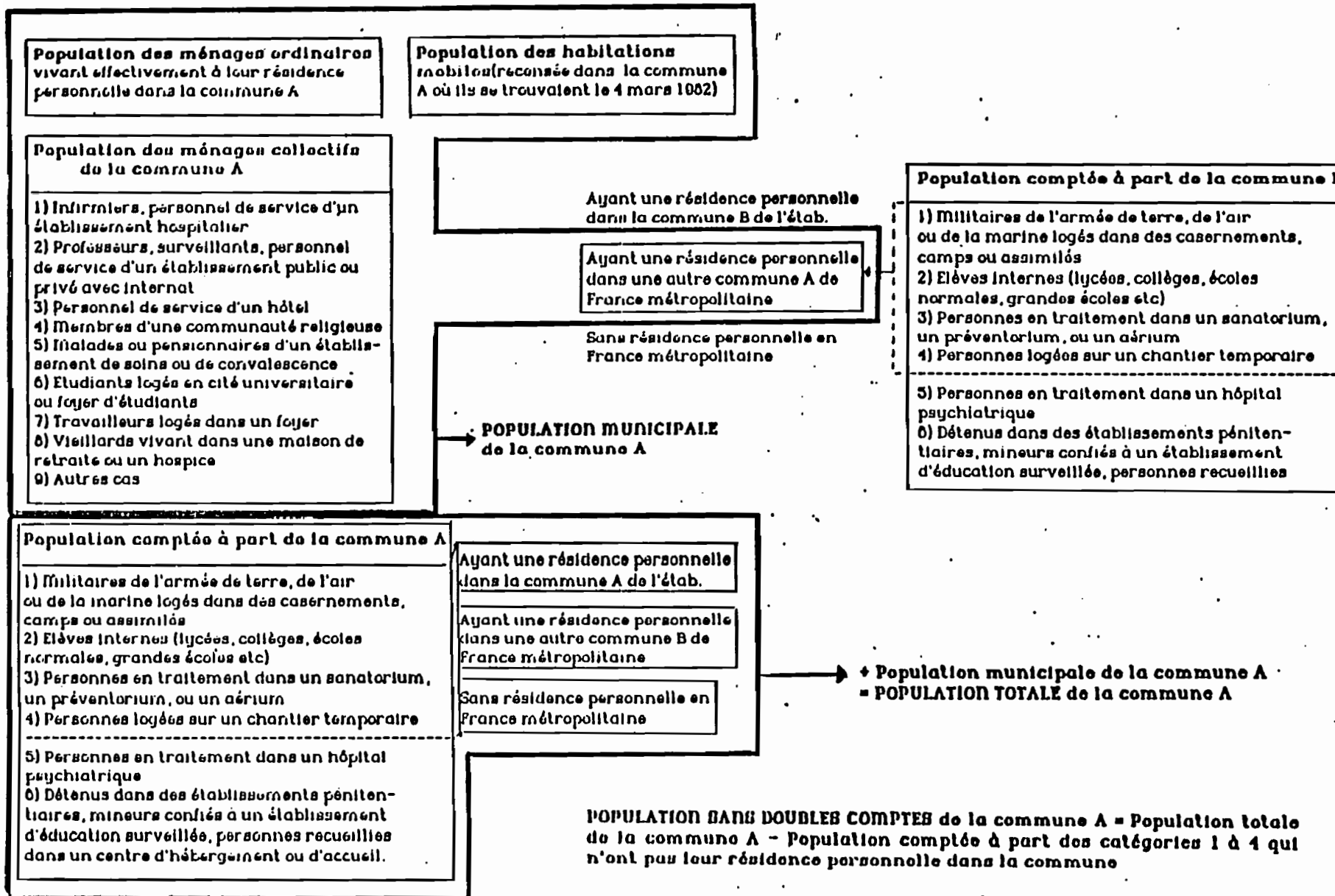
On appelle " *Logement*", tout local qui satisfait aux trois conditions suivantes :

- "séparé, c'est-à-dire complètement fermé par des murs et cloisons sans communication avec un autre local, si ce n'est sur les parties communes de l'immeuble"
- "indépendant, c'est-à-dire qu'il doit posséder une entrée indépendante où l'on a directement accès sur l'extérieur ou sur les parties communes de l'immeuble sans avoir à traverser un autre local"
- "utilisé pour l'habitation, en tenant compte de sa destination actuelle et non de sa destination primitive".

L'INSEE distingue plusieurs catégories de logements :

- les *résidences principales* : "logements occupés de façon permanente et à titre principal par un ménage"
- les *logements vacants* : "logements sans occupants"
- les *résidences secondaires* : logements "soit utilisés par le propriétaire, sa famille ou ses amis, soit loués à des particuliers, pour des séjours temporaires, en général pendant les vacances, fins de semaine, etc."

Tableau A.2.2.1 - Définition des différentes catégories de population utilisées par l'INSEE.



BORDEREAU DE MAISON

REPUBLIQUE FRANÇAISE
RECENSEMENT GÉNÉRAL DE LA POPULATION DE 1982

PAGE 4

CARACTÉRISTIQUES DE L'IMMEUBLE

4 Ce bordereau sera rempli par l'agent recenseur

Numéro du district de recensement: _____

Numéro d'ordre de l'immeuble: _____

Pour connaître ce numéro d'ordre voir le schéma page 5 du Manuel de l'agent recenseur

Coteur de la Mairie: _____

Établissez un bordereau de maison:

- Pour toute construction habitée, quelle que soit sa nature, quels que soient les matériaux utilisés, y compris les immeubles vétustes, les habitations de fortune; y compris également les immeubles en cours de construction partiellement habités.
- Pour toute construction à usage d'habitation, même sans occupants à l'époque du recensement (logements vacants et résidences secondaires) à l'exception des immeubles en démolition ou devenus totalement inhabitables par vétusté.
- Pour tout ensemble de bâtiments administratifs, industriels ou commerciaux, même inhabités.

Ne remplissez pas ce bordereau pour les caravanes, roulottes et autres habitations mobiles.

Adresse précise: N° _____ Rue (ou haud): _____

Commune: _____ Département: _____

Pour Paris: 15^e arrondissement

Donnez toutes autres indications permettant de repérer l'immeuble: _____

LISTE DES LOGEMENTS D'HABITATION

Cette liste sera établie après achèvement de la collecte, à raison d'une ligne pour chaque logement d'habitation des catégories 1 à 7 énumérées sur la première page de la feuille de logement (reportez-vous au paragraphe 5.1 du Manuel de l'agent recenseur)

S'il n'y a aucun logement d'habitation dans l'immeuble, inscrivez "neant" sur la première ligne.

N'inscrivez pas les locaux occupés par de la population comptée à part ou par des ménages collectifs (les imprimés n° 3 et n° 1 bis correspondants ne seront pas classés dans les bordereaux de maison)

Nombre de logements d'habitation (catégories 1 à 5): _____

logements vacants (catégorie 6): _____

résidences secondaires (catégorie 7): _____

Nombre total de logements d'habitation: _____

A remplir par l'agent recenseur				A remplir par la Mairie			
Numéro du logement	Localisation dans l'immeuble (escalier, étage, situation sur la parcelle, numéro du logement ou de la chambre, etc.)	Nom de l'occupant	Catégorie de logement (1 à 7), voir FL	Nombre d'imprimés n° 2		Nombre de votes B réintégrés	
				supprimés (liste A seulement)	ajoutés (liste B seulement)		
01							
02							
03							
04							
05							
06							
07							
08							
Total							
partiel à reporter en haut de la page 2)							
général							

1) TYPE D'IMMEUBLE

- 1 Ferme, bâtiments agricoles.
- 2 Habitation de fortune (halte pour wagon, baraque de "bidonvilles", bâtiment habité dans un terrain, etc.) ou construction provisoire à usage d'habitation.
- 3 Hôtel, pension de famille, garni (occupant la totalité ou la plus grande partie de l'immeuble)
- 4 Maison individuelle ou immeuble collectif entièrement ou principalement utilisé pour l'habitation
- 5 Immeuble principalement à usage industriel, commercial, administratif ou public; un tel immeuble peut comprendre un ou plusieurs logements d'habitation pour personnel logé (c'est souvent le cas du directeur, du gardien, etc.). Précisez la nature exacte de l'immeuble: _____

Exemples: usine, atelier, magasin, dépôt, grand magasin, banque, immeuble aménagé en bureaux, hôpital, clinique, maison de repos, sanatorium, dispensaire, école, internat, collège, immeuble administratif, mairie, caserna, gara, bureau de poste, église, stade, etc.

Si vous hésitez sur un cas particulier, décrivez-le: _____

2) Cet immeuble appartient-il à un organisme N.L.M. (société ou office)?

OUI 1
NON 0

Si oui, précisez le nom de l'organisme propriétaire: _____

3) ANNÉE D'ACHÈVEMENT DE LA CONSTRUCTION

- Si les différentes parties ne sont pas de la même époque, indiquez l'année d'achèvement de la partie habitée, ou de la partie habitée la plus importante.
- 1 Avant 1871
 - 2 de 1871 à 1914
 - 3 de 1915 à 1948
 - 4 de 1949 à 1980
 - 5 de 1981 à 1987
 - 6 de 1988 à 1974
 - 7 1975 ou après
 - 8 Immeuble en cours de construction partiellement habité
- Dans ce cas, précisez l'année d'achèvement: **19** _____

4) NOMBRE D'ÉTAGES AU-DESSUS DU REZ-DE-CHAUSSÉE:

Y compris les étages mansardés habitables. Ne comptez pas les caves et les sous-sols; le rez-de-chaussée surélevé est assimilé à un rez-de-chaussée ordinaire, mais l'entresol compte pour un étage.

S'il y a plusieurs corps de bâtiment, indiquez le nombre d'étages du corps de bâtiment le plus élevé.

5) ASCENSEUR

Y a-t-il un ascenseur dans l'immeuble? OUI 1
NON 0

6) ORIGINE DE L'EAU POUR USAGES DOMESTIQUES

- 1 Immeuble raccordé à un réseau collectif de distribution (ville, commune, service des eaux, usine, etc.)
- 2 Immeuble raccordé par une canalisation à un point d'eau (puits, puits, source, etc.)
- 3 L'eau courante n'est pas installée dans l'immeuble

7) GAZ

- 1 Immeuble raccordé à un réseau public de distribution
- 2 Immeuble alimenté par une chaudière de gaz.
- 3 Immeuble alimenté par une installation fixe de bouteilles de gaz.
- 4 Ni raccordement à un réseau public ni alimentation par une installation fixe (chaudière ou bouteilles)

8) ÉVACUATION DES CABINETS D'AISANCES (W.C.) ET DES EAUX MÉNAGÈRES

- 1 Cabinets d'aisances:
 - 1 Raccordement à l'égout.
 - 2 Autres cas (fosse septique, fosse fixe, etc.)
- 2 Eaux ménagères (évier, lessive, toilette, etc.):
 - 1 Raccordement direct à l'égout.
 - 2 Autres cas (cuvette, fossé, puits, etc.)

9) CHAUFFAGE CENTRAL

- 1 Chauffage urbain
- 2 Chauffage collectif pour un groupe d'immeubles.
- 3 Installation collective de chauffage central propre à l'immeuble (desservant la totalité ou la plupart des logements de l'immeuble), maison individuelle disposant du chauffage central
- 4 Pas d'installation de chauffage central (même si certains logements disposent d'installations individuelles)

Si vous avez marqué la case 2 ou la case 3, indiquez le combustible utilisé.

Charbon 1
Bois 2
Marquage (fuel) 3
Gaz 4
Électricité 5
Autres 6

Précisez: _____

10) EXPLOITATIONS AGRICOLES

- Cet immeuble est-il le siège d'une exploitation agricole? OUI 1
NON 0
- Si oui:
- 1 Superficie agricole utilisée: _____ hectares, _____ ares
 - 2 Orientation des productions agricoles (autres que céréales):
 - 1 Polyculture (cultures de terres labourables)
 - 2 Maraîchage ou horticulture
 - 3 Vignes ou arbres fruitiers
 - 4 Élevage d'herbivores (bovins, ovins...)
 - 5 Élevage de granivores (porcs, volailles...)
 - 6 Polyculture - élevage
 - 7 Autre cas (Élevage d'herbivores et de granivores)
 - 8 Autres
 - 3 Une personne ou moins travaille-t-elle à plein-temps sur l'exploitation (chef d'exploitation, membre de la famille ou salarié agricole)? OUI 1
NON 0

Figure A.2.2.2 - Documents utilisés par l'INSEE pour la collecte des données du recensement (I)

Mod. n° 6001 du Ministère de l'Économie et des Finances. Émis par décret n° 81-711 du 7 juin 1981 modifié. Ce document est distribué gratuitement. Il est interdit de reproduire sans autorisation écrite de l'INSEE.

FEUILLE DE LOGEMENT

REPUBLIQUE FRANÇAISE
RECENSEMENT GÉNÉRAL DE LA POPULATION DE 1982

1	Cet imprimé sera rempli pour tout logement d'habitation, occupé ou non.	Cadre à remplir par l'agent recenseur	Canton de la Mairie
		Numéro du district de recensement	Numéro de la Mairie
		Numéro d'ordre de l'immeuble	Numéro du logement (ou de l'habitation mobile)

POURQUOI LE RECENSEMENT ?

Le recensement général de la population est une opération à laquelle procèdent régulièrement tous les pays. Il a pour but de déterminer le nombre d'habitants de chaque commune, de connaître la répartition de la population selon l'âge, la profession, la branche d'activité, d'analyser les migrations, la composition et l'équipement des immeubles et logements, etc. Ces informations sont indispensables à l'étude de tout problème démographique, économique ou social dans le cadre national, régional ou communal.

Prescrit par le décret du 28 avril 1981 (Journal Officiel du 30 avril 1981), le recensement général de 1982 — le 31^e depuis 1801 — est obligatoire pour toutes les personnes résidant en France.

La loi sur le secret statistique garantit que les renseignements individuels figurant sur les questionnaires du recensement seront utilisés uniquement à des fins statistiques. Ces questionnaires confidentiels sont destinés à l'Institut National de la Statistique et des Études Économiques (I.N.S.E.E.); il est interdit d'en prendre copie sous peine des sanctions prévues à l'article 44 de la loi n° 78-17 du 6 janvier 1978 (voir la notice explicative).

COMMENT REMPLIR LES IMPRIMÉS ?

Vous remplirez d'abord le présent imprimé en commençant par les pages 2 et 3 (liste des personnes habitant dans le logement), puis en continuant par la page 4 (composition et équipement du logement). Ensuite, pour chacune des personnes inscrites dans la liste A de la page 2, vous remplirez un bulletin individuel (imprimé n° 2).

Avant de répondre à une question, lisez les explications qui figurent sur les imprimés; lorsque ces explications ne suffisent pas pour répondre à une situation particulière, reportez-vous à la notice explicative. En cas de difficulté, adressez-vous à l'agent recenseur.

En observant ces quelques recommandations et en suivant les indications qui vous seront fournies par l'agent recenseur, vous faciliteriez le travail de l'I.N.S.E.E. qui vous en remercie.

CADRE A REMPLIR PAR L'AGENT RECENSEUR

Logement occupé par M. _____ (nom des occupants et prénom usité)

Adresse très précise: _____

N° _____ Rue (ou boulevard) _____

Commune _____ (pour Paris Lyon Marseille préciser l'arrondissement)

Département _____

NUMÉRO D'IMPRIMÉS N° 2
RECUELLIS
(liste A seulement)

(Case à remplir
après vérification
des bulletins collectés)

Précisez la localisation du logement dans l'immeuble (escalier, étage, situation sur le balcon, numéro du logement ou de la chambre, etc.):

CATÉGORIE DE LOGEMENT

- 1 Logement ordinaire
- 2 Logement-foyer pour personnes âgées
- 3 Pièces indépendantes (exemple: chambre de domestique) louées, sous-louées ou prêtées à des particuliers.
Indiquez le logement principal auquel elles se rattachent: _____
- 4 Chambres meublées dans un hôtel, une pension de famille, un garni, etc.
- 5 Habitation de fortune ou construction provisoire à usage d'habitation. Précisez: _____
- 6 Logement vacanciers occupants, disponible ou non pour la location. } pas de bulletins individuels; remplir la page 4 du présent imprimé seul les questions 4, 5 et 12.
- 7 Résidence secondaire ou logement loué (ou à louer) pour des séjours touristiques.
- 8 Caravane, roulotte, habitation mobile (ne pas remplir la page 4, ne pas établir de bordereau de maison; à classer à part).

Si vous êtes en présence d'un cas que vous hésitez à classer dans la liste ci-dessus, décrivez-le:

Dans le cas d'un ménage collectif, établissez une feuille de ménage collectif (imprimé n° 3bis)

COMPOSITION ET ÉQUIPEMENT DU LOGEMENT

PAGE 4

1 AVEZ-VOUS UNE CUISINE ?

- OUI → Quelle est sa surface? moins de 7 m² 1
de 7 à 12 m² 2
plus de 12 m² 3

- NON → Avez-vous cependant dans votre logement une installation pour faire la cuisine? OUI 4
NON 5

Ne répondez "OUI" que si cette installation comprend un évier permettant l'évacuation des eaux usées.

2 Indiquez, s'il y a lieu, le nombre de pièces du logement ayant un usage exclusivement professionnel.

Par exemple: bureau d'homme d'affaires, cabinets de médecin ou d'avocat, atelier de tailleur en appartement, etc.

3 NOMBRE DE PIÈCES D'HABITATION

non compris la cuisine, ni les pièces à usage exclusivement professionnel.

Comptez comme pièces d'habitation les pièces telles que: chambre à coucher, salle à manger, salle de séjour, etc. quelle que soit leur surface, ainsi que les chambres de domestique et les mansardes habitables.

Ne comptez pas comme pièces d'habitation les pièces telles que: cuisine, couloir, salle de bains, écuvé, W.-C., buanderie, etc.

Cas particuliers des pièces indépendantes (par exemple, chambres de domestique séparées du logement proprement dit):

Comptez ces pièces parmi vos pièces d'habitation si vous en disposez vous-même (soit pour y loger un membre de votre ménage: enfant, domestique, etc.; soit comme débarras, etc.)

Ne les comptez pas si elles sont louées, sous-louées ou prêtées à d'autres personnes. Dans ce cas, elles constituent, au sens du recensement, un logement distinct et leurs occupants rempliront une feuille de logement distincte.

4 ÊTES-VOUS:

- Propriétaire de votre logement ou de la maison où se trouve votre logement? (y compris les différentes formes d'accès à la propriété, dont la location-vente et la location-attribution) 1
- Logé par votre employeur (à titre gratuit ou onéreux pendant la durée de votre fonction ou de votre contrat de travail)? 2
- Logé à titre gracieux, par exemple par des parents? (y compris le cas des personnes occupant un logement qu'elles ont vendu en viager ou dont elles ont la jouissance par usufruit) 3
- Locataire ou sous-locataire d'un local loué vide? 4
- Locataire ou sous-locataire d'un local loué meublé, d'une chambre d'hôtel, d'un garni? 5

Si vous êtes dans un cas non prévu ci-dessus, décrivez-le:

5 SI VOUS ÊTES PROPRIÉTAIRE DE VOTRE LOGEMENT (réponse 1 à la question 4)

- Pour l'achat de ce logement, avez-vous fait un emprunt que vous n'avez pas encore achevé de rembourser? OUI 1
NON 2
- Répondez "OUI" en cas de location-vente ou de location attribution.

6 ALIMENTATION EN EAU

- Eau courante dans le logement 1
Eau chaude par installation individuelle (propre au logement) 2
Eau chaude par installation collective 3
Eau froide seulement 4
- Pas d'eau courante dans le logement. 4

7 INSTALLATIONS SANITAIRES

- Avez-vous une baignoire et une douche installées avec eau courante et évacuation des eaux usées? 1
Baignoire 2
Douche 3
Ni baignoire ni douche 4

8 CABINETS D'AISANCES (W.-C.)

- Situés à l'intérieur du logement 1
- Situés hors du logement mais réservés à ses seuls occupants 2
avec chasse d'eau 3
sans chasse d'eau 4
- Autres cas 4

9 MODE DE CHAUFFAGE

Pour les maisons individuelles disposant du chauffage central, cochez le case 1.

- Chauffage central individuel avec une chaudière propre à votre logement (y compris le chauffage électrique mixte et le chauffage par air pulsé) 1
- Chauffage central collectif (commun à la totalité ou à la plupart des logements de l'immeuble) 2
- Autres modes de chauffage (poêle, cuisinière, radiateur électrique à accumulation, etc.) quel que soit le combustible utilisé 3

10 COMBUSTIBLE UTILISÉ POUR LE CHAUFFAGE

- Quel est le combustible (ou la source d'énergie) utilisé pour le chauffage de votre logement? 1
Charbon 2
Bois 3
Mélange fluide 4
Gaz 5
Électricité 6
Autres 7
Préciser 8

11 TÉLÉPHONE

- Est-il installé dans le logement? OUI 1
NON 2

12 Les habitants du logement disposent-ils d'une ou plusieurs voitures de tourisme?

- OUI — Combien? 1 1
2 ou plus 2
 NON 2

Figure A.2.2.2 - Documents utilisés par l'INSEE pour la collecte des données du recensement (II)

LISTE DES PERSONNES HABITANT DANS LE LOGEMENT

Chaque personne habitant normalement dans le logement, même si elle est absente à l'époque du recensement, doit être inscrite soit dans la liste A, soit dans la liste B ci-dessous. N'oubliez pas les enfants en bas âge.

NOTA. Si vous avez une résidence secondaire (par exemple une maison de campagne ou de vacances), vous ne devez remplir les listes A et B ci-dessous, que pour votre résidence principale. Vos bulletins individuels doivent être remplis une seule fois, dans la commune de votre résidence principale.

Liste A Inscrivez ci-dessous les personnes habitant dans le logement.

Y compris les personnes en voyage d'affaires ou d'agrément, les malades faisant un séjour à l'hôpital ou en clinique, les pêcheurs en mer, le personnel navigant des compagnies aériennes ou maritimes; non compris les personnes énumérées dans l'en-tête de la liste B ci-contre (lisez cet en-tête avant de remplir la liste A)

- Si vous logez chez vous des gens de maison, apprentis ou salariés à votre service, inscrivez-les ci-dessous (liste A).
- Si vous avez des pensionnaires (ou des enfants en nourrice) qui logent chez vous, inscrivez-les ci-dessous (liste A).
- Si vous sous-louez (ou si vous prêtez une partie de votre logement, inscrivez ci-dessous vos sous-locataires (ou les personnes que vous hébergez). Toutefois, si ces sous-locataires (ou personnes hébergées) habitent des pièces totalement indépendantes c'est-à-dire qui ont accès sur l'extérieur, soit directement, soit par l'intermédiaire des locaux communs de l'immeuble (escalier, vestibule, etc.) — par exemple une chambre de domestique entièrement séparée du logement proprement dit — vous devez considérer que ces pièces indépendantes forment un logement distinct, pour lequel une feuille de logement distincte doit être établie. Dans ce cas, n'inscrivez pas les sous-locataires dans la liste ci-dessous et ne comptez pas les pièces qu'ils occupent dans la composition du logement (page 4).

REMPLEZ UN BULLETIN INDIVIDUEL N°2 POUR CHAQUE PERSONNE DE LA LISTE A

NOM DE FAMILLE <small>(pour une femme, ajoutez le nom de jeune fille)</small>	PRÉNOM USUEL	PARENTÉ OU RELATION avec la personne inscrite sur la 1 ^{re} ligne <small>Indiquez par exemple époux ou épouse, fils, fille, beau-fils, belle-fille, mère, belle-mère, beau-père, etc. ou bien une pensionnaire sous-locataire, etc. (page 4)</small>
Inscrivez sur la 1 ^{re} ligne l'un des conjoints d'un couple (et, sur la 2 ^e ligne, l'autre conjoint) ou, à défaut, l'un des adultes habitant dans le logement		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		

Si vous avez plus de 16 personnes à inscrire, utilisez une feuille de logement supplémentaire

Liste B Inscrivez ci-dessous les personnes qui font partie de votre ménage et qui se trouvent actuellement dans l'un des cas suivants:

- Mères du contingent;
- Mères de carrière servant hors métropole;
- Enfants placés dans un internat, un pensionnat; Étudiants logés ailleurs pendant l'année scolaire (sauf les enfants et étudiants placés dans un établissement hors de la métropole, qui doivent être inscrits dans la liste A);
- Enfants placés en nourrice ailleurs;
- Enfants placés comme gens de maison, salariés ou apprentis et logés chez leur employeur (ou sur le lieu de leur travail);
- Malades en sanatorium, préventorium, asile;
- Vieilles ou infirmes placées dans une maison de retraite ou un établissement spécialisé, personnes en traitement dans un hôpital psychiatrique;
- Autres personnes en absence de longue durée (plus de 8 mois)

NE REMPLISSEZ PAS DE BULLETIN INDIVIDUEL POUR LES PERSONNES DE LA LISTE B
(Leurs bulletins seront établis à l'endroit où elles séjournent actuellement)

NOM DE FAMILLE et prénom d'usage	PARENTÉ OU RELATION avec la personne inscrite sur la 1 ^{re} ligne de la liste A	DATE ET LIEU DE NAISSANCE	DESIGNATION ET ADRESSE DE L'ÉTABLISSEMENT (ou du logement) où l'intéressé séjournait actuellement
1		Né(e) le : _____ A } Commune : _____ A } Département : _____	
2		Né(e) le : _____ A } Commune : _____ A } Département : _____	
3		Né(e) le : _____ A } Commune : _____ A } Département : _____	
4		Né(e) le : _____ A } Commune : _____ A } Département : _____	

CAS DES PERSONNES DE PASSAGE

chez vous au moment du recensement mais ayant ailleurs leur résidence habituelle.
(Parents, amis, voyageurs, etc.)

- N'inscrivez en aucun cas ces personnes dans les listes A et B ci-dessus;
- Ne remplissez pas de bulletin individuel pour elles.

Toutefois, si elles ont été absentes de leur résidence habituelle pendant toute la durée des opérations de recensement et si l'un d'eux a une personne à cette résidence pour répondre à leur place:

1. Établissez leur bulletin individuel (sans les inscrire dans les listes A et B), en mentionnant leur adresse actuelle dans le cadre réservé à cet effet en bas du bulletin;
2. Remettez ces bulletins séparément à l'agent recenseur; ils seront adressés à la mairie de leur résidence actuelle.

P Si vous avez des hésitations sur le cas d'une personne (vous vous demandez si tout l'inscrire dans la liste A ou dans la liste B), veuillez indiquer quelle est cette personne

NOM : _____
Prénom : _____
Né(e) le : _____
A } Commune : _____
A } Département : _____
Indiquez dans quel cas particulier se trouve actuellement cette personne : _____

2

A remplir après la feuille de logement (Imprimé n° 1). Cet imprimé sera rempli pour TOUTE PERSONNE inscrite dans la liste A de la feuille de logement.

Case à remplir par l'agent recenseur. Numéro du carnet de recensement, Numéro d'ordre de l'annuaire, Numéro du logement ou de l'habitation mobile, ou du ménage collectif.

Cochez de la main!

1 NOM ET PRÉNOMS. Cochez le nom en capitales, une femme ajouter son nom de jeune fille. Exemple: MAURIN née ALLARD, Marie, Lucie.

2 ADRESSE

3 SEXE. Masculin, Féminin

4 SITUATION DE FAMILLE. Célibataire, Marié(e), Veuf(ve), Divorcé(e)

5 DATE ET LIEU DE NAISSANCE. Né(e) le, à commune, Département.

6 NATIONALITÉ. Français de naissance, Devenu français, Indiquez votre nationalité antérieure, Étranger.

7 OÙ HABITEZ-VOUS LE 1er JANVIER 1976? Dans le même logement, Dans la même commune, Dans une autre commune, Indiquez cette autre commune, Département.

8 POUR TOUT ENFANT DE 2 A 14 ANS: L'enfant va-t-il à l'école? OUI, NON

POUR TOUTE PERSONNE DE 14 ANS OU PLUS

9 ÊTES-VOUS ACTUELLEMENT: Élu(e) ou étudiant(e), Étudiant(e) rémunéré(e), Dans une autre situation?

10 INDIQUEZ TOUTS LES DIPLÔMES QUE VOUS POSSÉDEZ: Enseignement général primaire ou secondaire, Enseignement technique et formation professionnelle de niveau secondaire, Enseignement supérieur.

11 A QUEL ÂGE avez-vous cessé de suivre régulièrement les cours d'un établissement scolaire? OUI, NON

12 Indiquez la profession ou le métier que vous exercez actuellement. Aidez-vous un membre de votre famille dans son travail?

13 Exercez-vous cette profession comme: Employeur ou travailleur indépendant, Aide familial non salarié, Apprenti sous contrat, Salarié.

14 OÙ TRAVAILLEZ-VOUS? Adresse de votre lieu de travail, Commune et dépt., NOM (ou raison sociale) de l'établissement, ACTIVITÉ de cet établissement.

15 POUR LES SALARIÉS: Indiquez la catégorie professionnelle de votre emploi actuel, Si vous êtes agent de l'État, Si vous êtes ingénieur, cadre, agent de maîtrise ou technicien.

16 SI VOUS N'EXERCEZ PAS ACTUELLEMENT D'ACTIVITÉ PROFESSIONNELLE OU SI VOUS ÊTES EN CHÔMAGE: Êtes-vous: Membre de famille, Retraité, Élève ou étudiant, Chômeur, Cherchez-vous actuellement du travail?

Figure A.2.22 - Documents utilisés par l'INSEE pour la collecte des données du recensement (IV)

POUR LES PERSONNES DE PASSAGE (voir le cadre C, page 3 de l'imprimé n° 1), adresse de la résidence habituelle; N° Rue (ou houblé); Commune; Département.

2. Stratification morphologique et caractéristiques socio-démographiques de la population marseillaise

Le Tableau A.2.2.3 fournit les moyennes arithmétiques des indicateurs de structure démographique intégrés dans la base de données gérée par le système SAVANE sur les districts de recensement (à ne pas confondre avec la moyenne du caractère dans chaque strate) ; il confirme l'intérêt de la stratification en types morphologiques pour l'appréhension des structures de population à Marseille.

- L'immigration étrangère

La ville de Marseille est bien connue pour sa forte proportion d'étrangers ; la division de l'agglomération en types morphologiques montre de très fortes différences dans la répartition des étrangers dans la ville. Le phénomène de concentration de la population étrangère dans le centre ville dense et les quartiers industriels du nord de la ville ressort clairement de l'examen des deux premières lignes du tableau : le pourcentage d'étrangers (hommes et femmes confondus) oppose le centre ville dense (12 %) et les quartiers industriels (9 %) au reste de la ville, où il est partout inférieur à 6 %. Les taux d'étrangers les plus faibles (entre 3 et 4,5 %) se rencontrent dans les strates "petits immeubles" et "pavillonnaire".

- Le chômage

Pour le taux de chômeurs parmi les hommes de plus de 19 ans, les différences entre moyennes par strate morphologique et moyenne générale sur la zone d'étude sont beaucoup moins importantes que précédemment : pour une moyenne générale de 7,55 %, les valeurs par strate s'échelonnent entre 5,31 et 9,67 %. La répartition des chômeurs à Marseille semble beaucoup plus homogène que celle des étrangers. La faible sensibilité de cet indicateur de chômage aux types de morphologie urbaine pourrait s'expliquer par la nature même du phénomène : tandis qu'une caractéristique individuelle permanente telle que la nationalité a un effet ségrégoire certain, le chômage constitue une caractéristique temporaire qui, si elle ne se prolonge pas trop longtemps, n'a pas d'effet systématique et immédiat sur le lieu et type de résidence des personnes.

En dépit de cette relative homogénéité des taux de chômeurs par strate, on note toutefois que les taux les plus élevés se rencontrent dans le centre ville dense et les grands immeubles d'habitation.

- La taille des ménages

Taille moyenne des ménages et taux de ménages d'une personne apparaissent négativement corrélés et créent quatre sous-ensembles dans l'espace marseillais :

- le centre ville et la strate 8 (immeubles élevés denses) sont caractérisés par une faible taille moyenne des ménages (2,2 personnes par ménage), avec un fort pourcentage de ménages d'une personne (37 %). L'examen des photos a confirmé cette observation : les îlots comptant un fort pourcentage de ménages d'une personne se rencontrent surtout dans les bâtiments à toit foncé, c'est-à-dire le vieux centre ville.

- dans les quartiers industriels, le pavillonnaire et la strate 7, les indicateurs de taille des ménages sont semblables à ceux observés sur l'ensemble de la zone d'étude : 2,5 personnes par ménage, et 28 % de ménages d'une personne.

Tableau A.2.2.3 - Indicateurs de structure de population selon les strates de morphologie urbaine (ensemble des districts INSEE de la base de sondage, moyennes des valeurs observées par district)

Strate : TOTAL	1	2	3	4	5	6	7	8	9	
%Etrangers (H + F)	12,18	9,11	5,76	3,29	3,50	4,48	5,05	4,36	3,70	7,27
%Etrangers (H >19 ans)	13,79	9,77	5,87	3,94	3,94	4,75	5,88	5,17	4,09	8,09
%Chômeurs (H > 19 ans)	9,67	6,72	7,92	6,28	5,65	5,31	7,10	5,85	6,22	7,55
Taille moy. des ménages	2,18	2,48	2,86	2,68	2,55	2,69	2,35	2,11	3,03	2,48
%Ménages d' 1 personne	37,68	28,75	20,51	20,36	23,70	20,50	29,54	36,69	13,60	28,28
%Ménages >= 6 personnes	2,13	3,11	5,38	3,03	2,55	3,65	1,64	0,88	3,40	2,93
% 0-4 ans révolus	3,93	4,02	4,93	4,20	3,31	3,28	3,46	3,52	3,33	3,89
% >=60 ans	29,56	24,94	19,94	22,05	29,64	28,33	26,92	32,85	19,32	27,01

(Source : Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

- les grands ensembles, les petits immeubles et l'espace péri-urbain regroupent des ménages un peu plus grands que la moyenne (2,7 personnes par ménage), et comptent peu de ménages d'une personne (20 %).

Le découpage induit par le pourcentage de ménages de 6 personnes et plus, s'il présente des similitudes avec le précédent, ne lui est pourtant pas strictement équivalent. Les grands ensembles s'isolent avec un pourcentage de familles très nombreuses (5,4 %) presque deux fois supérieur à la moyenne générale. Les strates péri-urbain, quartiers industriels et petits immeubles d'habitation ont des valeurs très proches de la moyenne (3 %) ; enfin, le centre ville et les zones pavillonnaires comptent peu de ménages nombreux, respectivement 2,1 et 2,5 % de ménages de 6 personnes et plus.

La structure par âge

Si les deux indicateurs de structure par âge retenus (% de moins de 5 ans et % de plus de 60 ans) ne varient que relativement peu d'une strate à l'autre (respectivement 3,3 à 5 %, et 19 à 33 %), ces variations semblent toutefois significatives, surtout lorsqu'il s'agit d'ensembles où les effectifs de population sont importants.

On observe une bonne corrélation négative entre ces deux indicateurs de la structure par âge. Les grands ensembles d'habitation sont les quartiers où les jeunes enfants sont les plus présents, suivis des petits immeubles d'habitation, puis du centre ville dense et des quartiers industriels, où les pourcentages sont comparables à ceux de l'ensemble de la zone d'étude. A l'opposé, les strates pavillonnaire et péri-urbain comptent peu de jeunes enfants et beaucoup de personnes âgées.

L'intérêt des quelques lignes qui précèdent ne réside pas dans l'embryon d'analyse socio-démographique de la ville de Marseille faite à partir d'une division en neuf strates morphologiques. Par contre, l'analyse du tableau A.2.2.3 montre que les variations des indicateurs de structure sont suffisamment significatives du point de vue numérique, et cohérentes avec ce que l'on connaît sur ce sujet, pour rendre *intéressante l'introduction d'une stratification morphologique dans les plans de sondage d'enquêtes ne visant pas seulement l'estimation d'effectifs totaux de population, mais aussi l'analyse des structures et comportements démographiques.*

3. Etude statistique préliminaire

Préliminaire

Dans la deuxième partie de la thèse (chapitre 4 à 8), on s'intéresse à l'estimation de statistiques très simples, attachées aux variables démographiques étudiées : totaux de caractères comme la population sans doubles comptes ou la population étrangère ou ratios entre deux caractères, qui décrivent la structure de ces populations comme, par exemple, le ratio population des hommes à la recherche d'un emploi / population des hommes de plus de 19 ans¹.

Cette étude décrit "l'univers" dans lequel se déroulera le test : la base de sondage, la première stratification élaborée. On donne un résumé des distributions de certains caractères tant sur la base entière qu'à l'intérieur de chacune des strates; enfin, on tente de fournir une première indication sur les gains à attendre de cette stratification en calculant les variances inter et intra-strate de ces caractères. Comme, d'une part, il serait fastidieux de répéter ici la description de tous les caractères considérés dans la phase marseillaise de l'expérience et que, d'autre part, de fortes ressemblances existent entre les différentes distributions, nous nous limitons à l'examen de trois variables définies sur l'ensemble des districts INSEE : la population sans doubles comptes, la surface² et la densité de population sans doubles comptes.

¹ : Voir la définition des caractères démographiques au sens de l'INSEE au début de cet annexe.

² : Bien que n'étant pas un caractère démographique, la surface des districts jouera par la suite un rôle très important dans la définition des plans de sondages utilisés.

3.1. Rappel sur la stratification et la base de sondage

La stratification considérée ici est une classification des districts INSEE qui composent la base de sondage d'après des critères de morphologie urbaine. Le travail d'analyse typologique n'a pas porté directement sur les districts INSEE, sur lesquels nous ne disposons pas de l'information morphologique exhaustive et précise. Cette information a été définie sur un ensemble de zones morphologiquement homogènes, déterminées lors de la photo-interprétation d'une mosaïque de photographies aériennes au 1/23 000¹.

Après avoir effectué une classification de cet ensemble de zones, le type morphologique de chaque zone a été affecté aux districts qu'elle contenait, ou, plus exactement, à ceux dont elle contenait la majorité de la surface. Bien sûr, l'homogénéité attribuée à une zone à l'échelle d'observation du 1/23 000 ne se retrouve pas nécessairement lorsque l'on considère des échelles plus grandes. Ceci explique, par exemple, que l'on trouve dans des zones où la densité du bâti est dans l'ensemble très faible (classées dans la strate "non bâti"), certains districts de petite taille très peuplés.

La stratification des districts INSEE doit donc être manipulée avec prudence. On doit plutôt la considérer comme une information d'appartenance à une zone plus vaste que le district, aux caractéristiques morphologiques globalement homogènes, que comme une donnée décrivant la morphologie du district lui-même. Ces précautions étant prises, on va constater que cette stratification opère des regroupements de districts intéressants du point de vue démographique.

Enfin, pour obtenir la base de sondage qui va nous intéresser ici, on a ramené le nombre des districts de la zone d'étude de 3534 à 3461, ceci à cause des exclusions suivantes :

1. 68 districts INSEE qui ne sont pas entièrement inclus dans la zone d'étude et dont on ne connaît que partiellement les caractéristiques morphologiques, qui constituent la strate 10.

2. cinq districts correspondant à des zones non décrites morphologiquement (chemin de fer, eau, digue, port), qui constituent la strate 11.

L'ensemble des résultats présentés dans cette annexe ainsi que ceux du chapitre 6 ont été établis sur cet ensemble de 3461 districts. Par la suite, dans les chapitres 7 et 8, la base de sondage subira une nouvelle modification lors de l'exclusion des districts de caractérisation morphologique mixte (c'est-à-dire ceux que leur situation géographique place à cheval sur des zones de morphologie différente) et des districts constitués de rond-points, terre-plein au milieu de voies larges ou autres zones pouvant être identifiées comme innocupées dès l'interprétation de l'image satellite (on n'a donc supprimé, au cours de cette phase, que des districts dont la surface dépassait l'équivalent de quatre pixels T.M.). Ces suppressions, qui ramènent le nombre de districts de la base à 2619, ont été effectuées parce qu'il n'y a pas de raison de penser que la base de sondage, constituée d'après l'image lors de la phase d'application, présente ce genre d'imperfections.

¹ : Pour plus de détails sur le travail ayant permis l'intégration des données et l'obtention de la stratification, se reporter aux chapitres 5 et 6.

3.2. Description statistique des caractères suivant les strates

Dans les tableaux A.2.2.4 a, b et c qui suivent, on donne successivement, pour la population sans doubles comptes, la surface et la densité des districts INSEE, les principaux paramètres statistiques des variables, calculés sur l'ensemble de la base de sondage et pour chacune des strates morphologiques.

De l'examen des trois tableaux on peut tirer les conclusions suivantes quant à l'intérêt de la stratification obtenue et à son emploi dans la suite du programme.

La répartition de la population sans doubles comptes dans les différentes strates et les variations de la densité d'une strate à l'autre nous laissent espérer une certaine efficacité de la stratification sur la précision des estimations ; en effet, d'une part 77,7 % de la population est concentrée dans les strates 3,1 et 5, c'est-à-dire dans la réunion des grands ensembles d'habitation, du centre ville dense et du pavillonnaire. Ces trois strates ne totalisent que 58 % de la surface totale de la zone d'étude. Compte tenu des intitulés qui décrivent synthétiquement la morphologie des strates, ces résultats peuvent paraître triviaux, mais il est déjà rassurant, concernant l'efficacité supposée de la stratification, que la population soit bien majoritairement là où on s'y attendait.

D'autre part, exception faite des strates 7 et 8, qui sont très marginales en surface (respectivement 3,3 et 0,9 % de l'ensemble de la zone) et qui regroupent à elles deux moins de 9% de la population totale, le sous-ensemble "peuplé" décrit plus haut regroupe bien les zones de densités les plus fortes : 3,02 h/are pour la strate 1, 1,30 h/are dans la strate 3 et 0,61 h/are dans la strate 5.

Tableau A.2.2.4 a : Surface des districts INSEE (statistiques d'ensemble et par strate.)

Strate	Somme (km ²)	part du total (%)	Minimum (km ²)	Maximum (m ²)	Moyenne (m ²)	Ecart type (m ²)
1 Centre ville dense	6,879	9,4	90	75573	6163	6764
2 Quartiers industriels	9,283	12,6	303	309038	3375	48576
3 Grands ensembles d'habitation	20,347	27,7	458	292264	33192	42330
4 Petits immeubles d'habitation	5,022	6,8	175	334615	34160	52976
5 Pavillonnaire	15,410	21,0	37	382529	16684	28879
6 Péri-urbain peu dense	6,710	9,1	1265	368221	75387	83243
7 Reliquat morphologique	2,432	3,3	747	113218	14306	15561
8 Dense: grands bâtiments élevés	0,664	0,9	341	76370	9486	13256
9 Non bâti	6,811	9,3	728	689231	119489	139181
Total	73,563	100,00	37	689231	21255	41682

(Source : Calculs effectués sur les chiffres du Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

Tableau A.2.2.4 b : Population sans doubles comptes des districts INSEE (statistiques d'ensemble et par strate.)

Strate	Somme	part du total (%)	Minimum	Maximum	Moyenne	Ecart type
1 Centre ville dense	207739	28,5	0	1034	186,15	156,31
2 Quartiers industriels	40680	5,6	0	1909	147,93	260,09
3 Grands ensembles d'habitation	264764	36,3	0	2964	431,92	416,40
4 Petits immeubles d'habitation	28750	3,9	0	1931	195,58	287,15
5 Pavillonnaire	93937	12,9	0	1460	101,66	141,33
6 Péri-urbain peu dense	11648	1,6	0	941	130,88	209,58
7 Reliquat morphologique	40534	5,6	0	1319	238,44	255,00
8 Dense: grands bâtiments élevés	23788	3,3	0	1358	339,83	287,36
9 Non bâti	17032	2,3	0	3407	298,80	549,40
Total	728872	100	0	3407	211,00	277,94

(Source : Calculs effectués sur les chiffres du Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement.)

Tableau A.2.2.4 c : Densité de population sans doubles comptes dans les districts INSEE (statistiques d'ensemble et par strate.)

Strate	Minimum (h/are)	Maximum (h/are)	Moyenne (h/are)	Densité moyenne dans la (h/are)	Ecart type	Coef. de Variation (%)
1 Centre ville dense	0	261,12	4,338	3,020	8,516	196
2 Quartiers industriels	0	17,78	1,291	0,438	2,442	189
3 Grands ensembles d'habitation	0	102,66	2,313	1,301	4,668	202
4 Petits immeubles d'habitation	0	22,48	1,050	0,573	2,059	196
5 Pavillonnaire	0	22,32	0,944	0,609	1,195	126
6 Péri-urbain peu dense	0	3,64	0,416	0,174	0,601	144
7 Reliquat morphologique	0	10,40	2,172	1,667	2,067	95
8 Dense: grands bâtiments élevés	0	12,60	4,695	3,583	2,838	60
9 Non bâti	0	3,01	0,342	0,250	0,576	168
Total	0	261,12	2,426	0,991	5,559	229

(Source : Calculs effectués sur les chiffres du Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

En résumé, la stratification morphologique permet donc d'isoler trois sous-ensembles principaux de densités bien différentes qui, chacun, regroupe une part importante de la population :

- 1. Les grands ensembles d'habitation : 36,6 % de la population sur 27,7 % de la surface. $d_3 = 1,301$ h/are.*
- 2. Le centre ville dense : 28,5 % de la population sur 9,4 % de la surface. $d_1 = 3,020$ h/are.*
- 3. Le pavillonnaire : 12,9 % de la population sur 21 % de la surface. $d_5 = 0,609$ h/are.*

Il faut également remarquer que du point de vue de la densité de population, les strates morphologiques sont assez homogènes puisque la totalité des coefficients de variation à l'intérieur des strates sont nettement inférieurs au même coefficient sur l'ensemble de la base ($cdv = 229$). Cette différence est particulièrement importante dans le pavillonnaire ($cdv_5 = 126$), les strates peu denses (péri-urbain, $cdv_6 = 144$, et non bâti, $cdv_9 = 168$) et surtout dans les deux strates "marginales" 7 et 8 : respectivement, $cdv_7 = 95$ et $cdv_8 = 60$. En revanche, dans les quatre premières strates, la densité de population est plus variable.

Du point de vue de la stratégie de sondage, un tel résultat nous porte à espérer un gain important de précision dû à l'emploi d'un plan stratifié, ceci grâce à la variance inter-strate des caractères démographiques.

Par ailleurs, en considérant les colonnes qui donnent l'écart-type des caractères "population sans doubles comptes" et "densité de population sans doubles comptes", on note des différences importantes suivant les strates : en ce qui concerne l'effectif, les strates "pavillonnaire" et "centre ville dense" ont des écarts-type très faibles (respectivement 141 et 156,3) alors que les strates "non bâti" et "grands ensembles d'habitation" sont celles où la variation du caractère est la plus forte (respectivement 549,4 et 416,4). Pour la densité, les différences sont plus importantes encore entre, d'une part, les strates peu denses, où les écarts types sont faibles : "non bâti", $\sigma_9 = 0,58$, et "péri-urbain peu dense", $\sigma_6 = 0,60$, et, d'autre part, les strates de densité plus forte où la variance est importante : "centre ville dense", $\sigma_1 = 8,5$ et "grands ensembles d'habitation", $\sigma_3 = 4,7$.

Ainsi, il est probable qu'une optimisation de l'allocation à chaque strate soit, elle aussi, payante grâce aux différences importantes de variance intra-strate des caractères d'une strate à l'autre (l'allocation optimale de Neyman est en effet proportionnelle à l'écart-type intra-strate¹).

On le voit, les gains de précision qu'on peut attendre de la stratification d'abord, puis de l'optimisation de l'allocation, peuvent être prévus et quantifiés grâce au calcul des variances inter et intra-strate du caractère à estimer. Rappelons d'abord rapidement les définitions et les principaux résultats qui sous-tendent ces calculs.

¹ : Voir J.DESABIE : Théorie et pratique des sondages (DUNOD 1966) P. 145 : L'échantillon de NEYMAN ou échantillon optimal.

3.3. Variance inter-strate et intra-strate des caractères, tentative de prévision du gain apporté par les stratifications

Il est classique en taxinomie mathématique ou en classification automatique¹ de caractériser une partition Q d'un ensemble I par le taux d'inertie, t(Q), ou de variance, t(V), dont elle rend compte. Pour préciser ces notions, nous adoptons les notations suivantes qui sont celles de J.P. BENZECRI :

- I est une partie finie d'un espace euclidien E
- $\|i-i'\|^2$ est le carré de la distance euclidienne entre deux éléments i et i' de I
- $m_I = (m_i / i \in I)$ est un système de masses positives sur I
- $m = \sum_{i \in I} m_i$ est la masse totale de I
- On note $g(I) = g = 1/m \sum_{i \in I} m_i i$, le centre de gravité du système.

Le moment centré d'ordre 2 ou moment d'inertie du système est :

$$M^2(I) = \sum_{i \in I} m_i \|i-g\|^2$$

Soit Q une partition de I en classes non vides deux à deux disjointes, et q le centre de gravité de la classe Cq, tandis que $m_{Cq} = \sum_{i \in Cq} m_i$ est la masse totale de la classe Cq. le moment d'ordre 2 :

$M^2(Q) = \sum_{Cq \in Q} m_{Cq} \|q-g\|^2$, est l'inertie inter-classe de la partition. La variance inter-classe est :

$$V^2(Q) = \sum_{Cq \in Q} \frac{m_{Cq}}{m} \|q-g\|^2$$

¹ : Voir par exemple : J.P. BENZECRI et Collaborateurs : L'analyse des données, 1. La taxinomie, P. 186 et 187.

Pour la classe C_q , on a de même : $M^2(C_q) = \sum_{i \in C_q} m_i \|i - q\|^2$ et

$$V^2(C_q) = \sum_{i \in C_q} \frac{m_i}{m_{C_q}} \|i - q\|^2$$

On montre alors facilement (théorème de Huygens) que :

$$M^2(I) = M^2(Q) + \sum_{C_q \in Q} M^2(q)$$

ou, ce qui revient au même, que :

$$V^2(I) = V^2(Q) + \sum_{C_q \in Q} \frac{m_q}{m} V^2(q)$$

Cela signifie que la variance totale $V^2(I)$ est égale à la somme de la variance entre les classes $V^2(Q)$ (ou variance inter-classe), et de la moyenne pondérée par les m_q des variances intérieures aux classes (variance intra-classe).

Le taux de variance qui caractérise la partition est alors :

$$t(Q) = V^2(Q)/V^2(I) = \text{var. inter-classe} / \text{var. totale}$$

L'application au problème qui nous intéresse est simple. L'espace euclidien E est la droite réelle sur laquelle l'ensemble I est l'ensemble des valeurs prises par le caractère dont on cherche à estimer le total ou la moyenne (population sans doubles comptes ou densité) sur l'ensemble des districts INSEE. Le système de masses quant à lui, est défini par la probabilité de tirage de chaque district dans le plan de sondage considéré. La partition Q est pour nous la stratification.

Bien sûr l'éventail des plans de sondage utilisables pour l'estimation d'effectifs de population est assez vaste, compte tenu des possibilités de choix dans la définition des unités spatiales à enquêter, des probabilités d'inclusion dans l'échantillon et de la stratification ; le formulaire des variances inter et intra-classe s'écrit cependant très facilement dans le cas du *plan équiprobable sans remise*. C'est dans ce cadre que nous nous plaçons maintenant.

On a alors $m_i = 1$ pour tout i et, si N est le cardinal de la base :

$$m=N, \quad g(I) = 1/N \sum_{i=1}^N y_i = \bar{y}, \quad V^2(I) = 1/N \sum_{i=1}^N (y_i - \bar{y})^2 = \sigma_y^2$$

où y_i est la valeur du caractère y pour le district i , \bar{y} la moyenne de y sur la base, σ_y^2 est donc la variance de y sur la base.

De même, le centre de gravité d'une strate h, de cardinal N_h , est la moyenne de y dans la strate :

$$\bar{y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_i$$

et la variance interne à la strate :

$$V_y^2(h) = \sigma_{yh}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_i - \bar{y}_h)^2$$

L'égalité entre variance totale et somme des variances inter-strate et intra-strate s'écrit, pour un caractère y donné :

$$V_y^2(I) = \sigma_y^2 = \sum_{h=1}^K \frac{N_h}{N} (\bar{y}_h - \bar{y})^2 + \sum_{h=1}^K \frac{N_h}{N} \sigma_{yh}^2$$

où K est le nombre de strates. Les résultats des calculs, effectués pour les deux caractères, population sans doubles comptes et densité de cette même population, sont consignés dans les tableaux A.2.2.5 a et b.

L'observation des deux tableaux permet de faire les remarques suivantes concernant la suite du test, c'est à dire la comparaison des variances d'estimation proprement dite.

1. Pour chacun des deux caractères, on calculera donc les variances des estimateurs, avec et sans stratification (estimateur du total pour l'effectif et de la moyenne pour la densité de population). On devrait, d'après les deux colonnes qui donnent les variances inter-strates, obtenir un gain significatif en stratifiant l'estimateur du total de population; mais, en revanche, l'estimateur de la moyenne des densités ne profitera que très peu de la stratification : 20 % de la variance de l'effectif est due à la variance inter-strates alors qu'elle ne représente que 5 % de la variance totale de la densité de population.

Soulignons que cette conclusion n'est valable que dans le cadre du plan de sondage équiprobable sans remise, avec un estimateur sans biais qui ne fait pas intervenir la surface des districts INSEE, et seulement dans le cas de l'allocation proportionnelle.

2. On recommencera les calculs de variances dans le cas stratifié en adoptant l'allocation optimale de Neyman ; le taux dans chaque strate sera proportionnel à l'écart-type du caractère dans la strate.

Comme on pouvait le prévoir, la contribution des strates à la somme des variances intra-strate est très variable (voir les colonnes " % de variance intra-strate" des tableaux 4 et 5), et ceci, particulièrement dans le cas de la densité de population sans doubles comptes. Donc ici, on s'attend à ce que l'allocation optimale profite beaucoup à l'estimation de la moyenne de cette densité.

L'ensemble de ces hypothèses ont trouvées leur confirmation au chapitre 7.

Tableau A.2.2.5 a : Variances inter-strate et intra-strate de la population sans doubles comptes des districts INSEE.

Strate	$\frac{N_h}{N}$	\bar{p}_h	$\frac{N_h}{N} (\bar{p}_h - \bar{p})^2$ (var. inter.)	σ_{ph}^2	$\frac{N_h}{N} \sigma_{ph}^2$ (var.intra.)	% de var.intra.
1 Centre ville dense	0,322	186,15	199,18	24431	7877,96	12,3
2 Quartiers industriels	0,079	147,93	316,09	67648	5375,14	8,4
3 Grands ensembles d'habitation	0,177	431,92	8643,90	173387	30709,81	7,8
4 Petits immeubles d'habitation	0,042	195,58	10,10	82457	3502,24	5,5
5 Pavillonnaire	0,267	101,66	3191,56	21705	5794,89	9,0
6 Péri-urbain peu dense	0,026	130,88	165,09	43923	1129,49	1,8
7 Reliquat morphologique	0,049	238,44	36,97	65025	3193,95	5,0
8 Dense: grands bâtiments élevés	0,020	339,83	335,68	82574	1670,09	2,6
9 Non bâti	0,016	298,81	126,98	301840	4971,06	7,7
Total	1,000		13025,54 16,86%		64224,62 83,14%	100,0

(Source : Calculs effectués sur les chiffres du Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

Tableau A.2.2.5 b : Variances inter-strate et intra-strate de la densité de population sans doubles comptes des districts.

Strate	$\frac{N_h}{N}$	\bar{d}_h	$\frac{N_h}{N} (\bar{d}_h - \bar{d})^2$ (var. inter.)	σ_{dh}^2	$\frac{N_h}{N} \sigma_{dh}^2$ (var.intra.)	% de var.intra.
1 Centre ville dense	0,322	3,020	1,328	72,522	23,385	81,57
2 Quartiers industriels	0,079	0,438	0,024	5,963	0,474	1,65
3 Grands ensembles d'habitation	0,177	1,301	0,017	21,791	3,860	13,46
4 Petits immeubles d'habitation	0,042	0,573	0,007	4,239	0,180	0,63
5 Pavillonnaire	0,267	0,609	0,039	1,427	0,381	1,33
6 Péri-urbain peu dense	0,025	0,174	0,017	0,361	0,009	0,03
7 Reliquat morphologique	0,049	1,667	0,022	4,274	0,210	0,73
8 Dense: grands bâtiments élevés	0,020	3,582	0,136	8,054	0,163	0,57
9 Non bâti	0,016	0,250	0,009	0,331	0,005	0,02
Total	1,000		1,600 (5,29%)		28,667 (94,71%)	100

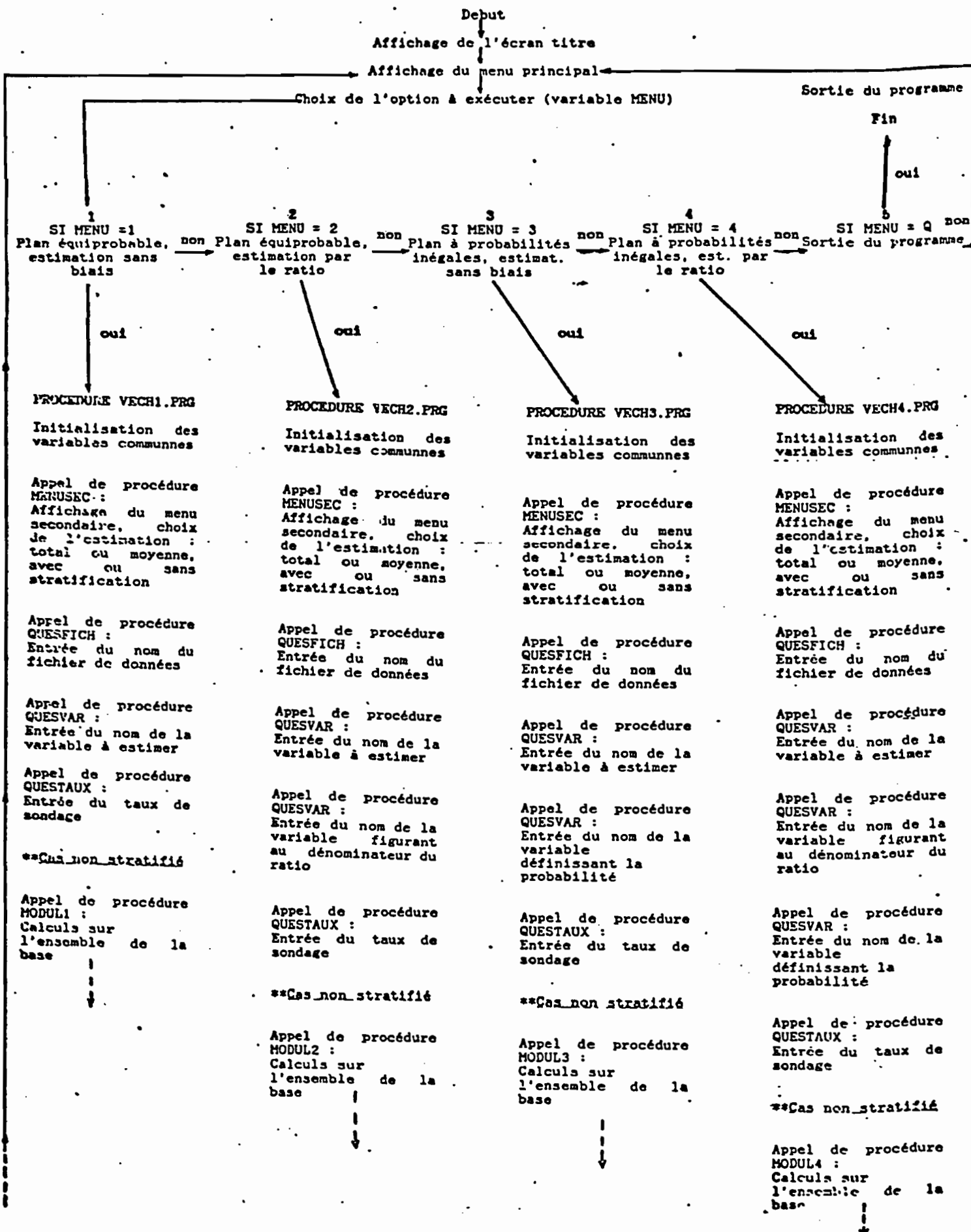
(Source : Calculs effectués sur les chiffres du Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

Annexe A.2.3

**L'OUTIL INFORMATIQUE MIS AU POINT DURANT LA PHASE
D'EXPERIENCE A MARSEILLE**

Organisation générale du programme VARECH.

PROGRAMME PRINCIPAL J. VARECH.PRG



Organisation générale du programme VARECH (suite).

Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat
**Cas_stratifié	**Cas_stratifié	**Cas_stratifié	**Cas_stratifié
Appel de procédure QUESNOMB : Entrée du nombre de strates	Appel de procédure QUESNOMB : Entrée du nombre de strates	Appel de procédure QUESNOMB : Entrée du nombre de strates	Appel de procédure QUESNOMB : Entrée du nombre de strates
Appel de procédure QUESVAR : Entrée du nom de la variable de stratification	Appel de procédure QUESVAR : Entrée du nom de la variable de stratification	Appel de procédure QUESVAR : Entrée du nom de la variable de stratification	Appel de procédure QUESVAR : Entrée du nom de la variable de stratification
* Allocation proportionnelle	* Allocation proportionnelle	* Allocation proportionnelle	* Allocation proportionnelle
Boucle sur h	Boucle sur h	Boucle sur h	Boucle sur h
Appel de procédure MODUL1 : Calculs sur la strate h	Appel de procédure MODUL2 : Calculs sur la strate h	Appel de procédure MODUL3 : Calculs sur la strate h	Appel de procédure MODUL4 : Calculs sur la strate h
Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat
Fin de la boucle sur h	Fin de la boucle sur h	Fin de la boucle sur h	Fin de la boucle sur h
Appel de procédure MODUL1 : Calculs sur l'ensemble de la base	Appel de procédure MODUL2 : Calculs sur l'ensemble de la base.	Appel de procédure MODUL3 : Calculs sur l'ensemble de la base	Appel de procédure MODUL4 : Calculs sur l'ensemble de la base
Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat
* Allocation optimale	* Allocation optimale	* Allocation optimale	* Allocation optimale
Boucle sur h	Boucle sur h	Boucle sur h	Boucle sur h
Calculs sur la strate h	Calculs sur la strate h	Calculs sur la strate h	Calculs sur la strate h
Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat
Fin de la boucle sur h	Fin de la boucle sur h	Fin de la boucle sur h	Fin de la boucle sur h
Calculs sur l'ensemble de la base	Calculs sur l'ensemble de la base	Calculs sur l'ensemble de la base	Calculs sur l'ensemble de la base
Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat	Appel de procédure ECRIT : Ecriture dans le fichier résultat

Liste des procédures du programme VARECH.

Niveaux	Noms	Fonctions
-1- Programme principal	VARECH	Programme principal : affichage de la page de titre, menu principal, appel de procédures principales (VECH1, VECH2, VECH3, VECH4), sortie du programme.
-2- Procédures principales calcul des variances	VECH1	Procédure de calcul des variances : cas du plan équiprobable, estimation sans biais. Appel de la procédure MODUL1 et des procédures utilitaires.
	VECH2	Procédure de calcul des variances : cas du plan équiprobable, estimation par le ratio. Appel de la procédure MODUL2 et des procédures utilitaires.
	VECH3	Procédure de calcul des variances : cas du plan à probabilités inégales, estimation sans biais. Appel de la procédure MODUL3 et des procédures utilitaires.
	VECH4	Procédure de calcul des variances : cas du plan à probabilités inégales, estimation par le ratio. Appel de la procédure MODUL4 et des procédures utilitaires.
-3.1- Procédures de calculs pour les procédures principales	MODUL1	Calcul des variances, écarts-type, coefficients de variation pour VECH1. Appel de CALVECH1.
	CALVECH1	Sommation pour le calcul de l'effectif de la base et de la variance du caractère Y sur la base, pour MODUL1. Probabilités uniformes.
	MODUL2	Calcul des variances, écarts-type, coefficients de variation pour VECH2. Appel de CALVECH2.
	CALVECH2	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y et X sur la base, pour MODUL2. Probabilités uniformes.

Liste des procédures du programme VARECH (suite).

Niveaux	Noms	Fonctions
	MODUL3	Calcul des variances, écarts-type, coefficients de variation pour VECH3. Appel de CALVECH3.
	CALVECH3	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y et Z sur la base, pour MODUL3. Probabilités inégales, proportionnelles à Z.
	MODUL4	Calcul des variances, écarts-type, coefficients de variation pour VECH4. Appel de CALVECH4.
	CALVECH4	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y, X et Z sur la base, pour MODUL4. Probabilités inégales, proportionnelles à Z.
-3.2- Procédures utilitaires pour les procédures principales	MENUSEC	Procédure du menu secondaire pour les procédures principales : affichage du menu secondaire, choix de l'estimation (stratifiée, non stratifiée, estimation d'un total ou d'une moyenne).
	QUESFICH	Procédure d'entrée d'un nom de fichier dans le programme VARECH.
	QUESVAR	Procédure d'entrée d'un nom de variable dans le programme VARECH.
	QUESNOMB	Procédure d'entrée du nombre de strates dans le programme VARECH.
	QUESTAUX	Procédure d'entrée du taux de sondage dans le programme VARECH.
	QUESNEYM	Procédure de choix de l'allocation optimale de NEYMAN dans le programme VARECH.
	ECRIT	Procédure d'écriture des résultats de calcul du programme dans le fichier résultat.

Déroulement du programme VARECH.

Varech, version 5.

ECRAN 1

Jeudi 25 Février 1988

MENU PRINCIPAL

- <1> : Plan équiprobable, estimateur sans biais.
- <2> : Plan équiprobable, estimateur d'un/par le ratio.
- <3> : Plan à probabilités inégales, estimateur sans biais.
- <4> : Plan à probabilités inégales, estimateur d'un/par le ratio
- <Q> : Quitter le programme

Choisissez une option :1

Jeudi 25 Février 1988

ECRAN 2

MENU SECONDAIRE (2)

- <1> : Estimation non stratifiée.
- <2> : Estimation stratifiée.

Choisissez une option :2

Jeudi 25 Février 1988

ECRAN 3

MENU SECONDAIRE (1)

- <1> : Estimation d'un total.
- <2> : Estimation d'une moyenne.
- <R> : Retour au menu principal.

Choisissez une option :1

Déroulement du programme VARECH (suite).

ECRAN 4

FICHER DE DONNEES

Entrez le nom du fichier, sans extension ni chemin d'accès.
(8 caractères maximum).

il doit être dans le repertoire dbaseiii du disque dur.

nom : QMANEST1

VARIABLE A ESTIMER

Entrez le nom de la variable. (12 caractères maximum) : EFFECTRES

VARIABLE EXOGENE AU DENOMINATEUR DU RATIO

Entrez le nom de la variable. (12 caractères maximum) : NBMENAGE

VARIABLE EXOGENE DEFINISSANT LA PROBABILITE

Entrez le nom de la variable. (12 caractères maximum) : SURFACE

PLAN EQUIPROBABLE, ESTIMATION BIAISEE :

taux 5

Estimation stratifiée

fichier :

ECRAN 5

QMANEST1.DBF

Estimation du quotient : EFFECTRES/NBMENAGE

STRATE N° : 1

Calcul en cours, ... patientez.

Listage du programme VARECH

. Programme principal

```
*****
*
* VARECH.PRG
*
* Auteur : O.BARBARY
* Version 5,
*
* Programme de calcul, d'après données exhaustives sur une base de sondage, des
* variances réelles d'estimations stratifiées ou non, sous divers plan de sondages,
* à divers taux
* 4 options :
*
*
* 1. Plan équiprobable, estimation sans biais.
* 2. Plan équiprobable, estimation d'un/par le ratio.
* 3. Plan à probabilités inégales, estimation sans biais.
* 4. Plan à probabilités inégales, estimation d'un/par le ratio.
*
* Programme principal
*
*****
```

```
SET DECIMALS TO 0
SET TALK OFF
SET SCOREBOARD OFF
SET SAFETY OFF
SET BELL OFF
SET DATE FRENCH
SET COLOR TO W/B,B/W,G
CLEAR
TEXT
```

VARIANCE ECHANTILLONNAGE

Diagram showing a grid of boxes representing data points or variables, arranged in a structured pattern.

Version 5

O.BARBARY Aout 87

Header line of boxes

ENDTEXT
WAIT

* Ouverture du fichier de procedures

SET PROCEDURE TO VARPROC5.PRG

* Creation des variables communes

TAUX=0
BIAIS1=0
BIAIST=0
BIAIS2T=0
NT=0
VART=0
VARTOTT=0
ECRTT=0
ECRTOTT=0
COEFT=0
COEFTOTT=0
NBSTRAT=0
NOMFICH=""
CPT =1
RECNCB = 1

* Entrée du nom de fichier résultat

CLEAR
IF CPT >1
ACCEPT " Voulez vous changer de fichier résultat ? (O/N)" TO REP
ELSE
REP = "O"
ENDIF
IF REP = "O"
CLEAR
@ 2,5 SAY "FICHER DES RESULTATS"
@ 3,5 SAY "*****"

```
@ 5,5 SAY "Faut il creer un nouveau fichier de résultats (1) ou existe t'il déjà  
(2)"  
ACCEPT " Réponse :" TO REP2  
@ 8,5 SAY " "  
DO QUESFICH  
NOMFICR = NOMFICH  
IF REP2 = "1"  
    USE VARECHRE  
    COPY STRUCTURE TO &NOMFICR  
    CLOSE DATABASES VARECHRE  
ENDIF  
ENDIF  
* Affichage menu principal  
WAIT ' Appuyez sur une touche pour afficher le menu.'  
ECRMENU = .T.  
DO WHILE ECRMENU  
    CLEAR  
    MENU = "  
    DATE = CDOW( DATE() ) + ' ' + STR( DAY( DATE() ), 2, 0 ) + ' '  
    + CMONTH( DATE() ) + ' ' + STR( YEAR( DATE() ), 4, 0 )  
    @ 1,3 SAY 'Varech, version 5.'  
    @ 3,3 SAY DATE  
    @ 5,15 SAY '          MENU PRINCIPAL'  
    @ 6,15 SAY '          *****'  
    @ 8,15 SAY '<1> : Plan équiprobable, estimateur sans biais.'  
    @ 10,15 SAY '<2> : Plan équiprobable, estimateur d"un/par le ratio.'  
    @ 12,15 SAY '<3> : Plan à probabilités inégales, estimateur sans biais.'  
    @ 14,15 SAY '<4> : Plan à probabilités inégales, estimateur d"un/par le ratio.'  
    @ 16,15 SAY '<Q> : Quitter le programme'  
    @ 17,1 SAY ''  
    BOUCLE1 = .T.  
    DO WHILE BOUCLE1  
        ACCEPT ' Choisissez une option :' TO MENU  
        MENU = UPPER( MENU )  
        IF MENU = '1' .OR. MENU = '2' .OR. MENU = '3' .OR. MENU = ;  
        '4' .OR. MENU = 'Q'  
            BOUCLE1 = .F.  
        ENDIF  
    ENDDO  
    CLEAR  
*  
    DO CASE  
*  
        CASE MENU = '1'  
            DO VECH1  
        CASE MENU = '2'  
            DO VECH2  
        CASE MENU = '3'  
            DO VECH3  
        CASE MENU = '4'  
            DO VECH4  
        CASE MENU = 'Q'  
            SET COLOR TO  
            CLEAR  
            CLOSE PROCEDURE  
            RETURN  
    ENDCASE
```

CPT = CPT +1
ENDDO

. Procédures

```

*****
*
* VARPROC5.PRG: FICHIER DE PROCEDURES POUR LE PROGRAMME VARECH
* (Version n°5)
*
*****
*
* 1: PROCEDURE VECH1
* Procédure de calcul de la variance de l'estimateur sans biais sous un plan de
* sondage équiprobable avec remise.
*****
*
  PROCEDURE VECH1
  CLEAR
* Initialisation des variables en commun
  CHSTRAT=0
  CHEST=0
  N=0
  SY=0
  EY=0
  VY=0
  SNSTR=0
  SVYSTR=0
  SSYSTR=0
  VAR=0
  NOMFICH = " "
  REPVAR = " "
  VARY = " "
  NOMVARSTR = " "
  NBSTRAT = 0
  NT=0
  NB=0
  NOSTR =0
  OPT = "N"
* Choix du plan de sondage, stratifié ou non, et du type d'estimation, total ou moyenne.
  BOUCLEM = .T.
  INDBOUCLE =0
  DO WHILE BOUCLEM
    DO MENUSEC
    CLEAR
* Sortie de vech1
    IF CHSTRAT = 2
      RETURN
    ENDIF
    REPONSE = "0"
    IF INDBOUCLE >0
      ACCEPT "Voulez vous changer de fichier ou de variable (0/N)" TO REPONSE
    ENDIF
    INDBOUCLE = 1

```

```
IF REPONSE = "0"
  CLEAR
  * Entrée du nom de fichier
    @ 2,5 SAY "FICHIER DE DONNEES"
    @ 3,5 SAY "*****"
    DO QUESFICH
  * Entrée du nom de variable
    @ 15,5 SAY "VARIABLE A ESTIMER"
    @ 16,5 SAY "*****"
    DO QUESVAR
    VARY = REPVAR
  ENDIF
  USE &NOMFICH
  * Entrée du taux de sondage
    CLEAR
    @ 2,5 SAY "TAUX DE SONDAGE"
    @ 3,5 SAY "*****"
    DO QUESTAUX
    DO CASE
  *
  * 1: Cas non stratifié
  *
    CASE CHSTRAT = 0
    CLEAR
    @ 2,5 SAY "PLAN EQUIPROBABLE, ESTIMATION SANS BIAIS : "
    @ 3,5 SAY "taux : "
    @ 3,9 SAY TAUX
    @ 4,5 SAY "Sans stratification fichier : "
    @ 4,60 SAY NOMFICH
    DO MODUL1
  * Sauvegarde : ecriture dans le fichier résultats.
    DO ECRIT
  *
  * 2: Cas stratifié
  *
  * a/ Allocation proportionnelle
  *
    CASE CHSTRAT = 1
    CLEAR
  * Entrée du nombre de strates
    DO QUESNOMB
    NBSTRAT=NB
  * Entrée du nom de la variable de stratification
    @ 3,1 SAY "Le fichier de données doit contenir une variable indiquant la
    strate,"
    DO QUESVAR
    NOMVARSTR = REPVAR
    CLEAR
    @ 2,5 SAY "PLAN EQUIPROBABLE, ESTIMATION SANS BIAIS : "
    @ 3,5 SAY "taux : "
    @ 3,9 SAY TAUX
    @ 4,5 SAY "Estimation stratifiée fichier : "
    @ 4,60 SAY NOMFICH
  * Boucle sur les strates
    NOSTR = 1
    SNSTR = 0
    SSYSTR = 0
    SVYSTR = 0
```

```
SSQRTVY = 0
DO WHILE NOSTR <= NBSTRAT
  @ 8,5 SAY "STRATE N° :"
  @ 8,17 SAY NOSTR
  DO MODUL1
  SNSTR = SNSTR + N
  SSYSTR = SSYSTR + SY
  SVYSTR = SVYSTR + N^2*VAR/NT
  IF N^2*VAR >10^-9
    SSQRTVY = SSQRTVY + SQRT(N^2*VAR)
  ENDIF
* Sauvegarde : ecriture dans le fichier résultats.
  DO ECRIT
  NOSTR = NOSTR + 1
  ENDDO
* Calculs sur l'ensemble de la base
  @ 8,5 SAY " ENSEMBLE DE LA BASE"
  N = SNSTR
  SY = SSYSTR
  VY = SVYSTR
  EY = SY/N
* Effectifs d'échantillons.
  NT = N*TAUX/100
* Variances
  IF CHEST = 1
    VART = VY
  ELSE
    VART = VY/N^2
  ENDIF
  IF VART < 10^-9
    VART=0
  ENDIF
* Ecart-types et coefficients de variation.
  ECRTT = SQRT(VART)
  IF CHEST = 1
    IF SY=0
      COEFT=0
    ELSE
      COEFT = 100*ECRTT/SY
    ENDIF
  ELSE
    IF EY=0
      COEFT=0
    ELSE
      COEFT = 100*ECRTT/EY
    ENDIF
  ENDIF
  @ 20,5 SAY 'Calculs terminés.'
* Sauvegarde : ecriture dans le fichier résultats.
  DO ECRIT
*
* b/ Allocation optimale de NEYMAN
*
  CLEAR
  DO QUESNEYM
  IF OPT = "0"
    SVYSTROP = 0
    SNST = 0
```

```
BOUCLEOP = 1
* Boucle sur les strates
DO WHILE BOUCLEOP <= NBSTRAT
  USE &NOMFICR
  NENREG = RECNB
  NENREG = NENREG - NBSTRAT
  GOTO NENREG
  N = RCARD
  SNST = SNST + N
  IF RVARIANCE*REFFECH >10^-9
    IF CHEST=1
      EXP = SQRT(RVARIANCE*REFFECH)
    ELSE
      EXP = RCARD*SQRT(RVARIANCE*REFFECH)
    ENDIF
  ELSE
    EXP=0
  ENDIF
  NTOP = NT * EXP/SSQRTVY
  IF NTOP=0
    VAROP =0
  ELSE
    VAROP = REFFECH*RVARIANCE/NTOP
  ENDIF
  ECRTOP = SQRT(VAROP)
  IF RTOTALY =0
    COEFOP=0
  ELSE
    IF CHEST = 1
      COEFOP = ECRTOP*100/RTOTALY
    ELSE
      COEFOP = ECRTOP*100*RCARD/RTOTALY
    ENDIF
  ENDIF
  IF CHEST = 1
    SVYSTROP = SVYSTROP + VAROP
  ELSE
    SVYSTROP = SVYSTROP + N^2*VAROP
  ENDIF
  CLOSE DATABASES &NOMFICR
  DO ECRIT
  BOUCLEOP = BOUCLEOP + 1
ENDDO
* Résultats sur l'ensemble de la base
N = SNST
NTOP = NT
IF CHEST = 1
  VAROP = SVYSTROP
ELSE
  VAROP = SVYSTROP/N^2
ENDIF
IF VAROP <10^-9
  VAROP =0
ENDIF
ECRTOP = SQRT(VAROP)
IF CHEST = 1
  IF SY=0
    COEFOP=0
```



```

ELSE
  COEFOP = ECRTOP*100/SY
ENDIF
ELSE
  IF EY=0
    COEFOP =0
  ELSE
    COEFOP = ECRTOP*100/EY
  ENDIF
ENDIF
DO ECRIT
ENDIF
OPT = "N"
ENDCASE
ENDDO
RETURN

```

```

*
*
*****
*
*
* 2: PROCEDURE VECH2
* Procédure de calcul de la variance de l'estimateur d'un/par le ratio, sous un plan
* de sondage équiprobable avec remise.
*
*****
*
*

```

```

PROCEDURE VECH2
CLEAR
* Initialisation des variables en commun
CHSTRAT=0
CHEST=0
N=0
NT=0
R=0
SY=0
SX=0
EY=0
EX=0
VY=0
VX=0
VXY=0
BIAIS=0
SNSTR=0
SVYSTR=0
SSYSTR=0
SBIAISTR=0
VAR=0
NOMFICH = " "
REPVAR = " "
VARY = " "
NOMVARSTR = " "
NBSTRAT = 0
NB=0
NOSTR =0
OPT = "N"

```

```
* Choix du plan de sondage, stratifié ou non, et du type d'estimation, total, moyenne ou
quotient.
BOUCLEM = .T.
INDBOUCLE =0
DO WHILE BOUCLEM
DO MENUSEC
CLEAR
* Sortie de vech2
IF CHSTRAT = 2
RETURN
ENDIF
REPONSE = "0"
IF INDBOUCLE >0
ACCEPT "Voulez vous changer de fichier ou de variables (O/N)" TO REPONSE
ENDIF
INDBOUCLE = 1
IF REPONSE = "0"
CLEAR
* Entrée du nom de fichier
@ 2,5 SAY "FICHER DE DONNEES"
@ 3,5 SAY "*****"
DO QUESFICH
* Entrée des noms de variables
@ 12,5 SAY "VARIABLE AU NUMERATEUR"
@ 13,5 SAY "*****"
DO QUESVAR
VARY = REPVAR
@ 18,5 SAY "VARIABLE AU DENOMINATEUR"
@ 19,5 SAY "*****"
DO QUESVAR
VARDEN = REPVAR
ENDIF
USE &NOMFICH
* Entrée du taux de sondage
CLEAR
@ 2,5 SAY "TAUX DE SONDAGE"
@ 3,5 SAY "*****"
DO QUESTAUX
DO CASE
*
* 1: Cas non stratifié
*
CASE CHSTRAT = 0
CLEAR
@ 2,5 SAY "PLAN EQUIPROBABLE, ESTIMATION PAR LE RATIO :"
@ 3,5 SAY "taux :"
@ 3,9 SAY TAUX
@ 4,5 SAY "Sans stratification fichier :"
@ 4,60 SAY NOMFICH
DO MODUL2
* Sauvegarde : ecriture dans le fichier résultats.
DO ECRIT
```

```
*
* 2: Cas stratifié
*
*
* a/ Allocation proportionnelle
*
      CASE CHSTRAT = 1
      CLEAR
* Entrée du nombre de strates
      DO QUESNOMB
      NBSTRAT=NB
* Entrée du nom de la variable de stratification
      @ 3,1 SAY "Le fichier de données doit contenir une variable indiquant la
      strate,"
      DO QUESVAR
      NOMVARSTR = REPVAR
      CLEAR
      @ 2,5 SAY "PLAN EQUIPROBABLE, ESTIMATION PAR LE RATIO : "
      @ 3,5 SAY "taux : "
      @ 3,9 SAY TAUX
      @ 4,5 SAY "Estimation stratifiée fichier : "
      @ 4,60 SAY NOMFICH
* Boucle sur les strates
      NOSTR = 1
      SNSTR = 0
      SSYSTR = 0
      SSXSTR = 0
      SVYSTR = 0
      SVTOTSTR = 0
      SBIAISTR = 0
      SSQRTVY = 0
      DO WHILE NOSTR <= NBSTRAT
        @ 8,5 SAY "STRATE N' : "
        @ 8,17 SAY NOSTR
        DO MODUL2
        SNSTR = SNSTR + N
        SSYSTR = SSYSTR + SY
        SSXSTR = SSXSTR + SX
        SVYSTR = SVYSTR + N^2*VAR/NT
        IF SX * 0
          SBIAISTR = SBIAISTR + BIAIS*N/SX
        ENDIF
        IF VAR >10^-9
          SSQRTVY = SSQRTVY + N*SQRT(VAR)
        ENDIF
        IF CHEST = 1
          SVTOTSTR = SVTOTSTR + VARTOTT
        ENDIF
        IF CHEST = 2
          SVTOTSTR = SVTOTSTR + VARTOTT*N^2
        ENDIF
        IF CHEST = 3
          SVTOTSTR = SVTOTSTR + VARTOTT*SX^2
        ENDIF
* Sauvegarde : ecriture dans le fichier résultats.
      DO ECRIT
      NOSTR = NOSTR + 1
      ENDDO
```

```
* Calculs sur l'ensemble de la base
@ 8,5 SAY " ENSEMBLE DE LA BASE"
N = SNSTR
SY = SSYSTR
SX = SSXSTR
R = SY/SX
EY = SY/N

* Effectifs d'échantillons.
NT = N*TAUX/100

* Biais aux divers taux
BIAIST = N*SBIAISTR/NT
IF CHEST = 2
  BIAIST = BIAIST/N
ENDIF
IF CHEST = 3
  IF SX = 0
    BIAIST = 0
  ELSE
    BIAIST = BIAIST/SX
  ENDIF
ENDIF

* Variances
VART = SVYSTR
IF CHEST = 2
  VART = VART/N^2
ENDIF
IF CHEST = 3
  IF SX = 0
    VART = 0
  ELSE
    VART = VART/SX^2
  ENDIF
ENDIF

* Variances + carré du biais
VARTOTT = SVTOTSTR
IF CHEST = 2
  VARTOTT = SVTOTSTR/N^2
ENDIF
IF CHEST = 3
  IF SX = 0
    VARTOTT = 0
  ELSE
    VARTOTT = SVTOTSTR/SX^2
  ENDIF
ENDIF
IF VART < 10^-9
  VART = 0
ENDIF
IF VARTOTT < 10^-9
  VARTOTT = 0
ENDIF

* Ecart-types et coefficients de variation.
ECRTT = SQRT(VART)
ECRTOTT = SQRT(VARTOTT)
IF CHEST = 1
  IF SY = 0
    COEFT = 0
```

```
      COEFTOTT=0
    ELSE
      COEFT = 100*ECRTT/SY
      COEFTOTT = 100*ECRTOTT/SY
    ENDIF
  ENDIF
IF CHEST = 2
  IF EY=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/EY
    COEFTOTT = 100*ECRTOTT/EY
  ENDIF
ENDIF
IF CHEST = 3
  IF R=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/R
    COEFTOTT = 100*ECRTOTT/R
  ENDIF
ENDIF
```

@ 20,5 SAY 'Calculs terminés.'

* Sauvegarde : ecriture dans le fichier résultats.

DO ECRIT

*

* b/ Allocation optimale de NEYMAN

*

```
CLEAR
DO QUESNEYM
IF OPT = "0"
  SVYSTROP = 0
  SBIAISTROP = 0
  SVTOTOP = 0
  BOUCLEOP = 1
  SNST = 0
  SSYSTR = 0
  SSXSTR = 0
```

* Boucle sur les strates

```
DO WHILE BOUCLEOP <= NBSTRAT
  USE &NOMFICR
  NENREG = RECNCB
  NENREG = NENREG - NBSTRAT
  GOTO NENREG
  N = RCARD
  SNST = SNST + N
  SY = RTOTALY
  SX = RTOTALX
  IF RTOTALX = 0
    R = 0
  ELSE
    R = RTOTALY/RTOTALX
  ENDIF
  SSYSTR = SSYSTR + SY
  SSXSTR = SSXSTR + SX
```

```
IF CHEST=1
  EXP = RECRTYPE*SQRT(REFFECH)
ENDIF
IF CHEST = 2
  EXP = N*RECRTYPE*SQRT(REFFECH)
ENDIF
IF CHEST = 3
  EXP = SX*RECRTYPE*SQRT(REFFECH)
ENDIF
NTOP = NT * EXP/SSQRTVY
IF NTOP=0
  VAROP=0
  BIAISTOP=0
ELSE
  VAROP = REFFECH*RVARIANCE/NTOP
  IF CHEST = 3
    VAROP = REFFECH*(RECRTYPE^2)/NTOP
  ENDIF
  BIAISTOP = REFFECH*RBIAISTOT/NTOP
ENDIF
VARTOTOP = VAROP + BIAISTOP^2
ECRTOP = SQRT(VAROP)
ECRTOTOP = SQRT(VARTOTOP)
IF RTOTALY=0
  COEFOP=0
  COEFTOTOP=0
ELSE
  IF CHEST = 1
    COEFOP = ECRTOP*100/RTOTALY
    COEFTOTOP = ECRTOTOP*100/RTOTALY
  ELSE
    IF CHEST =2
      COEFOP = ECRTOP*100*N/RTOTALY
      COEFTOTOP = ECRTOTOP*100*N/RTOTALY
    ELSE
      IF R = 0
        COEFOP = 0
        COEFTOTOP = 0
      ELSE
        COEFOP = ECRTOP*100/R
        COEFTOTOP = ECRTOTOP*100/R
      ENDIF
    ENDIF
  ENDIF
ENDIF
IF CHEST = 1
  SVYSTROP = SVYSTROP + VAROP
  SBIAISTROP = SBIAISTROP + BIAISTOP
  SVTOTOP = SVTOTOP + VARTOTOP
ENDIF
IF CHEST =2
  SVYSTROP = SVYSTROP + N^2*VAROP
  SBIAISTROP = SBIAISTROP + N*BIAISTOP
  SVTOTOP = SVTOTOP + N^2*VARTOTOP
ENDIF
IF CHEST =3
  SVYSTROP = SVYSTROP + SX^2*VAROP
  SBIAISTROP = SBIAISTROP + SX*BIAISTOP
```

```
SVTOTOP = SVTOTOP + SX^2 * VARTOTOP
ENDIF
CLOSE DATABASES &NOMFICR
DO ECRIT
BOUCLEOP = BOUCLEOP + 1
ENDDO
```

* Résultats sur l'ensemble de la base

```
N = SNST
SX = SSXSTR
SY = SSYSTR
IF SX = 0
  R = 0
ELSE
  R = SY/SX
ENDIF
NTOP = NT
IF CHEST = 1
  VAROP = SVYSTROP
  BIAISTOP = SBIAISTROP
  VARTOTOP = SVTOTOP
ENDIF
IF CHEST = 2
  VAROP = SVYSTROP/N^2
  BIAISTOP = SBIAISTROP/N
  VARTOTOP = SVTOTOP/N^2
ENDIF
IF CHEST = 3
  VAROP = SVYSTROP/SX^2
  BIAISTOP = SBIAISTROP/SX
  VARTOTOP = SVTOTOP/SX^2
ENDIF
IF VAROP < 10^-9
  VAROP = 0
ENDIF
IF VARTOTOP < 10^-9
  VARTOTOP = 0
ENDIF
ECRTOP = SQRT(VAROP)
ECRTOTOP = SQRT(VARTOTOP)
IF CHEST = 1
  IF SY = 0
    COEFOP = 0
    COEFTOTOP = 0
  ELSE
    COEFOP = ECRTOP * 100 / SY
    COEFTOTOP = ECRTOTOP * 100 / SY
  ENDIF
ENDIF
IF CHEST = 2
  IF EY = 0
    COEFOP = 0
    COEFTOTOP = 0
  ELSE
    COEFOP = ECRTOP * 100 / EY
    COEFTOTOP = ECRTOTOP * 100 / EY
  ENDIF
ENDIF
ENDIF
```

```
IF CHEST =3
  IF R=0
    COEFOP=0
    COEFTOTOP=0
  ELSE
    COEFOP = ECRTOP*100/R
    COEFTOTOP = ECRTOTOP*100/R
  ENDIF
ENDIF
DO ECRIT
ENDIF
OPT = "N"
ENDCASE
ENDDO
RETURN
```

```
*
*
*****
*
* 3: PROCEDURE VECH3
* Procédure de calcul de la variance de l'estimateur sans biais sous un plan de
* sondage à probabilités proportionnelles à la variable X (ex: surface des districts).
* Plan avec remise
*
*****
*
```

```
PROCEDURE VECH3
CLEAR
* Initialisation des variables en commun
CHSTRAT=0
CHEST=0
N=0
SY=0
SX=0
EY=0
VY=0
SNSTR=0
SVYSTR=0
SSYSTR=0
VAR=0
NOMFICH = " "
REPVAR = " "
VARY = " "
NOMVARSTR = " "
NBSTRAT = 0
NB=0
NOSTR =0
OPT = "N"
* Choix du plan de sondage, stratifié ou non, et du type d'estimation, total ou moyenne.
BOUCLEM = T.
INDBOUCLE =0
DO WHILE BOUCLEM
  DO MENUSEC
  CLEAR
* Sortie de vech1
  IF CHSTRAT =2
    RETURN
  ENDIF
```



```
REPONSE = "0"
IF INDBOUCLE >0
  ACCEPT "Voulez vous changer de fichier ou de variables (O/N)" TO REPONSE
ENDIF
INDBOUCLE = 1
IF REPONSE = "0"
  CLEAR
* Entrée du nom de fichier
  @ 2,5 SAY "FICHER DE DONNEES"
  @ 3,5 SAY "*****"
  DO QUESFICH
* Entrée des noms de variables
  @ 12,5 SAY "VARIABLE A ESTIMER"
  @ 13,5 SAY "*****"
  DO QUESVAR
  VARY = REPVAR
  @ 18,5 SAY "VARIABLE EXOGENE DEFINISSANT LA PROBABILITE"
  @ 19,5 SAY "*****"
  DO QUESVAR
  VARZ = REPVAR
ENDIF
USE &NOMFICH
* Entrée du taux de sondage
  CLEAR
  @ 2,5 SAY "TAUX DE SONDAGE"
  @ 3,5 SAY "*****"
  DO QUESTAUX
  DO CASE
*
* 1: Cas non stratifié
*
  CASE CHSTRAT = 0
  CLEAR
  @ 2,5 SAY "PLAN A PROBA. INEGALES, ESTIMATION SANS BIAIS :"
  @ 3,5 SAY "taux :"
  @ 3,9 SAY TAUX
  @ 4,5 SAY "Sans stratification fichier :"
  @ 4,60 SAY NOMFICH
  DO MODUL3
* Sauvegarde : ecriture dans le fichier résultats.
  DO ECRIT
*
* 2: Cas stratifié
*
*
* a/ Allocation proportionnelle
*
  CASE CHSTRAT = 1
  CLEAR
* Entrée du nombre de strates
  DO QUESNOMB
  NBSTRAT=NB
* Entrée du nom de la variable de stratification
  @ 3,1 SAY "Le fichier de données doit contenir une variable indiquant la
strate,"
  DO QUESVAR
  NOMVARSTR = REPVAR
  CLEAR
```

```
@ 2,5 SAY "PLAN A PROBA. INEGALES, ESTIMATION SANS BIAIS :"  
@ 3,5 SAY "taux :"  
@ 3,9 SAY TAUX  
@ 4,5 SAY "Estimation stratifiée fichier :"  
@ 4,60 SAY NOMFICH  
* Boucle sur les strates  
  NOSTR = 1  
  SNSTR = 0  
  SSYSTR = 0  
  SVYSTR = 0  
  SSQRTVY = 0  
  DO WHILE NOSTR <= NBSTRAT  
    @ 8,5 SAY "STRATE N' :"  
    @ 8,17 SAY NOSTR  
    DO MODUL3  
      SNSTR = SNSTR + N  
      SSYSTR = SSYSTR + SY  
      SVYSTR = SVYSTR + VAR/NT  
      IF VAR >10^-9  
        SSQRTVY = SSQRTVY + SQRT(VAR)  
      ENDIF  
    ENDDO  
  * Sauvegarde : ecriture dans le fichier résultats.  
    DO ECRIT  
      NOSTR = NOSTR + 1  
    ENDDO  
  * Calculs sur l'ensemble de la base  
    @ 8,5 SAY " ENSEMBLE DE LA BASE "  
    N = SNSTR  
    SY = SSYSTR  
    EY = SY/N  
  * Effectifs d'échantillons.  
    NT = N*TAUX/100  
  * Variances  
    IF CHEST = 1  
      VART = SVYSTR  
    ELSE  
      VART = SVYSTR/N^2  
    ENDIF  
    IF VART <10^-9  
      VART=0  
    ENDIF  
  * Ecart-types et coefficients de variation.  
    ECRTT = SQRT(VART)  
    IF CHEST = 1  
      IF SY=0  
        COEFT=0  
      ELSE  
        COEFT = 100*ECRTT/SY  
      ENDIF  
    ELSE  
      IF EY=0  
        COEFT =0  
      ELSE  
        COEFT = 100*ECRTT/EY  
      ENDIF  
    ENDIF  
  @ 20,5 SAY 'Calculs terminés.'
```

* Sauvegarde : ecriture dans le fichier résultats.
DO ECRIT

*
* b/ Allocation optimale de NEYMAN
*

```
CLEAR
DO QUESNEYM
IF OPT = "0"
  SVYSTROP = 0
  BOUCLEOP = 1
  SNST = 0
* Boucle sur les strates
  DO WHILE BOUCLEOP <= NBSTRAT
    USE &NOMFICR
    NENREG = RECNB
    NENREG = NENREG - NBSTRAT
    GOTO NENREG
    N = RCARD
    SNST = SNST + N
    IF RVARIANCE*REFFECH >10^-9
      IF CHEST=1
        EXP = SQRT(RVARIANCE*REFFECH)
      ELSE
        EXP = RCARD*SQRT(RVARIANCE*REFFECH)
      ENDIF
    ELSE
      EXP=0
    ENDIF
    NTOP = NT * EXP/SSQRTVY
    IF NTOP=0
      VAROP=0
    ELSE
      VAROP = REFFECH*RVARIANCE/NTOP
    ENDIF
    ECRTOP = SQRT(VAROP)
    IF RTOTALY=0
      COEFOP=0
    ELSE
      IF CHEST = 1
        COEFOP = ECRTOP*100/RTOTALY
      ELSE
        COEFOP = ECRTOP*100*RCARD/RTOTALY
      ENDIF
    ENDIF
    IF CHEST = 1
      SVYSTROP = SVYSTROP + VAROP
    ELSE
      SVYSTROP = SVYSTROP + VAROP*N^2
    ENDIF
    CLOSE DATABASES &NOMFICR
    DO ECRIT
    BOUCLEOP = BOUCLEOP + 1
  ENDDO
* Résultats sur l'ensemble de la base
  N = SNST
  NTOP = NT
  IF CHEST = 1
    VAROP = SVYSTROP
```

```

ELSE
  VAROP = SVYSTROP/N^2
ENDIF
IF VAROP < 10^-9
  VAROP=0
ENDIF
ECRTOP = SQRT(VAROP)
IF CHEST = 1
  IF SY=0
    COEFOP =0
  ELSE
    COEFOP = ECRTOP*100/SY
  ENDIF
ELSE
  IF EY=0
    COEFOP=0
  ELSE
    COEFOP = ECRTOP*100/EY
  ENDIF
ENDIF
DO ECRIT
ENDIF
OPT = "N"
ENDCASE
ENDDO
RETURN

```

```

*
*
*****
*
* 4: PROCEDURE VECH4
* Procedure de calcul de la variance de l'estimateur biaisé :
*
*  $T_4(Y) = \bar{X} \cdot \sum (y_i/z_i) / (\sum x_i/z_i)$ , sous un plan de sondage à probabilités proportion-
* nelles à la variable Z (ex: surface des districts) Plan avec remise
*
*****
*

```

```

PROCEDURE VECH4
CLEAR
* Initialisation des variables en commun
CHSTRAT=0
CHEST=0
N=0
R=0
SY=0
SX=0
SZ=0
BIAIS=0
BIAISS=0
EPY=0
EPX=0
VPY=0
VPX=0
VPXY=0
SNSR=0
SVYSTR=0

```

```
SSYSTR=0
SSXSTR=0
SSZSTR=0
SBIAISTR=0
SBIAI2STR=0
VAR=0
NOMFICH = " "
REPVAR = " "
VARY = " "
NOMVARSTR = " "
NBSTRAT = 0
NB=0
NOSTR =0
OPT = "N"
* Choix du plan de sondage, stratifié ou non, et du type d'estimation, total ou moyenne.
BOUCLEM = T.
INDBOUCLE =0
DO WHILE BOUCLEM
  DO MENUSEC
  CLEAR
* Sortie de vech1
  IF CHSTRAT = 2
    RETURN
  ENDIF
  REponse = "0"
  IF INDBOUCLE >0
    ACCEPT "Voulez vous changer de fichier ou de variables (0/N)" TO REponse
  ENDIF
  INDBOUCLE = 1
  IF REponse = "0"
    CLEAR
* Entrée du nom de fichier
    @ 2,5 SAY "FICHER DE DONNEES"
    @ 3,5 SAY "*****"
    DO QUESFICH
* Entrée des noms de variables
    @ 9,5 SAY "VARIABLE A ESTIMER"
    @ 10,5 SAY "*****"
    DO QUESVAR
    VARY = REPVAR
    @ 13,5 SAY "VARIABLE EXOGENE AU DENOMINATEUR DU RATIO"
    @ 14,5 SAY "*****"
    DO QUESVAR
    VARX = REPVAR
    @ 18,5 SAY "VARIABLE EXOGENE DEFINISSANT LA PROBABILITE"
    @ 19,5 SAY "*****"
    DO QUESVAR
    VARZ = REPVAR
  ENDIF
  USE &NOMFICH
* Entrée du taux de sondage
  CLEAR
  @ 2,5 SAY "TAUX DE SONDAGE"
  @ 3,5 SAY "*****"
  DO QUESTAUX
  DO CASE
```

*

* 1: Cas non stratifié

*

```
CASE CHSTRAT = 0
CLEAR
@ 2,5 SAY "PLAN A PROBA. INEGALES, ESTIMATION BIAISEE : "
@ 3,5 SAY "taux : "
@ 3,9 SAY TAUX
@ 4,5 SAY "Sans stratification fichier : "
@ 4,60 SAY NOMFICH
DO MODUL4
```

* Sauvegarde : ecriture dans le fichier résultats.

```
DO ECRIT
```

*

* 2: Cas stratifié

*

*

* a/ Allocation proportionnelle

*

```
CASE CHSTRAT = 1
CLEAR
```

* Entrée du nombre de strates

```
DO QUESNOMB
NBSTRAT=NB
```

* Entrée du nom de la variable de stratification

```
@ 3,1 SAY "Le fichier de données doit contenir une variable indiquant la
strate,"
```

```
DO QUESVAR
NOMVARSTR = REPVAR
CLEAR
```

```
@ 2,5 SAY "PLAN EQUIPROBABLE, ESTIMATION BIAISEE : "
```

```
@ 3,5 SAY "taux : "
```

```
@ 3,9 SAY TAUX
```

```
@ 4,5 SAY "Estimation stratifiée fichier : "
```

```
@ 4,60 SAY NOMFICH
```

* Boucle sur les strates

```
NOSTR = 1
SNSTR = 0
SSXSTR = 0
SSYSTR = 0
SVYSTR = 0
SBIAISTR = 0
SBIAI2STR = 0
SSQRTVY = 0
DO WHILE NOSTR <= NBSTRAT
```

```
@ 8,5 SAY "STRATE N' : "
```

```
@ 8,17 SAY NOSTR
```

```
DO MODUL4
```

```
SNSTR = SNSTR + N
```

```
SSXSTR = SSXSTR + SX
```

```
SSYSTR = SSYSTR + SY
```

```
SVYSTR = SVYSTR + VAR
```

```
SBIAISTR = SBIAISTR + BIAISS
```

```
SBIAI2STR = SBIAI2STR + BIAISS^2
```

```
IF NT*VAR > 10^-9
```

```
SSQRTVY = SSQRTVY + SQRT(NT*VAR)
```

```
ENDIF
```

* Sauvegarde : ecriture dans le fichier résultats.

```
DO ECRIT
NOSTR = NOSTR + 1
ENDDO
```

* Calculs sur l'ensemble de la base

```
@ 8,5 SAY " ENSEMBLE DE LA BASE"
N = SNSTR
SY = SSYSTR
SX = SXXSTR
EY = SY/N
EX = SX/N
IF SX = 0
  R = 0
ELSE
  R = SY/SX
ENDIF
```

* Effectifs d'échantillons.

```
NT = N*TAUX/100
```

* Biais

```
BIAIST = SBIAISTR
IF CHEST = 2
  BIAIST = BIAIST/N
ENDIF
IF CHEST = 3
  IF SX = 0
    BIAIST = 0
  ELSE
    BIAIST = BIAIST/SX
  ENDIF
ENDIF
```

* Variances

```
VART = SVYSTR
IF CHEST = 2
  VART = VART/N^2
ENDIF
IF CHEST = 3
  IF SX = 0
    VART = 0
  ELSE
    VART = VART/SX^2
  ENDIF
ENDIF
```

* Variances + carré du biais

```
VARTOTT = SVYSTR + SBIAI2STR
IF CHEST = 2
  VARTOT = VARTOT/N^2
ENDIF
IF CHEST = 3
  IF SX = 0
    VARTOTT = 0
  ELSE
    VARTOTT = VARTOTT/SX^2
  ENDIF
ENDIF
```

* Ecart-types et coefficients de variation.

```
ECRTT = SQRT(VART)
ECRTOTT = SQRT(VARTOTT)
```

```
IF CHEST = 1
  IF SY=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/SY
    COEFTOTT = 100*ECRTOTT/SY
  ENDIF
ENDIF
IF CHEST = 2
  IF EY=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/EY
    COEFTOTT = 100*ECRTOTT/EY
  ENDIF
ENDIF
IF CHEST = 3
  IF R=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/R
    COEFTOTT = 100*ECRTOTT/R
  ENDIF
ENDIF
```

@ 20,5 SAY 'Calculs terminés.'

* Sauvegarde : ecriture dans le fichier résultats.

DO ECRIT

*

* b/ Allocation optimale de NEYMAN

*

```
CLEAR
DO QUESNEYM
IF OPT = "0"
  SSYSTR = 0
  SSXSTR = 0
  SVYSTROP = 0
  SBIAISTROP = 0
  SBIAIS2OP = 0
  SVTOTOP = 0
  BOUCLEOP = 1
  SNST = 0
```

* Boucle sur les strates

```
DO WHILE BOUCLEOP <= NBSTRAT
  USE &NOMFICR
  NENREG = RECNB
  NENREG = NENREG - NBSTRAT
  GOTO NENREG
  N = RCARD
  SX = RTOTALX
  SY = RTOTALY
  IF SX = 0
    R = 0
  ELSE
    R = SY/SX
  ENDIF
```



```
SNST = SNST + N
SSYSTR = SSYSTR + SY
SSXSTR = SSXSTR + SX
EXP = SQRT(REFFECH)
IF CHEST = 1
  NTOP = NT*EXP*RECRTYPE/SSQRTVY
ENDIF
IF CHEST = 2
  NTOP = RCARD*NT*EXP*RECRTYPE/SSQRTVY
ENDIF
IF CHEST = 3
  NTOP = SX*NT*EXP*RECRTYPE/SSQRTVY
ENDIF
IF NTOP=0
  VAROP=0
  BIAISTOP=0
ELSE
  VAROP = REFFECH*RVARIANCE/NTOP
  IF CHEST = 3
    VAROP = REFFECH*(RECRTYPE^2)/NTOP
  ENDIF
  BIAISTOP = REFFECH*RBIAISTOT/NTOP
ENDIF
VARTOTOP = VAROP + BIAISTOP^2
IF VAROP <10^-9
  VAROP=0
ENDIF
IF VARTOTOP <10^-9
  VARTOTOP=0
ENDIF
ECRTOP = SQRT(VAROP)
ECRTOTOP = SQRT(VARTOTOP)
IF RTOTALY=0
  COEFOP=0
  COEFTOTOP=0
ELSE
  IF CHEST = 1
    COEFOP = ECRTOP*100/RTOTALY
    COEFTOTOP = ECRTOTOP*100/RTOTALY
  ENDIF
  IF CHEST = 2
    COEFOP = ECRTOP*100*RCARD/RTOTALY
    COEFTOTOP = ECRTOTOP*100*RCARD/RTOTALY
  ENDIF
  IF CHEST = 3
    COEFOP = ECRTOP*100*RTOTALX/RTOTALY
    COEFTOTOP = ECRTOTOP*100*RTOTALX/RTOTALY
  ENDIF
ENDIF
IF CHEST =1
  SVYSTROP = SVYSTROP + VAROP
  SBIAIS2OP = SBIAIS2OP + BIAISTOP^2
  SBIAISTROP = SBIAISTROP + BIAISTOP
ENDIF
IF CHEST = 2
  SVYSTROP = SVYSTROP + VAROP*RCARD^2
  SBIAIS2OP = SBIAIS2OP + BIAISTOP^2*RCARD^2
  SBIAISTROP = SBIAISTROP + BIAISTOP*RCARD
```

```
ENDIF
IF CHEST = 3
  SVYSTROP = SVYSTROP + VAROP*RTOTALX^2
  SBIAIS2OP = SBIAIS2OP + BIAISTOP^2*RTOTALX^2
  SBIAISTROP = SBIAISTROP + BIAISTOP*RTOTALX
ENDIF
CLOSE DATABASES &NOMFICR
DO ECRIT
BOUCLEOP = BOUCLEOP + 1
ENDDO
* Résultats sur l'ensemble de la base
N = SNST
NTOP = NT
SY = SSYSTR
SX = SSXSTR
IF SX = 0
  R = 0
ELSE
  R = SY/SX
ENDIF
IF CHEST = 1
  VAROP = SVYSTROP
  BIAISTOP = SBIAISTROP
  VARTOTOP = SVYSTROP + SBIAIS2OP
ENDIF
IF CHEST = 2
  VAROP = SVYSTROP/N^2
  BIAISTOP = SBIAISTROP/N
  VARTOTOP = (SVYSTROP + SBIAIS2OP)/N^2
ENDIF
IF CHEST = 3
  VAROP = SVYSTROP/SX^2
  BIAISTOP = SBIAISTROP/SX
  VARTOTOP = (SVYSTROP + SBIAIS2OP)/SX^2
ENDIF
IF VAROP < 10^-9
  VAROP=0
ENDIF
IF VARTOTOP < 10^-9
  VARTOTOP=0
ENDIF
ECRTOP = SQRT(VAROP)
ECRTOTOP = SQRT(VARTOTOP)
IF CHEST = 1
  IF SY=0
    COEFOP=0
    COEFTOTOP=0
  ELSE
    COEFOP = ECRTOP*100/SY
    COEFTOTOP = ECRTOTOP*100/SY
  ENDIF
ENDIF
IF CHEST = 2
  IF EY=0
    COEFOP=0
    COEFTOTOP=0
```

```

ELSE
  COEFOP = ECRTOP*100/EY
  COEFTOTOP = ECRTOTOP*100/EY
ENDIF
ENDIF
IF CHEST = 3
  IF R=0
    COEFOP=0
    COEFTOTOP=0
  ELSE
    COEFOP = ECRTOP*100/R
    COEFTOTOP = ECRTOTOP*100/R
  ENDIF
ENDIF
DO ECRIT
ENDIF
OPT = "N"
ENDCASE
ENDDO
RETURN

```

```

*
*
*****
*
* 5: PROCEDURE CALVECH1
* Calcul de l'effectif de la base et de la variance du caractère Y sur la base ou dans
* les strates (probabilité uniforme). La procedure doit être appelée avec deux noms
* de variables.
*
*****
*

```

```

PROCEDURE CALVECH1
PARAMETERS Y, ST, NOSTR, CHSTRAT
IF CHSTRAT = 1
  SET FILTER TO &ST = NOSTR
ELSE
  SET FILTER
ENDIF
COUNT TO N
SUM &Y, &Y^2 TO SY, SY2
EY = SY/N
VY = SY2/N - EY^2
RETURN

```

```

*
*
*****
*
* 6: PROCEDURE CALVECH2
* Calcul de l'effectif de la base et des variances des
* caractères X et Y sur la base ou dans les strates (probabilité uniforme).
* La procedure doit être appelée avec trois noms de variables.
*
*****
*

```

```

PROCEDURE CALVECH2
PARAMETERS Y, X, ST, NOSTR, CHSTRAT
IF CHSTRAT = 1
  SET FILTER TO &ST = NOSTR

```

```

ELSE
  SET FILTER
ENDIF
COUNT TO N
SUM &Y, &X, &Y^2, &X^2, &X*&Y TO SY, SX, SY2, SX2, SXY
IF SX=0
  R=0
ELSE
  R = SY/SX
ENDIF
EY = SY/N
EX = SX/N
VY = SY2/N - EY^2
VX = SX2/N - EX^2
VXY = SXY/N - EX*EY
RETURN

```

```

*
*
*****
*
* 7: PROCEDURE CALVECH3
* Calcul de l'effectif de la base et des variances et covariances, sous le plan  $\pi$ , des
* caractères Z et Y sur la base ou dans les strates (probabilité inégales proportion-
* nelles a Z). La procedure doit être appelée avec trois noms de variables.
*
*****
*

```

```

PROCEDURE CALVECH3
PARAMETERS Y, Z, ST, NOSTR, CHSTRAT
IF CHSTRAT = 1
  SET FILTER TO &ST = NOSTR
ELSE
  SET FILTER
ENDIF
COUNT TO N
SUM &Y, &Z, &Y^2/&Z TO SY, SZ, VYZ
RETURN

```

```

*
*
*****
*
* 8: PROCEDURE CALVECH4
* Calcul de l'effectif de la base et des variances et covariances, sous le plan  $\pi$ , des
* caractères X et Y sur la base ou dans les strates (probabilité inégales proportion-
* nelles a Z). La procedure doit être appelée avec quatre noms de variables.
*
*****
*

```

```

PROCEDURE CALVECH4
PARAMETERS Y, X, Z, ST, NOSTR, CHSTRAT
IF CHSTRAT = 1
  SET FILTER TO &ST = NOSTR
ELSE
  SET FILTER
ENDIF
COUNT TO N
SUM &Y, &X, &Z, (&Y^2)/&Z, (&X^2)/&Z, (&X*&Y)/&Z;
TO SY, SX, SZ, SPY2, SPX2, SPXY

```

```

IF SX=0
  R=0
ELSE
  R = SY/SX
ENDIF
VPYZ = SPY2/SZ - (SY^2)/(SZ^2)
VPXZ = SPX2/SZ - (SX^2)/(SZ^2)
VPXYZ = SPXY/SZ - (SX*SY)/(SZ^2)
RETURN

```

```

*
*
*****
*
* 9 : PROCEDURE ECRIT
*
* Procedure d'ecriture des résultats du programme VARECH dans le fichier de
* résultats.
*
*****
*
  PROCEDURE ECRIT
* Ouverture du fichier
  USE &NOMFICR
  APPEND BLANK
  TITRE = ""
  IF CHSTRAT = 0
    CLEAR
* Ecran conversationnel
  @ 2,15 SAY 'ECRITURE DANS LE FICHER DE RESULTATS'
  @ 3,15 SAY '*****'
* Entrée du titre de session
  BOUCLE = .T.
  DO WHILE BOUCLE
    ACCEPT 'Voulez vous entrer ou modifier le titre de session (O/N)' TO REP
    IF REP = 'O' .OR. REP = 'N'
      BOUCLE = F.
    ENDIF
  ENDDO
  IF REP = 'O'
    BOUCLE = .T.
    DO WHILE BOUCLE
      ACCEPT 'Entrez votre titre (60 caractères maximum)' TO TITRE
      IF LEN(TITRE) <= 60
        BOUCLE = F.
      ENDIF
    ENDDO
  ENDIF
* Ecriture du titre, date, etc...
  REPLACE RTITRE WITH TITRE
  REPLACE RDATE WITH DATE
  REPLACE RFICHER WITH NOMFICH
  REPLACE RMENU WITH MENU
  REPLACE RCHSTRAT WITH CHSTRAT
  REPLACE RCHEST WITH CHEST
  REPLACE RVAREST WITH VARY
  REPLACE RCARD WITH N
  REPLACE ROPT WITH OPT

```

* Définition du champ : RPLANSOND.

```
IF MENU = '1'  
  CHAINE = 'Plan equiprobable, estimateur sans biais.'  
ENDIF  
IF MENU = '2'  
  CHAINE = 'Plan equiprobable, estimateur par le/ d"un ratio.'  
ENDIF  
IF MENU = '3'  
  CHAINE = 'Plan à probabilités inégales, estimateur sans biais.'  
ENDIF  
IF MENU = '4'  
  CHAINE = 'Plan à probabilités inégales, estimateur d"un ratio.'  
ENDIF  
REPLACE RPLANSOND WITH CHAINE
```

* Définition du champ : RESTIMATEUR.

```
IF CHEST = 1  
  CHAINE = 'Estimation du total de : ' + VARY  
ENDIF  
IF CHEST = 2  
  CHAINE = 'Estimation de la moyenne de : ' + VARY  
ENDIF  
IF CHEST = 3 .AND. MENU = '4'  
  CHAINE = 'Estimation du quotient : ' + VARY + '/' + VARX  
ENDIF  
IF CHEST = 3 .AND. MENU = '2'  
  CHAINE = 'Estimation du quotient : ' + VARY + '/' + VARDEN  
ENDIF  
REPLACE RESTIMAT WITH CHAINE
```

* Ecriture des variables et paramètres de la stratification

```
IF CHSTRAT = 1  
  IF OPT = "0"  
    NOSTR = BOUCLEOP  
  ENDIF  
  REPLACE RVARSTRAT WITH NOMVARSTR  
  REPLACE RNBSTRAT WITH NBSTRAT  
  REPLACE RNOSTRAT WITH NOSTR  
  IF OPT = "0"  
    REPLACE RALLOC WITH "Allocation optimale de NEYMAN"  
  ELSE  
    REPLACE RALLOC WITH "Allocation proportionnelle"  
  ENDIF  
ENDIF
```

* Ecriture des noms de variables exogènes et des résultats

* intermediaires.

```
IF MENU = '3' .OR. MENU = '4'  
  REPLACE RVARPROB WITH VARZ  
ENDIF  
IF MENU = '2'  
  REPLACE RVAREXO WITH VARDEN  
  REPLACE RRATIO WITH R  
  REPLACE RTOTALX WITH SX  
ENDIF  
IF MENU = '4'  
  REPLACE RVAREXO WITH VARX  
  REPLACE RRATIO WITH R  
  REPLACE RTOTALX WITH SX  
ENDIF
```

* Ecriture des résultats au taux de t%

```

IF OPT = '0'
  REPLACE RCARD WITH N
  REPLACE RTAUX WITH TAUX
  REPLACE REFFECH WITH NTOP
  REPLACE RTOTALY WITH SY
  REPLACE RVARIANCE WITH VAROP
  REPLACE RECRTYPE WITH ECRTOP
  REPLACE RCOEFVAR WITH COEFOP
  IF MENU = '2'
    REPLACE RBIAISTOT WITH BIAISTOP
    REPLACE RVARTOT WITH VARTOTOP
    REPLACE RECRTTOT WITH ECRTOTOP
    REPLACE RCOEFTOT WITH COEFTOTOP
  ENDIF
  IF MENU = '4'
    REPLACE RBIAISTOT WITH BIAISTOP
    REPLACE RVARTOT WITH VARTOTOP
    REPLACE RECRTTOT WITH ECRTOTOP
    REPLACE RCOEFTOT WITH COEFTOTOP
  ENDIF
ELSE
  REPLACE RCARD WITH N
  REPLACE RTAUX WITH TAUX
  REPLACE REFFECH WITH NT
  REPLACE RTOTALY WITH SY
  REPLACE RVARIANCE WITH VART
  REPLACE RECRTYPE WITH ECRT
  REPLACE RCOEFVAR WITH COEFT
  IF MENU = '2'
    REPLACE RBIAISTOT WITH BIAIST
    REPLACE RVARTOT WITH VARTOTT
    REPLACE RECRTTOT WITH ECRTOTT
    REPLACE RCOEFTOT WITH COEFTOTT
  ENDIF
  IF MENU = '4'
    REPLACE RBIAISTOT WITH BIAIST
    REPLACE RVARTOT WITH VARTOTT
    REPLACE RECRTTOT WITH ECRTOTT
    REPLACE RCOEFTOT WITH COEFTOTT
  ENDIF
ENDIF
IF CHSTRAT = 0
  WAIT "Fin de l'écriture, appuyez sur une touche pour revenir au menu."
ENDIF
RECNOB = RECNO()
USE &NOMFICH
RETURN

```

```

*****
*
* 10 : PROCEDURE MENUSEC
*
* Affichage du menu secondaire du programme VARECH et retour des valeurs
* CHSTRAT et CHEST : choix stratifié / non stratifié, estimation d'un total, d'une
* moyenne ou d'un quotient.
*
*****

```

PROCEDURE MENUSEC

* 1: Affichage du menu 1 à l'écran

```
CLEAR
@ 1,3 SAY 'Varech, version 4.'
@ 3,3 SAY DATE
@ 5,15 SAY '          MENU SECONDAIRE (1)'
@ 6,15 SAY '          *****'
@ 8,15 SAY '<1> : Estimation d'un total.'
@ 10,15 SAY '<2> : Estimation d'une moyenne.'
IF MENU = '2' .OR. MENU = '4'
  @ 12,15 SAY '<3> : Estimation d'un quotient.'
  @ 14,15 SAY '<R> : Retour au menu principal.'
  @ 17,1 SAY ''
ELSE
  @ 12,15 SAY '<R> : Retour au menu principal.'
  @ 17,1 SAY ''
ENDIF
```

* 2: Reponse de l'utilisateur au menu 1

```
REP1 = ''
BOUCLE = .T.
DO WHILE BOUCLE
  ACCEPT ' Choisissez une option : ' TO REP1
  REP1 = UPPER(REP1)
  IF REP1 = '1' .OR. REP1 = '2' .OR. (REP1 = '3' .AND. (MENU = ;
  '2' .OR. MENU = '4')) .OR. REP1 = 'R'
    BOUCLE = .F.
  ENDIF
ENDDO
CLEAR
IF REP1 = 'R'
  CHSTRAT = 2
  RETURN
ENDIF
```

* 3: Affichage du menu 2 à l'écran

```
@ 1,3 SAY 'Varech, version 4.'
@ 3,3 SAY DATE
@ 5,15 SAY '          MENU SECONDAIRE (2)'
@ 6,15 SAY '          *****'
@ 8,15 SAY '<1> : Estimation non stratifiée.'
@ 10,15 SAY '<2> : Estimation stratifiée.'
@ 17,1 SAY ''
```

* 4: Reponse de l'utilisateur au menu 2

```
REP2 = ''
BOUCLE = .T.
DO WHILE BOUCLE
  ACCEPT ' Choisissez une option : ' TO REP2
  REP2 = UPPER(REP2)
  IF REP2 = '1' .OR. REP2 = '2'
    BOUCLE = .F.
  ENDIF
ENDDO
CLEAR
```

* 5: Traduction de la réponse en CHSTRAT et CHEST
DO CASE

```
*
CASE REP2 = '1'
  CHSTRAT = 0
```



```

IF REP1 = '1'
  CHEST = 1
ENDIF
IF REP1 = '2'
  CHEST = 2
ENDIF
IF REP1 = '3'
  CHEST = 3
ENDIF
*
CASE REP2 = '2'
CHSTRAT = 1
IF REP1 = '1'
  CHEST = 1
ENDIF
IF REP1 = '2'
  CHEST = 2
ENDIF
IF REP1 = '3'
  CHEST = 3
ENDIF
ENDCASE
RETURN
*
*
*****
*
* 11 : PROCEDURE QUESFICH
*
* Entree d'un nom de fichier dans le programme VARECH.
*
*****
*
PROCEDURE QUESFICH
BOUCLE = .T.
DO WHILE BOUCLE
  ? 'Entrez le nom du fichier, sans extention ni chemin d"accés.'
  ? '(8 caractères maximum).'
  ? 'il doit être dans le repertoire dbaseiii du disque dur.'
  ? "
  ACCEPT ' nom : ' TO NOMFICH
  NOMFICH = TRIM(NOMFICH) + '.DBF'
  NOMFICH = UPPER(NOMFICH)
  IF LEN(NOMFICH) >0 .AND. LEN(NOMFICH) <13
    BOUCLE = F.
  ENDIF
ENDDO
* PUBLIC NOMFICH
RETURN
*
*****
*
* 12 : PROCEDURE QUESVAR
*
* Entree d'un nom de variable dans le programme VARECH.
*
*****

```

```
PROCEDURE QUESVAR
BOUCLE = .T.
DO WHILE BOUCLE
? "
ACCEPT 'Entrez le nom de la variable.(12 caractères maximum) :' TO REPVAR
REPVAR = TRIM(REPVAR)
REPVAR = UPPER(REPVAR)
IF LEN(REPVAR) >0 .AND. LEN(REPVAR) <12
BOUCLE = .F.
ENDIF
ENDDO
* PUBLIC REPVAR
RETURN
*
*
*****
*
* 13 : PROCEDURE QUESNOMB
*
* Entree du nombre de strates dans le programme VARECH.
*
*****
*
PROCEDURE QUESNOMB
INPUT 'Entrez le nombre de strates :' TO NB
PUBLIC NB
RETURN
*
*
*****
*
* 14 : PROCEDURE QUESTAUX
*
* Entree du taux de sondage dans le programme VARECH.
*
*****
*
PROCEDURE QUESTAUX
BOUCLE = .T.
DO WHILE BOUCLE
? "
INPUT 'Entrez le taux de sondage (en %) désiré (T>0) :' TOTAU
IF TAUX >0
BOUCLE = .F.
ENDIF
ENDDO
RETURN
*
*****
*
* 15 : PROCEDURE QUESNEYM
*
* Choix de l'allocation de NEYMAN dans le programme VARECH.
*
*****
*
PROCEDURE QUESNEYM
BOUCLE = .T.
```

```

DO WHILE BOUCLE
  ? "
  ACCEPT 'Voulez vous passer à l"allocation optimale de NEYMAN (O/N) :' TO OPT
  OPT = UPPER(OPT)
  IF OPT = 'O' .OR. OPT = 'N'
    BOUCLE = .F.
  ENDIF
ENDDO
RETURN

```

```

*
*
*****
*
* 16 : MODUL1.
*
* Module de calcul des variance, ecart type, coefficient de variation pour VECH1.
*
*****
*

```

```

PROCEDURE MODUL1
IF CHEST = 1
  @ 6,5 SAY 'Estimation du total de la variable .'
  @ 6,42 SAY VARY
ELSE
  @ 6,5 SAY 'Estimation de la moyenne de la variable .'
  @ 6,47 SAY VARY
ENDIF
@ 20,5 SAY 'Calcul en cours, ... patientez.'

```

```

* Calculs
N=0
VY=0
SY=0
SY2=0
EY=0
DO CALVECH1 WITH VARY, NOMVARSTR, NOSTR, CHSTRAT
* Effectif d'échantillon.
NT = N*TAUX/100
* Variance
VAR = VY
IF CHEST = 1
  VART = N^2*VAR/NT
ELSE
  VART = VAR/NT
ENDIF
IF VART < 10^-9
  VART=0
ENDIF
* Ecart-type et coefficient de variation.
ECRTT = SQRT(VART)
IF CHEST = 1
  IF SY=0
    COEFT =0
  ELSE
    COEFT = 100*ECRTT/SY
  ENDIF

```

```

ELSE
  IF EY=0
    COEFT=0
  ELSE
    COEFT = 100*ECRTT/EY
  ENDIF
ENDIF
* @ 10,5 SAY 'Calculs terminés.'
RETURN
*
*
*****
* 17: PROCEDURE MODUL2.
*
* Module de calcul des variance, ecart type, coefficient de variation et biais pour
* VECH2.
*
*****
*
PROCEDURE MODUL2
IF CHEST = 1
  @ 6,5 SAY 'Estimation du total de la variable : '
  @ 6,42 SAY VARY
ENDIF
IF CHEST = 2
  @ 6,5 SAY 'Estimation de la moyenne de la variable : '
  @ 6,47 SAY VARY
ENDIF
IF CHEST = 3
  @ 6,5 SAY 'Estimation du quotient : '
  TEXTEC = VARY + '/' + VARDEN
  @ 6,32 SAY TEXTEC
ENDIF
@ 20,5 SAY 'Calcul en cours, ... patientez.'
* Calculs
N=0
R=0
VY=0
VX=0
VXY=0
SY=0
SX=0
EY=0
DO CALVECH2 WITH VARY, VARDEN, NOMVARSTR, NOSTR, CHSTRAT
* Effectif d'échantillon.
NT = N*TAUX/100
* Biais
BIAIS = - (VXY - R*VX)
IF SX=0
  BIAIST=0
ELSE
  IF CHEST = 1
    BIAIST = (BIAIS*N^2)/(NT*SX)
  ENDIF
  IF CHEST = 2
    BIAIST = (BIAIS*N)/(NT*SX)
  ENDIF
ENDIF

```

```
      IF CHEST = 3
        BIAIST = (BIAIS*N^2)/(NT*SX^2)
      ENDIF
    ENDIF
  * Variance
  VAR = VY + VX*R^2 - 2*R*VXY
  IF CHEST = 1
    VART = N^2*VAR/NT
  ENDIF
  IF CHEST = 2
    VART = VAR/NT
  ENDIF
  IF CHEST = 3
    VART = N^2*VAR/(NT*SX^2)
  ENDIF
  IF VART < 10^-9
    VART=0
  ENDIF
  * Variance + carré du biais
  VARTOTT = VART + BIAIST^2
  IF VARTOTT < 10^-9
    VARTOTT=0
  ENDIF
  * Ecart-type et coefficient de variation.
  ECRTT = SQRT(VART)
  ECRTOTT = SQRT(VARTOTT)
  IF CHEST = 1
    IF SY=0
      COEFT=0
      COEFTOTT=0
    ELSE
      COEFT = 100*ECRTT/SY
      COEFTOTT = 100*ECRTOTT/SY
    ENDIF
  ENDIF
  IF CHEST = 2
    IF EY=0
      COEFT=0
      COEFTOTT=0
    ELSE
      COEFT = 100*ECRTT/EY
      COEFTOTT = 100*ECRTOTT/EY
    ENDIF
  ENDIF
  IF CHEST = 3
    IF R=0
      COEFT=0
      COEFTOTT=0
    ELSE
      COEFT = 100*ECRTT/R
      COEFTOTT = 100*ECRTOTT/R
    ENDIF
  ENDIF
  * @ 10,5 SAY 'Calculs terminés.'
  RETURN
```

```
*
*****
* 18: PROCEDURE MODUL3.
*
* Module de calcul des variances, ecarts type, coefficients de variation pour VECH3.
*
*****
*
  PROCEDURE MODUL3
  IF CHEST = 1
    @ 6,5 SAY 'Estimation du total de la variable :'
    @ 6,42 SAY VARY
  ELSE
    @ 6,5 SAY 'Estimation de la moyenne de la variable :'
    @ 6,47 SAY VARY
  ENDIF
  @ 20,5 SAY 'Calcul en cours, ... patientez.'
* Calculs
  N=0
  VYZ=0
  SY=0
  SZ=0
  EY=0
  DO CALVECH3 WITH VARY, VARZ, NOMVARSTR, NOSTR, CHSTRAT
  EY=SY/N
* Effectif d'échantillon.
  NT = N*TAUX/100
* Variance
  VAR = SZ*VYZ -SY^2
  IF CHEST = 1
    VART = VAR/NT
  ELSE
    VART = VAR/(N^2*NT)
  ENDIF
  IF VART < 10^-9
    VART=0
  ENDIF
* Ecart-type et coefficient de variation.
  ECRTT = SQRT(VART)
  IF CHEST = 1
    IF SY=0
      COEFT=0
    ELSE
      COEFT = 100*ECRTT/SY
    ENDIF
  ELSE
    IF EY=0
      COEFT=0
    ELSE
      COEFT = 100*ECRTT/EY
    ENDIF
  ENDIF
  @ 10,5 SAY 'Calculs terminés.'
  RETURN
```

```
*
*****
* 19: PROCEDURE MODUL4.
*
* Modul de calcul des variance, ecarts type, coefficient de variation pour VECH4.
*
*****
*
  PROCEDURE MODUL4
  IF CHEST = 1
    @ 6,5 SAY 'Estimation du total de la variable : '
    @ 6,42 SAY VARY
  ENDIF
  IF CHEST = 2
    @ 6,5 SAY 'Estimation de la moyenne de la variable : '
    @ 6,47 SAY VARY
  ENDIF
  IF CHEST = 3
    @ 6,5 SAY 'Estimation du quotient : '
    TEXTEC = VARY + '/' + VARX
    @ 6,32 SAY TEXTEC
  ENDIF
  @ 20,5 SAY 'Calcul en cours, ... patientez.'
* Calculs
  N=0
  VPYZ=0
  VPXZ=0
  VPXYZ=0
  SY=0
  SX=0
  SZ=0
  R=0
  EY=0
  EX=0
  EZ=0
  DO CALVECH4 WITH VARY, VARX, VARZ, NOMVARSTR, NOSTR, CHSTRAT
  EY=SY/N
  EX=SX/N
  EZ=SZ/N
* Effectif d'échantillon.
  NT = N*TAUX/100
* Biais asymptotiquement nul
  IF SX=0
    BIAIST=0
  ELSE
    EXP = R*VPXZ - VPXYZ
    BIAISS = EXP*(SZ^2/(NT*SX))
    BIAIST = EXP*(SZ^2/(NT*SX^2))
  ENDIF
  IF CHEST = 1
    BIAIST = BIAISS
  ENDIF
  IF CHEST = 2
    BIAIST = BIAISS/N
  ENDIF
* Variance
```

```
IF SX=0
  VART=0
ELSE
  VAR = (SZ^2/NT)*(VPYZ + VPXZ*R^2 - 2*R*VPXYZ)
  VART = (SZ^2/(NT*SX^2))*(VPYZ + VPXZ*R^2 - 2*R*VPXYZ)
ENDIF
IF CHEST = 1
  VART = VAR
ENDIF
IF CHEST = 2
  VART = VAR/N^2
ENDIF
IF VART < 10^-9
  VART=0
ENDIF
* Variance + carré du biais
VARTOTT = VART + BIAIST^2
IF VARTOTT < 10^-9
  VARTOTT=0
ENDIF
* Ecart-type et coefficient de variation.
ECRTT = SQRT(VART)
ECRTOTT = SQRT(VARTOTT)
IF CHEST = 1
  IF SY=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/SY
    COEFTOTT = 100*ECRTOTT/SY
  ENDIF
ENDIF
IF CHEST = 2
  IF EY=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/EY
    COEFTOTT = 100*ECRTOTT/EY
  ENDIF
ENDIF
IF CHEST = 3
  IF R=0
    COEFT=0
    COEFTOTT=0
  ELSE
    COEFT = 100*ECRTT/R
    COEFTOTT = 100*ECRTOTT/R
  ENDIF
ENDIF
* @ 10,5 SAY 'Calculs terminés.'
RETURN
```


Annexe A.3.1

LES ENQUETES PRELIMINAIRES : METHODE DE COLLECTE, QUESTIONNAIRE, CODIFICATION

- 1) Enquête Morphologie urbaine : instructions aux enquêteurs**
- 2) Enquête Morphologie urbaine : codification**
- 3) Enquête Bâti / population : instructions aux enquêteurs**
- 4) Enquête Bâti / population : codification**

1) Enquête Morphologie urbaine : instructions aux enquêteurs

QUITO - Enquête morphologie urbaine
Octobre / Décembre 1986

INSTRUCTIONS AUX ENQUETEURS

SOMMAIRE

1. OBJECTIFS SCIENTIFIQUES

- 1. 1. Le programme de recherche**
- 1. 2. L'enquête morphologie urbaine**

2. TECHNIQUE D'ENQUETE

3. ORGANISATION DU TRAVAIL

4. COLLECTE DE L'INFORMATION

- 4. 1. Principes de base**
- 4. 2. Description des batiments**
- 4. 3. Description des éléments non batis**

5 . EXEMPLE D' ILOT A ENQUETER

1. OBJECTIFS SCIENTIFIQUES

1. 1. Le programme de recherche

L'objectif du programme de recherche est de mettre au point un système d'observation permanent des populations urbaines, en intégrant l'information satellitaire sur la morphologie urbaine : suivi des effectifs et caractéristiques générales des populations, et mise en oeuvre de sondages spatiaux stratifiés sur des images satellites SPOT ou THEMATIC MAPPER, pour des investigations approfondies sur le comportement des citadins.

Deux idées sont à la base de ce programme :

- Seul un système basé sur la technique des sondages peut remplir les conditions de souplesse et rapidité nécessaires dans les villes à croissance rapide
- Une meilleure connaissance de l'espace urbain constitue un facteur important d'amélioration d'un système d'enquête démographique par sondage : pour l'établissement d'une base de sondage correcte et à jour, composée d'ilôts clairement définis, pour l'optimisation du sondage par stratification sur la morphologie urbaine, et pour l'obtention de résultats localisés, selon un découpage significatif de la ville enquêtée.

Les satellites renseignent de manière continue sur l'occupation du sol en milieu urbain : il s'agit donc, dans la méthode que nous développons, d'utiliser cette information exhaustive sur la morphologie urbaine pour recueillir rapidement, par sondage, des données relatives aux populations urbaines. Les travaux réalisés (par VERNIERE notamment) sur photographies aériennes ont montré que l'application de ce principe permet un allègement de l'échantillon d'enquête (donc, pour un même coût global et un même niveau de précision, une possibilité d'approfondir l'observation) et une spatialisation des résultats.

La réalisation de l'objectif général du programme implique des travaux, en parallèle, dans quatre directions :

- L'interprétation des images satellite à haute résolution en milieu urbain
- Les relations bâti/population et les sondages spatiaux en démographie
- L'intégration des données de télédétection dans des bases de données

localisées

-La télédétection comme système autonome d'observation démographique

Nous avons débuté nos travaux en 1985 sur la ville de Marseille (FRANCE), qui possédait dès cette date toutes les données nécessaires à une première phase de développement méthodologique.

Sur la ville de Quito, trois types d'enquête devront être réalisées :

- Enquête morphologie urbaine
- Enquête morphologie/démographie
- Enquête démographie : effectifs de population et systèmes résidentiels

1. 2. L'enquête morphologie urbaine

La première enquête que nous réalisons à Quito dans le cadre de ce programme de recherche a pour objectif de collecter des informations précises sur l'occupation du sol, c'est à dire la morphologie urbaine, de la ville de Quito.

L'observation au sol de la morphologie urbaine sur un échantillon d'îlots de Quito est indispensable pour interpréter les images satellite : ce n'est qu'en confrontant les données de terrain relevées au cours de l'enquête morphologie, et les données radiométriques enregistrées par les satellites, que l'on pourra interpréter correctement les images satellites de Quito. La connaissance de la morphologie urbaine sur l'échantillon d'îlots sert à établir les "clefs d'interprétation" qui seront appliquées aux images satellites sur l'ensemble de la zone urbanisée de Quito.

L'échantillon à enquêter du point de vue morphologie urbaine est constitué de 189 îlots, tirés sur les cartes INEC de 1982 au 1/10000, à l'aide d'une grille de 750 mètres de côté. L'information sur l'occupation du sol doit être recueillie précisément et exhaustivement sur ces 189 îlots, durant les mois d'Octobre, Novembre et Décembre 1986.

2. TECHNIQUE D'ENQUETE

Chaque îlot de l'échantillon est décrit par trois types de documents :

- **Carte** : sur le fond de plan INEC agrandi au 1/1000, sur lequel ne figurent au départ de l'enquête que les limites de chaque îlot à enquêter, les enquêteurs établiront la cartographie de l'occupation du sol de l'îlot. Tous les bâtiments ainsi que tous les éléments non bâtis devront figurer sur ces plans.
- **Fiche de description des bâtiments** : chacun des édifices dessinés sur la carte est décrit sur cette fiche. Il est repéré par un numéro de 1 à n.
- **Photographies** : pour chaque îlot, les chercheurs de l'Orstom feront des photographies depuis un point surplombant l'îlot (étage élevé d'un immeuble ou relief), et au sol, en bordure et/ou à l'intérieur de l'îlot.

Pour chaque îlot de l'échantillon, il doit y avoir concordance complète de l'information contenue dans ces trois documents, à partir desquels seront codés les descripteurs de la morphologie urbaine, sur une fiche de synthèse qui servira aux traitements informatiques.

3. ORGANISATION DU TRAVAIL

Tous les lundi matin, chaque enquêteur se verra attribuer la liste d'îlots à enquêter dans la semaine : il s'agit d'une liste bien précise d'îlots appartenant à l'échantillon.

La liste sera établie en fonction des principes suivants :

- un îlot, en moyenne, nécessite une demi-journée (4 heures) de travail pour être enquêté correctement.
- les îlots seront regroupés géographiquement, afin de minimiser les temps de déplacement.

Pour chacun des îlots de la liste hebdomadaire, l'enquêteur recevra une carte au 1/1000, ainsi que des fiches de description.

Des modifications du tissu urbain ayant pu se produire depuis 1982 (date d'établissement de la cartographie INEC), un chercheur accompagnera l'enquêteur pour vérifier la délimitation de tous les îlots que ce dernier doit enquêter durant la semaine.

Afin de rechercher collectivement des solutions aux problèmes rencontrés au cours de la collecte d'informations et de vérifier la cohérence des cartes et fiches de description par rapport aux photographies, deux rendez-vous sont fixés chaque semaine : le mercredi et le vendredi en fin d'après midi.

En cas d'incomplétude ou d'incohérence des cartes et fiches de description, il sera demandé aux enquêteurs de retourner sur le terrain pour corriger ou compléter l'information.

4. COLLECTE DE L'INFORMATION

4. 1. Principes de base

La première opération consiste à **rechercher une vision d'ensemble de l'îlot, depuis un point surplombant l'îlot** : étage élevé d'un immeuble voisin de l'îlot, ou relief. Ainsi, l'enquêteur pourra procéder plus aisément à :

- la cartographie des éléments bâtis et non bâtis composant l'îlot
- la numérotation de ces éléments
- au remplissage des questions relatives aux toits des bâtiments

D'autre part, certains éléments de l'occupation du sol ne sont visibles qu'en surplombant l'îlot, et non depuis les rues jouxtant l'îlot : cette première phase d'observation depuis un point élevé est donc particulièrement importante.

Dans un second temps, l'enquêteur procédera au remplissage des fiches de description, et complétera le plan d'après ce qui est **visible depuis les rues bordant l'îlot, et les cours ou passages intérieurs**, s'ils sont accessibles.

Enfin, si certaines réponses demeurent impossibles à obtenir par une vision directe, l'enquêteur questionnera les résidents de l'îlot pour obtenir l'information.

4. 2. Description des bâtiments

Pour chaque manzana, l'enquêteur remplit une ou plusieurs fiches de description des bâtiments, selon le nombre de ceux-ci ; chacun des bâtiments est dessiné sur le plan de l'îlot, et décrit dans une colonne de la fiche descriptive.

Définition du bâtiment

On appelle bâtiment toute construction en élévation, en cours ou achevée, à usage d'habitation ou non, qui comporte (ou comportera quand elle sera totalement achevée) au moins un toit supporté par des poteaux.

On inclue donc dans cette catégorie les hangars ou les garages couverts non fermés par des murs, les passages couverts entre bâtiments, les grandes citernes ou silos.

Les bâtiments en construction sont inclus à partir du moment où les fondations sont recouvertes d'un matériau de construction.

Limites du bâtiment

Les limites de chaque bâtiment doivent figurer sur le plan. Dans le cas où il s'agit de **constructions jointives**, il peut être délicat de déterminer les limites. On adoptera les **conventions** suivantes :

a) Tissu urbain continu (type centre historique de Quito) : chaque immeuble est individualisé. L'observation des toitures peut permettre de lever les ambiguïtés sur les limites d'immeubles.

b) Immeuble élevé avec garage attenant (1 seul niveau) : on distingue deux bâtiments, l'immeuble et le garage, qui ont des caractéristiques architecturales bien différentes.

c) Villa composée de blocs de constructions de différents niveaux : on ne compte qu'un seul bâtiment pour l'ensemble (y compris le garage), si ces blocs appartiennent visiblement à une même unité d'habitation.

d) Immeuble élevé composé de différents blocs de constructions de différents niveaux présentant une unité architecturale : on ne compte qu'un seul bâtiment.

Numéro du bâtiment

Tout bâtiment situé dans l'îlot est identifié par un numéro de 1 à n. A chaque îlot, la numérotation redémarre à 1 ; au sein d'un même îlot, deux bâtiments ne doivent pas porter le même numéro. Sur la carte, ce numéro est indiqué au centre du bâtiment ; sur la fiche, il est noté en haut de chacune des colonnes.

Stade de construction

Il s'agit de distinguer :

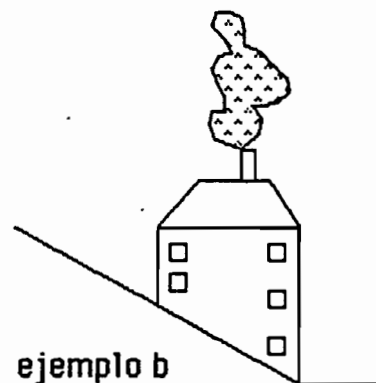
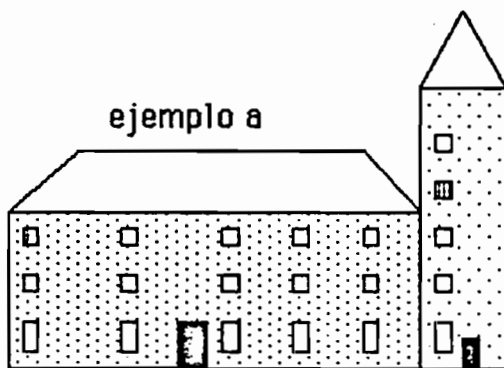
- En construction : bâtiment en chantier (à partir du moment où les fondations sont recouvertes d'un matériau de construction) non encore habitable (par exemple, toiture non terminée, vitres non posées...etc)
- Terminé : bâtiment habitable, dont l'étanchéité est assurée (toit, portes et fenêtres), même si les finitions intérieures (peinture, carrelage...etc) ne sont pas faites.

Un bâtiment habitable dans les étages inférieurs, dont le dernier niveau est constitué d'une dalle de béton avec des poteaux en béton, ou des fers à béton qui dépassent, est considéré comme "terminé".

Nombre d'étages

Nombre d'étages que comporte le bâtiment, y compris le rez de chaussée, et non compris les penthouses.

Si, en terrain plat, un même bâtiment complexe comporte différents nombres d'étages selon les parties du bâtiment, indiquer le nombre minimal et le nombre maximal (exemple a : 3/5, pour décrire un bâtiment comptant 3 étages dans une partie, et 5 dans l'autre).

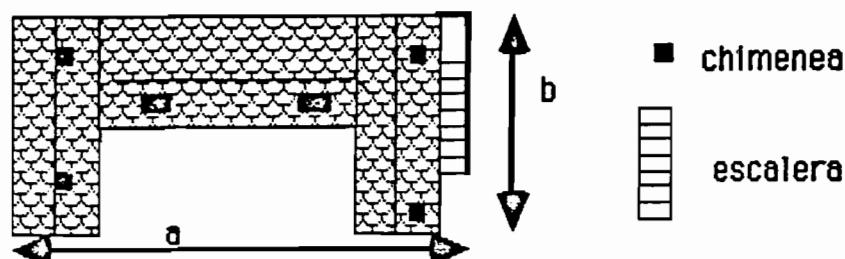


Si en raison d'une forte pente, un bâtiment compte 2 étages sur un des côtés et 3 étages sur un autre, en contrebas, indiquer le nombre d'étages minimal et maximal (2/3 dans l'exemple b), et ne pas oublier de noter la pente du sol sur le plan d'ilot.

Dimensions au sol

Les côtés a et b désignent les côtés du plus petit rectangle incluant le bâtiment (y compris les escaliers extérieurs menant aux étages du bâtiment).

Ces côtés a et b sont indiqués sur le plan de la façon suivante :

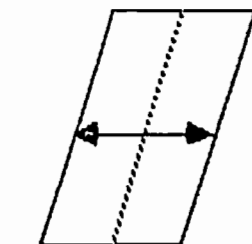
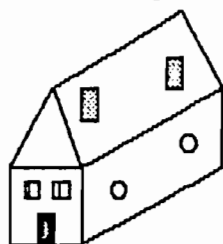


Sur la fiche de description des bâtiments, les mesures des côtés sont exprimées en mètres.

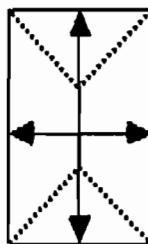
Nombre de pans du toit

Toit plat : Nombre de pans = 0. Ne rien indiquer sur le plan.

Toit avec pente(s) : Indiquer le nombre de pans sur la fiche, et le sens de la pente pour chacun des pans sur le croquis de l'îlot par une flèche dirigée vers le bas de la pente.



Toit à 2 pans



Toit à 4 pans

Matériau prédominant du toit

Il s'agit du matériau qui occupe la plus grande surface sur le toit.

Dans le cas où le toit est composé de plusieurs matériaux occupant des surfaces sensiblement égales, indiquer ces différents matériaux sur la fiche, et les délimiter sur le plan.

En dehors des matériaux les plus courants cités sur la fiche, pour lesquels il suffit de mettre une croix dans la case correspondante, on peut rencontrer à Quito d'autres types de matériaux pour les toitures : shingle, verre, carrelage ...etc. Dans ce cas, il faut indiquer en clair la dénomination du matériau, si elle est connue, ou, sinon, décrire précisément ce matériau sur la fiche "observations complémentaires".

Couleur prédominante du toit


Indiquer la couleur occupant la plus grande surface sur le toit du bâtiment, en faisant attention d'être cohérent avec le matériau prédominant : c'est la couleur du(des) matériau(x) prédominant(s) que l'on doit indiquer.

Forme du matériau du toit

Sur la fiche, dans la colonne relative au bâtiment, faire un croquis, en coupe, du matériau prédominant du toit.

Exemples de formes possibles selon les matériaux :

Dalle de béton, carrelage, paille : _____

Eternit :  ou _____

Tuiles :  ou _____

Nombre de cheminées

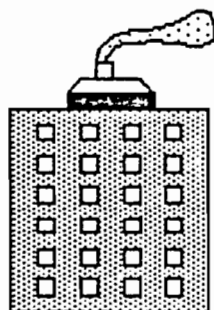
Indiquer le nombre de cheminées présentes sur le toit du bâtiment.

Présence de poteaux

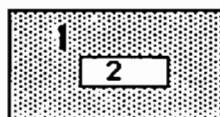
Nous retrouvons ici le cas des bâtiments habitables dans les étages inférieurs et dont le dernier étage est constitué d'une dalle en béton, d'où émergent des fers à béton ou des poteaux en bétons.

Présence de penthouse

S'il y a une penthouse (c'est à dire un petit bâtiment sur la terrasse supérieure d'un immeuble) sur un bâtiment, la dessiner sur le plan, lui attribuer un numéro et reporter celui-ci à la fois sur le plan et dans la colonne du bâtiment sur lequel elle est située (en face de la ligne "présence de penthouse") et la décrire dans une autre colonne de la fiche, comme si c'était un bâtiment indépendant.



Edificio de 6 pisos
con un penthouse



Croquis sobre el plano

Nombre de citernes

Indiquer le nombre de citernes couvertes présentes sur le toit du bâtiment : on désigne sous le terme de "citernes" les réservoirs d'eau qui servent à alimenter en eau le bâtiment en cas de coupure ou insuffisance du réseau public de distribution d'eau potable.

Présence de végétation

Indiquer s'il y a "aucune", "peu", "beaucoup" de végétation :

- sur les balcons saillants : balcons dépassant du mur extérieur du bâtiment
- sur la terrasse du toit

Autres observations sur les bâtiments

Dans cette rubrique, préciser rapidement une information incomplète dans les rubriques prévues par le questionnaire, ou indiquer la présence éventuelle d'éléments non répertoriés (capteurs solaires sur le toit par exemple). Si nécessaire, utiliser la fiche "observations complémentaires", en spécifiant bien le numéro du bâtiment dont il s'agit.

4. 3. Description du non bâti

Définition

On appelle "non-bâti" toute la surface de l'îlot où il n'y a pas de bâtiment.

Chaque zone homogène (du point de vue de l'occupation du sol) constitue un "élément non bâti".

La somme des surfaces occupées par des bâtiments, et des surfaces occupées par des éléments non-bâtis, doit correspondre à la surface totale de l'îlot (ce dernier incluant les rues qui bordent l'îlot jusqu'à la moitié de la chaussée).

Pente du sol

Indiquer sur le plan de l'îlot la pente du sol par une flèche :

- Pente faible ou nulle —————>
- Pente moyenne —————>>
- Pente forte —————>>>

Identification

Chaque élément non bâti et homogène de l'îlot doit être dessiné sur le plan : l'enquêteur lui attribue une lettre de A à Z (lettres majuscules), qui indique le type de non bâti. Sur le plan de l'îlot, cette lettre doit être inscrite au centre de la zone occupée par cet élément non bâti.

Sur la nomenclature (page 14), sont répertoriés 20 types de non bâti : si un élément correspond à un de ces types, l'enquêteur utilisera toujours ce codage. Pour toute autre catégorie de non bâti, il utilisera une autre lettre (U à Z), et indiquera sa signification sur la fiche "Observations complémentaires"

Si plusieurs éléments non bâtis d'un îlot sont du même type, la lettre qui correspond à ce type sera répétée dans chacune des zones occupées par ces éléments.

Végétation

Délimiter et identifier la végétation à l'intérieur de l'îlot, ainsi que sur la voirie qui le borde. Ne pas oublier la végétation sur les trottoirs (pelouse, arbres..)

Ne délimiter les arbres un par un que s'ils sont réellement isolés.

Si il y a une rangée d'arbres ou un groupe d'arbres, ne considérer la rangée ou le groupe que comme un seul élément non bâti (lettre G)

Si l'espèce est connue, l'indiquer sur la fiche "Observations complémentaire".

Voirie

On appelle voirie : les rues y compris les trottoirs et le terre-plein, les parkings, les cours, les passages intérieurs (sans toit) de l'îlot, les allées, les escaliers au sol.

QUITO - Morphologie- Oct./Dec. 1986

Description du non bâti

Indiquer sur le plan la pente du sol et la présence de quebrada

EAU

- A - rivière
- B - canal
- C - réservoir d'eau, bassin (sans toit)
- D - piscine

VEGETATION

- E - pelouse
- F - fleurs
- G - arbres (noter l'espèce)
- H - arbustes sans fleurs
- I - arbustes avec fleurs
- J - potager
- K - pâturage
- L - champ cultivé

VOIRIE

- M - bitume
- N - pavés ou dalles en béton
- O - pierres
- P - terre ou sable

VOITURES

- Q - avec des voitures en circulation
- R - avec des voitures en stationnement
- S - sans voitures

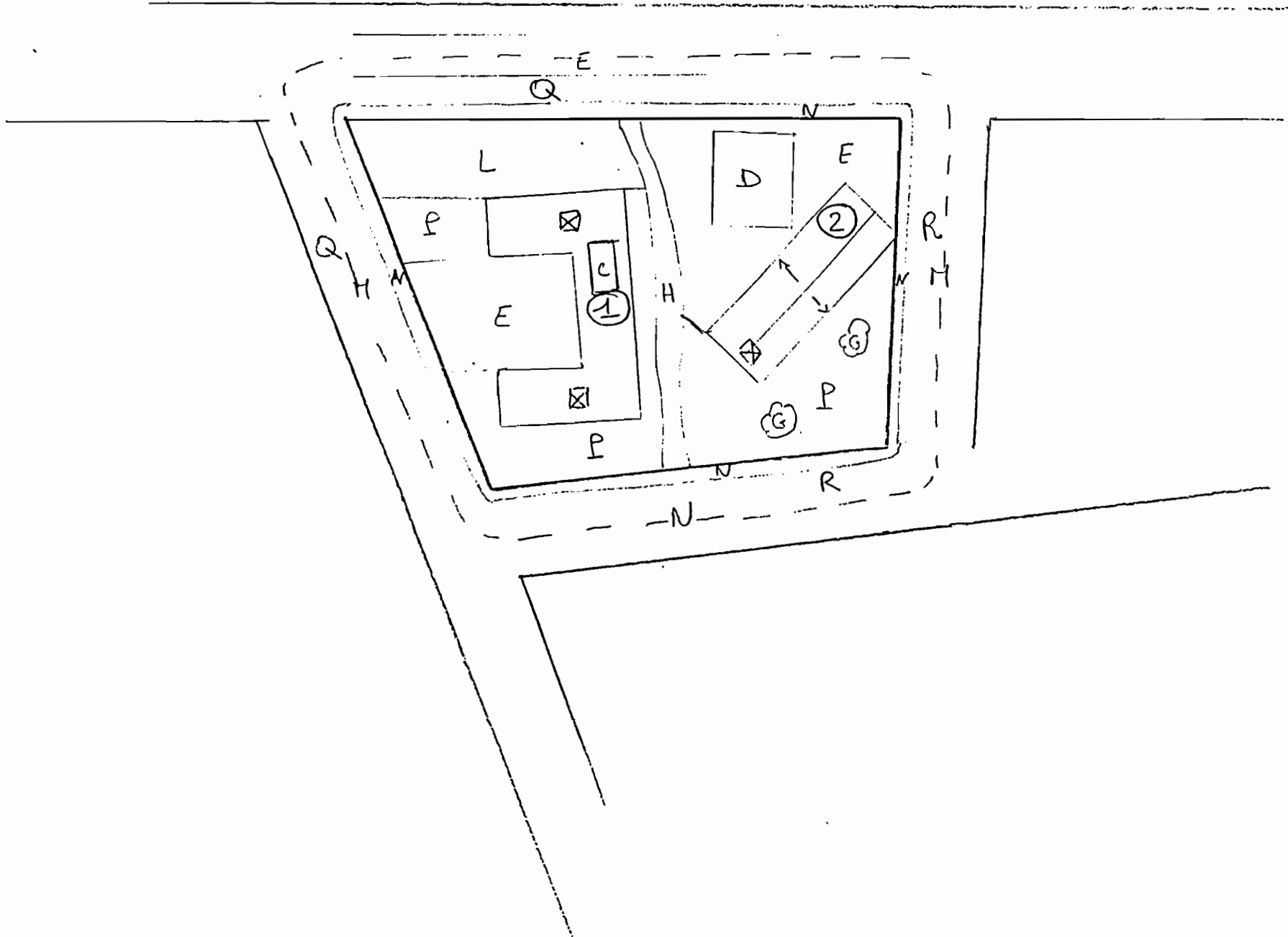
AUTRES

- T - mur de clôture

Pour toute autre catégorie de non bâti (par ex., carrelage..etc.) utiliser une lettre de U à Z, et noter sa signification sur la fiche "autres observations".

Si il y a beaucoup de matériaux entreposés sur le sol, distinguer la zone qui est couverte par ces matériaux de la zone qui n'est pas couverte, chacune de ces zones correspondant à un élément non bâti.

Utiliser une lettre de U à Z encore non utilisée pour les matériaux, et indiquer la signification de cette lettre sur la fiche "Observations complémentaires.



QUITO-Morphologie-Oct./Dec 1986

Numero d'îlot :		DESCRIPTION DES BATIMENTS				
Date:	Heure:					
Numero du batiment correspondant au repère porté sur le plan						
D	<u>Bâtiment:</u> Terminé	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	En construction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I	<u>Nombre d'étages (y compris RDC)</u>	-----	-----	-----	-----	-----
	<u>Dimensions au sol :</u> Côté a	-----	-----	-----	-----	-----
M	Côté b	-----	-----	-----	-----	-----
	(Indiquer ces côtés sur le plan)					
E	<u>Nombre de pans du toit (0 si plat)</u>	-----	-----	-----	-----	-----
	<u>Materiau prédominant du toit</u>					
T	Dalle de béton	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Eternit ou similaires	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Zinc ou similaires	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Tuiles vernies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Tuiles sans vernis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Paille ou similaires	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Autres (shingle, carrelage..)	-----	-----	-----	-----	-----
<u>Couleur prédominante du toit</u> (Blanc, Marron foncé, Marron clair, Gris, Brique, Noir, Rouge ou Vert)	-----	-----	-----	-----	-----	
<u>Forme du matériau du toit</u> (Faire un croquis en coupe)						
<u>Autres éléments sur le toit</u>						
A	Nombre de cheminées	-----	-----	-----	-----	-----
	Presence de poteaux en béton	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Presence de penthouse (à décrire)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nombre de tanques	-----	-----	-----	-----	-----	
<u>Presence de végétation</u>						
Sur les balcons saillants	-----	-----	-----	-----	-----	
Sur le toit	-----	-----	-----	-----	-----	
<i>Autres observations sur le batiment</i>		-----	-----	-----	-----	-----

QUITO-Morphologie-Oct./Dec 1986

Numero d'îlot: 48

OBSERVATIONS COMPLEMENTAIRES

Date: 1/10/86 Heure: 15h

Description du type d'îlot (notamment le standing de l'habitat): Résidentiel haut standing

Détails sur la description des bâtiments ou du non bâti de l'îlot :

2) Enquête Morphologie urbaine : codification

GENERALITES SUR LA CODIFICATION

NOMBRE D'ILOTS SAISIS ET CODES

* 199 îlots ont été codés et saisis, c'est à dire :

- îlots 101 à 166 inclus, sauf l'îlot 155 (n'a fait l'objet que de photos)
- îlots 201 à 276 inclus, sauf les îlots 243 (parc à côté de la Patria) et 265 (parc de la Caroline)
- îlots 302 à 362 inclus, sauf l'îlot 351 (usine qui n'a fait l'objet que de photos).

Le descriptif de ces 199 îlots (recto de la fiche de codification) correspond au fichier QMANZANS.

* A ces 199 îlots, correspondent 2133 genres de bâtiments, décrits dans le fichier QBATIMENT (saisie des données du verso de la fiche de codification)

1 . RECTO DE LA FICHE DE CODIFICATION

IDENTIFIANTS

N° Manzana

Numéro à 3 chiffres de 101 à 166, de 201 à 276 et de 302 à 362 affectés avant l'enquête morphologie aux îlots composant l'échantillon.

N° INEC

* Identifiant INEC en 7 chiffres (Zone en 3 chiffres, Secteur en 2 chiffres et Manzana en 2 chiffres) tel qu'il peut être lu sur les cartes au 1/10.000, suivi d'un code à 1 chiffre précisant si :

- 1 = la manzana INEC a été conservée entièrement pour l'enquête morphologique
- 2 = la manzana n'a été enquêtée qu'en partie
- 3 = la manzana enquêtée est plus grande que la manzana INEC

* *Dans le cas où la manzana enquêtée correspond à 2 manzanas INEC entières, deux fiches de codification ont été établies ; sur chacune, a été décrite morphologiquement la portion de "manzana 1986" correspondant à la manzana INEC. C'est le cas de la manzana 237, qui correspond en fait à 2 manzanas INEC ; il y a donc deux fiches, l'une notée 237, l'autre notée 275.*

* Si la manzana enquêtée morphologiquement en octobre-novembre 1986 *n'existe pas sur les plans INEC* de 1982 au 1/10000 (hors limites INEC de Quito, ou non existante en 1982), elle reçoit le code 999.99.99.9.

N° enquêteur

Codes :

- 1 = Ligia
- 2 = Xavier
- 3 = Olivier
- 9 = îlot ayant uniquement fait l'objet de photos.

SURFACES

Une fois connues les surfaces des îlots de l'échantillon par TIGRE, les surfaces mesurées sur carte devraient être corrigées ; il ne faut utiliser les chiffres de surface notés sur la fiche de codification qu'avec prudence. Ils ne sont destinés qu'à permettre le calcul des ratios (ex : % non bâti) n'ayant qu'une valeur *relative*, % à rassembler en classes.

Surface îlot

Surface de l'îlot en m², non compris la voirie bordant l'îlot. Quand l'îlot est représenté sur la carte au 1/15000 de l'IGM, il est mesuré sur cette carte ; dans le cas contraire, la mesure est faite sur les plans INEC au 1/10.000 (Voir note dans le coin haut droite des fiches de codification des manzanas). Dans tous les cas, la mesure est faite avec une grille millimétrée sur support stable.

La voirie interne à l'îlot est incluse dans ce calcul de surface.

Surface 1/2 voirie

Moitié de la surface en m² de la voirie encadrant l'îlot.

Mêmes principes de mesure que pour la surface de l'îlot.

Surface totale

Surface totale = surface îlot + surface 1/2 voirie. Exprimée en m².

TYPE MORPHOLOGIQUE

Carte Caza/Godard

Il s'agit des modalités de la légende de la carte "Esbozo de tipologia del habitat", établie par P. CAZAMAJORE et H. GODARD en 1984, soit :

- 1 = Zone industrielle
- 2 = Zone de végétation
- 3 = Centre des affaires et de commerces de luxe
- 4 = Zone non urbanisée en 1984 (le code 16 des fiches de codification a été recodé 4 dans le fichier, car même signification)
- 5 = Zone urbaine "diverse" jusqu'en 1983
- 6 = Habitat classe haute
- 7 = Habitat mixte classes haute et moyenne
- 8 = Habitat classe moyenne
- 9 = Programmes de vivienda de l'Etat classe moyenne
- 10 = Programmes de vivienda de l'Etat classe populaire
- 11 = Habitat mixte classes haute et populaire
- 12 = Habitat mixte classes moyenne et populaire
- 13 = Habitat populaire consolidé ou en voie de consolidation
- 14 = Habitat populaire non consolidé
- 15 = Taudification verticale (centre) et horizontale (quartiers périphériques)

Ont été ajoutées les modalités :

- 17 = Commerce
- 99 = Type non défini

Enquêteur et Photo

Les mêmes codes sont utilisés pour le type défini par l'enquêteur et pour celui noté au moment de la prise de photo et au vu des diapos.

Après examen des modalités retenues par chacun des enquêteurs, ont été retenus les types suivants (les classes 1 à 6 se rapportant aux catégories d'habitat) :

- 1 = Alta ou Burgesia alta
- 2 = Media alta
- 3 = Media media
- 4 = Media baja
- 5 = Media y Baja
- 6 = Baja
- 7 = Industriel
- 8 = Scolaire
- 9 = Commercial
- 10 = Equipements sociaux
- 11 = Végétation (hors terrains agricoles)
- 12 = Agricole
- 13 = Bureaux

Un îlot dont seulement une partie est consacrée à l'habitation est classé avec les codes 1 à 6 selon le standing de l'habitat ; les codes 7 à 13 sont utilisés exclusivement pour les îlots ne comptant aucun bâtiment d'habitation.

BATI

Surface batie

Il s'agit de la surface correspondant à la somme des surfaces de tous les bâtiments de la manzana, qu'ils soient terminés ou non.

Densité du bâti

Cinq niveaux de densité, établis d'après le plan de l'enquêteur et les diapos :

- 1 = Non bâti (pas un seul bâtiment sur la manzana)
- 2 = Densité très faible
- 3 = Densité faible
- 4 = Densité moyenne
- 5 = Densité forte
- 6 = Densité très forte

Implantation du bâti dans l'îlot

Codée d'après le plan établi par l'enquêteur et les diapos :

- 1 = Pas de type d'implantation particulier. Bâti homogène sur tout l'îlot, sans directions particulières.
- 2 = Bâtiments concentrés au centre de l'îlot, le bord des rues étant non bâti
- 3 = Bâtiments non alignés, concentrés sur le pourtour de l'îlot, près des rues, le centre de l'îlot étant vide
- 4 = Bâtiments concentrés dans un coin de l'îlot, le reste étant à peu près non bâti
- 5 = Bâtiments concentrés dans deux angles opposés de l'îlot, le reste étant à peu près non bâti
- 6 = Bâtiments alignés régulièrement près des rues, le centre de l'îlot étant vide
- 7 = Répartition des bâtiments homogène, selon un plan régulier sur tout l'îlot (en semis, et non en barre)
- 9 = Bâtiments concentrés le long de la rue principale
- 10 = Maisons accolées formant des barres continues (dans ce cas, la barre constitue *un seul* bâtiment ; c'est cette barre qui est décrite dans le tableau descriptif des bâtiments terminés, au dos de la fiche de codification)
- 11 = A rassembler avec code 10. Pas de différence significative entre ces 2 classes
- 12 = Idem 10, mais les maisons n'étant pas strictement accolées, la "barre" est divisée en n bâtiments semblables, qui sont tous décrits dans le tableau du dos de la fiche de codification.

Nombre de bâtiments

Le nombre de bâtiments total (= terminés + en construction), le nombre de bâtiments terminés et le nombre de bâtiments en construction correspondent aux bâtiments décrits par les enquêteurs sur les questionnaires.

Les pent-houses sont comptabilisés et décrits comme tout autre bâtiment de plein pied.

Répartition des bâtiments terminés par genre

Dans la colonne correspondant à chaque genre, est porté le nombre de bâtiments correspondant à ce genre ; ne sont comptabilisés et donc décrits au verso de la fiche de codification que les bâtiments terminés.

Pente du toit

S'il n'existe pas de direction majoritaire des faîtes, le code est 5 (cette variable ayant été recodée après saisie, considérer les modalités inscrites dans le fichier QMANZANS et non celles notées dans les fiches de codification).

S'il existe une direction majoritaire, le code est :

1 = Nord / sud

2 = Est / ouest

3 = Nord est / sud ouest

4 = Nord ouest / sud est

S'il n'existe pas de bâtiments terminés dans la manzana, le code est 9.

Penthouse, Poteaux fer/béton et Cheminées

Pour chacune de ces trois variables, est noté le nombre de bâtiments ^{terminés} présentant l'un de ces éléments sur son toit.

Végétation sur balcons saillants

1 = Aucune végétation sur des balcons saillants

2 = Végétation sur balcons saillants peu importante

3 = Végétation sur balcons saillants très importante

* Dans le cas d'un *ilot ne comptant aucun bâtiment*, toutes les variables de la section "Bâti" sont codées 0 pour les nombres de bâtiments, et 9 ou 99 pour les autres variables de la section.

NON BATI

Toute la description du NON BATI ne concerne que l'intérieur de l'îlot, à l'exclusion de la voirie entourant l'îlot.

Pente du sol

Direction (mêmes codes pour dominante et sous-dominante) :

- 1 = Nord
- 2 = Nord est
- 3 = Est
- 4 = Sud est
- 5 = Sud
- 6 = Sud ouest
- 7 = Ouest
- 8 = Nord ouest

Niveau : 1 = aucune pente (sol quasiment plat)

- 2 = pente faible
- 3 = pente forte
- 4 = pente très forte

Dans le cas où il n'existe pas de pente du tout, les codes sont :

Dominante : Direction = 9, Niveau = 1

Sous-dominante : Direction = 9, Niveau = 9

Eau

Le type de surface en eau reprend les rubriques prévues dans le manuel d'enquête, et il est codé :

- 1 = Rivière
- 2 = Canal
- 3 = Réservoir d'eau, bassin (sans toit)
- 4 = Piscine

Si pas de surface en eau dans la manzana, le code est 9.

La surface en m² est estimée d'après le plan établi par l'enquêteur et les diapos de la manzana ; elle correspond au total des surfaces en eau, tous types 1 à 4 confondus, de la manzana.

Végétation dans l'îlot

Le type de surface occupée par de la végétation reprend les rubriques prévues dans le manuel d'enquête, et il est codé :

- 1 = Pelouse (en fait, toute surface en herbe non consacrée spécifiquement à du pâturage)
- 2 = Fleurs
- 3 = Arbres
- 4 = Arbustes sans fleurs

- 5 = Arbustes avec fleurs
- 6 = Potager
- 7 = Pâturage
- 8 = Champ cultivé

Si pas de surface occupée par de la végétation dans la manzana, le code est 9.
La surface en m² est estimée d'après le plan établi par l'enquêteur et les diapos de la manzana ; elle correspond au total des surfaces occupées par de la végétation, tous types 1 à 8 confondus, de la manzana.
Dans le cas où il n'existe pas de végétation sous-dominante (pas du tout de végétation dans l'îlot ou végétation d'un seul type, noté "dominant"), le code du type de végétation sous-dominante est 9.

Voirie dans l'îlot

On appelle "voirie" toute surface non bâtie et non occupée par de la végétation.
Le type de surface occupée par de la voirie reprend les rubriques prévues dans le manuel d'enquête, et il est codé :

- 1 = Bitume
- 2 = Pavés ou dalles en béton
- 3 = Pierres
- 4 = Terre ou sable
- 5 = Carrelage
- 6 = Matériaux de bâtiment entreposés sur le sol
- 8 = Terre battue (tennis)

Si pas de surface occupée par de la voirie dans la manzana, le code est 9.
La surface en m² est estimée d'après le plan établi par l'enquêteur et les diapos de la manzana ; elle correspond au total des surfaces occupées par de la voirie, tous types 1 à 8 confondus, de la manzana.
Dans le cas où il n'existe pas de voirie sous-dominante (pas du tout de voirie dans l'îlot ou voirie d'un seul type, noté "dominant"), le code du type de voirie sous-dominant est 9.

Murs de cloture

Il s'agit de murs de séparation à l'exclusion des murs constitutifs des bâtiments (soutenant leur toiture).

A l'intérieur de l'îlot, les codes sont :

- 1 = Aucun mur de cloture
- 2 = Murs de cloture peu nombreux
- 3 = Murs de cloture très nombreux

Sur le périmètre de l'îlot, le code correspond au % du périmètre de l'îlot occupé par des murs.

1/2 VOIRIE BORDANT L'ÎLOT

Trottoir

Dans le cas où la chaussée de circulation est bordée par un trottoir pour le cheminement des piétons, on utilise les codes suivants pour décrire les matériaux dominants et sous-dominants des trottoirs :

- 1 = Bitume
- 2 = Pavés ou dalles en béton
- 3 = Pierres
- 4 = Terre ou sable
- 5 = Carrelage
- 6 = Matériaux de bâtiment entreposés sur le sol
- 7 = Pelouse
- 8 = Arbres

Dans le cas où tous les trottoirs bordant l'îlot sont construits dans le même matériau, le matériau dominant est codé d'après les codes cités ci-dessus, et le matériau sous-dominant est codé 9.

Dans le cas où il n'existe pas de trottoir, le code des matériaux dominants et sous-dominants est 9.

Chaussée

Dans le cas où il y a une chaussée de circulation, on utilise les codes suivants pour décrire les matériaux dominants et sous-dominants des chaussées :

- 1 = Bitume
- 2 = Pavés ou dalles en béton
- 3 = Pierres
- 4 = Terre ou sable
- 5 = Carrelage

Dans le cas où toutes les chaussées bordant l'îlot sont construites dans le même matériau, le matériau dominant est codé d'après les codes cités ci-dessus, et le matériau sous-dominant est codé 9.

Dans le cas où il n'existe pas de chaussée (limite d'îlot = quebrada, mur ou végétation) le code des matériaux dominants et sous-dominants est 9.

Le terre-plein central n'est pas considéré comme faisant partie de la chaussée ; c'est pourquoi il n'est pas prévu de code désignant de la végétation pour les chaussées.

Pelouse, Arbres et Arbustes

Ces trois variables décrivent le % de la longueur de voirie bordant l'îlot occupé par chacun de ces types de végétation, que ce soit sur les trottoirs ou sur les terre-plein centraux.

Dans la quasi-totalité des cas, cette végétation est intermittente, c'est à dire que la pelouse alterne avec le béton sur les trottoirs, et que les arbres ne forment pas une ligne continue de végétation mais sont espacés entre eux de quelques mètres ; c'est néanmoins la longueur totale occupée par cet agencement (pelouse/béton, par exemple) qui est prise en compte pour le calcul du %.

Voitures

En stationnement : % de longueur de bordure d'îlot occupé par des voitures en stationnement (au maximum)

En circulation : % de longueur de bordure d'îlot occupé par des voitures en circulation. Ce qui correspond, en fait, au % du périmètre de l'îlot disposant d'une chaussée où peuvent circuler des voitures.

2 . VERSO DE LA FICHE DE CODIFICATION

* **N° Manzana, N° INEC et N° Enquêteur** : idem recto de la fiche de codification.

* Ne sont décrits que les bâtiments considérés comme terminés par l'enquêteur.






* Le nombre de bâtiments de chaque genre est noté sur le recto de la fiche de codification ; il peut y avoir au maximum 17 genres pour une seule manzana. Sont regroupés dans chaque genre les bâtiments de la manzana qui ont des caractéristiques du bâti semblables ou très voisines.

* Les codes utilisés pour les descripteurs des bâtiments sont ceux notés sur le verso de la fiche de codification, soit :

Matériau du toit

- 1 Dalle de béton
- 2 Eternit ou similaires
- 3 Zinc ou similaires
- 4 Tuiles vernies
- 5 Tuiles non vernies
- 6 Paille ou similaires
- 7 Carrelage
- 8 Autres
- 9 Non observé

Forme du matériau du toit

1	
2	
3	
4	
5	
6	Autre
9	Non observé

Couleur du toit

1 Blanc	9 Bleu clair
2 Orange	10 Bleu foncé
3 Marron clair	11 Gris
4 Marron foncé	12 Noir
5 Brique	13 Autre
6 Rouge	99 Non observé
7 Vert clair	
8 Vert foncé	

* Les codes 9 ou 99 n'ont en fait pas été utilisés pour les descripteurs des bâtiments, puisqu'il n'est pas apparu de cas "non observé".

* Couleur du toit

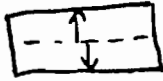
A l'issue de la codification, il s'est avéré nécessaire de regrouper des modalités pour le descripteur "couleur du toit", les qualificatifs "clair" et "foncé" pour les couleurs bleu et vert n'ayant pas la même signification selon les enquêteurs. On a donc finalement :

- Couleur code 7 : vert clair ou vert foncé (code 8 supprimé)
- Couleur code 9 : bleu clair ou bleu foncé (code 10 supprimé)

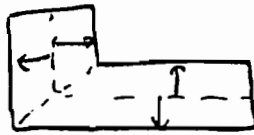
* Le code 14 = jaune a été ajouté, afin de rendre compte de quelques cas d'Eternit jaune.

*** Nombre de pans du toit**

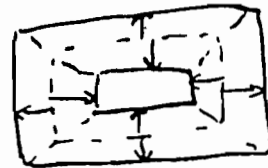
Dans le cas de bâtiments en angle, la codification a été fait comme suit :



2 pans



2 pans



2 pans

*** Forme du matériau du toit**

A été ajoutée la modalité 6 =

(ex : marché national, dans le sud de Quito)

*** Surface au sol**

Il s'agit de la surface moyenne des bâtiments composant le genre de bâtiment dans la manzana.

L'analyse des surfaces de bâtiments selon les enquêteurs (histogrammes, tableaux de fréquences et indicateurs statistiques) a conduit à augmenter les surfaces mesurées par l'enquêtrice de 20 %, afin d'homogénéiser ses mesures avec celles de l'enquêteur ; au vu des tests effectués, il était net que l'enquêtrice avait systématiquement sous-estimé les dimensions des bâtiments, alors que celles de l'enquêteur étaient correctes.

N° Manzana N° INEC N° enquêteur (Ligia=1, Xavier=2)
 (Entière=1, en partie=2)

Surface îlot Surface 1/2 voirie Surface totale
 (En m² - Mesure sur IGM 1/15000)

TYPE Carte Caza/Godard Enquêteur Photo

BATI Surface (m²) (D'après quest/plan)

Densité du bâti (D'après plan/quest./photo) Implantation du bâti dans l'îlot

Nombre de bâtiments : Total Terminés En construction

Répartition des bâtiments terminés par genre :

Genre	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Nbre																	

Pente du toit (direction majoritaire des faîtes) (Sans pente 0, N/S1, E/W2, NE/SW3, NW/SE4)

Nombre de bâtiments terminés avec penthouse

Nombre de bâtiments terminés avec poteaux fer/béton

Nombre de bâtiments terminés avec cheminées

Présence de végétation sur balcons saillants (Aucune 1, Peu nomb. 2, Très nomb. 3)

NON BATI

Pente du sol : Dom. : Direction (Ss pente=0, N=1, NE=2, E=3, Niveau (Aucune 1, Faible 2, Es-dom. : Direction SE=4, S=5, SW=6, W=7, NW=8) Niveau Forte 3, Très forte 4)

Eau : Type Surface (m²)

Végétation dans l'îlot : Hors bâti : Type dom. Type ss-dom. Surface (m²)
 Couvrant le bâti : Type dom. Type ss-dom. Surface (m²)

Voirie dans l'îlot : Matériau dom. Matériau ss-dom. Surface (m²)

Présence de murs de clôture : dans l'îlot (Aucun 1, Peu nomb. 2, Très nomb. 3)
 sur le périmètre de l'îlot (En % du périmètre de l'îlot)

1/2 VOIRIE BORDANT L'ÎLOT

Trottoir : Mat. dom. Mat. ss-dom. Chaussée : Mat. Dom. Mat. Ss-dom.
 Pelouse Arbres Arbustes (en % longueur de bordure)
 Voitures : En stationnement En circulation

CODIFICATION QUITO - FICHE MANZARA

N° Manzana n° INEC :






N° enquêteur (Ligia=1, Xavier=2)

N° du genre de bâtiment dans la manzana	Descripteurs des bâtiments terminés					
	Matériau du toit	Forme mat. du toit	Couleur mat. du toit	Nbre pans du toit	Nbre étages (Rdc inclus)	Surface au sol (m2)
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						

Matériau du toit

- 1 Dalle de béton
- 2 Eternit ou similaires
- 3 Zinc ou similaires
- 4 Tuiles vernies
- 5 Tuiles non vernies
- 6 Paille ou similaires
- 7 Carrelage
- 8 Autres
- 9 Non observé

Forme du matériau du toit

- 1 
- 2 
- 3 
- 4 
- 5 
- 6 Autre
- 8 Non observé

Couleur du toit

- 1 Blanc
- 2 Orange
- 3 Marron clair
- 4 Marron foncé
- 5 Brique
- 6 Rouge
- 7 Vert clair
- 8 Vert foncé
- 9 Bleu clair
- 10 Bleu foncé
- 11 Gris
- 12 Noir
- 13 Autre
- 99 Non observé

3) Enquête Bâti / population : instructions aux enquêteurs

QUITO - Enquête bâti/population
Avril 1987

INSTRUCTIONS AUX ENQUETEURS

Orstom - IPGH

F. DUREAU,
Quito, mars 1987

SOMMAIRE

1. OBJECTIFS SCIENTIFIQUES

1. 1. Le programme de recherche

1. 2. L'enquête bâti/population

2. TECHNIQUE D'ENQUETE

3. ORGANISATION DU TRAVAIL

4. COLLECTE DE L'INFORMATION

4. 1. Principes de base

4. 2. Recensement, délimitation et numérotation des logements

4. 3. Définition de la population à enquêter. Statut de résidence

4. 4. Renseignements relatifs aux personnes

5 . LA FICHE RECAPITULATIVE

6. EXEMPLE DE MANZANA ET DE LOGEMENT

1. OBJECTIFS SCIENTIFIQUES

1. 1. Le programme de recherche

L'objectif du programme de recherche est de mettre au point un système d'observation permanent des populations urbaines, en intégrant l'information apportée par les images satellite sur la morphologie urbaine : suivi des effectifs et caractéristiques générales des populations, et mise en oeuvre de sondages spatiaux stratifiés sur des images satellites SPOT ou THEMATIC MAPPER, pour des investigations approfondies sur le comportement des citoyens.

Deux idées sont à la base de ce programme :

- Seul un système basé sur la technique des sondages peut remplir les conditions de souplesse et rapidité nécessaires dans les villes à croissance rapide
- Une meilleure connaissance de l'espace urbain constitue un facteur important d'amélioration d'un système d'enquête démographique par sondage.

Les satellites renseignent de manière continue sur l'occupation du sol en milieu urbain : il s'agit donc, dans la méthode que nous développons, d'utiliser cette information exhaustive sur la morphologie urbaine pour recueillir rapidement, par sondage, des données relatives aux populations urbaines. Les travaux réalisés (par VERNIERE notamment) sur photographies aériennes ont montré que l'application de ce principe permet un allègement de l'échantillon d'enquête (donc, pour un même coût global et un même niveau de précision, une possibilité d'approfondir l'observation) et une spatialisation des résultats.

La réalisation de l'objectif général du programme implique des travaux, en parallèle, dans quatre directions :

- L'interprétation des images satellite à haute résolution en milieu urbain
- Les relations bâti/population et les sondages spatiaux en démographie
- L'intégration des données de télédétection dans des bases de données localisées
- La télédétection comme système autonome d'observation démographique

Nous avons débuté nos travaux en 1985 sur la ville de Marseille (FRANCE), qui possédait dès cette date toutes les données nécessaires à une première phase de développement méthodologique.

Sur la ville de Quito, trois types d'enquête sont réalisées :

- Enquête morphologie urbaine (Octobre-Décembre 1986)
- Enquête bâti/population (Avril 1987)
- Enquête démographie : effectifs de population et systèmes résidentiels (Octobre 1987)

L'enquête morphologie urbaine à Quito a été réalisée sur le terrain du 1er octobre au 9 décembre 1986 : cette première enquête a permis de collecter des informations précises sur l'occupation du sol, c'est à dire la morphologie urbaine, de 199 îlots de la ville de Quito, tirés sur les cartes INEC de 1982 au 1/10000, à l'aide d'une grille de 750 mètres de côté.

1. 2. L'enquête bâti/population

Par l'enquête bâti/population, il s'agit de recueillir des données démographiques et socio-économiques sur un sous-échantillon des 199 îlots ayant fait l'objet de l'enquête morphologie urbaine fin 1986 ; ainsi, nous disposerons à la fois des données sur la morphologie urbaine et sur les habitants de ce sous-échantillon d'îlots.

Nous pourrions donc tenter de répondre aux deux questions suivantes :

- Quels sont les descripteurs de la morphologie urbaine les plus pertinents pour l'observation démographique à Quito?
- Quelles sont l'unité spatiale de collecte et la stratégie de sondage pertinentes pour une enquête démographique à Quito?

Ce sont ces descripteurs et stratégie de sondage qui seront retenus pour l'enquête démographique générale d'octobre 1987.

L'échantillon à enquêter en avril 1987 du point de vue socio-démographique est constitué de 55 îlots, choisis de façon raisonnée parmi les 199 îlots de l'enquête morphologie urbaine ; les îlots ont été sélectionnés de manière à rassembler les différents types de bâtiments, les différents types d'îlots, et les différentes localisations au sein de la ville de Quito. La définition des types de bâtiments et types d'îlots est issue de l'exploitation des données recueillies au cours de l'enquête morphologie urbaine de fin 1986.

2. TECHNIQUE D'ENQUETE

Chaque îlot de l'échantillon de l'enquête bâti/population a déjà été enquêté en octobre-décembre 1986, lors de l'enquête morphologie urbaine ; dès le début de l'enquête bâti/population, l'enquêteur reçoit donc pour chaque îlot :

- **un plan au localisant tous les bâtiments de l'îlot** : sur ce plan, chaque bâtiment est identifié par un numéro de 1 à n. Ce numéro est indiqué au centre du bâtiment qu'il concerne. (Voir exemple de plan page 23)
- **un jeu de fiches décrivant tous les bâtiments de l'îlot**: chaque colonne de la fiche correspond à un bâtiment, dont le numéro (cohérent avec celui noté sur le plan) est indiqué en haut de la colonne. (Voir exemple de fiche page 24)

Ce plan et ces fiches permettent à l'enquêteur de repérer et identifier sans ambiguïté les bâtiments de l'îlot : l'enquêteur doit **conserver la même délimitation et la même numérotation de bâtiments** que celle apparaissant sur les documents de l'enquête morphologie urbaine (plan et fiches).

Pour collecter les données de l'enquête bâti/population, l'enquêteur reçoit deux types de fiches :

- **La fiche récapitulative** : il y a une fiche récapitulative par îlot. Sur cette fiche, l'enquêteur récapitule des informations générales relatives aux logements appartenant à l'îlot.
- **Les fiches logement** : ces fiches sont utilisées pour décrire les personnes résidant dans les logements des bâtiments faisant partie de l'îlot.

Il doit y avoir concordance complète de l'information entre ces quatre types de documents (Plan et fiches de l'enquête morphologie, fiches récapitulatives et fiches logement de l'enquête bâti/population).

3. ORGANISATION DU TRAVAIL

L'enquête débute par une **réunion d'information** le Jeudi 9 Avril, au cours de laquelle sont exposés les objectifs de l'enquête et la technique de collecte de l'information.

Le lendemain, chaque enquêteur **visitera** avec le superviseur et les chercheurs de l'Orstom **tous les ilots** qu'il aura à enquêter pendant les deux semaines, afin qu'il n'y ait pas d'ambiguïté sur la localisation et la délimitation des ilots et de leurs bâtiments ; il **recevra les documents d'enquête** relatifs à ces ilots (cartes et fiches descriptives des bâtiments et questionnaires vierges en nombre suffisant), afin de pouvoir démarrer le travail de collecte dès le Samedi 11 Avril. Enfin, on procédera avec lui à une **expérience de collecte sur quelques logements**, pour que les concepts et définitions utilisés pour l'enquête soient bien assimilés, et de façon homogène, par tous les enquêteurs.

La liste d'ilots de chaque enquêteur est établie en fonction des principes suivants :

- les ilots de chaque enquêteur sont, dans la mesure du possible, regroupés géographiquement, afin de minimiser les temps de déplacement.
- un enquêteur peut enquêter environ 400 logements en deux semaines de travail à temps complet

Chaque enquêteur organise lui-même ses horaires de travail pour réaliser la collecte des informations sur le terrain, en fonction des contraintes d'horaires (de présence à leur domicile) propres à la population à enquêter.

Tous les deux jours, chaque enquêteur aura un **rendez-vous avec le superviseur** de l'enquête. Cette rencontre permettra à l'enquêteur d'exposer les problèmes rencontrés au cours de la collecte, afin de rechercher ensemble une solution à ces problèmes de collecte de l'information ; c'est également à cette occasion que chaque enquêteur remettra au superviseur les questionnaires déjà remplis.

Entre deux rendez-vous d'un enquêteur avec le superviseur, celui-ci contrôlera l'avancée du travail, ainsi que la complétude et la cohérence des questionnaires ; en cas d'incomplétude ou d'incohérence des questionnaires, il sera demandé à l'enquêteur, lors du rendez-vous suivant, de retourner sur le terrain pour corriger ou compléter l'information.

L'enquête se terminera par un **rendez-vous final** dans les bureaux de l'IGM le Lundi 27 Avril ; tous les enquêteurs devront remettre lors de cette réunion les derniers questionnaires remplis.

4. COLLECTE DE L'INFORMATION

4. 1. Principes de base

* La première opération consiste à **identifier, sur le terrain, les bâtiments dessinés sur le plan au 1/500 et décrits dans les fiches** de l'enquête morphologie : il s'agit de retrouver, sur le terrain, à quels bâtiments correspondent les bâtiments identifiés lors de l'enquête morphologie urbaine, et quelles sont leurs limites. Cette opération a pour but de réutiliser les mêmes numéros de bâtiments, afin de pouvoir mettre en correspondance les résultats de l'enquête morphologie et ceux de l'enquête bâti/population.

* Démarrer l'enquête par le bâtiment N°1, et poursuivre, une fois ce premier bâtiment enquêté, par le bâtiment N°2...etc, jusqu'au dernier bâtiment de l'îlot. Cette façon de procéder, **en suivant l'ordre des numéros de bâtiments** facilitera les travaux de contrôle et de saisie de l'information collectée.

* Pour les **bâtiments ne comptant pas de logement**, aucune fiche logement n'est remplie ; l'enquêteur note uniquement l'usage du bâtiment (garage individuel, tienda, industrie...etc) dans la colonne "Usage principal" de la fiche récapitulative, dans la ligne relative à ce bâtiment.

* **Tous les logements** faisant partie des bâtiments situés dans l'îlot doivent être enquêtés. Pour chaque logement, il est rempli autant de fiches que nécessite le nombre de personnes vivant dans le logement.

* Pour remplir les fiches logement, il n'est pas nécessaire que l'enquêteur rencontre chaque personne vivant dans le logement : les questions doivent être posées soit au **chef de ménage**, s'il est présent, soit à **sa femme ou à toute autre personne du ménage pouvant répondre aux questions** se rapportant aux différents membres du ménage.

* Après avoir enquêté un logement, et donc rempli les fiches logement correspondantes, l'enquêteur reporte quelques informations sur la **fiche récapitulative relative à la manzana** (Voir mode de remplissage de cette fiche page 21).

4. 2. Recension, délimitation et numérotation des logements

Définition du logement

On appelle "logement" tout local utilisé pour l'habitation : un local non destiné originellement à l'habitation, mais qui est habité au moment de l'enquête est assimilé à un logement.

Délimitation des logements

Plusieurs situations peuvent se rencontrer :

- Bâtiment à étages, divisé en appartements : chaque appartement (c'est à dire ensemble de pièces ayant une entrée commune, indépendante, qui occupe une partie d'un bâtiment) constitue un logement.

S'il existe des chambres de service avec des entrées indépendantes, directes, sur les parties communes de l'immeuble (couloirs), ces chambres sont considérées comme des logements indépendants.

Si une chambre de service (ou une pièce louée) n'a pas d'entrée indépendante de celle desservant l'appartement dont elles font partie, on considère qu'elle est incluse dans cet appartement, et ne constitue pas un logement indépendant.

- Villa ou maison : un bâtiment non divisé en appartements (unités indépendantes d'habitation) constitue un seul logement.

S'il existe des chambres de service avec des entrées indépendantes, permettant de rejoindre directement la rue sans passer par la maison, ces chambres sont considérées comme des logements indépendants.

Si une chambre de service (ou une pièce louée) n'a pas d'entrée indépendante de celle desservant la maison dont elle fait partie, on considère qu'elle est incluse dans cette maison, et ne constitue pas un logement indépendant.

Numérotation des logements

Chaque logement est repéré par trois numéros :

- Un numéro de manzana
- Un numéro de bâtiment, qui va de 1 à Π dans chaque manzana. C'est le numéro attribué au cours de l'enquête morphologie urbaine.
- Un numéro de logement. Pour chaque bâtiment de l'îlot, la numérotation des logements va de 1 à Π , Π étant le nombre de logements dans le bâtiment.

Ces trois numéros sont notés dans la partie supérieure de la fiche logement, et doivent être reportés sur chaque ligne de la fiche récapitulative.

Type de logement

Noter, en clair, le type du logement **individuel**, en utilisant les catégories suivantes:

Maison ou villa, appartement, pièce louée... etc.

S'il s'agit d'un logement **collectif**, indiquer en clair de quel type de logement collectif il s'agit, en utilisant les catégories suivantes :

Caserne, hôtel, hôpital, internat, couvent ...etc.

Condition d'occupation du logement

Quatre situations peuvent se produire :

- Logement en réparation ou en construction, donc inhabitable au moment de l'enquête
- Logement habitable, mais innocupé
- Logement habitable, occupé, dont on a pu enquêter les habitants
- Logement habitable, occupé, dont on n'a pas pu enquêter les habitants

Nombre de pièces utilisées pour dormir

Noter, dans les deux cases prévues à cet effet, le nombre de pièces généralement utilisées pour dormir, qu'elles soient ou non effectivement occupées au moment de l'enquête. Une pièce dont ce n'est pas la vocation première, mais qui est également utilisée pour dormir (exemple : un salon) doit être comptée dans le calcul du nombre total de pièces utilisées pour dormir au même titre qu'une chambre à coucher.

4. 3. Définition de la population à enquêter. Statut de résidence.

La population décrite dans les colonnes de la fiche logement comprend deux catégories, suivant le statut de résidence des personnes :

- **présents** : toutes les personnes ayant dormi dans le logement pendant au moins huit nuits, consécutives ou non, pendant le mois précédant le passage de l'enquêteur (qu'elles soient ou non présentes au moment du passage de l'enquêteur ou la nuit précédant son passage);

Si ces personnes résident la majeure partie de l'année dans le logement enquêté, on les appelle "**Résidents présents**".

Si ces personnes résident la majeure partie de l'année dans un autre logement que celui enquêté, on les appelle "**Non résidents présents**".

et - **Résidents absents** : toutes les personnes qui résident habituellement (c'est à dire la majeure partie de l'année) dans le logement, mais en sont absentes depuis plus d'un mois (tourisme, voyage d'affaires, travail temporaire ou saisonnier à l'extérieur de Quito), pour quelle que raison que ce soit (sauf un départ définitif du logement ou les cas particuliers cités ci-dessous) et qui doivent revenir vivre dans le logement dans un délai inférieur à 6 mois.

Cas particuliers

Ne font pas partie de la population à enquêter dans les logements individuels (car sont comptabilisés dans les logements collectifs) :

- Militaires, élèves internes,
- Emprisonnés ou hospitalisés pour une longue durée (supérieure à un mois)

Font partie de la population à enquêter :

- Emprisonnés ou hospitalisés pour une courte durée (inférieure à 1 mois)
- Travailleurs exerçant leur activité en dehors de Quito, vivant donc en dehors de la capitale une partie de l'année, mais revenant régulièrement habiter dans leur logement de Quito (chaque fin de semaine, chaque mois)
- Personnes venant d'emménager dans le logement pour y résider de façon définitive après un déménagement (transfert du mobilier...etc. dans le nouveau logement).

4. 3. Renseignements relatifs aux personnes

Le plus simple est de réaliser l'interview en **deux temps** :

1) d'abord, demander de donner une **liste des personnes** habitant dans le logement, et noter leur nom et leur statut de résidence.

Demander à la personne interrogée de commencer par le chef de ménage, et poursuivre par les autres personnes vivant avec lui dans le logement, en respectant l'**ordre général** suivant :

Chef de ménage, épouse, enfants célibataires, enfants mariés avec les membres de leur famille, autres parents, personnes non apparentées au chef de ménage.

Numéroter ces personnes de 1 à N dans chaque logement.

2) puis, questionner pour avoir les **différents renseignements sur chacune de ces personnes**, en redonnant un par un les noms des personnes de cette liste.

Les renseignements relatifs à chaque personne occupent une colonne **de la fiche logement**.

Numéro de la fiche logement

Quatre cases sont prévues dans la partie supérieure de la fiche logement pour indiquer son numéro : dans les deux cases de gauche, l'enquêteur indique le numéro d'ordre de la fiche pour le logement considéré ; dans les deux cases de droite, il indique le nombre total de fiches que comporte le logement.

Exemples

- 1ère fiche d'un logement ne comptant qu'une seule fiche :

0	1
---	---

 /

0	1
---	---

- 2ème fiche d'un logement comptant trois fiches :

0	2
---	---

 /

0	3
---	---

Nom et prénom

Ces renseignements sur l'identité des personnes enquêtées ne feront pas partie du fichier exploité sur ordinateur ; il faut donc insister auprès des enquêtés sur le fait que l'exploitation de l'enquête est faite de façon **anonyme**. Ces renseignements ne sont demandés que pour faciliter la collecte de l'information, afin d'éviter que des personnes du logement soient omises ou au contraire comptées deux fois, ainsi que pour faciliter la collecte de l'information sur les liens de parenté au sein du logement.

Quel lien de parenté (ou autre relation) avez-vous avec le chef de ménage?

S'il existe un lien de parenté avec le chef de ménage, noter, en clair, le lien de parenté existant entre la personne et le chef de ménage.

Exemples :

- Frère de 1, épouse de 1, neveu de 1... etc. (Rappel : le chef de ménage est toujours noté 1)
- Nièce de 1, employée domestique : pour rendre compte du cas où une personne de la famille du chef de ménage a également une relation de "service" avec le chef de ménage.

S'il n'existe pas de lien de parenté avec aucune personne du logement, noter en clair la nature de cet autre lien.

Exemple : employée domestique, nurse, gardien.

S'il n'existe pas de lien de parenté avec le chef de ménage, mais qu'il **existe un lien de parenté avec une autre personne du logement** (elle même non apparentée au chef de ménage), noter en clair ce lien, de la façon suivante :

Exemple : la personne N° 12 est la fille de l'employée domestique, qui porte le N° 11.

On note donc pour la personne 12 : fille de 11.

Si **l'on ne connaît pas le lien (de parenté ou autre)**, noter "INCONNU".

Sexe de la personne

Indiquer le sexe de la personne, en notant dans la case prévue à cet effet :

- 1, pour les hommes
- 2, pour les femmes
- 0 = inconnu

Quelle est votre date de naissance ou quel âge avez-vous?

Si la personne **connait précisément sa date de naissance**, la noter dans les six cases prévues à cet effet (ne rien noter dans les deux cases situées en dessous, pour l'âge):

Exemple : 6 Août 1938

1	0	6	0	8	3	8
---	---	---	---	---	---	---

Si la personne **ne connait pas précisément sa date de naissance**, noter son âge dans les deux cases prévues à cet effet (ne rien noter dans les six cases situées au dessus, pour la date de naissance). Il s'agit de l'âge atteint au dernier anniversaire, et non de l'âge qu'aura la personne à son prochain anniversaire. Ainsi, pour un enfant de moins d'un an, on note 00 :

Exemple : 23 ans

2	3
---	---

Dans le cas où l'âge de la personne est inconnu, noter, dans les deux cases prévues pour l'âge :

Exemple : date de naissance et âge inconnus

9	9
---	---

Où êtes-vous né?

Si la personne est née à **Quito** (stricto-sensu, non compris les villes de la vallée de Los Chillos, San Antonio, Cumbaya, Tumbaco), noter "Ici" sur la ligne inférieure. Ne rien indiquer sur les autres lignes.

Si la personne est née en **Equateur en dehors de Quito**, indiquer en clair, sur chacune des trois lignes prévues (ne rien indiquer sur la ligne "pays" :

- le nom de la province
- le nom de la paroisse
- le caractère rural ou urbain du lieu de naissance : urbain s'il s'agit du chef-lieu de la paroisse, rural s'il s'agit d'une autre localité que le chef-lieu de la paroisse.

Si la personne est née à l'étranger, indiquer seulement le pays de naissance.

Si le lieu de naissance est inconnu, noter "inconnu" sur la ligne supérieure, et mettre un trait pour les autres lignes.

Si le lieu de naissance n'est connu que de façon imprécise (par exemple, on connaît la province, mais pas la paroisse), indiquer l'élément de réponse qui est connu, et noter "inconnu" pour les autres éléments, non connus.

Exemples :

- Personne née à l'étranger :

Pays C.H.I.L.E.....
Province /
Paroisse /
Rural/Urbain /

- Personne née à Quito :

Pays /
Province /
Paroisse /
Rural/Urbain A.Q.U.I.....

- Personne née à Cuenca, dans la ville même :

Pays /
Province A.Z.U.A.Y.....
Paroisse C.U.E.N.C.A.....
Rural/Urbain Urbana.....

- Personne née dans la paroisse de Cuenca, en dehors de la ville de Cuenca :

Pays /
Province A.Z.U.A.Y.....
Paroisse C.U.E.N.C.A.....
Rural/Urbain Rural.....

- Personne dont on ignore totalement le lieu de naissance :

Pays D.E.S.C.....
Province /
Paroisse /
Rural/Urbain /

- Personne dont on ne connaît que le pays de naissance :

Pays /
Province P.I.C.H.I.N.C.H.A.....
Paroisse D.E.S.C.....
Rural/Urbain D.E.S.C.....

Depuis quelle date ou depuis combien d'années vivez-vous dans ce logement?

Avec cette question, il faut saisir la date d'arrivée ou la durée d'occupation dans le logement actuel :

- pour les **résidents présents ou absents** : il s'agit de la **date à laquelle la personne en a fait sa résidence principale**. Par résidence principale, on entend, si la personne habite dans plusieurs logements au cours de l'année, la résidence où elle habite le plus de temps au cours de l'année.
- pour les **présents** : il s'agit de la dernière date d'arrivée de ces personnes dans le logement.

Si la personne **a toujours vécu dans ce logement**, noter "TOUJOURS".

Si la personne **a habité autrefois dans ce logement**, en est partie pour résider ailleurs, et y est revenue s'installer, noter la date du début de la dernière installation.

Si la personne **connait sa date d'arrivée** dans le logement, la noter dans les 6 cases prévues à cet effet (ne rien noter dans les deux cases situées en dessous, pour la durée).

Si la personne **connait la durée écoulée depuis son arrivée**, noter la durée en années dans les deux cases prévues à cet effet (ne rien noter dans les six cases situées au dessus, pour la date d'arrivée). Il s'agit de la durée en années entières ; ainsi, si la personne occupe ce logement depuis moins d'un an, noter 00.

Dans le cas où **la date d'arrivée est totalement inconnue**, tout comme la durée, noter **99** dans les cases prévues pour noter la durée d'occupation du logement.

Exemples :

- Personne occupant le logement depuis le 3 Septembre 1983 :

1	0	5	0	9	8	3
---	---	---	---	---	---	---
- Personne occupant le logement depuis 1952 (jour et mois inconnus) :

9	9	9	9	5	2
---	---	---	---	---	---
- Personne occupant le logement depuis 38 ans :

3	8
---	---
- Personne dont on ignore depuis quelle date ou depuis combien de temps elle occupe le logement :

9	9
---	---

Depuis quelle date ou depuis combien d'années vivez-vous à Quito?

Le mode de remplissage de la fiche pour cette question est similaire à celui concernant la date d'arrivée dans le logement.

Avec cette question, il faut saisir la date d'arrivée ou la durée de séjour à Quito (stricto-sensu, non compris les villes de la vallée de Los Chillos, San Antonio, Cumbaya, Tumbaco) :

- pour les **résidents présents ou absents et pour les présents ayant leur résidence principale à Quito**, mais dans un autre logement : il s'agit de la **date à laquelle la personne a eu sa résidence principale à Quito**, qu'il s'agisse de ce logement ou non. Par résidence principale, on entend, si la personne habite alternativement à Quito et en dehors de Quito au cours de l'année, la résidence où elle habite le plus de temps au cours de l'année.
- pour les **présents n'ayant pas leur résidence principale à Quito** : il s'agit de la dernière date d'arrivée de ces personnes à Quito.

Si la personne a **toujours vécu à Quito**, noter "TOUJOURS".

Si la personne a **habité autrefois à Quito**, en est partie pour résider ailleurs, et y est revenue s'installer, noter la date du début de la dernière installation.

Si la personne **connait sa date d'arrivée** à Quito, la noter dans les 6 cases prévues à cet effet (ne rien noter dans les deux cases situées en dessous, pour la durée).

Si la personne **connait la durée écoulée depuis son arrivée**, noter la durée en années dans les deux cases prévues à cet effet (ne rien noter dans les deux cases situées au dessus, pour la date d'arrivée). Il s'agit de la durée en années entières ; ainsi, si la personne occupe ce logement depuis moins d'un an, noter 00.

Dans le cas où la **date d'arrivée est totalement inconnue**, tout comme la durée, noter 99 dans les cases prévues pour noter la durée d'occupation du logement.

Où habitiez-vous avant de résider à Quito?

Pour les résidents et les présents ayant leur lieu de résidence principal actuel à Quito, il s'agit du lieu de résidence principal de la personne juste avant qu'elle n'ait son lieu de résidence principal à Quito.

Pour les présents ayant leur lieu de résidence principal actuel en dehors de Quito, il s'agit du lieu de résidence principal actuel.

Si la personne a toujours eu sa résidence principale à Quito, noter "Ici" sur la ligne inférieure. Ne rien indiquer sur les autres lignes.

Si la personne résidait auparavant en Equateur en dehors de Quito, indiquer en clair, sur chacune des trois lignes prévues (ne rien indiquer sur la ligne "pays") :

- le nom de la province
- le nom de la paroisse
- le caractère rural ou urbain du lieu de naissance : urbain s'il s'agit du chef-lieu de la paroisse, rural s'il s'agit d'une autre localité que le chef-lieu de la paroisse.

Si la personne résidait auparavant à l'étranger, indiquer seulement le pays de naissance.

Si le lieu de résidence antérieur est inconnu, noter "inconnu" sur la ligne supérieure, et mettre un trait pour les autres lignes.

Si le lieu de résidence antérieur n'est connu que de façon imprécise (par exemple, on connaît la province, mais pas la paroisse), indiquer l'élément de réponse qui est connu, et noter "inconnu" pour les autres éléments, non connus.

Exemples :

Voir question sur le lieu de naissance

De quel type de propriétaire s'agit-il?

Indiquer, en clair, dans la colonne du chef de ménage, si le propriétaire est :

- Une personne physique (naturelle)
- Une société privée
- Une organisme semi-public
- Un organisme public
- Un organisme municipal
- Un organisme religieux
- Un organisme diplomatique
- Un organisme de bienfaisance
- Une organisation syndicale
- Une coopérative ...etc.

Si la réponse est inconnue, noter "INCONNU".

Quel lien de parenté ou autre relation avez-vous avec le propriétaire?

Cette question ne concerne que les chefs de ménage, et seulement dans les cas où le propriétaire est une personne physique et ne réside pas dans le logement (que ce soit le chef de ménage ou une autre personne qui soit propriétaire).

S'il existe un lien de parenté entre le chef de ménage et le propriétaire

noter, en clair, ce lien de parenté, dans la colonne Chef de ménage :

Exemples :

- Frère de P, petit-fils de P, neveu de P...etc. (Le propriétaire est toujours noté P)

Si le chef de ménage n'a pas de lien de parenté avec le propriétaire du logement, noter sans lien familial (abrégié SLP)

Si le chef de ménage n'a pas de lien de parenté avec le propriétaire, mais qu'il existe un lien de parenté entre une autre personne du logement et le propriétaire (elle même non apparentée au chef de ménage), noter en clair ce lien, dans la colonne relative à cette personne, de la façon suivante :

Exemple : la personne N° 12 est la fille du propriétaire.

On note donc pour la personne 12 : fille de P.

Activité

Cette question concerne chaque personne vivant dans le logement, **sans condition d'âge minimum**.

Il s'agit de la situation au mois d'Avril 1987.

Profession

Plusieurs cas peuvent se produire.

Si la personne ne travaille pas, pour décrire les différentes situations, utiliser les termes suivants :

- | | |
|---|---|
| - Etudiant | - Travaux ménagers |
| - Retraité | |
| - Ne travaille pas, recherche un emploi | - Ne travaille pas, ne recherche pas d'emploi |

Si la personne travaille, indiquer sa profession en détail :

Exemples

- | | |
|-------------------------|--------------------------------|
| - Professeur de musique | - Professeur de l'Université |
| - Menuisier | - Vendeur d'assurances |
| - Chauffeur de camions | - Mécanicien automobile...etc. |

Une personne peut avoir **plusieurs professions pendant l'année**. Noter **toutes** ses professions, et **souligner la profession principale** (c'est à dire celle qui lui rapporte le plus de revenus de l'année).

Donner un numéro à chacune de ces professions. Ces numéros seront utilisés pour indiquer la branche d'activité et le statut professionnel pour chaque profession de la personne.

Branche d'activité

Pour obtenir une réponse correcte, demander :

- "Quel est l'activité principale de l'établissement où vous travaillez?"

L'enquêteur doit indiquer exactement le type d'activité de l'établissement. Ne pas accepter une réponse vague, et insister pour avoir plus de détails.

Si une personne a plusieurs professions, utiliser les numéros des professions, pour décrire la branche d'activité de chaque profession.

Exemples

- | | |
|----------------------------|---------------------------|
| - Education cycle primaire | - Consultations médicales |
|----------------------------|---------------------------|

- Usine de textile
- Boulangerie

- Menuiserie
- Ministère de l'Agriculture...etc.

Statut professionnel

Indiquer la situation dans la profession (définitions du recensement de l'INEC) :

- **Patron** : personne qui dirige sa propre entreprise, ou qui exerce une profession pour son propre compte, et qui a un ou plusieurs employés rémunérés (ex : directeur d'une usine).

- **Travailleur indépendant** : personne qui dirige sa propre entreprise, ou qui exerce pour son propre compte un travail, mais qui n'a aucun employé rémunéré (ex : vendeur ambulant).

- **Employé ou salarié** : employé qui travaille pour l'Etat ou dans une entreprise privée, et qui reçoit une rémunération, une solde, un salaire, une commission ou une somme forfaitaire.

- **Travailleur familial sans rémunération** : personne exerçant une activité non rémunérée dans une entreprise ou un commerce d'un membre de sa famille, la majeure partie de la semaine.

- **Apprenti** : personne travaillant dans une entreprise pour apprendre un métier, ne recevant pas de rémunération, ou une rémunération minime.

- **Autres** (préciser)

Si une personne a plusieurs emplois, utiliser les numéros de ses professions, pour décrire le statut professionnel dans chacun de ces emplois.

Dans la mesure du possible, noter le **montant** des revenus annuels ou mensuels (préciser la période de référence), procurés par chaque emploi.

Autres sources de revenus

Noter si la personne a une autre source de revenus que les emplois décrits plus haut : indiquer la source de revenus et son montant annuel approximatif.

Exemples : mise en location d'un appartement (250.000 sucres/mois). propriétaire d'un taxi (X sucres/mois)...etc

5. FICHE RECAPITULATIVE

L'enquêteur remplit une fiche récapitulative par manzana (ou plusieurs fiches récapitulatives si le nombre de logements dans la manzana est supérieur à 21). Sur ces fiches, sont inscrites des informations globales concernant la manzana.

Chaque ligne de la fiche correspond à un logement.

Nombre de bâtiments, de fiches récapitulatives, de fiches logement

Dans ces trois cas, il s'agit du nombre total dans la manzana.

Ordre des logements dans la fiche

Commencer par le bâtiment n°1, logement n°1.

Puis continuer en prenant les bâtiments par ordre de numéro, et, dans chaque bâtiment, les logements par ordre de numéro : (Bâtiment 1, Logement 1), puis (Bâtiment 1, Logement 2), puis (Bâtiment 2, Logement 2).....etc.

Nombre de personnes dans le logement

Il s'agit du nombre de total de personnes du logement, réparties selon leur statut de résidence :

- Nombre de Résidents Présents
- Nombre de Résidents Absents
- Nombre de Non Résidents Présents

Usage du bâtiment

Noter en clair quel est l'usage principal du bâtiment en utilisant de préférence les catégories suivantes :

- | | |
|---|-----------------------|
| - Habitation | - Service domestique |
| - Commerce (préciser le type de commerce) | - Bureaux privés |
| - Entrepôt | - Education |
| - Industrie | - Culture, spectacles |
| - Artisanat (préciser garage, menuiserie...etc) | - Sport |
| - Bureaux publics | - Santé |

Si un bâtiment a plusieurs usages, noter ces différents usages, en soulignant l'usage principal.

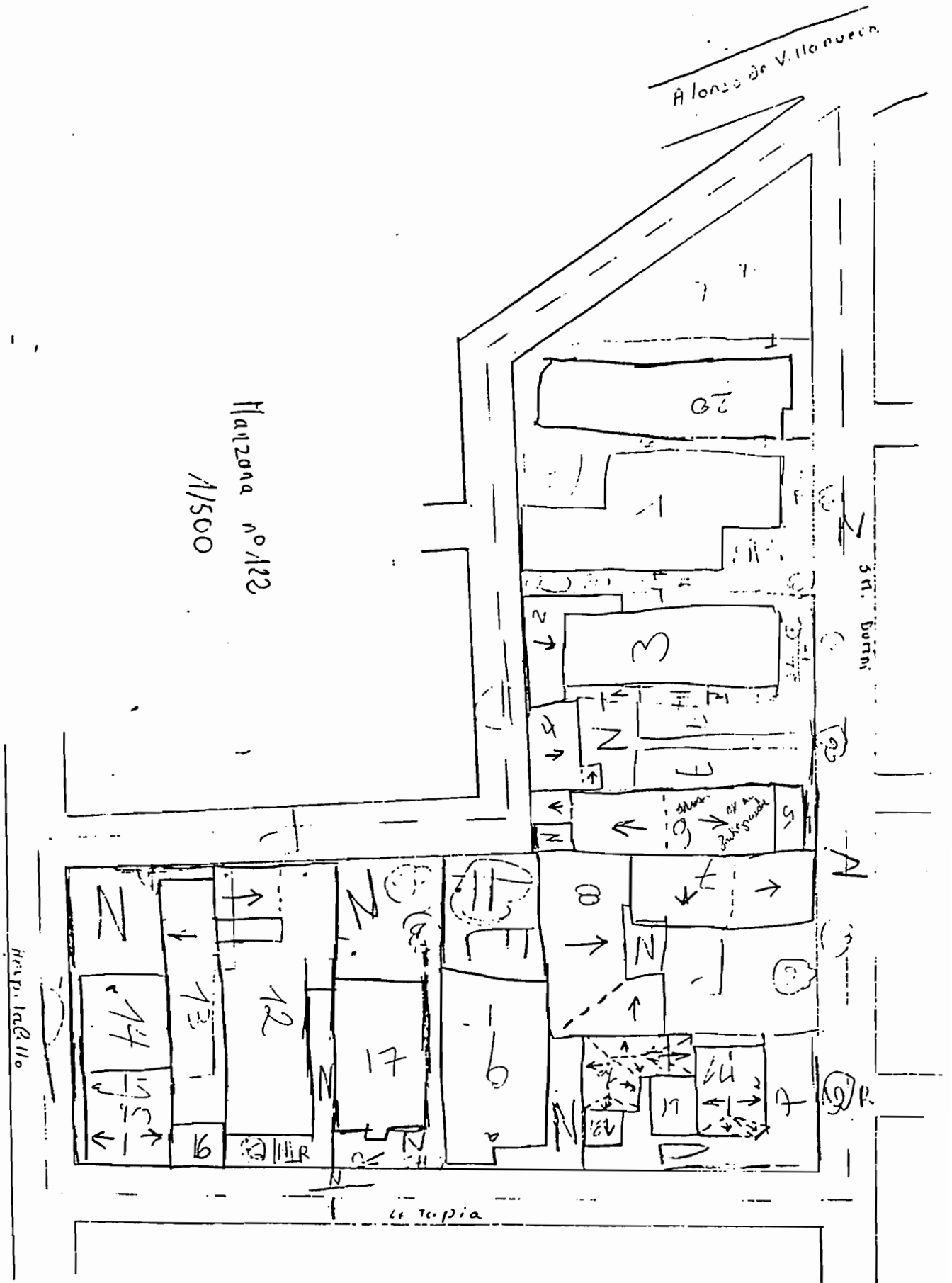
Si un bâtiment ne sert qu'à l'habitation, indiquer "Seul. Habitation".

Observations

Indiquer dans cette case :

- les problèmes éventuels rencontrés au cours de l'enquête (refus de réponse, absence prolongée des habitants...etc),
- des remarques sur l'état du bâtiment, si cet état explique la non-utilisation du bâtiment (donc, pas d'habitants, ou d'usage du bâtiment),
- toute autre observation pour décrire une situation non prévue dans les questionnaires.

6. EXEMPLE DE MANZANA ET DE LOGEMENT



ENQUETE BATI/POPULATION - FICHE LOGEMENT

N° Manzana Nom Enquêteur N° Bâtiment N° Logement N° de la fiche logement /

Type de logement..... Condition d'occup. du log. :
 Inhabitable
 Habit. innoc.
 Habit. oc. non enq.
 Habit oc. enq.
 Nombre de pièces util. pour dormir

N° page	RUBRIQUES	N° personne <input type="text"/> <input type="text"/> <input type="text"/>	N° personne <input type="text"/> <input type="text"/> <input type="text"/>	N° personne <input type="text"/> <input type="text"/> <input type="text"/>	N° personne <input type="text"/> <input type="text"/> <input type="text"/>	N° personne <input type="text"/> <input type="text"/> <input type="text"/>	N° personne <input type="text"/> <input type="text"/> <input type="text"/>
11	Nom et prénom						
10	Statut de résidence	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>	Résid. présent <input type="checkbox"/> Résid. absent <input type="checkbox"/> Non res. prés. <input type="checkbox"/> <input type="checkbox"/>
11	Lien de parenté avec le chef de ménage <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
11	Sexe	Homme =1 Femme =2 <input type="checkbox"/>	Homme =1 Femme =2 <input type="checkbox"/>	Homme =1 Femme =2 <input type="checkbox"/>	Homme =1 Femme =2 <input type="checkbox"/>	Homme =1 Femme =2 <input type="checkbox"/>	Homme =1 Femme =2 <input type="checkbox"/>
12	Date de naissance ou age	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
12	Lieu de naissance : Pays Province Paroisse Rural/ Urbain <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
14	Date d'entrée dans le logement ou durée d'occupation du logement	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
15	Date d'arrivée à Quito ou durée de séjour à Quito	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
16	Lieu de résidence antérieur : Pays Province Paroisse Rural/ Urbain <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

N° pa.	RUBRIQUES	N° personne <input type="text"/>	N° personne <input type="text"/>	N° personne <input type="text"/>	N° personne <input type="text"/>	N° personne <input type="text"/>	N° personne <input type="text"/>
17	Statut d'occupation du logement <i>(préciser, selon le cas, le montant du loyer, ou de l'antichresis)</i> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
17	Si propriétaire, coût construction achat, héritage ou donation <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
18	Type de propriétaire <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
18	Lien de parenté (ou autre) avec le propriétaire du logement <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
19	Profession Activité : Branche d'activité Statut professionnel <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
20	Autres sources de revenu (agricole, foncier...etc) <i>(indiquer la source et le montant annuel approximatif de chaque source de revenu)</i> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

QUITO-Morfología -Oct./Dec1986

Número de manzana : 122		DESCRIPCION DE LOS EDIFICIOS					
Fecha: Hora:							
Número del edificio que corresponde a la señal localizada en el plano		1 ₈	2 ₂	3 ₉	4 ₂	5 ₈	
DESCRIPCIONES	Edificio: Acabado	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	En construccion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Número de pisos (incluido P.B.)	2	1	2	1	2	
	Dimensiones en el suelo: Lado a	15	15	10	15	12	
		Lado b	20	7	14	7	20
	(Indicar este(os) lado(s) en el plano)						
	Número de pendientes del techo (0 si plano)	0	1	0	1	0	
	Material predominante en el techo						
	Losa de hormigón	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
	Eternit o similares	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Zinc o similares	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
Teja con barniz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Teja sin barniz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Paja o similares	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Otros (chova, embaldosado, techoline...)	---	---	---	---	---		
Color predominante en el techo							
(Blanco, Café oscuro, Café claro, Gris, color ladrillo, Negro, Rojo, o Verde)	gris	gris	gris	gris	gris		
Forma del material del techo	—	~	—	~	—		
(Hacer un croquis de la sección)							
Otros elementos sobre el techo							
Número de chimeneas	---	---	---	---	---		
Presencia de pilares de hormigón	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Presencia de penthouse (a describir)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
Número de tanques de reserva	---	---	2	---	---		
Presencia de vegetación :							
Sobre balcones sobresalientes	---	---	---	---	---		
Sobre el techo	---	---	---	---	---		
Otras observaciones sobre el edificio	---	---	---	---	---		

TECHO

OTROS

4) Enquête Bâti / population : codification

NOMBRE DEL ENCUESTADOR

1	Ramiro VASQUEZ	9	DESCONOCIDO
2	Marcelo AREVALO		
3	Henri SANCHEZ ALBAN		
4	Cesar ALVAREZ MONTUFAR		
5	Katia VALENCIA		
6	Jorge ROJAS ALMEIDA		
7	O. BARBARY et F. DUREAU		

ESTATUTO DE RESIDENCIA DE LOS HABITANTES

1	Residente presente	9	DESCONOCIDO
2	Residente ausente		
3	No residente presente		

SEXO

1	Hombre	9	DESCONOCIDO
2	Mujer		

LAZO DE PARENTESCO (O OTRA RELACION) CON EL JEFE DEL HOGAR

PERSOAS QUE TIENEN SOLO UN LAZO DE PARENTESCO CON EL JEFE DEL HOGAR

- | | |
|----------------------------------|------------------------------------|
| 01 Jefe del hogar | 07 Primo(a) del jefe del hogar |
| 02 Conyuge del jefe del hogar | 08 Cunado(a) del jefe del hogar |
| 03 Hijo(a) del jefe del hogar | 09 Suegros del jefe del hogar |
| 04 Hermano(a) del jefe del hogar | 10 Yerno(nuera) del jefe del hogar |
| 05 Padres del jefe del hogar | 11 Nieto(a) del jefe del hogar |
| 06 Tio(a) del jefe del hogar | 12 Sobrino(a) del jefe del hogar |
| | 13 Otros parientes |

PERSOAS QUE TIENEN UN LAZO DE PARENTESCO Y UNA RELACION DE SERVICIO CON EL JEFE DEL HOGAR

- | | |
|----------------------------------|------------------------------------|
| | 27 Primo(a) del jefe del hogar |
| 22 Conyuge del jefe del hogar | 28 Cunado(a) del jefe del hogar |
| 23 Hijo(a) del jefe del hogar | 29 Suegros del jefe del hogar |
| 24 Hermano(a) del jefe del hogar | 30 Yerno(nuera) del jefe del hogar |
| 25 Padres del jefe del hogar | 31 Nieto(a) del jefe del hogar |
| 26 Tio(a) del jefe del hogar | 32 Sobrino(a) del jefe del hogar |
| | 33 Otros parientes |

PERSOAS QUE NO TIENEN UN LAZO DE PARENTESCO CON EL JEFE DEL HOGAR, PERO TIENEN UN LAZO DE PARENTESCO CON OTRA PERSONA DE LA VIVIENDA (ella misma no emparentada con el jefe del hogar)

- | | |
|---------------|----------------------------------|
| | 47 Primo(a) |
| 42 Conyuge | 48 Cunado(a) |
| 43 Hijo(a) | 49 Suegros |
| 44 Hermano(a) | 50 Yerno(nuera) |
| 45 Padres | 51 Nieto(a) |
| 46 Tio(a) | 52 Sobrino(a) del jefe del hogar |
| | 53 Otros parientes |

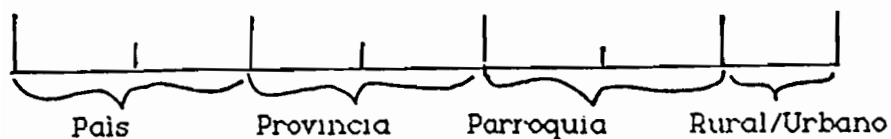
PERSOAS QUE NO TIENEN LAZO DE PARENTESCO CON NINGUNA PERSONA DE LA VIVIENDA

- | | |
|--|---------------------------------|
| 60 Amigo(a) | 64 Empleado(a) domestico |
| 61 Colega de trabajo | 65 Guardián |
| 62 Nurse | 66 Otras relaciones de servicio |
| 63 Jardinero | 67 Otro tipo de relacion |
| 69 Tipo de relación desconocido (salvo lazo de parentesco) | |

99- DESCONOCIDO

CODIGO GEOGRAFICO

El lugar de nacimiento y el lugar de residencia anterior estan descritos de la manera siguiente :



Caso de un pais extranjero

Anotar el codigo del pais.

Provincia = 00

Parroquia = 00

Rural/urbano = 0

Caso de un lugar totalmente desconocido

Provincia = 99

Parroquia = 99

Rural/urbano = 9

Caso de un lugar parcialmente desconocido

Utilizar los codigos para las informaciones conocidas.

Para las informaciones desconocidas, anotar 99 o 9.

Rural/urbano

- 1 Rural (si se trata de una localidad que no sea la cabecera parroquial)
- 2 Urbano (si se trata de la cabecera parroquial)
- 3 DESCONOCIDO

CODIGOS DE PAISES

AMERICA

- 01 Ecuador
- 11 Argentina
- 12 Bolivia
- 13 Brasil
- 14 Colombia
- 15 Costa Rica
- 16 Cuba
- 17 Chile
- 18 Guatemala
- 19 Haiti
- 20 Honduras
- 21 Jamaica
- 22 Malvinas
- 23 México
- 24 Nicaragua
- 25 Panamá
- 26 Paraguay
- 27 Perú
- 28 Puerto Rico
- 29 Republica Dominicana
- 30 El Salvador
- 31 Trinidad Tobago
- 32 Uruguay
- 33 Venezuela
- 34 Curazao
- 35 Otros paises de America
del Sur o Central
- 40 Estados Unidos
- 41 Canada
- 42 Otros paises de America del Norte

EUROPA

- 51 Alemania
- 52 Espana
- 53 Francia
- 54 Gran Bretana
- 55 Union Soviética
- 58 Otros paises de Europa

ASIA

- 61 Corea
- 62 China
- 63 India
- 64 Japón
- 65 Otros paises de Asia

70 AFRICA

80 OCEANIA

99 DESCONOCIDO

CODIGOS DE PROVINCIAS DEL ECUADOR

- 01 Azuay
- 02 Bolivar
- 03 Canar
- 04 Carchi
- 05 Chimborazo
- 06 Cotopaxi
- 07 El oro
- 08 Esmeraldas
- 09 Galapagos
- 10 Guayas
- 11 Imbabura
- 12 Loja
- 13 Los rios
- 14 Manabi
- 15 Morona Santiago
- 16 Napo
- 17 Pastaza
- 18 Pichincha
- 19 Tungurahua
- 20 Zamora Chinchipe

- 99 DESCONOCIDO (en Ecuador)

- 00 Si se trata de un pais extranjero

CODIGOS DE PARROQUIAS

Codigo INEC de las parroquias (Censo de 1982)

99 DESCONOCIDO

00 Si se trata de un país extranjero

CALIDAD DE OCUPACION DE LA VIVIENDA

(solo en la columna del jefe del hogar)

- 01 Proprietario (jefe del hogar)
- 02 Proprietario (otra persona que vive en la vivienda)
- 03 Inquilino
- 04 Ocupación de favor (préstamo de un amigo, pariente, colega...etc)
- 05 Por servicios (portero, guardián, jardinero...etc del edificio)
- 06 Anticresis
- 07 Ocupación ilegal de un local teóricamente abandonado
- 08 Otro tipo de ocupacion

- 99 DESCONOCIDO

MONTO DEL ALQUILER MENSUAL O DEL ANTICRESIS

(solo en la columna del jefe del hogar)


En el caso 03 (Inquilino), indicar el monto del alquiler mensual de la manera siguiente (la unidad es de 1000 sucres):

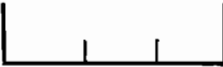
monto < 1000 sucres	000
1000 - 1999 sucres	001
2000 - 2999 sucres	002
.	
.	
10.000 - 10999 sucres	010 ...etc
DESCONOCIDO	999

En el caso 06 (anticresis), indicar el monto mensual del anticresis de la misma manera, y el monto del anticresis mismo en tres casilleros situados encima.

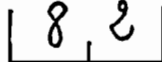
Si se trata de otra calidad de ocupacion de la vivienda (por ejemplo, propietario o por servicios), anotar 998

COSTO DE LA CONSTRUCCION O DEL COMPRA DE LA VIVIENDA


Año de la construcción
o del compra


Costo de la construcción
o del compra

Año de la construcción o del compra (o de la herencia o donación)

Ejemplo : año 1982 

Si la fecha esta desconocida : 99

Costo de la construcción o de la vivienda

Indicar el costo de la manera siguiente (la unidad es de 100.000 sucres) :

costo < 100.000 sucres	000
100.000 - 199.999 sucres	001
200.000 - 299.999 sucres	002
.	.
1.000.000 - 1.099.999 sucres	010
1.100.000 - 1.199.999 sucres	011etc
Donación	996
Herencia	997
Proprietario no vive en la vivienda	998 (Y fecha de compra = 98)
DESCONOCIDO	999

CLASE DE PROPIETARIO

- | | |
|--|------------------------------|
| 00 Persona que vive en la vivienda (jefe del hogar o otra persona) | |
| 01 Persona física (natural), que no vive en la vivienda | |
| 02 Sociedad privada | |
| 03 Organismo semi-público | 08 Organización sindical |
| 04 Organismo público | 09 Cooperativa |
| 05 Organismo municipal | 10 Organismo religioso |
| 06 Organismo diplomático | 11 Otra clase de propietario |
| 07 Organización de beneficencia | 99 DESCONOCIDO |

LAZO DE PARENTESCO CON EL PROPIETARIO

Esta pregunta concierne solamente a los jefes del hogar. Para las otras personas, utilizar el código 98 (salvo el caso si la persona no tiene parentesco con el jefe del hogar, pero tiene una relación de parentesco con el propietario).

- 00 El jefe del hogar no tiene lazo de parentesco con el propietario (que está una persona física)
- 01 El propietario es el jefe del hogar
- 02 El propietario no es una persona física

EL JEFE DEL HOGAR TIENE UN LAZO DE PARENTESCO CON EL PROPIETARIO

Indicar, en los casilleros del jefe del hogar :

- | | |
|-------------------------------|---------------------------------|
| 03 Primo(a) del propietario | |
| 04 Conyuge del propietario | 09 Cunado(a) del propietario |
| 05 Hijo(a) del propietario | 10 Suegros del propietario |
| 06 Hermano(a) del propietario | 11 Yerno(nuera) del propietario |
| 07 Padres del propietario | 12 Nieto(a) del propietario |
| 08 Tío(a) del propietario | 13 Sobrino(a) del propietario |
| | 14 Otros parientes |

UN OTRA PERSONA DE LA VIVIENDA TIENE UN LAZO DE PARENTESCO CON EL PROPIETARIO

Indicar en los casilleros del jefe del hogar : 00

Indicar en los casilleros de la persona que vive en la vivienda y que tiene un lazo de parentesco con el propietario :

- | | |
|-------------------------------|---------------------------------|
| 15 Primo(a) del propietario | |
| 16 Conyuge del propietario | 21 Cunado(a) del propietario |
| 17 Hijo(a) del propietario | 22 Suegros del propietario |
| 18 Hermano(a) del propietario | 23 Yerno(nuera) del propietario |
| 19 Padres del propietario | 24 Nieto(a) del propietario |
| 20 Tío(a) del propietario | 25 Sobrino(a) del propietario |
| | 26 Otros parientes |

99 DESCONOCIDO

ACTIVIDAD

La actividad está descrita de la manera siguiente :



Si una persona tiene **varios ocupaciones**, anotar los códigos de la primera ocupación en los casilleros. Anotar los códigos de la segunda ocupación por encima de estos casilleros.

Si una persona **no trabaja**, indicar el código de su tipo de ocupación. Para la rama de actividad y la categoría de ocupación, escribir 000 y 0.

OCUPACION

Si la persona **trabaja**, utilizar los códigos del INEC (Censo 1982).

Si la persona **no trabaja**, utilizar los códigos siguientes :

001 Estudiante

002 Jubilado

003 Quehaceres domésticos

004 Pensionista

005 No trabaja, busca un trabajo

006 No trabaja, no busca un trabajo (los pequeños niños están en este caso)

Si la ocupación está **DESCONOCIDA** : 999

- 27 Industrias metálicas básicas (de hierro, acero, metales no ferrosos)
- 28 Fabricación de productos metálicos, maquinaria y equipo, material de transporte
- 29 Otras industrias manufactureras (joyas, instrumentos de música, artículos de deporte ...etc)

Rama 3 : Electricidad, gas y agua

- 31 Generación, transmisión y distribución de electricidad
- 32 Producción y distribución de gas
- 33 Suministro de vapor y agua caliente
- 34 Obras hidráulicas y suministro de agua (captación, purificación y distribución)
- 39 Otras actividades de la rama 3

Rama 4 : Construcción y materiales de construcción

- 41 Fabricación de materiales de construcción a base de arcilla
- 42 Fabricación de materiales de construcción a base de cemento (cemento, bloques...etc)
- 43 Fabricación de otros materiales de construcción
- 44 Obras públicas
- 45 Paredes maestras
- 46 Tubería, electricidad
- 47 Pintura, vidrios, embaldosado, decoración
- 48 Carpintería de obra
- 49 Otras actividades de la rama 4

Rama 5 : Comercio por mayor y por menor

- 51 Comercios ambulantes
- 52 Comercios generales fijos
- 53 Comercios especializados fijos (librería, ferretería, papelería, farmacia, venta de muebles...etc)
- 54 Comercios de carburante, combustible y materiales de construcción
- 55 Tienda
- 56 Comercios de alimentación
- 57 Restaurantes, bar y hoteles
- 58 Import-export
- 59 Otro comercio

Rama 6 : Transportes, almacenamiento y comunicaciones

- 60 Taxímetros
- 61 Transporte urbano colectivo de pasajeros

- 62 Transporte terrestre interurbano de pasajeros por carretera
- 63 Transporte terrestre de carga por carretera
- 64 Transporte ferroviario
- 65 Transporte aéreo
- 66 Transporte por agua (marítimo, por ríos, ...etc)
- 67 Depósito y almacenamiento (como servicio independiente)
- 68 Comunicaciones (correo, telégrafo, teléfono, radio...etc)
- 69 Otros servicios relacionados con el transporte o las comunicaciones

Rama 7 : Establecimientos financieros, seguros, bienes inmuebles y servicios prestados a las empresas

- 71 Establecimientos financieros (bancos, establecimientos de cambio...etc)
- 72 Seguros
- 73 Arrendamiento y explotación de bienes inmuebles
- 74 Servicios prestados a las empresas (jurídicos, contabilidad, publicidad, asesoría técnica...etc)
- 79 Otros establecimientos financieros y otros servicios prestados a las empresas

Rama 8 : Servicios comunales, sociales y personales

- 80 Defensa
- 81 Administración pública
- 82 Instrucción pública (escuelas primarias y secundarias)
- 83 Instrucción pública (universidades, escuelas profesionales) y investigaciones
- 84 Servicios médicos, odontológicos y otros servicios de sanidad
- 85 Servicios de diversión y esparcimiento y servicios culturales (cine, teatro, biblioteca, museo, jardín botánico o zoológico, sala de baile, estadio...etc)
- 86 Servicios de reparación (de cuero, eléctricas, de automóviles, de relojes...etc)
- 87 Servicios domésticos para el mantenimiento de los hogares (cocineras, nineras, jardineros...etc)
- 88 Servicios personales directos (peluquería, fotografía, lavandería ...etc)
- 89 Otros servicios comunales, sociales y personales

Rama 9 : Actividades no bien especificadas

- 91 Persona que no trabaja (ejemplo : estudiante, quehaceres domésticos) y no tiene otras fuentes de ingresos (alquiler, pensión...etc)
- 92 Persona que no trabaja y vive de otras fuentes de ingresos (alquiler, pensión...etc) *Para los jubilados, no utilizar este código 92 : en este caso, indicar la rama de actividad de su ex-trabajo.*
- 99 DESCONOCIDO (rama de actividad desconocida para una persona que trabaja o que está jubilado)

CATEGORIA DE OCUPACION

- 1 **Patrono** (o socio activo) : persona que dirige su propia empresa o negocio, o que ejerce una profesión u oficio y que tiene uno o más empleados remunerados (ej: el dueño de una fábrica).
- 2 **Trabajador por cuenta propia** : si la persona explota su empresa o negocio, o ejerce por cuenta propia una profesión u oficio, pero no tiene ningún empleado remunerado (ej: vendedor ambulante).
- 3 **Empleado o asalariado del sector publico** : cuando la persona es un funcionario, empleado u obrero del Estado, que recibe una remuneración, sueldo, salario o pago a destajo, bajo nombramiento, contrato o jornal.
- 4 **Empleado o asalariado del sector privado** : cuando la persona es un empleado u obrero de una empresa particular, que recibe una remuneración, sueldo, salario o pago a destajo, bajo nombramiento, contrato o jornal.
- 5 **Empleado o asalariado del sector semi-publico** : cuando la persona es un funcionario, empleado u obrero de una empresa del sector semi-publico, que recibe una remuneración, sueldo, salario o pago a destajo, bajo nombramiento, contrato o jornal.
- 6 **Trabajador familiar sin remuneración** : la persona realiza una ocupación no remunerada en una empresa o negocio de un familiar, en la mayor parte de la semana.
- 7 **Aprendiz** : se trata de una persona que trabaja sin remuneración (o mínima) en una empresa para aprender una práctica profesional.
- 8 **Otros**
- 9 **DESCONOCIDO**
- 0 **Si la persona no trabaja**

FUENTES Y MONTO DE LOS INGRESOS

Fuentes de ingresos

- 01 Sin fuente de ingresos
- 02 Sólo actividad
- 11 Actividad y fuente de ingresos ligada a los bienes inmuebles (ej. : alquiler)
- 12 Actividad y fuente de ingresos ligada a un comercio (ej. : propietario de una tienda)
- 13 Actividad y fuente de ingresos ligada a los transportes (ej. : propietario de un taxi)
- 14 Actividad y pensión
- 15 Actividad y jubilación
- 16 Actividad y otro tipo de fuente de ingresos
- 21 Sólo fuente de ingresos ligada a los bienes inmuebles (ej. : alquiler)
- 22 Sólo fuente de ingresos ligada a un comercio (ej. : propietario de una tienda)
- 23 Sólo fuente de ingresos ligada a los transportes (ej. : propietario de un taxi)
- 24 Sólo pensión
- 25 Sólo jubilación
- 26 Sólo pensión y jubilación
- 27 Sólo otro tipo de fuente de ingresos
- 28 Jubilación y fuente de ingresos ligada a los bienes inmuebles

- 97 Otras fuentes de ingresos conocidas, actividad desconocida
- 98 Actividad conocida, otras fuentes de ingresos desconocidas
- 99 Actividad y otras fuentes de ingresos desconocidos

Monto de los ingresos

Se trata del monto total de los ingresos por mes : ingresos ligados a la(s) actividad(des), y ingresos de otras fuentes (ejemplo : alquiler de un departamento, propietario de un taxímetro...etc)

Indicar el monto de la manera siguiente (la unidad es de 1000 sucres por mes) :

monto < 1000 sucres	000
1.000 - 1.999 sucres	001
2.000 - 2.999 sucres	002

10.000 - 10.999 sucres	010
11.000 - 11.999 sucres	011etc

997 Monto de las otras fuentes de ingresos conocido, ingresos ligados a la actividad desconocidos

998 Monto de los ingresos ligados a la actividad conocido, otras fuentes de ingresos desconocidas

999 DESCONOCIDO (ingresos ligados a la actividad y a las otras fuentes de ingresos desconocidos)

DATOS RECAPITULATIVOS SOBRE LAS VIVIENDAS

Véase el ejemplo de ficha de codificación en la página 19.

Numeración de los edificios

Nuevo número : se trata del número en la encuesta habitat/poblacion

Viejo número : se trata del número de la encuesta Morfología urbana

Nuevo edificio : número en la encuesta Morfología = 98

99 DESCONOCIDO

Numeración de las viviendas

Número anotado en las fichas de encuesta.

96 Edificio no encuestado. Número de viviendas en el edificio desconocido.

Tipo de vivienda

- 1 Casa o villa
- 2 Departamento
- 3 Cuarto
- 4 Vivienda colectiva
- 6 Edificio que no existe en el momento de la encuesta habitat/poblacion (pero que existaba en el momento de la encuesta morfología urbana)
- 8 Si no se trata de una vivienda
- 9 DESCONOCIDO

Condición de ocupación de la vivienda

- 1 Si no se trata de una vivienda
- 2 Vivienda inhabitable en el momento de la encuesta
- 3 Vivienda habitable, ocupada en el momento de la encuesta
- 4 Vivienda habitable, desocupada en el momento de la encuesta
- 5 Vivienda en construcción
- 6 Edificio que no existe en el momento de la encuesta habitat/poblacion (pero que existaba en el momento de la encuesta morfología urbana)
- 9 DESCONOCIDO

Numero de piezas utilizadas para dormir

Se trata del número de piezas anotado en la ficha de la encuesta.

99 DESCONOCIDO

Numero de personas en la vivienda

Se trata del número total de personas en la vivienda (residentes presentes, residentes ausentes, y no residentes presentes)

99 DESCONOCIDO

Uso general del edificio

- 1 Solo vivienda
- 2 Vivienda y comercio
- 3 Vivienda y taller
- 4 Solo comercios, oficinas, equipamientos
- 5 Solo artesanal o industrial
- 6 Bodega, banos ... etc de una vivienda
- 7 Utilizado para los animales
- 8 Garage
- 9 DESCONOCIDO

Encuesta

- 1 Vivienda ocupada encuestada
- 4 Vivienda ocupada no encuestada (no quiere contestar)
- 5 Vivienda ocupada no encuestada (en viaje)
- 8 No ocupada (sin residentes)

QUITO - ENQUETE BATI/POPULATION - RECAPITULATIF LOGEMENT

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° Manzana	Nouv. N° Bat.	Anc.N° Bat.princ.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° vivienda	Type de vivienda	Type d'occ. de la viv.	Nbre de pièces	Nbre de personnes	Usage géné. du bat.	Déroule enquête		

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° Manzana	Nouv. N° Bat.	Anc.N° Bat.princ.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° vivienda	Type de vivienda	Type d'occ. de la viv.	Nbre de pièces	Nbre de personnes	Usage géné. du bat.	Déroule enquête		

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° Manzana	Nouv. N° Bat.	Anc.N° Bat.princ.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° vivienda	Type de vivienda	Type d'occ. de la viv.	Nbre de pièces	Nbre de personnes	Usage géné. du bat.	Déroule enquête		

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° Manzana	Nouv. N° Bat.	Anc.N° Bat.princ.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° vivienda	Type de vivienda	Type d'occ. de la viv.	Nbre de pièces	Nbre de personnes	Usage géné. du bat.	Déroule enquête		

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° Manzana	Nouv. N° Bat.	Anc.N° Bat.princ.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.	Anc.N° Bat.
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
N° vivienda	Type de vivienda	Type d'occ. de la viv.	Nbre de pièces	Nbre de personnes	Usage géné. du bat.	Déroule enquête		

Annexe A.3.2

L'ENQUETE DEMOGRAPHIE / MIGRATIONS : COLLECTE, CODIFICATION ET SAISIE

1) Manuel d'instructions aux enquêteurs

2) Conditions de collecte, codification et saisie des données

1) Manuel d'instructions aux enquêteurs

QUITO - Enquête démographique/migrations

Novembre-Décembre 1987

INSTRUCTIONS AUX ENQUETEURS

**ATLAS INFORMATISE DE QUITO
ORSTOM-IGM-IPGH-MAIRIE DE QUITO**

SOMMAIRE

GENERALITES

1. OBJECTIFS SCIENTIFIQUES
 1. 1. Le programme de recherche
 1. 2. L'enquête Démographie/Migrations
2. TECHNIQUE D'ENQUETE
3. ORGANISATION DU TRAVAIL

COLLECTE DE L'INFORMATION

Principes de base

1. VARIABLES SUR CHACUNE DES PERSONNES VIVANT DANS LE LOGEMENT
2. VARIABLES SUR LES CONDITIONS D'OCCUPATION DU LOGEMENT
3. LIEUX DE RESIDENCE ET ACTIVITES DU CHEF DE MENAGE DURANT LES DEUX DERNIERES ANNEES
4. BIOGRAPHIE MIGRATOIRE. PROFESSIONNELLE ET FAMILIALE DU CHEF DE MENAGE DEPUIS SA NAISSANCE
5. VARIABLES SUR LES PARENTS DU CHEF DE MENAGE ET DE SON CONJOINT NE VIVANT PAS DANS LE LOGEMENT ENQUETE

ANNEXES

Questionnaire

Feuille d'information

A. GENERALITES

1. OBJECTIFS SCIENTIFIQUES

1. 1. Le programme de recherche

L'objectif du programme de recherche est de mettre au point un système d'observation permanent des populations urbaines, intégrant l'information apportée par les images satellite sur la morphologie urbaine : suivi des effectifs et caractéristiques générales de la population, et réalisation de sondages spatiaux stratifiés sur les images satellite SPOT ou THEMATIC MAPPER, pour des investigations approfondies sur le comportement des citadins.

Deux idées sont à la base de ce programme :

- Seul un système basé sur la technique des sondages peut remplir les conditions de souplesse et de rapidité nécessaires dans les villes à croissance rapide.
- Une meilleure connaissance de l'espace urbain constitue un facteur important d'amélioration d'un système d'enquête par sondage.

Les satellites renseignent de manière continue sur l'occupation du sol en milieu urbain : il s'agit donc, dans la méthode que nous développons, d'utiliser cette information exhaustive sur la morphologie urbaine pour collecter rapidement, par sondage, des données relatives aux populations urbaines. Les travaux réalisés (par VERNIERE notamment) sur photographies aériennes ont montré que l'application de ce principe permet un allègement de l'échantillon d'enquête (donc, pour un même coût global et un même niveau de précision, une possibilité d'approfondir l'observation) et une spatialisation des résultats.

La réalisation de l'objectif général du programme implique des travaux, en parallèle, dans quatre directions :

- l'interprétation des images satellite à haute résolution en milieu urbain
- les relations entre données morphologiques et données démographiques
- les sondages spatiaux
- l'intégration des données de télédétection dans les bases de données localisées.

Nous avons débuté nos travaux en 1985 sur la ville de Marseille (FRANCE), qui disposait dès cette date de toutes les données nécessaires pour une première phase de développement méthodologique.

Sur la ville de Quito, trois types d'enquête sont réalisées :

- Enquête Morphologie urbaine (octobre-décembre 1986)
- Enquête Bâti/population (avril 1987)
- Enquête Démographie/migrations (novembre-décembre 1987)

L'enquête Morphologie urbaine à Quito a été réalisée sur le terrain entre le 1er octobre et le 9 décembre 1986 ; cette première enquête a permis de collecter des informations précises sur l'occupation du sol, c'est à dire la morphologie urbaine, de 199 îlots de la ville de Quito, sélectionnés sur des plans INEC au 1/ 10 000 à l'aide d'une grille de 750 mètres de côté .

L'enquête Bâti/population, réalisée en avril 1987, a permis de collecter des données démographiques et socio-économiques sur un sous-échantillon des 199 îlots de l'enquête Morphologie urbaine ; l'échantillon enquêté en avril 1987 comprenait 53 îlots.

1. 2. L'enquête Démographie/migrations

Par l'enquête Démographie/migrations, il s'agit de collecter des informations démographiques et socio-économiques auprès d'un échantillon de ménages de Quito (environ 3500 ménages). Cette enquête a un double objectif : effectifs de population et analyse des systèmes de résidence et d'activité.

La sélection des îlots à enquêter provient de l'exploitation d'une image satellite SPOT (résolution 10 mètres), qui nous fournit deux types d'informations : les limites de la ville de Quito, et une classification de l'aire urbaine selon le niveau de densité de bâti (6 classes). La technique de sélection de l'échantillon consiste, pour chaque niveau de densité du bâti, en une sélection systématique sans remise à l'aide d'une grille superposée sur une image satellite au 1/ 15 000. La fraction de sondage, donc la distance entre les lignes de la grille utilisée pour la sélection des îlots, est variable selon les classes de densité du bâti.

2. TECHNIQUE D'ENQUETE

L'enquêteur reçoit :

- **un plan de localisation des ilots à enquêter**

et, pour chaque îlot :

- **une liste des logements de l'îlot** : sur cette fiche, figurent le nombre de logements de l'îlot, la liste et la localisation des logements à enquêter, la description de la méthode de sélection des logements dans l'îlot.

Pour collecter les données de l'enquête Démographie/migrations, l'enquêteur reçoit deux types de fiches :

- **Les questionnaires logement** : ces fiches sont utilisées pour décrire les personnes résidant dans les logements de l'échantillon. Ces questionnaires sont divisés en 5 fiches (voir exemple page 27) :

- 1. Variables sur chacune des personnes vivant dans le logement
- 2. Variables sur les conditions d'occupation du logement
- 3. Lieux de résidence et activités du chef de ménage durant les deux dernières années
- 4. Biographie migratoire et professionnelle du chef de ménage depuis sa naissance
- 5. Variables sur les parents du chef de ménage et de son conjoint ne vivant pas dans le logement enquêté

- **Les fiches supplémentaires** : ces fiches doivent être utilisées lorsque le logement compte plus de 6 personnes (Fiches supplémentaires de type 1), ou que le chef de ménage a eu de nombreux lieux de résidence ou beaucoup d'activités différentes depuis sa naissance (Fiches supplémentaires de type 4).

3. ORGANISATION DU TRAVAIL

L'enquête débute par une **période de formation (10/11/12 novembre)**, au cours de laquelle seront expliqués les objectifs de l'enquête et la technique de collecte de l'information. A l'issue du premier jour de formation, 18 enquêteurs seront sélectionnés parmi tous les enquêteurs présents. Le 11 Novembre, on effectuera une **expérience de collecte dans quelques logements**, afin que les concepts et définitions utilisées pour l'enquête soient bien assimilés, et de façon homogène par tous les enquêteurs. L'enquête proprement dite débutera le 23 novembre, et durera jusqu'au 23 décembre. Les 5, 6, 7 décembre seront considérés comme fériés.

Le 23 novembre, chaque enquêteur **visitera** avec le superviseur **tous les ilots** que l'enquêteur devra enquêter les jours suivants, pour qu'il n'y ait pas d'ambiguïté sur la localisation et la délimitation des ilots et de ses bâtiments. Pour chaque ilot, le superviseur indiquera précisément à l'enquêteur la méthode de sélection des logements à enquêter. L'enquêteur **recevra les documents d'enquête** relatifs à ces ilots (plans de localisation des ilots, listes de logements et questionnaires) afin de pouvoir démarrer le travail de collecte.

La **liste d'ilots** de chaque enquêteur est établie en fonction des principes suivants :

- les ilots de chaque enquêteur sont, dans la mesure du possible, regroupés géographiquement, afin de minimiser les temps de déplacement.
- un enquêteur peut enquêter environ 220 logements en 27 jours de travail à temps complet (c'est-à-dire une moyenne de 8-9 logements par jour).

Chaque enquêteur organise lui-même ses horaires de travail pour réaliser la collecte des informations sur le terrain, en fonction des contraintes d'horaires (de présence à leur domicile) propres à la population à enquêter.

Tous les deux jours, chaque enquêteur aura un **rendez-vous avec le superviseur** de l'enquête. Cette entrevue permettra à l'enquêteur d'exposer les problèmes rencontrés au cours de la collecte, afin de rechercher ensemble une solution à ces problèmes de recueil de l'information ; de plus, à cette occasion, chaque enquêteur remettra au superviseur les questionnaires remplis. Entre deux rendez-vous d'un enquêteur avec le superviseur, celui-ci contrôlera l'avancée du travail, ainsi que la complétude et la cohérence des questionnaires ; en cas d'incomplétude ou d'incohérence des questionnaires, il sera demandé à l'enquêteur, lors du rendez-vous suivant, de retourner sur le terrain pour **corriger ou compléter** l'information. L'enquête se terminera par une **réunion finale** dans les bureaux de l'IGM le 23 décembre ; tous les enquêteurs devront remettre lors de cette réunion les derniers questionnaires remplis.

B. COLLECTE DE L'INFORMATION

. Principes de base

* Le **nombre de logements** que l'enquêteur devra enquêter dans un ilot dépend du nombre total de logements de l'ilot :

- **Si un ilot compte moins de 10 logements (10 inclus)**, l'enquêteur devra enquêter **tous les logements** de l'ilot.

- **Si un ilot compte plus de 10 logements**, l'enquêteur devra enquêter un échantillon de **10 logements**. Le superviseur indiquera précisément à l'enquêteur les 10 logements à enquêter. Si, après plusieurs passages, parmi ces 10 logements, il demeure des logements non enquêtés, l'enquêteur enquêtera d'autres logements, sélectionnés selon une méthode définie par le superviseur.

* Pour remplir les fiches de logement, il n'est pas nécessaire que l'enquêteur rencontre chacune des personnes vivant dans le logement : les questions doivent être posées soit au **chef de ménage**, s'il est présent, soit à **son épouse ou à toute autre personne du ménage pouvant répondre aux questions** se rapportant aux différents membres du ménage et à celles relatives aux migrations et activités du chef de ménage.

* **Si il n'y a personne** (ou aucune personne qui puisse répondre) dans un logement de l'échantillon, laisser une **feuille d'information** (voir exemple page 33) sur l'enquête (remplir les lignes relatives au rendez-vous), continuer l'enquête dans les autres logements, et revenir plus tard pour ce rendez-vous (demander au voisin à quelle heure les habitants de ce logement rentrent chez eux).

1. VARIABLES SUR CHACUNE DES PERSONNES VIVANT DANS LE LOGEMENT

Recensement, délimitation et numérotation des logements

Définition du "logement"

On appelle logement tout local indépendant utilisé pour l'habitation : un local non destiné originellement à l'habitation, mais qui est habité au moment de l'enquête est assimilé à un logement.

Délimitation des logements

Plusieurs situations peuvent se rencontrer :

- Bâtiment à étages, divisé en appartements : chaque appartement (c'est-à-dire, ensemble de pièces ayant une entrée commune, indépendante, qui occupe une partie d'un bâtiment) constitue un logement.

S'il existe des chambres avec des entrées indépendantes, directes, sur les parties communes de l'immeuble (couloirs), ces chambres sont considérées comme des logements indépendants.

Si une chambre n'a pas d'entrée indépendante de celle desservant l'appartement dont elle fait partie, on considère qu'elle est incluse dans cet appartement et ne constitue pas un logement indépendant.

- Maison : un bâtiment non divisé en appartements (unités d'habitation d'habitation indépendantes) constitue un seul logement.

S'il existe des chambres de service (ou pièces louées) avec des entrées indépendantes, permettant de rejoindre directement la rue sans passer par la maison, ces chambres sont considérées comme des logements indépendants.

Si une chambre n'a pas d'entrée indépendante de celle desservant la maison dont elle fait partie, on considère qu'elle est incluse dans cette maison et ne constitue pas un logement indépendant.

Numérotation des logements

Chaque logement est identifié par trois numéros :

- Un numéro d'ilot

- Un numéro de bâtiment, qui va de 1 à N1 dans chaque ilot

- Un numéro de logement. Pour chaque bâtiment de l'ilot, la numérotation des logements va de 1 à N2, si le bâtiment compte N2 logements.

Ces trois numéros sont notés dans la partie supérieure de la première feuille du questionnaire logement.

. Type de logement

Marquer d'une croix la case qui correspond au type de logement.

Le logement est **individuel** s'il est occupé par une personne ou un groupe de personnes qui vivent sous un régime familial. Il peut s'agir d'une **maison**, d'un **appartement**, d'une **chambre**.

Le logement est **collectif** quand il est habité par un groupe de personnes qui se le partagent pour raison de santé, discipline, religion...etc. S'il s'agit d'un logement collectif, noter le nombre de personnes qui vivent dans le logement, et indiquer précisément le type de logement collectif, en utilisant les catégories suivantes : caserne, hôtel, hôpital, internat, couvent...etc.

. Conditions d'occupation du logement

Mettre une croix dans la case qui correspond à la situation du logement.

Quatre situations peuvent se produire :

- Logement en réparation ou en construction, donc **inhabitable** au moment de l'enquête
- Logement **habitable, mais non occupé**
- Logement habitable, **occupé**, mais où l'enquête n'a pas pu être réalisée
- Logement habitable, **occupé**, où on a pu enquêter les habitants.

. Nombre de pièces utilisées pour dormir

Noter, dans les deux cases prévues à cet effet, le nombre de pièces généralement utilisées pour dormir, qu'elles soient ou non effectivement occupées au moment de l'enquête. Une pièce dont ce n'est pas la vocation première, mais qui est également utilisée pour dormir (exemple : un salon) doit être comptée dans le calcul du nombre total de pièces utilisées pour dormir au même titre qu'une chambre à coucher.

Définition de la population à enquêter. Statut de résidence.

La population décrite dans les colonnes de la fiche logement comprend trois catégories, suivant le statut de résidence des personnes :

- **présents** : toutes les personnes ayant dormi dans le logement pendant au moins huit nuits, consécutives ou non, pendant le mois précédant le passage de l'enquêteur (qu'elles soient ou non présentes au moment du passage de l'enquêteur ou la nuit précédant son passage);

Si ces personnes résident la majeure partie de l'année dans le logement enquêté, on les appelle "**Résidents présents**".

Si ces personnes résident la majeure partie de l'année dans un autre logement que celui enquêté, on les appelle "**Non résidents présents**".

- **Résidents absents** : toutes les personnes qui résident habituellement (c'est à dire la majeure partie de l'année) dans le logement, mais en sont absentes depuis plus d'un mois (tourisme, voyage d'affaires, travail temporaire ou saisonnier à l'extérieur de Quito), pour quelle que raison que ce soit (sauf un départ définitif du logement ou les cas particuliers cités ci-dessous) et qui doivent revenir vivre dans le logement dans un délai inférieur à 6 mois.

Cas particuliers

Ne font pas partie de la population à enquêter dans les logements individuels (car sont comptabilisés dans les logements collectifs) :

- Militaires, élèves internes,
- Emprisonnés ou hospitalisés pour une longue durée (supérieure à un mois)

Font partie de la population à enquêter :

- Emprisonnés ou hospitalisés pour une courte durée (inférieure à 1 mois)
- Travailleurs exerçant leur activité en dehors de Quito, vivant donc en dehors de la capitale une partie de l'année, mais revenant régulièrement habiter dans leur logement de Quito (chaque fin de semaine, chaque mois)
- Personnes venant d'emménager dans le logement pour y résider de façon définitive après un déménagement. (transfert du mobilier...etc dans le nouveau logement).

Collecte des renseignements relatifs aux personnes vivant dans le logement

Le plus simple est de réaliser l'interview en **deux temps** :

1) D'abord, demander de donner une **liste des personnes** habitant dans le logement, et noter leur **nom** et leur statut de résidence.

Demander à la personne interrogée de commencer par le chef de ménage, et poursuivre par les autres personnes vivant avec lui dans le logement, en respectant l'ordre général suivant :

Chef de ménage, épouse, enfants célibataires, enfants mariés avec les membres de leur famille, autres parents, personnes non apparentées au chef de ménage.

Numéroter ces personnes de 1 à N dans chaque logement.

2) Puis, questionner pour avoir les **différents renseignements sur chacune de ces personnes**, en redonnant un par un les noms des personnes de cette liste.

Les renseignements relatifs à chaque personne occupent **une colonne** de la première fiche du questionnaire.

Si un logement compte plus de 6 personnes, utiliser une seconde fiche pour les personnes n° 7 à 12, et, si nécessaire, une troisième fiche pour les personnes n° 13 à 18 ...etc. Ne pas oublier de noter le numéro de l'ilot, du bâtiment et du logement sur la fiche supplémentaire.

NOM ET PRÉNOM

Ces renseignements sur l'identité des personnes enquêtées ne feront pas partie du fichier exploité sur ordinateur ; il faut donc insister auprès des enquêtés sur le fait que l'exploitation de l'enquête est faite de façon **anonyme**. Ces renseignements ne sont demandés que pour faciliter la collecte de l'information, afin d'éviter que des personnes du logement soient omises ou au contraire comptées deux fois, ainsi que pour faciliter la collecte de l'information sur les liens de parenté au sein du logement.

QUEL LIEN DE PARENTÉ (OU AUTRE RELATION) AVEZ-VOUS AVEC LE CHEF DE MÉNAGE?

S'il existe un lien de parenté avec le chef de ménage, noter en clair le lien de parenté existant entre la personne et le chef de ménage.

Exemples :

- Frère de 1, épouse de 1, neveu de 1... etc. (Rappel : le chef de ménage est toujours noté 1)
- Nièce de 1, employée domestique : pour rendre compte du cas où une personne de la famille du chef de ménage a également une relation de "service" avec le chef de ménage.

S'il n'existe pas de lien de parenté avec aucune personne du logement, noter en clair la nature de cet autre lien.

Exemple : employée domestique, nurse, gardien.

S'il n'existe pas de lien de parenté avec le chef de ménage, mais qu'il existe un lien de parenté avec une autre personne du logement (elle même non apparentée au chef de ménage), noter en clair ce lien de la façon suivante :

Exemple : la personne N° 12 est la fille de l'employée domestique, qui porte le N° 11.

On note donc pour la personne 12 : fille de 11.

Si l'on ne connaît pas le lien (de parenté ou autre), noter "INCONNU".

SEXE DE LA PERSONNE

Marquer d'une croix la case qui correspond au sexe de la personne.

QUEL EST VOTRE SITUATION MATRIMONIALE ?

Marquer d'une croix la case qui correspond à la situation matrimoniale de la personne.

QUELLE EST VOTRE DATE DE NAISSANCE OU QUEL AGE AVEZ-VOUS?

Si la personne connaît précisément sa date de naissance, la noter dans les six cases prévues à cet effet (ne rien noter dans les cases situées en dessous, pour l'âge):

Exemple : 6 Août 1938

0	6	0	8	3	8
---	---	---	---	---	---

Si la personne ne connaît pas précisément sa date de naissance, noter son âge dans les cases prévues à cet effet (ne rien noter dans les cases situées au dessus, pour la date de naissance). Il s'agit de l'âge atteint au dernier anniversaire, et non de l'âge qu'aura la personne à son prochain anniversaire. Ainsi, pour un enfant de moins d'un an, on note 00 et le nombre de mois.

Pour les personnes qui ont un an ou plus, ne pas noter le nombre de mois.

Exemples : personne qui a 23 ans

2	3
année	mois

enfant qui a 6 mois

0	0
année	mois

Dans le cas où l'âge de la personne est inconnu, noter, dans les deux cases prévues pour l'âge (années) : "D" (Desconocido)

Exemple : date de naissance et âge inconnus

D
année

mois	

OU ETES-VOUS NE?

Si la personne est née à Quito (stricto-sensu, non compris les villes de la vallée de Los Chillos, San Antonio, Cumbaya, Tumbaco), noter "Ici" sur la ligne inférieure. Ne rien indiquer sur les autres lignes.

Si la personne est née en Equateur en dehors de Quito, indiquer en clair, sur chacune des trois lignes prévues (ne rien indiquer sur la ligne "pays") :

- le nom de la province
- le nom de la paroisse

- le nom de la localité-même où est née la personne (qu'il s'agisse d'une ville ou d'un village).

Si la personne est née à l'étranger, indiquer seulement le pays de naissance.

Si le lieu de naissance est inconnu, noter "D" (Desconocido) sur la ligne supérieure.

Si le lieu de naissance n'est connu que de façon imprécise (par exemple, on connaît la province, mais pas la paroisse), indiquer l'élément de réponse qui est connu, et noter "D" (Desconocido) pour les autres éléments, non connus.

Exemples :

- Personne née à l'étranger :

Pays *CHILE*
Province
Paroisse
Nom de la localité
localité

Personne née à San Rafael (Vallée de Los Chillos):

Pays
Province *PICHINCHA*
Paroisse *SAN GOLOM*
Nom de la localité *San Rafael*

- Personne née à Quito (stricto sensu) :

Pays
Province *PICHINCHA*
Paroisse
Nom de la localité *Quito*
localité

Personne née dans la paroisse de Cuenca, dans le village NNN:

Pays
Province *AZUAY*
Paroisse *CUENCA*
Nom de la localité *NNN*

- Personne née à Cuenca, dans la ville-même :

Pays
Province *AZUAY*
Paroisse *CUENCA*
Nom de la localité *Cuenca*
localité

Personne à la province de naissance connue, mais paroisse inconnue :

Pays
Province *CHIMBORAZO*
Paroisse *D*
Nom de la localité *D*

- Personne au lieu de naissance inconnu :

Pays *D*
Province *-*
Paroisse *-*
Nom de la localité *-*

DEPUIS QUELLE DATE VIVEZ-VOUS DANS CE LOGEMENT OU DEPUIS COMBIEN D'ANNEES ?

Avec cette question, il s'agit de saisir la date d'arrivée ou la durée d'occupation dans le logement actuel :

- pour les **résidents présents** et les **résidents absents** : il s'agit de la date à laquelle la personne a fait de ce logement sa résidence principale. Par résidence principale, on entend, si la personne habite dans plusieurs logements au cours de l'année, la résidence où elle habite le plus de temps au cours de l'année.

- pour les **non résidents présents** : il s'agit de la dernière date d'arrivée de ces personnes dans le logement.

- Si la personne a toujours vécu dans ce logement, noter "TOUJOURS".

Si la personne a habité autrefois dans ce logement, en est partie pour résider ailleurs, et est revenue s'y installer, noter la date du début de la dernière installation.

Si la personne connaît sa date d'arrivée dans le logement, la noter dans les 6 cases prévues à cet effet (ne rien noter dans les cases situées en dessous, pour la durée).

Si la personne connaît la durée écoulée depuis son arrivée, noter la durée en années dans les cases prévues à cet effet (ne rien noter dans les six cases situées au dessus, pour la date d'arrivée). Il s'agit de la durée en années entières ; ainsi, si la personne occupe ce logement depuis moins d'un an, noter 00 (années) et le nombre de mois.

Ne pas indiquer le nombre de mois pour les durées d'un an ou plus.

Dans le cas où la date d'arrivée est totalement inconnue, tout comme la durée, noter "D" (Desconocido) dans les cases prévues pour noter la durée d'occupation (années) du logement.

Exemples :

- Personne occupant le logement depuis le 5 Septembre 1983 :

0	5	0	9	8	3
---	---	---	---	---	---

- Personne occupant le logement depuis 1952 (jour et mois inconnus) :

D	D	5	2
---	---	---	---

- Personne occupant le logement depuis 38 ans :

38	1
année	mois

- Personne occupant le logement depuis 2 mois :

00	02
année	mois

- Personne dont on ignore depuis quelle date ou depuis combien de temps elle occupe le logement :

?	
année	mois

OU VIVIEZ-VOUS AVANT DE RESIDER DANS CE LOGEMENT ?

Deux situations peuvent se produire (il s'agit de Quito stricto sensu, sans prendre en compte les villes de la vallée de Los Chillos, San Antonio, Cumbayá, Tumbaco) :

- la personne **vivait à Quito**, avant de résider dans ce logement actuel:
 - marquer d'une croix la case qui correspond à cette situation,
 - noter l'adresse du logement où vivait la personne : quartier, et rues (par exemple : LA FLORESTA, Toledo y Madrid).

- la personne **ne vivait pas à Quito**, avant de résider dans le logement actuel :
 - marquer d'une croix la case qui correspond à cette situation ("Hors de Quito")

DEPUIS QUELLE DATE VIVEZ-VOUS A QUITO OU DEPUIS COMBIEN D'ANNEES ?

Le mode de remplissage de la fiche pour cette question est similaire à celui concernant la date d'arrivée dans le logement.

Avec cette question, il faut saisir la date d'arrivée ou la durée de séjour à Quito (stricto-sensu, non compris les villes de la vallée de Los Chillos, San Antonio, Cumbaya, Tumbaco):

- pour les résidents présents ou les résidents absents: et pour les non résidents présents ayant leur résidence principale à Quito, mais dans un autre logement : il s'agit de la date à laquelle la personne a eu sa résidence principale à Quito, qu'il s'agisse de ce logement ou non. Par résidence principale,

Par résidence principale, on entend, si la personne habite alternativement à Quito et en dehors de Quito au cours de l'année, la résidence où elle habite le plus de temps au cours de l'année.

- pour les non résidents présents n'ayant pas leur résidence principale à Quito : il s'agit de la dernière date d'arrivée de ces personnes à Quito.

Si la personne a toujours vécu à Quito, noter "TOUJOURS".

Si la personne a habité autrefois à Quito, en est partie pour résider ailleurs, et est revenue s'y installer, noter la date du début de la dernière installation.

Si la personne connaît sa date d'arrivée à Quito, la noter dans les 6 cases prévues à cet effet (ne rien noter dans les cases situées en dessous, prévues pour la durée).

Si la personne connaît la durée écoulée depuis son arrivée à Quito, noter la durée en années dans les deux cases prévues à cet effet (ne rien noter dans les cases situées au dessus, pour la date d'arrivée). Il s'agit de la durée en années entières ; ainsi, si la personne vit à Quito depuis moins d'un an, noter 00 (années) et le nombre de mois. Ne pas indiquer le nombre de mois pour les durées supérieures à un an.

Dans le cas où la date d'arrivée est totalement inconnue, tout comme la durée, noter "D" dans les cases prévues pour noter la durée de séjour à Quito.

OÙ HABITIEZ-VOUS AVANT DE RÉSIDER À QUITO?

Pour les résidents présents, les résidents absents, et les non résidents présents ayant leur lieu de résidence principal actuel à Quito, il s'agit du lieu de résidence principal de la personne juste avant qu'elle ait son lieu de résidence principal à Quito.

Pour les non résidents présents ayant leur lieu de résidence principal actuel en dehors de Quito, il s'agit du lieu de résidence principal actuel.

Si la personne résidait auparavant en **Equateur en dehors de Quito**, indiquer en clair, sur chacune des trois lignes prévues (ne rien indiquer sur la ligne "pays") :

- le nom de la province
- le nom de la paroisse
- le nom de la localité-même où est née la personne (que ce soit une ville ou un village).

Si la personne résidait auparavant à l'**étranger**, indiquer seulement le pays de naissance.

Si le **lieu de résidence est inconnu**, noter "D" (desconocido) sur la ligne supérieure, et mettre un trait pour les autres lignes.

Si le **lieu de résidence n'est connu que de façon imprécise** (par exemple, on connaît la province, mais pas la paroisse), indiquer l'élément de réponse qui est connu, et noter "D" (desconocido) pour les autres éléments, non connus.

QUEL EST LE DERNIER COURS QUE VOUS AVEZ SUIVI DANS L'ENSEIGNEMENT ?

Marquer d'une croix la case qui correspond à la situation de la personne, et noter le numéro du cours le plus élevé qu'a suivi complètement la personne. Ne pas prendre en compte les cours d'enseignement accéléré.

Si le **niveau scolaire est totalement inconnu**, noter "D" (Desconocido), et ne marquer aucune case avec une croix.

Si le **niveau scolaire est connu de façon imprécise** (cycle connu, mais cours inconnu), marquer d'une croix la case qui correspond au cycle suivi et noter "D" pour le niveau dans le cycle.

ACTIVITÉ

Cette question concerne chaque personne vivant dans le logement, sans condition d'âge minimum.

Il s'agit de la situation durant les mois de novembre et décembre 1937. Une personne peut avoir deux activités (exemple : professeur d'économie à l'université, et comptable à son

compte). Noter dans la case "activité principale" l'activité qui procure les revenus les plus importants à la personne.

Profession

Plusieurs cas peuvent se produire.

Si la personne ne travaille pas, pour décrire les différentes situations, utiliser les termes suivants :

- Etudiant
- Retraité
- Ne travaille pas, recherche un emploi
- Travaux ménagers
- Ne travaille pas, ne recherche pas d'emploi

Si la personne travaille, indiquer sa profession en détail :

Exemples

- Professeur de musique
- Menuisier
- Chauffeur de camions
- Professeur de l'Université
- Vendeur d'assurances
- Mécanicien automobile...etc.

Branche d'activité

Cette question ne concerne que les personnes qui travaillent actuellement, ou qui sont retraitées (dans ce cas, indiquer la branche de la dernière activité).

Pour obtenir une réponse correcte, demander :

"Quelle est l'activité principale de l'établissement où vous travaillez?"

L'enquêteur doit indiquer exactement le type d'activité de l'établissement. Ne pas accepter une réponse vague, et insister pour avoir plus de détails.

Exemples de réponses correctes

- Education cycle primaire
- Fabrication de tissus
- Boulangerie
- Consultations médicales
- Menuiserie
- Ministère de l'Agriculture...etc.

Statut professionnel

Cette question ne concerne que les personnes qui travaillent actuellement, ou qui sont retraitées (dans ce cas, indiquer le statut professionnel lors de la dernière activité).

Indiquer le statut professionnel :

- **Patron** : personne qui dirige sa propre entreprise, ou qui exerce une profession pour son propre compte, et qui a un ou plusieurs employés rémunérés (ex : directeur d'une usine).
- **Travailleur indépendant** : personne qui dirige sa propre entreprise, ou qui exerce pour son propre compte un travail, mais qui n'a aucun employé rémunéré (ex : vendeur ambulant, cordonnier).
- **Employé ou salarié** : employé qui travaille pour l'Etat, la Municipalité ou dans une entreprise privée, et qui reçoit une rémunération, une solde, un salaire, une commission ou une somme forfaitaire.
- **Travailleur familial sans rémunération** : personne exerçant une activité non rémunérée dans une entreprise ou un commerce d'un membre de sa famille, la majeure partie de la semaine.
- **Apprenti** : personne travaillant dans une entreprise pour apprendre un métier, ne recevant pas de rémunération, ou une rémunération minime.
- **Autres (préciser)**

Noter le **montant des revenus** procurés par chaque activité (préciser la période de référence : mois, semaine, ou autre durée).

AVEZ VOUS D'AUTRES SOURCES DE REVENUS ?

Il s'agit des revenus autres que ceux procurés par les activités décrites ci-dessus dans le questionnaire.

Pour chaque type de source de revenus, noter :

- "NON", si la personne n'a pas de revenus de ce type
- "D", si la réponse relative à ce type de revenus est inconnue
- si la personne a des revenus de ce type : la **source précise** (exemple, pour les transports : "propriétaire d'un taxi dans Quito") et le **montant mensuel** en sucres.

Si la personne ne veut pas répondre à cette question sur les sources de revenus, marquer d'une croix la case qui correspond à cette situation.

2. VARIABLES SUR LES CONDITIONS D'OCCUPATION DU LOGEMENT

Ces questions ne concernent que le chef de ménage.

A quel titre occupez-vous ce logement?

Marquer d'une croix la case qui correspond à la situation et remplir les cases correspondantes.

Si le chef du ménage ou une autre personne vivant dans le logement est propriétaire du logement :

- dans le deuxième cas, indiquer le **numéro de la personne** qui est propriétaire : il s'agit du numéro de personne annoté sur la première page du questionnaire ;
- dans les deux cas, noter :
 - la **date** (jour, mois, année) et
 - le **coût** (en sucres) :
 - de l'achat du logement (exemple : achat d'un appartement dans un immeuble, ou d'une maison dans un "condominio"),
 - de l'achat du terrain, et de la construction du bâtiment, s'il s'agit d'un bâtiment que le propriétaire a fait construire lui-même.

Si la date ou le coût est inconnu , noter "D" dans les cases correspondantes.

Si le logement a été acquis par don ou héritage, indiquer seulement la date d'acquisition du logement.

S'il s'agit d'une occupation pour services, préciser le type de service (exemples : portier, jardinier d'un immeuble) et le montant du loyer mensuel en sucres.

S'il s'agit d'un autre type d'occupation (qui n'est pas prévu dans la liste), noter en clair le type d'occupation.

Si le propriétaire ne vit pas dans le logement : de quel type de propriétaire s'agit-il ?

Marquer d'une croix la case qui correspond à la situation.

S'il s'agit d'une personne physique qui ne vit pas dans le logement, noter en clair le lien de parenté du chef de ménage avec le propriétaire du logement.

Exemples : Le chef de ménage est le fils du propriétaire. Noter : "Fils de P."

Le chef de ménage est l'époux de la fille du propriétaire. Noter : "Gendre de P".

3. LIEUX DE RESIDENCE ET ACTIVITES DU CHEF DE MENAGE DURANT LES DEUX DERNIERES ANNEES

Il s'agit de noter dans le tableau chaque lieu de résidence, chaque activité et la (les) raison(s) du changement de résidence ou d'activité du chef de ménage durant les deux dernières années. Commencer par la situation actuelle du chef de ménage (lieu de résidence et d'activité), et continuer vers le bas en décrivant les situations antérieures.

Tous les lieux de résidence et toutes les activités doivent figurer dans ce tableau, **quelque soit la durée de la résidence ou de l'activité** (par exemple, un voyage d'une semaine doit être considéré comme un changement de résidence) ; un changement de résidence dans Quito doit être noté.

La manière de remplir la colonne "**Lieu de résidence**" est semblable à celle utilisée pour la première fiche du questionnaire ; noter l'adresse précise pour les résidences à l'intérieur de Quito.

La manière de remplir la colonne "**Activité principale**" est semblable à celle utilisée pour la première fiche du questionnaire.

Pour la colonne "**Raison du changement de résidence ou d'activité**", noter en clair et avec précision (au niveau de la date du changement) la (les) raisons du changement.

Exemples :

- pour un changement de résidence : "J'ai hérité d'une maison", "Pour avoir un logement plus grand", "Pour vivre plus près de mon travail"...
- pour un changement d'activité : "L'usine où je travaillais a licencié tout le personnel", "Pour améliorer mes revenus"...etc

Pour chacun des lieux de résidence et chacune des activités, noter sa durée de la manière suivante :

Exemple : Depuis le 15 avril 1987 jusqu'au 12 Août 1987, la personne vivait à la Floresta ; le 12 août, elle a déménagé pour San Juan. Depuis le 1er juin 1987, après une période de salariat, elle travaille dans sa propre entreprise de cuir.

Calendrier	Date du changem. de résid.	LIEU DE RESIDENCE (Cuelque soit la durée) <i>Pays, province, paroisse, localité, adresse</i>	Date du changem. d'activité	ACTIVITE PRINCIPALE <i>Profession, branche d'activité, statut professionnel, revenus mensuels</i>
Septem. 87		Quito, San Juan, Buga... y. Haïti		
Août 87	12. Août			Cordonnier, Réparations cuir
Juillet 87		Quito, La Floresta, Toledo y Madrid		A son compte, 18.000 s/c
Juin 87			1 ^{er} Juin	
Mai 87				Cordonnier, Réparations cuir
Avril 87	15. Avr.	Quito, San Carlos, P... y Carlos Quinto		Employé, 12.000 = 1 mois
Mars 87				

S'il s'agit de migrations ou d'activités périodiques (c'est-à-dire qui se répètent régulièrement), remplir le tableau de la façon suivante :

Exemple : Depuis le 1er décembre 1986 jusqu'au 30 avril 1987, la personne vivait pendant la semaine à Quito, et retournait chaque fin de semaine à Ibarra, où vivait sa famille.

Calendrier	Date du changem. de résid.	LIEU DE RESIDENCE (Cuelque soit la durée) <i>Pays, province, paroisse, localité, adresse</i>	Date du changem. d'activité	ACTIVITE PRINCIPALE <i>Profession, branche d'activité, statut professionnel, revenus mensuels</i>
Mai 87				
Avril 87	30 ^e Avr.	Pendant la semaine : Quito, La Floresta, Toledo y Madrid		Cordonnier, Réparations à son compte, 13.000
Mars 87				
Février 87		Chaque fin de semaine :		
Janvier 87		Imbabura, Ibarra, Ibarra		
Décem. 86	1 ^{er} Dé.			

4. BIOGRAPHIE MIGRATOIRE, PROFESSIONNELLE ET FAMILIALE DU CHEF DE MENAGE DEPUIS SA NAISSANCE

Chaque ligne du tableau correspond à une période définie par deux âges ou deux dates. Chaque fois qu'il y a eu un changement de résidence ou d'activité, changer de ligne. Le tableau commence par la naissance du chef de ménage, et se termine par la situation actuelle.

Si la fiche 4 n'est pas suffisante pour noter la biographie complète du chef de ménage, utiliser une fiche supplémentaire n° 4 ; ne pas oublier de noter en haut de cette fiche supplémentaire les numéros d'ilot, bâtiment et logement.

On appelle "**changement de résidence**" (seulement si durée > 6 mois):

- tout changement de localité de résidence, s'il ne s'agit pas de Quito
- tout changement de logement, s'il s'agit d'un changement à l'intérieur de Quito (changement d'adresse dans Quito).

On appelle "**changement d'activité**" :

- tout changement de profession (y compris un changement au sein d'une même entreprise)
- tout changement de statut professionnel
- tout changement d'entreprise

La manière de remplir la colonne "**Lieu de résidence**" est semblable à celle utilisée pour la première fiche et le tableau 3 du questionnaire (lieux de résidence et d'activité durant les deux dernières années) ; ne noter l'adresse précise que pour les résidences dans Quito.

La manière de remplir la colonne "**Activité principale**" est semblable à celle utilisée pour la première fiche du questionnaire et pour le tableau 3;

Pour la colonne "**Raison du changement de résidence ou d'activité**", noter en clair et avec précision la (les) raisons du changement.

Exemples :

- pour un changement de résidence : "J'ai hérité d'une maison", "Pour avoir un logement plus grand", "Pour vivre plus près de mon travail"...
- pour un changement d'activité : "L'usine où je travaillais a licencié tout le personnel", "Pour améliorer mes revenus" ...etc

4. BIOGRAPHIE MIGRATOIRE ET PROFESSIONNELLE DU CHEF DE MENAGE DEPUIS SA NAISSANCE

Age (ans)	Date (jo. mo. an)	LIEU DE RESIDENCE (> 6 mois) Pays, province, paroisse, localité ou localité, quartier et adresse, s'il s'agit de Quito	ACTIVITE Profession, titre, niche d'activité, statut professionnel, revenus mensuels (sucres)	RAISON DU CHANGEMENT DE RESID
Naiss. à	01/08/45 à 11/11/53	Catapaxi, Santa Ana, Salceda	Prof Eran (Miño) Statut Revenus	Exemple: Pour apprendre un métier avec son oncle
à	à 11/11/57	Quito, San Carlos, Pensa y Carlos Quiato	Prof Cordonnier Eran Réparation Cuir Statut Apprenti /mois Revenus 200	Fin apprentissage Cherche un emploi près de ses parents
à	à 11/11/70	Pichincha, Santa Dominga, Santa Dominga	Prof Cordonnier Eran Réparation Cuir Statut Employé /mois Revenus 2000	Déménagement de l'entre-prise. Devient chef d'atelier
à	à 11/08/82	Esmeraldas, Esmeraldas, Esmeraldas	Prof Cordonnier (chef d'atelier) Eran Réparation Cuir Statut Employé /mois/ Revenus 3000	A eu l'occasion de louer un bon atelier à Quito
à	à Maintenant	Quito, Floresta, Madrid y Toledo	Prof Cordonnier Eran Réparation cuir Statut /mois Revenus 10000	
à	à		Prof Eran Statut Revenus	
à	à		Prof Eran Statut Revenus	
à	à		Prof Eran Statut Revenus	
à	à		Prof Eran Statut Revenus	
à	à		Prof Eran Statut Revenus	
à	à		Prof Eran Statut Revenus	

5. VARIABLES SUR LES PARENTS DU CHEF DE MENAGE ET DE SON CONJOINT NE VIVANT PAS DANS LE MENAGE ENQUETE

A chaque parent du chef de ménage ou de son conjoint qui ne vit pas dans le logement enquêté correspond une ligne du tableau.

Pour toutes les variables (sexe, âge ou année de naissance, année de décès, lieu de résidence actuel et occupation actuelle), la manière de remplir la fiche est semblable à celle utilisée pour la première fiche du questionnaire.

Si le parent est déjà décédé, il s'agit de son dernier lieu de résidence et de sa dernière activité, au moment du décès.

Pour la question sur "la relation économique actuelle avec ce parent", noter dans chaque case de la colonne le code qui correspond au type de relation économique actuel. Avec un même parent peuvent exister plusieurs types de relation économique ; noter dans la case du tableau tous les codes qui correspondent à ces différents types de relations.

Exemple :

Age ou année de naissance	Si décédé(e) année du décès	Lieu de résidence actuel Pays, province, paroisse, localité ou localité, quartier et adresse, s'il s'agit de Guiso	activité principale actuelle	Relation économique 1.2 avec ce parent (voir index)
1 22	1	Quito, La Florida, Toledo y Madrid	Pro Retraité Enm Boulangerie statut A son compte	5 7
1 40	8,3	Pichincha, Sangolquí, San Rafaël	Pro Ménagère Enm Statut	1 1
27 1	1	PERU	Pro Standardiste Enm Mia. Agriculteur statut Employée	5 1

2) Conditions de collecte, codification et saisie des données¹

2.1 : La collecte des données sur le terrain

Seize enquêteurs, encadrés par trois superviseurs, ont réalisé le travail de collecte des données sur le terrain entre le 23 novembre 1987 et la fin du mois de décembre 1987 ; rappelons que les superviseurs avaient à charge, outre l'organisation et le contrôle du travail des enquêteurs, le recensement des logements occupés dans les îlots de l'échantillon.

- Un contexte social plutôt favorable. Peu de non-réponses.

Les refus de réponse ont été nettement moins nombreux que lors de l'enquête démographique précédente que nous avons réalisée à Quito, en avril 1987 ; il régnait alors en Equateur, et à Quito en particulier, un climat social très tendu, lié au séisme du 5 mars 1987. Du fait de la rupture de l'oléoduc (le pétrole est la principale source de revenus de l'Etat équatorien), l'économie nationale s'est trouvée complètement désorganisée après le tremblement de terre : augmentation immédiate du prix du carburant (multiplié par deux), rations de carburant, forte augmentation des prix de nombreux produits de base...etc. Six mois plus tard, au moment de l'enquête Migrations, la situation économique de l'Equateur n'était certes pas meilleure ; mais les esprits étaient alors mobilisés par la campagne électorale (élections législatives, municipales et premier tour des présidentielles de janvier 1988). Aucun événement ou mesure brutale ne s'étant produit dans les jours précédant l'enquête, on peut donc admettre que le contexte social était plus favorable que lors de l'enquête précédente, d'où un meilleur accueil réservé aux enquêteurs. Si l'on considère les îlots comptant moins de 10 logements occupés, où l'enquête devait être exhaustive, le taux moyen de non-réponses (nombre de logements enquêtés/ nombre de logements total de l'îlot) est de 10 %, incluant à la fois les refus de réponse et les logements non enquêtés pour cause d'absences répétées des habitants ; dans les îlots comptant plus de 10 logements, les enquêteurs pratiquaient une règle systématique de remplacement des logements de l'échantillon impossibles à enquêter.

- L'impossibilité d'enquêter dans l'extrême nord-ouest de la ville

Contrairement à de nombreuses autres villes d'Amérique Latine, l'agglomération quiténienne n'est pas marquée par la multiplication des invasions illégales de terres ; certes, le phénomène existe, mais il reste peu massif, et ces invasions en périphérie de la ville, souvent réalisées par des "coopératives", sont peu à peu légalisées par les autorités municipales ou étatiques. Enquêter dans les quartiers populaires tels que ceux du "Comité de Lucha de los pobres" au sud de la ville, ou du "Comité del pueblo" au nord, ne pose pas de problèmes particuliers.

Seul l'extrême nord-ouest de Quito échappe à cette règle : depuis 1984 un conflit de terres oppose deux coopératives, l'une "Jaime Roldos" soutenue par le gouvernement et disposant d'un niveau d'équipement (voirie, eau, électricité, écoles...etc) tout à fait décent, l'autre "Pisuli" ne bénéficiant d'absolument aucune infrastructure de base. Dans les jours précédant l'enquête Migrations, les affrontements armés ont repris, faisant plusieurs morts parmi la population de la coopérative Pisuli ; après avoir recensé les logements dans quelques îlots de Pisuli et eu un bref entretien, sur place, avec l'un des dirigeants de la coopérative, nous avons préféré annuler l'enquête dans les deux coopératives. Les 6 îlots de l'échantillon appartenant à ces deux quartiers ont donc été exclus de l'échantillon, et l'ensemble des deux quartiers a été éliminé de la base de sondage. Les résultats de l'enquête concernent donc finalement l'ensemble de l'aire urbanisée de Quito à la fin 1987, hormis Pisuli et Jaime Roldos.

¹ : Cette partie de l'annexe est reprise de la communication faite par F. DUREAU aux Journées Démographiques de l'ORSTOM 1988 : "A propos de l'analyse des systèmes résidentiels, présentation de l'enquête migrations réalisée à Quito (Equateur) en décembre 87.", Quito, 17p.

Si cette exclusion ne pose pas de problèmes majeurs pour l'estimation de la population totale de Quito, car les effectifs de ces deux coopératives sont relativement bien connus de leurs responsables, il est par contre certain que nous éliminons une catégorie de population aux origines géographiques bien particulières (la coopérative de Pisuli compte une très forte proportion d'immigrés venus de la côte), et aux comportements spécifiques.

- Les points faibles du questionnaire

La fiche 1 du questionnaire, déjà employée lors de l'enquête réalisée en avril 1987 auprès de plus de 1000 ménages, n'a pas posé de problèmes particuliers ; les quelques aménagements de la fiche (introduction de deux nouvelles questions sur le lieu de résidence antérieur dans Quito et le niveau scolaire, reformulation des questions sur l'activité et les autres sources de revenus) se sont révélés efficaces. La fiche 2, relative aux conditions d'occupation du logement, a été bien comprise et a permis la collecte d'informations intéressantes.

Le recueil des biographies migratoires et professionnelles des chefs de ménage (fiche 4), opération la plus longue de l'interview, s'est déroulé dans de bonnes conditions ; les lieux de résidence et les activités ont été collectés de façon précise. Seules les données sur les revenus manquent dans une bonne partie des questionnaires pour les périodes anciennes, et ce quelque soit l'enquêteur ; l'information sur les raisons des changements de résidence ou d'activité varie, elle, selon l'enquêteur et la capacité de la personne à formuler une argumentation "rationnelle" sur un événement de son passé.

Quant à la fiche 5, d'emploi aisé pour le recueil des caractéristiques générales relatives aux parents du chef de ménage et de son conjoint, elle s'est révélée un peu succincte pour rendre compte de la complexité des relations économiques dans les cas de relations multiples et étroites entre les différents membres de la parentèle

Très logiquement, c'est sur la fiche 3 que s'est concentré l'essentiel des problèmes, non du fait des enquêtés mais surtout des enquêteurs eux-mêmes. Pour les enquêteurs de métier, tout particulièrement ceux ayant participé aux opérations de recensement, il fut difficile d'admettre que le terme "résidence" noté sur le questionnaire faisait en fait référence à tous les lieux de séjour, quelle que soit la durée de séjour ; les enquêteurs n'ayant pas ou peu de pratique du concept de "résidence" ont eu beaucoup plus de facilité à assimiler notre définition très extensive. Il est certain que le fait d'avoir conservé ce terme de "résidence", si chargé sémantiquement, sur la fiche 3 constitue une erreur de notre part ; nous avons dû insister très fortement auprès des enquêteurs durant la formation et les premiers jours d'enquête pour obtenir d'eux qu'ils notent bien *tous* les lieux de séjour.

2.2 : L'exploitation des données

L'organisation des données, en cinq fichiers de base

Le traitement de la masse d'informations collectées sur un IBM AT interdit un archivage dans un seul et même fichier ; nous avons donc recherché comment diviser l'information en différents fichiers. Parmi l'ensemble des critères techniques pris en considération pour définir la structuration de l'information, rappelons les principaux : faciliter le travail de codification en respectant la structure du questionnaire, regrouper les données se rapportant à un même type d'individu statistique (ilot, logement, tout individu du logement, chef de ménage..etc), éviter les redondances d'information (afin de ne coder, saisir et corriger qu'une seule fois chaque information élémentaire), respecter un format fixe d'enregistrement.

Outre ces considérations d'ordre pratique, la conception de l'organisation des fichiers a été guidée, bien évidemment, par les objectifs de l'enquête ; elle est cohérente avec les différentes approches, classiques ou plus nouvelles, exposées au début de cette communication. En ce qui concerne, par exemple, la prise en compte des différents types d'unités collectives, les fichiers 2 et 4 permettent des analyses au niveau du ménage au sens strict, tandis qu'avec le fichier 5 un groupe social plus large est pris en considération.

Le tableau 1 récapitule les caractéristiques des différents fichiers rassemblant l'information collectée durant l'enquête. Cette structuration des données, adéquate pour les phases de codification.

de saisie, de contrôle et de correction des données correspond également à l'essentiel des exploitations que nous avons prévues. Elle pourra toutefois faire l'objet de modifications ultérieures pour des exploitations particulières ; l'utilisation du logiciel DBASEIII rend possibles toutes les opérations de gestion de fichiers pour une éventuelle réorganisation des données.

Tableau 1 - Caractéristiques des cinq fichiers de l'enquête Migrations

	Contenu du fichier	Nombre enregist.	Nombre variables par enreg.	Taille du fichier
Fichier 1 ILOTS	1 enregistrement par îlot Données servant à l'extrapolation : nb total de logements et surface de l'îlot	426	5	6 Ko
Données relatives aux conditions LOGEMENTS	1 enregistrement par logement enquêté 3 169 d'occupation du logement (début fiche 1 et fiche 2 du quest.)	19	160 Ko	Fichier 2
ménage ou de son PARENTS	1 enregistrement par parent du chef de conjoint ne vivant pas dans le logement enquêté Données relatives à l'activité, au lieu de résidence et aux relations économiques (fiche 5 du questionnaire)	13 462	18	500 Ko Fichier 3
dans le logement INDIVIDUS	1 enregistrement par personne vivant 14 553 Caractéristiques démographiques générales et données relatives à l'activité et aux migrations (fiche 1 du questionnaire)	43	1,5 Mo	Fichier 4
par 2 changements MIGRATIONS	1 enregistrement par période définie de résidence et/ou d'activité du chef de ménage Données décrivant le lieu de séjour et l'activité durant la période, le type et les raisons du changement (fiches 3 et 4 du questionnaire)	20 601	26	1,4 Mo Fichier 5

- Codification, saisie, contrôles et corrections

Cinq des enquêteurs ayant participé aux opérations de collecte ont eu la charge de la codification des données ; ils ont achevé ce travail en trois mois de travail à temps partiel, de février à avril 1988. La saisie sur IBM AT, sous-traitée auprès d'une petite entreprise quiténienne dirigée par un statisticien travaillant avec le logiciel SURVEY MATE, a été réalisée au fur et à mesure de

l'avancement du travail de codification ; la saisie a donc été terminée peu après la codification, en mai 1988.

L'étape de contrôle des données a débuté en mars 1988, dès la réception du premier fichier saisi, c'est à dire le fichier concernant les conditions d'occupation des 3169 logements de l'échantillon, maintenant complètement apuré et prêt à être exploité. Les corrections sont effectuées à l'aide du logiciel SPSS PC+, principalement le module Data Entry.

La phase de correction du fichier individus, qui comprend 14553 enregistrements de 99 caractères (soit un fichier de 1,5 Mo) fait clairement apparaître les limites du matériel utilisé, un IBM AT avec disque dur de 40 Mo ; les temps de traitement rendent cette étape de correction très longue. A l'heure actuelle, seule la moitié des variables ont pu être corrigées.

- Le plan d'exploitation des données

Trois grands types d'exploitation des données de l'enquête ont été définis :

- *la production de données pour les décideurs*, principalement la Direction de la Planification de la Mairie de Quito ; estimations, à l'aide du programme d'extrapolation, *d'indicateurs démographiques majeurs* (effectifs, grandes structures ...etc), dans des découpages géographiques définis par les besoins de la planification ;

- *l'analyse des résultats au niveau de l'échantillon* : traitements statistiques réalisés directement sur l'échantillon, sans calcul d'extrapolation, pour des analyses particulières, non localisées. Notre propre travail sur les systèmes résidentiels et pratiques économiques des quiténiens s'inscrit principalement dans ce second type d'exploitation, au même titre que celles que pourront mener les collègues de l'Orstom sur leur propre champ de recherche ;

- *l'exploitation localisée des données* en relation avec toutes les autres informations qui seront intégrées dans la base de données localisées gérée par le système SAVANE¹.

La phase de production de données par extrapolation a commencé en juillet 1988, et progresse au fur et à mesure de l'avancement des travaux de correction des fichiers. La nouvelle équipe de la Direction de la Planification de la Mairie, mise en place très récemment, a un certain nombre de besoins immédiats en données quantitatives sur la population quiténienne, afin d'asseoir sa politique de gestion de Quito ; nous nous efforçons donc de satisfaire ces demandes très concrètes, avant d'aborder, début 1988, notre propre travail d'analyse.

¹ : Système de gestion de bases de données relationnelle développé par l'Unité d'Infographie de l'ORSTOM, utilisé pour la base de données urbaines de l'Atlas Informatisé de Quito.

VARIABLES	Enregistr.	Enregistr.	Enregistr.	Enregistr.	Enregistr.
V1 : 1-3 <i>N° Ilot</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V2 : 4-6 <i>N° Bâtim.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V3 : 7 <i>N° Logem.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V4 : 8-9 <i>N° Perso.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V5 : 10 <i>Statut de résidence</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
V6 : 11-12 <i>Lien Par.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V7 : 13-14 <i>Personne</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V8 : 15 <i>Sere</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
V9 : 16 <i>Etat matrimonial</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
V10 : 17 - 22 <i>Date de naissance</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V11:23-24 <i>Age(années)</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V12:25-26 <i>Age(mois)</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V13:27-28 <i>Lieu de naissance</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V14:29-30 <i>Pays</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V15:31-32 <i>Province</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V16:33 <i>Paroisse</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V17:34-39 <i>Paroisse</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V18:40-41 <i>Date arrivée ds le log.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V19:42-43 <i>années</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V20:44-46 <i>mois</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V21:47-52 <i>Logement ant. Quito</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V22:53-54 <i>Date arrivée Quito</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V23:55-56 <i>années</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V24:57-58 <i>mois</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V25:59-60 <i>Lieu de resid. anter.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V26:61-62 <i>Pays</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V27:63 <i>Provinc.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V28:64 <i>Paroisse</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V29:65-66 <i>Paroisse</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V30:67 <i>Année</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V31:68-70 <i>Activité Principale</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V32:71-73 <i>Type act.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V33:74 <i>Profession</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V34:75-77 <i>Ebranche</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V35:78-80 <i>Statut prof.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V36:81 <i>Revenus enq.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V37:82-84 <i>Rev. esti.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V38:85-87 <i>Seconde Activité</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V39:88 <i>Type act.</i>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
V40:89-91 <i>Profession</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V41:92-94 <i>Ebranche</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V42:95-96 <i>Statut prof.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V43:97-99 <i>Revenus enq.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V44:92-94 <i>Rev. esti.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V45:95-96 <i>Sources rev.</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
V46:97-99 <i>Montant</i>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Annexe A.3.3

**L'OUTIL INFORMATIQUE POUR LA DEFINITION DU SECOND DEGRE DU
SONDAGE : LE PROGRAMME VAR2DEG (LISTAGE)**

```

*****
*
* VAR2DEG.PRG
*
*
* Auteur : O.PARBARY
* Version 1,
*
*
* Programme de calcul, d'après données sur échantillon, considéré
* comme une base de sondage, des deux termes, A et B de la vari-
* -ances d'estimation pour un plan de sondages à deux degrés.
*
*
* 2 options :
*
* 1. Plan équiprobable, estimation sans biais.
* 2. Plan à probabilités inégales, estimation sans biais.
*
* Programme principal.
*
*
* Note :
* Le programme utilise deux fichiers de données, un fichier d'un-
* -ités primaires et un fichier d'unité secondaire, la variable
* * identifiant les unités primaires dans ces deux fichier doit
* * porter le même noms dans les deux fichier.
*
*****

```

```

SET DECIMALS TO 4
SET TALK OFF
SET SCOREBOARD OFF
SET SAFETY OFF
SET BELL OFF
SET DATE FRENCH
SET COLOR TO W/P,B/W,G
CLEAR
TEXT

```

VARIANCEDEUX DEGRE

```

      F1      F2      F3      F4      F5      F6      F7      F8      F9      F10
      F11     F12     F13     F14     F15     F16     F17     F18     F19     F20
      F21     F22     F23     F24     F25     F26     F27     F28     F29     F30
      F31     F32     F33     F34     F35     F36     F37     F38     F39     F40
      F41     F42     F43     F44     F45     F46     F47     F48     F49     F50
      F51     F52     F53     F54     F55     F56     F57     F58     F59     F60
      F61     F62     F63     F64     F65     F66     F67     F68     F69     F70
      F71     F72     F73     F74     F75     F76     F77     F78     F79     F80
      F81     F82     F83     F84     F85     F86     F87     F88     F89     F90
      F91     F92     F93     F94     F95     F96     F97     F98     F99     F100

```

```

      ENDTEXT
*
* Ouverture du fichier de procedures
*
      SET PROCEDURE TO VAR2PRO.FRG
*
* Creation des variables communes
*
NOMFICH=''
REPVAR=''
NB=0
A1=0
A2=0
E1=0
E2=0
*
* Affichage menu
*
      WAIT ' Appuyez sur une touche pour afficher le menu.'
      ECRMENU = .T.
      DO WHILE ECRMENU
      CLEAR
      MENU = ''
      DATE = CDOW( DATE() ) + ' ' + STR( DAY( DATE() ), 2, 0 ) + ' ' ;
      + CMONTH( DATE() ) + ' ' + STR( YEAR( DATE() ), 4, 0 )
      @ 1,3 SAY 'Varédeg. version 1.'
      @ 3,3 SAY DATE
      @ 5,15 SAY '          MENU'
      @ 6,15 SAY '          ****'
      @ 12,15 SAY '<1> : Entré des parametres et calculs.'
      @ 14,15 SAY '<Q> : Quitter le programme'
      @ 17,1 SAY ' '
      BOUCLE1 = .T.
      DO WHILE BOUCLE1
          ACCEPT ' Choisissez une option : ' TO MENU
          MENU = UPPER( MENU )
          IF MENU = '1' .OR. MENU = 'Q'
              BOUCLE1 = .F.
          ENDIF
      ENDDO
      CLEAR
*
      DO CASE
*
      CASE MENU = 'Q'
          SET COLOR TO
          CLEAR
          CLOSE PROCEDURE
          RETURN
      CASE MENU = '1'
* Entrée des noms de fichiers
      CLEAR
      @ 2,5 SAY "FICHER DES UNITES PRIMAIRES"
      @ 3,5 SAY "*****"
      DO QUESFICH
      FICHPRIM=NOMFICH
      @ 10,5 SAY "FICHER DES UNITES SECONDAIRES"
      @ 11,5 SAY "*****"

```

```
DO QUESFICH
FICHSEC=NOMFICH
* Entrée des noms de variables
CLEAR
@ 2,5 SAY "VARIABLE A ESTIMER"
@ 3,5 SAY "*****"
DO QUESVAR
VARY = REPVAR
@ 10,5 SAY "VARIABLE EXOGENE DEFINISSANT LA PROBABILITE"
@ 11,5 SAY "*****"
DO QUESVAR
.VARZ = REPVAR
@ 18,5 SAY "VARIABLE IDENTIFIANT LES UNITES PRIMAIRES"
@ 19,5 SAY "*****"
DO QUESVAR
IDPRIM = REPVAR
* Entrée du code des valeurs manquantes
CLEAR
@ 2,5 SAY "VALEUR DE CODE AFFECTEE AUX DONNEES MANQUANTES"
@ 3,5 SAY "*****"
DO QUESVAL
VALMANQ = NB
CLEAR
*
* Ouverture des zones de travail
*
SELECT 1
USE &FICHPRIM
SELECT 2
USE &FICHSEC
WAIT
*
* Calculs
*
DO CALPRIM
*
* Edition
*
SET PRINTER ON
SET DEVICE TO PRINT
*
@ 2,5 SAY "VARIABLE ESTIMEE :"
@ 2,25 SAY VARY
@ 5,5 SAY "VARIABLE EXOGENE DEFINISSANT LA PROBABILITE :"
@ 5,53 SAY VARZ
*
@ 7,5 SAY "PLAN EQUIPROPABLE"
@ 8,5 SAY "*****"
@ 10,10 SAY "1° terme (inter-grappes)"
@ 10,40 SAY "2° terme (intra-grappes)"
@ 13,10 SAY "A ="
@ 13,13 SAY A1
@ 13,40 SAY "B ="
@ 13,43 SAY B1
```

@ 15,5 SAY "PLAN A PROBABILITES INEGALES"
@ 16,5 SAY "*****"
@ 18,10 SAY "1° terme (inter-grappes)"
@ 18,40 SAY "2° terme (intra-grappes)"
@ 21,10 SAY "A ="
@ 21,13 SAY A2
@ 21,40 SAY "B ="
@ 21,43 SAY B2

SET PRINTER OFF
SET DEVICE TO SCREEN
ENDCASE
ENDDO

```
* .....*
*
* VAR2PRO.PRO: FICHIER DE PROCEDURES POUR LE PROGRAMME VAR2DEG*
* (Version n°1)
*
*****
*
* 1) Procédures de calcul
*
-----
* 1 : PROCEDURE CALPRIM
*
* Sommation sur les unités primaires.
*-----
*          PROCEDURE CALPRIM
*
* Initialisation
*
*          A1=0
*          A2=0
*          P1=0
*          P2=0
*          N=0
*          Y=0
*          Z=0
*          SA1=0
*          SA2=0
*          SB1=0
*          SB2=0
*          NI=0
*          SYI=0
*          S2YI=0
*
* Boucle sur les unités primaires
*
*          SELECT 1
*          DO WHILE .NOT. EOF()
*          NOMANZ='DPRIM
*          DO CALSEC WITH VARY, NOMANZ
*          SELECT 1
*          ZI=VARY
*          SA1=SA1 + SYI*Z
*          N=N + 1
*          Y=Y + SYI
*          Z=Z + ZI
*          SA2=SA2 + SYI*Z/ZI
*          SB1=SB1 + NI*SYI
*          SB2=SB2 + NI*S2YI/ZI
*          ? RECD()
*          ?? NOMANZ, SA1, SB1, SA2, SB2
*          WAIT
*          SKIP
*          ENDDO
```

```
*
* Resultats
*
  A1=N^2*(SA1/N - (Y/N)^2)
  A2=Z*SA2 - Y^2
  B1=N*SB1
  B2=Z*SB2
  RETURN
```

```
*-----*
* 2 : PROCEDURE CALSEC
```

```
* Sommation sur les unités secondaires.
```

```
*-----*
  PROCEDURE CALSEC
  PARAMETERS Y, NO
  SELECT 2
  SET FILTER TO &IDPRIM=NO .AND. &Y#VALMANG
  COUNT TO NI
  SUM &Y, &Y^2 TO SY, SY2
  S2YI= (SY2 - SY^2/NI)/(NI-1)
  SYI=SY
* ?? NO, NI, SY, SY2, S2YI
* WAIT
  RETURN
```

```
* 2) Procédures utilitaires
```

```
*-----*
* 1 : PROCEDURE QUESFICH
```

```
* Entree d'un nom de fichier dans le programme VARSDMCI.
```

```
*-----*
  PROCEDURE QUESFICH
  BOUCLE = .T.
  DO WHILE BOUCLE
  ? 'Entrez le nom du fichier, sans extention ni chemin
d"accés.'
  ? '(8 caractères maximum).'
  ? 'il doit être dans le repertoire dbaseiii du disque dur.'
  ? ''
  ACCEPT ' nom :' TO NOMFICH
  NOMFICH = TRIM(NOMFICH) + '.DEF'
  NOMFICH = UPPER(NOMFICH)
  IF LEN(NOMFICH) >0 .AND. LEN(NOMFICH) <13
  BOUCLE = .F.
  ENDIF
  ENDDO
  RETURN
```


* 2 : PROCEDURE QUESVAR

* Entree d'un nom de variable dans le programme VAR2DEG.

```
PROCEDURE QUESVAR
BOUCLE = .T.
DO WHILE BOUCLE
? ''
ACCEPT 'Entrez le nom de la variable.(12 caractères
maximum) :' TO REPVAR
REPVAR = TRIM(REPVAR)
REPVAR = UPPER(REPVAR)
IF LEN(REPVAR) >0 .AND. LEN(REPVAR) <12
BOUCLE = .F.
ENDIF
ENDDO
RETURN
```

* 3 : PROCEDURE QUESVAL

* Entree de la valeur de code dans le programme VAR2DEG.

```
PROCEDURE QUESVAL
INPUT 'Entrez la valeur du code :' TO NB
RETURN
```

Annexe A.4.1

PROGRAMMES D'ESTIMATION POUR L'EXPLOITATION DE L'ENQUETE FINALE :

LISTAGE DES PROCEDURES SPSS

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration" .  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des .  
* fichiers originaux  
* préparation du fichier système unités secondaires pour l'estimation de totaux .  
* (fichier procédure ESTPUS.PR1)  
*****
```

```
SET LENGTH = 60 / WIDTH = 130.  
SET PRINTER = OFF.  
GET FILE = 'D:\USR\OLIV\DONNEES\CVIVIO3T.SPD'.  
COMPUTE IDMANZ = V1.  
IF ((V5 NE 8) AND (V7 LT 998)) YMANZ = V7.  
IF ((V5=8) OR (V7 GE 998)) YMANZ = 999.  
MISSING VALUE YMANZ (999).  
VARIABLE LABEL IDMANZ 'Identifiant de l"UP' / YMANZ 'Variable à estimer'.  
SET PRINTER = ON.  
DISPLAY VAR = IDMANZ, YMANZ.  
SET PRINTER = OFF.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD' / DROP = V1 TO V19.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation du fichier système unités primaires pour l'estimation de totaux  
* (fichier procédure ESTPUP.PR1)  
*****
```

SET PRINTER = OFF.

GET FILE = 'D:\USR\OLIV\DONNEES\CMANZ03T.SPD'.

* Correction des surface en fonction du type de mesure.

IF (V5 = 1) SURFCOR = 1.04889 * V4.

IF (V5 = 2) SURFCOR = V4.

IF (V5 = 3) SURFCOR = 1.14329 * V4.

VARIABLE LABEL SURFCOR 'Surface corrigée de la manzana'.

* Calcul du nombre de ménages ordinaires.

COMPUTE NBMENO = V6.

IF (V1=73) NBMENO = NBMENO-1.

IF (V1=53) NBMENO = NBMENO-1.

IF (V1=71) NBMENO = NBMENO-1.

IF (V1=32) NBMENO = NBMENO-1.

IF (V1=118) NBMENO = NBMENO-1.

IF (V1=126) NBMENO = NBMENO-3.

IF (V1=125) NBMENO = NBMENO-1.

IF (V1=200) NBMENO = NBMENO-1.

IF (V1=246) NBMENO = NBMENO-1.

IF (V1=298) NBMENO = NBMENO-1.

VARIABLE LABEL NBMENO 'Nombre de mén. ord. dans la manzana'.

* Calcul de la variable de stratification croisée ; regroup.

IF ((V2=1) AND (V3=1 OR V3=4)) STRC=1.

IF ((V2=1) AND (V3=2 OR V3=5)) STRC=2.

IF ((V2=1) AND (V3=3 OR V3=6)) STRC=3.

IF ((V2=1) AND (V3>6)) STRC=4.

IF ((V2=2) AND (V3=1 OR V3=4)) STRC=5.

IF ((V2=2) AND (V3=2 OR V3=5)) STRC=6.

IF ((V2=2) AND (V3=3 OR V3=6)) STRC=7.

IF ((V2=2) AND (V3>6)) STRC=8.

IF ((V2=3) AND (V3=1)) STRC=9.

IF ((V2=3) AND (V3=2 OR V3=5)) STRC=10.

IF ((V2=3) AND (V3=3)) STRC=11.

IF ((V2=3) AND (V3=4)) STRC=12.

IF ((V2=3) AND (V3=6)) STRC=13.

IF ((V2=3) AND (V3=7 OR V3=10)) STRC=14.

IF ((V2=3) AND (V3=8)) STRC=15.

IF ((V2=3) AND (V3=9)) STRC=16.

IF ((V2=3) AND (V3=11)) STRC=17.

IF ((V2=3) AND (V3=12)) STRC=18.

IF ((V2=4) AND (V3=1 OR V3=4)) STRC=19.

IF ((V2=4) AND (V3=2)) STRC=20.

IF ((V2=4) AND (V3=3 OR V3=6)) STRC=21.

IF ((V2=4) AND (V3=5)) STRC=22.

```
IF ((V2=4) AND (V3=8)) STRC=23.  
IF ((V2=4) AND (V3=9)) STRC=24.  
IF ((V2=4) AND (V3=10)) STRC=25.  
IF ((V2=4) AND (V3=11)) STRC=26.  
IF ((V2=4) AND (V3=12)) STRC=27.
```

```
IF ((V2=5) AND (V3=4 OR V3=5 OR V3=6)) STRC=28.  
IF ((V2=5) AND (V3=7 OR V3=8 OR V3=9)) STRC=29.  
VARIABLE LABEL STRC 'stratification croisée'.
```

```
COMPUTE IDMANZ = V1.  
COMPUTE ZMANZ = SURFCOR.  
COMPUTE STRDENS = V2.  
COMPUTE STRLOC = V3.  
COMPUTE STRDLOC =STRC.  
COMPUTE NBMENMAN = NBMENO.  
VARIABLE LABEL IDMANZ 'Identifiant de l"UP' / ZMANZ 'Surface de l"UP' /  
STRDENS 'Stratif densité du bâti' /  
STRLOC 'Stratif localisation' /  
STRDLOC 'Stratif densité/ localisation' /  
NBMENMAN 'Nombre d"US dans l"UP'.  
SORT CASES IDMANZ.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' / DROP = V1 TO V6,  
NBMENO, STRC, SURFCOR.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation des fichiers systèmes strates densité, densité x localisation  
* (fichier procédure ESTPSTR.PR1)  
*****
```

```
SET PRINTER = OFF.
```

```
* Fichier strates densité
```

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDENS.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDENS.SSD'.
```

```
* Fichier strates densité x localisation
```

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDLOC.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD'.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Estimation de totaux  
* Etape 1 : estimation sur les unités primaires, sommation sur les unités sec  
* ondaire (fichier procédure TOTUS.PR1)  
*****
```

SET PRINTER = OFF.

* Ouverture du fichier d'unités secondaires

CET FILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD'.

* Sommation sur le fichier d'US.

```
AGGREGATE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUP.TEM' /  
PRESORTED /  
BREAK = IDMANZ /  
NBENQMAN 'Nb de ménages ayant répondu' = NU(YMANZ) /  
MYMANZ 'Moyenne estimée de Y dans la manzana' = MEAN(YMANZ) /  
EYMANZ 'Ecart-type estimé dans la manzana' = SD(YMANZ).
```

* Merge des résultat avec le fichier d'unités primaires.

```
MATCH FILE = 'D:\USR\OLIV\DONNEES\TOTUP.TEM' /  
KEEP = IDMANZ, NBENQMAN, MYMANZ, EYMANZ /  
FILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' /  
KEEP = IDMANZ, NBMENMAN, ZMANZ, STRDENS, STRLOC, STRDLOC /  
BY = IDMANZ /  
MAP.
```

* Calcul des estimateurs et variances.

```
COMPUTE ESTTOTMA = NBMENMAN * MYMANZ.  
COMPUTE VARESTMA = (NBMENMAN * (NBMENMAN - NBENQMAN) / NBENQMAN)  
* EYMANZ ** 2.  
VARIABLE LABEL ESTTOTMA 'Estimation du total de Y dans L"UP' /  
VARESTMA 'Variance est. de l"estimation dans l"UP'.
```

* Traitement des UP vides.

RECODE NBENQMAN, MYMANZ, EYMANZ, ESTTOTMA, VARESTMA (SYSMIS = 0).

* Sauvegarde du fichier résultats d'UP.

SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.

* Edition des résultats.

```
SET PRINTER=ON / WIDTH=130 / LENGTH=60.  
REPORT FORMAT=LIST / VARIABLES = IDMANZ, STRDENS, STRLOC, STRDLOC,  
ZMANZ, NBMENMAN, NBENQMAN, MYMANZ, EYMANZ, ESTTOTMA, VARESTMA /  
BREAK = (NOBREAK).
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Estimation de totaux  
* Etape 2 : estimation sur les strates, sommation sur les unités primaires  
* Stratification croisée densité du bâti x localisation  
* (fichier procédure TOTUPDLO.PR1)  
*****
```

```
SET PRINTER = OFF.
```

```
* Ouverture du fichier d'unités primaires
```

```
GET FILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.
```

```
* Calcul des quantités à sommer
```

```
COMPUTE YDZ = ESTTOTMA / ZMANZ.  
COMPUTE Y2DZ2 = ESTTOTMA ** 2 / ZMANZ ** 2.  
COMPUTE VYDZ = VARESTMA / ZMANZ.  
COMPUTE VYDZ2 = VARESTMA / ZMANZ ** 2.
```

```
* Sommation sur le fichier d'UP.
```

```
SORT CASES STRDLOC.  
AGGREGATE OUTFILE = * /  
PRESORTED /  
BREAK = STRDLOC /  
STRDENS 'Stratification densité' = MAX(STRDENS) /  
NBMANZEQ 'Nb de manz. enquêt. dans la strate' = NU(YDZ) /  
MYDZ 'Moyenne de y/z dans la strate' = MEAN(YDZ) /  
MY2DZ2 'Moyenne de y2/z2 dans la strate' = MEAN (Y2DZ2) /  
MVYDZ 'Moyenne de V2(y)/z dans la strate' = MEAN (VYDZ) /  
MVYDZ2 'Moyenne de V2(y)/z dans la strate' = MEAN (VYDZ2) /  
SZENQ 'Somme des surfaces enquêtées' = SUM (ZMANZ).  
COMPUTE STRATE = STRDLOC.
```

```
* Merge des résultat avec le fichier des strates.
```

```
MATCH FILE = * /  
FILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD' /  
BY = STRATE /  
MAP.
```

```
* Calcul des estimateurs et variances.
```

```
COMPUTE ESTTOTST = ZSTRATE * MYDZ.  
COMPUTE T1 = MY2DZ2 * ZSTRATE ** 2 / (NBMANZEQ - 1).  
COMPUTE T2 = 2 * ZSTRATE * ESTTOTST * MYDZ / (NBMANZEQ - 1).  
COMPUTE T3 = ESTTOTST ** 2 / (NBMANZEQ - 1).  
COMPUTE T4 = ZSTRATE * MVYDZ.  
COMPUTE VARTOTST = T1 - T2 + T3 + T4.  
COMPUTE VAR2STR = (ZSTRATE ** 2) * MVYDZ2 / NBMANZEQ.  
COMPUTE VAR1STR = VARTOTST - VAR2STR.  
COMPUTE COEFTOT = 100 * SQRT(VARTOTST)/ESTTOTST.  
VARIABLE LABEL  
ESTTOTST 'Estimation du total de Y dans la strate' /  
VAR1STR 'Variance est., terme inter-UP' /  
VAR2STR 'Variance est., terme intra-UP' /  
VARTOTST 'Variance est. totale' /  
COEFTOT 'Coeff. de var. est.'.  
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.
```


* Sommatation sur les strates 1

```
AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = STRDENS /
ZSTRATE NBMANZEQ SZENQ ESTTOTST VAR1STR VAR2STR VARTOTST =
SUM (ZSTRATE NBMANZEQ SZENQ ESTTOTST VAR1STR VAR2STR VARTOTST).
COMPUTE STRATE = 29 + STRDENS.
COMPUTE COEFTOT = 100 * SQRT(VARTOTST) / ESTTOTST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.
```

* Sommatation sur les strates 2

```
GET FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
COMPUTE INDTOT = 1.
AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = INDTOT /
ZSTRATE NBMANZEQ SZENQ ESTTOTST VAR1STR VAR2STR VARTOTST =
SUM (ZSTRATE NBMANZEQ SZENQ ESTTOTST VAR1STR VAR2STR VARTOTST).
COMPUTE STRATE = 35.
COMPUTE COEFTOT = 100 * SQRT(VARTOTST) / ESTTOTST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.
```

* Sauvegarde du fichier résultat des strates.

```
FORMATS MYDZ (F7.6) MY2DZ2 (F9.8) MVDZ (F6.5) VAR1STR (F13.1) VAR2STR (F9.1)
VARTOTST (F13.1) COEFTOT (F7.2).
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.SSR'.
```

* Edition des résultats.

```
SORT CASES STRATE.
SET PRINTER=ON / WIDTH=130 / LENGTH=500.
REPORT FORMAT=LIST /
VARIABLES = STRATE (LABEL), ZSTRATE, NBMANZEQ, SZENQ, ESTTOTST,
VAR1STR, VAR2STR, VARTOTST ,COEFTOT /
BREAK = (NOBREAK) /
```

```
*****
Procédure SPSS : estimation à partir de l'enquête "démographie / migration"
Etape préli. : création des fichiers de données systèmes SPSS à partir des
fichiers originaux
préparation du fichier système unités secondaires pour l'estimation de moyenn.
(fichier procédure ESTPUS.PR1)
*****
```

```
SET LENGTH = 60 / WIDTH = 130.
SET PRINTER = OFF.
SET FILE = 'D:\USR\OLIV\DONNEES\CVIVIO3T.SPD'.
COMPUTE IDMANZ = V1.
IF ((V5 NE 8) AND (V7 LT 998)) YMANZ = V7.
IF ((V5=8) OR (V7 GE 998)) YMANZ = 999.
MISSING VALUE YMANZ (999).
VARIABLE LABEL IDMANZ 'Identifiant de l"UP"/ YMANZ 'Variable à estimer'.
SET PRINTER = ON.
DISPLAY VAR = IDMANZ, YMANZ.
SET PRINTER = OFF.
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD' / DROP = V1 TO V19.
```

```

*****
Procedure SPSS : estimation à partir de l'enquête "démographie / migration"
Etape préli. : création des fichiers de données systèmes SPSS à partir des
fichiers originaux
préparation du fichier système unités primaires pour l'estimation de moyennes
(fichier procédure ESTPUP.PR1)
*****

```

ET PRINTER = OFF.

ET FILE = 'D:\USR\OLIV\DONNEES\CMANZO3T.SPD'.

Correction des surface en fonction du type de mesure.

```

F (V5 = 1) SURFCOR = 1.04889 * V4.
F (V5 = 2) SURFCOR = V4.
F (V5 = 3) SURFCOR = 1.14329 * V4.
VARIABLE LABEL SURFCOR 'Surface corrigée de la manzana'.

```

Calcul du nombre de ménages ordinaires.

```

COMPUTE NBMENO = V6.
F (V1=73) NBMENO = NBMENO-1.
F (V1=53) NBMENO = NBMENO-1.
F (V1=71) NBMENO = NBMENO-1.
F (V1=32) NBMENO = NBMENO-1.
F (V1=118) NBMENO = NBMENO-1.
F (V1=126) NBMENO = NBMENO-3.
F (V1=125) NBMENO = NBMENO-1.
F (V1=200) NBMENO = NBMENO-1.
F (V1=246) NBMENO = NBMENO-1.
F (V1=298) NBMENO = NBMENO-1.
VARIABLE LABEL NBMENO 'Nombre de mén. ord. dans la manzana'.

```

Calcul de la variable de stratification croisée ; regroup.

```

IF ((V2=1) AND (V3=1 OR V3=4)) STRC=1.
IF ((V2=1) AND (V3=2 OR V3=5)) STRC=2.
IF ((V2=1) AND (V3=3 OR V3=6)) STRC=3.
IF ((V2=1) AND (V3>6)) STRC=4.

IF ((V2=2) AND (V3=1 OR V3=4)) STRC=5.
IF ((V2=2) AND (V3=2 OR V3=5)) STRC=6.
IF ((V2=2) AND (V3=3 OR V3=6)) STRC=7.
IF ((V2=2) AND (V3>6)) STRC=8.

IF ((V2=3) AND (V3=1)) STRC=9.
IF ((V2=3) AND (V3=2 OR V3=5)) STRC=10.
IF ((V2=3) AND (V3=3)) STRC=11.
IF ((V2=3) AND (V3=4)) STRC=12.
IF ((V2=3) AND (V3=6)) STRC=13.
IF ((V2=3) AND (V3=7 OR V3=10)) STRC=14.
IF ((V2=3) AND (V3=8)) STRC=15.
IF ((V2=3) AND (V3=9)) STRC=16.
IF ((V2=3) AND (V3=11)) STRC=17.
IF ((V2=3) AND (V3=12)) STRC=18.

IF ((V2=4) AND (V3=1 OR V3=4)) STRC=19.
IF ((V2=4) AND (V3=2)) STRC=20.
IF ((V2=4) AND (V3=3 OR V3=6)) STRC=21.
IF ((V2=4) AND (V3=5)) STRC=22.

```

```
IF ((V2=4) AND (V3=8)) STRC=23.  
IF ((V2=4) AND (V3=9)) STRC=24.  
IF ((V2=4) AND (V3=10)) STRC=25.  
IF ((V2=4) AND (V3=11)) STRC=26.  
IF ((V2=4) AND (V3=12)) STRC=27.
```

```
IF ((V2=5) AND (V3=4 OR V3=5 OR V3=6)) STRC=28.  
IF ((V2=5) AND (V3=7 OR V3=8 OR V3=9)) STRC=29.  
VARIABLE LABEL STRC 'stratification croisée'.
```

```
COMPUTE IDMANZ = V1.  
COMPUTE ZMANZ = SURFCOR.  
COMPUTE STRDENS = V2.  
COMPUTE STRLOC = V3.  
COMPUTE STRDLOC =STRC.  
COMPUTE NBMENMAN = NBMENO.  
VARIABLE LABEL IDMANZ 'Identifiant de l"UP' / ZMANZ 'Surface de l"UP' /  
STRDENS 'Stratif densité du bâti' /  
STRLOC 'Stratif localisation' /  
STRDLOC 'Stratif densité/ localisation' /  
NBMENMAN 'Nombre d"US dans l"UP'.  
SORT CASES IDMANZ.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' / DROP = V1 TO V6,  
NBMENO, STRC, SURFCOR.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation des fichiers systèmes strates densité, densité x localisation  
* (fichier procédure ESTPSTR.PR1)  
*****
```

SET PRINTER = OFF.

* Fichier strates densité

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDENS.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDENS.SSD'.
```

* Fichier strates densité x localisation

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDLOC.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD'.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Estimation de moyennes générales par unité secondaire  
* Etape 1 : estimation sur les unités primaires, sommation sur les unités sec  
* ondaire (fichier procédure MOYUS.PR1)  
*****
```

SET PRINTER = OFF.

* Ouverture du fichier d'unités secondaires

GET FILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD'.

* Sommation sur le fichier d'US.

AGGREGATE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUP.TEM' /
PRESORTED /
BREAK = IDMANZ /
NBENQMAN 'Nb de ménages ayant répondu' = NU(YMANZ) /
MYMANZ 'Moyenne estimée de Y dans la manzana' = MEAN(YMANZ).

* Merge des résultat avec le fichier d'unités primaires.

MATCH FILE = 'D:\USR\OLIV\DONNEES\TOTUP.TEM' /
KEEP = IDMANZ, NBENQMAN, MYMANZ /
FILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' /
KEEP = IDMANZ, NBMENMAN, ZMANZ, STRDENS, STRLOC, STRDLOC /
BY = IDMANZ /
MAP.

* Calcul des estimateurs et variances.

COMPUTE ESTTOTMA = NBMENMAN * MYMANZ.
VARIABLE LABEL ESTTOTMA 'Estimation du total de Y dans L"UP'.

* Traitement des UP vides.

RECODE NBENQMAN, MYMANZ, ESTTOTMA (SYSMIS = 0).

* Sauvegarde du fichier résultats d'UP.

SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.

* Edition des résultats.

SET PRINTER=ON / WIDTH=130 / LENGTH=60.
REPORT FORMAT=LIST / VARIABLES = IDMANZ, STRDENS, STRLOC, STRDLOC,
ZMANZ, NBMENMAN, NBENQMAN, MYMANZ, ESTTOTMA /
BREAK = (NOBREAK).

```

*****
Procedure SPSS : estimation à partir de l'enquête "démographie / migration"
Estimation de moyennes générales par unité secondaire
Etape 2 : estimation sur les strates, sommation sur les unités primaires,
sommation sur les strates Stratification croisée bâti x localisation
(fichier procédure MOYUPDLO.PR1)
*****

```

```

SET PRINTER = OFF.

```

```

* Ouverture du fichier d'unités primaires

```

```

GET FILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.

```

```

* Calcul des quantités à sommer

```

```

COMPUTE YDZ = ESTTOTMA / ZMANZ.
COMPUTE NDZ = NBMENMAN / ZMANZ.
COMPUTE Y2DZ2 = ESTTOTMA ** 2 / ZMANZ ** 2.
COMPUTE NYDZ2 = NBMENMAN * ESTTOTMA / ZMANZ ** 2.
COMPUTE N2DZ2 = NBMENMAN ** 2 / ZMANZ ** 2.

```

```

* Sommation sur le fichier d'UP.

```

```

SORT CASES STRDLOC.
AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = STRDLOC /
STRDENS 'Stratification densité' = MAX(STRDENS) /
NBMANZEQ 'Nb de manz. enquêt. dans la strate' = NU(YDZ) /
MYDZ 'Moyenne de y/z dans la strate' = MEAN(YDZ) /
MNDZ 'Moyenne de n/z dans la strate' = MEAN(NDZ) /
MY2DZ2 'Moyenne de y2/z2 dans la strate' = MEAN (Y2DZ2) /
MNYDZ2 'Moyenne de ny/z2 dans la strate' = MEAN (NYDZ2) /
MN2DZ2 'Moyenne de n2/z2 dans la strate' = MEAN (N2DZ2) /
SZENQ 'Somme des surfaces enquêtées' = SUM (ZMANZ).
COMPUTE STRATE = STRDLOC.

```

```

* Merge des résultat avec le fichier des strates.

```

```

MATCH FILE = * /
FILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD' /
BY = STRATE /
MAP.

```

```

* Calcul des estimateurs et variances.

```

```

COMPUTE ESTTOTY = ZSTRATE * MYDZ.
COMPUTE ESTTOTN = ZSTRATE * MNDZ.
COMPUTE ESTMOYST = MYDZ / MNDZ.
COMPUTE T1 = MY2DZ2.
COMPUTE T2 = 2 * ESTMOYST * MNYDZ2.
COMPUTE T3 = (ESTMOYST ** 2) * MN2DZ2.
COMPUTE VARA = (ZSTRATE ** 2) * (T1 - T2 + T3) / (NBMANZEQ - 1).
COMPUTE VARMOYST = VARA / ESTTOTN ** 2.
COMPUTE COEFTOT = 100 * SQRT(VARMOYST)/ESTMOYST.
VARIABLE LABEL
ESTTOTY 'Estimation du total de Y dans la strate' /
ESTTOTN 'Estimation du nombre d"up dans la strate' /
ESTMOYST 'Estimation de la moyenne de Y dans la strate' /
VARMOYST 'Variance est. de la moyenne' /

```

COEFTOT 'Coeff. de var. est.'
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.

* Sommmation sur les strates 1.

AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = STRDENS /
ZSTRATE NEMANZEQ SZENQ ESTTOTY ESTTOTN SVARA = SUM (ZSTRATE NEMANZEQ SZENQ
ESTTOTY ESTTOTN VARA).
COMPUTE STRATE = 29 + STRDENS.
COMPUTE ESTMOYST = ESTTOTY / ESTTOTN.
COMPUTE VARMOYST = SVARA / ESTTOTN ** 2.
COMPUTE COEFTOT = 100 * SQRT(VARMOYST)/ESTMOYST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.

* Sommmation sur les strates 2.

GET FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
COMPUTE INDTOT = 1.
AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = INDTOT /
ZSTRATE NEMANZEQ SZENQ ESTTOTY ESTTOTN SVARA = SUM (ZSTRATE NEMANZEQ SZENQ
ESTTOTY ESTTOTN SVARA).
COMPUTE STRATE = 35.
COMPUTE ESTMOYST = ESTTOTY / ESTTOTN.
COMPUTE VARMOYST = SVARA / ESTTOTN ** 2.
COMPUTE COEFTOT = 100 * SQRT(VARMOYST)/ESTMOYST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.

* Sauvegarde du fichier résultat des strates.

FORMATS MYDZ (F7.6) MNDZ (F7.6) MY2DZ2 (F9.8) MNYDZ2 (F9.8) MN2DZ2 (F9.8)
ESTTOTY (F8.0) ESTTOTN (F8.0) ESTMOYST (F6.4) VARMOYST (F8.6) COEFTOT (F6.4).
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.SSR'.

* Edition des résultats.

SORT CASES STRATE.
SET PRINTER=ON / WIDTH=130 / LENGTH=500.
REPORT FORMAT=LIST /
VARIABLES = STRATE (LABEL) ZSTRATE NEMANZEQ SZENQ ESTTOTY ESTTOTN ESTMOYST
VARMOYST COEFTOT /
BREAK = (NOBREAK) (SKIP(3)).


```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation du fichier système unités secondaires pour l'estimation de ratios  
* (fichier procédure RATPUS.PR1)  
*****
```

```
SET LENGTH = 60 / WIDTH = 130.  
SET PRINTER = OFF.  
GET FILE = 'D:\USR\OLIV\DONNEES\CVIVIO3R.SPD'.  
COMPUTE IDMANZ = V1.  
IF (V5 NE 8) XMANZ = 1.  
IF ((V5=8) OR (V7 GE 998)) XMANZ = 999.  
IF ((V5 NE 8) AND (MEN1 LT 998)) YMANZ = 1 - MEN1.  
IF ((V5=8) OR (MEN1 GE 998)) YMANZ = 999.  
MISSING VALUE XMANZ YMANZ (999).  
VARIABLE LABEL IDMANZ 'Identifiant de l"UP' / YMANZ 'Variable au numérateur' /  
XMANZ 'Variable au dénominateur'.  
SET PRINTER = ON.  
DISPLAY VAR = IDMANZ, YMANZ, XMANZ.  
SET PRINTER = OFF.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD' / DROP = V1 TO V19,  
MEN1, MEN7P, MENPROP.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation du fichier système unités primaires pour l'estimation de quotient.  
* (fichier procédure ESTPUP.PR1)  
*****
```

```
GET PRINTER = OFF.
```

```
GET FILE = 'D:\USR\OLIV\DONNEES\CMANZO3T.SPD'.
```

```
* Correction des surface en fonction du type de mesure.
```

```
IF (V5 = 1) SURFCOR = 1.04889 * V4.
```

```
IF (V5 = 2) SURFCOR = V4.
```

```
IF (V5 = 3) SURFCOR = 1.14329 * V4.
```

```
VARIABLE LABEL SURFCOR 'Surface corrigée de la manzana'.
```

```
* Calcul du nombre de ménages ordinaires.
```

```
COMPUTE NBMENO = V6.
```

```
IF (V1=73) NBMENO = NBMENO-1.
```

```
IF (V1=53) NBMENO = NBMENO-1.
```

```
IF (V1=71) NBMENO = NBMENO-1.
```

```
IF (V1=32) NBMENO = NBMENO-1.
```

```
IF (V1=118) NBMENO = NBMENO-1.
```

```
IF (V1=126) NBMENO = NBMENO-3.
```

```
IF (V1=125) NBMENO = NBMENO-1.
```

```
IF (V1=200) NBMENO = NBMENO-1.
```

```
IF (V1=246) NBMENO = NBMENO-1.
```

```
IF (V1=298) NBMENO = NBMENO-1.
```

```
VARIABLE LABEL NBMENO 'Nombre de mén. ord. dans la manzana'.
```

```
* Calcul de la variable de stratification croisée ; regroup.
```

```
IF ((V2=1) AND (V3=1 OR V3=4)) STRC=1.
```

```
IF ((V2=1) AND (V3=2 OR V3=5)) STRC=2.
```

```
IF ((V2=1) AND (V3=3 OR V3=6)) STRC=3.
```

```
IF ((V2=1) AND (V3>6)) STRC=4.
```

```
IF ((V2=2) AND (V3=1 OR V3=4)) STRC=5.
```

```
IF ((V2=2) AND (V3=2 OR V3=5)) STRC=6.
```

```
IF ((V2=2) AND (V3=3 OR V3=6)) STRC=7.
```

```
IF ((V2=2) AND (V3>6)) STRC=8.
```

```
IF ((V2=3) AND (V3=1)) STRC=9.
```

```
IF ((V2=3) AND (V3=2 OR V3=5)) STRC=10.
```

```
IF ((V2=3) AND (V3=3)) STRC=11.
```

```
IF ((V2=3) AND (V3=4)) STRC=12.
```

```
IF ((V2=3) AND (V3=6)) STRC=13.
```

```
IF ((V2=3) AND (V3=7 OR V3=10)) STRC=14.
```

```
IF ((V2=3) AND (V3=8)) STRC=15.
```

```
IF ((V2=3) AND (V3=9)) STRC=16.
```

```
IF ((V2=3) AND (V3=11)) STRC=17.
```

```
IF ((V2=3) AND (V3=12)) STRC=18.
```

```
IF ((V2=4) AND (V3=1 OR V3=4)) STRC=19.
```

```
IF ((V2=4) AND (V3=2)) STRC=20.
```

```
IF ((V2=4) AND (V3=3 OR V3=6)) STRC=21.
```

```
IF ((V2=4) AND (V3=5)) STRC=22.
```

```
IF ((V2=4) AND (V3=8)) STRC=23.  
IF ((V2=4) AND (V3=9)) STRC=24.  
IF ((V2=4) AND (V3=10)) STRC=25.  
IF ((V2=4) AND (V3=11)) STRC=26.  
IF ((V2=4) AND (V3=12)) STRC=27.
```

```
IF ((V2=5) AND (V3=4 OR V3=5 OR V3=6)) STRC=28.  
IF ((V2=5) AND (V3=7 OR V3=8 OR V3=9)) STRC=29.  
VARIABLE LABEL STRC 'stratification croisée'.
```

```
COMPUTE IDMANZ = V1.  
COMPUTE ZMANZ = SURFCOR.  
COMPUTE STRDENS = V2.  
COMPUTE STRLOC = V3.  
COMPUTE STRDLOC =STRC.  
COMPUTE NBMENMAN = NBMENO.  
VARIABLE LABEL IDMANZ 'Identifiant de l"UP' / ZMANZ 'Surface de l"UP' /  
STRDENS 'Stratif densité du bâti' /  
STRLOC 'Stratif localisation' /  
STRDLOC 'Stratif densité/ localisation' /  
NBMENMAN 'Nombre d"US dans l"UP'.  
SORT CASES IDMANZ.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' / DROP = V1 TO V6,  
NBMENO, STRC, SURFCOR.
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"  
* Etape préli. : création des fichiers de données systèmes SPSS à partir des  
* fichiers originaux  
* préparation des fichiers systèmes strates densité, densité x localisation  
* (fichier procédure ESTPSTR.PR1)  
*****
```

```
SET PRINTER = OFF.
```

```
* Fichier strates densité
```

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDENS.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDENS.SSD'.
```

```
* Fichier strates densité x localisation
```

```
INCLUDE 'D:\USR\OLIV\DONNEES\STRDLOC.SPL'.  
SAVE OUTFILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD'.
```

```
*****
* Procedure SPSS : estimation à partir de l'enquête "démographie / migration"
* Estimation de quotients
* Etape 1 : estimation sur les unités primaires, sommation sur les unités se
* ondairees (fichier procédure RATUS.PR1)
*****
```

SET PRINTER =OFF.

* Ouverture du fichier d'unités secondaires

GET FILE = 'D:\USR\OLIV\DONNEES\TOTUS.SSD'.

* Sommation sur le fichier d'US.

```
AGGREGATE OUTFILE = 'D:\USR\OLIV\DONNEES\TEMPUP.SSD' /
PRESORTED /
BREAK = IDMANZ /
NBENQMAN 'Nb de ménages ayant répondu' = NU(YMANZ) /
MYMANZ 'Moyenne estimée de Y dans la manzana' = MEAN(YMANZ) /
MXMANZ 'Moyenne estimée de X dans la manzana' = MEAN(XMANZ) /
```

* Merge des résultat avec le fichier d'unités primaires.

```
MATCH FILE = 'D:\USR\OLIV\DONNEES\TEMPUP.SSD' /
KEEP = IDMANZ, NBENQMAN, MYMANZ, MXMANZ /
FILE = 'D:\USR\OLIV\DONNEES\TOTUP.SSD' /
KEEP = IDMANZ, NBMENMAN, ZMANZ, STRDENS, STRLOC, STRDLOC /
BY = IDMANZ /
MAP.
```

* Calcul des estimateurs et variances.

```
COMPUTE ESTTOTY = NBMENMAN * MYMANZ.
COMPUTE ESTTOTX = NBMENMAN * MXMANZ.
VARIABLE LABEL ESTTOTY 'Estimation du total de Y dans L"UP' /
ESTTOTX 'Estimation du total de X dans L"UP'.
```

* Traitement des UP vides.

RECODE NBENQMAN, MYMANZ, ESTTOTY, ESTTOTX (SYSMIS = 0).

* Sauvegarde du fichier résultats d'UP.

```
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.
```

* Edition des résultats.

```
SET PRINTER=ON / WIDTH=130 / LENGTH=60.
REPORT FORMAT=LIST / VARIABLES = IDMANZ, STRDENS, STRLOC, STRDLOC, ZMANZ,
NBMENMAN, NBENQMAN, MYMANZ, MXMANZ, ESTTOTY, ESTTOTX /
BREAK = (NOBREAK).
```

```
*****  
* Procedure SPSS : estimation à partir de l'enquête "démographie / migr  
* Estimation de quotients  
* Etape 2 : estimation sur les strates, sommation sur les unités primai  
* sommation sur les strates Stratification croisée bâti x localis  
* (fichier procédure RATUPDLO.PR1)  
*****
```

SET PRINTER = OFF.

* Ouverture du fichier d'unités primaires

GET FILE = 'D:\USR\OLIV\RESULT\TOTUP.SSR'.

* Calcul des quantités à sommer

```
COMPUTE YDZ = ESTTOTY / ZMANZ.  
COMPUTE XDZ = ESTTOTX / ZMANZ.  
COMPUTE Y2DZ2 = ESTTOTY ** 2 / ZMANZ ** 2.  
COMPUTE XYDZ2 = ESTTOTX * ESTTOTY / ZMANZ ** 2.  
COMPUTE X2DZ2 = ESTTOTX ** 2 / ZMANZ ** 2.
```

* Sommation sur le fichier d'UP.

```
SORT CASES STRDLOC.  
AGGREGATE OUTFILE = * /  
PRESORTED /  
BREAK = STRDLOC /  
STRDENS 'Stratification densité' = MAX(STRDENS) /  
NBMANZEQ 'Nb de manz. enquêt. dans la strate' = NU(YDZ) /  
MYDZ 'Moyenne de y/z dans la strate' = MEAN(YDZ) /  
MXDZ 'Moyenne de x/z dans la strate' = MEAN(XDZ) /  
MY2DZ2 'Moyenne de y2/z2 dans la strate' = MEAN (Y2DZ2) /  
MXYDZ2 'Moyenne de xy/z2 dans la strate' = MEAN (XYDZ2) /  
MX2DZ2 'Moyenne de x2/z2 dans la strate' = MEAN (X2DZ2) /  
SZENQ 'Somme des surfaces enquêtées' = SUM (ZMANZ).  
COMPUTE STRATE = STRDLOC.
```

* Merge des résultat avec le fichier des strates.

```
MATCH FILE = * /  
FILE = 'D:\USR\OLIV\DONNEES\STRDLOC.SSD' /  
BY = STRATE /  
MAP.
```

* Calcul des estimateurs et variances.

```
COMPUTE ESTTOTY = ZSTRATE * MYDZ.  
COMPUTE ESTTOTX = ZSTRATE * MXDZ.  
COMPUTE ESTRATST = MYDZ / MXDZ.  
COMPUTE T1 = MY2DZ2.  
COMPUTE T2 = 2 * ESTRATST * MXYDZ2.  
COMPUTE T3 = (ESTRATST ** 2) * MX2DZ2.  
COMPUTE VARA = (ZSTRATE ** 2) * (T1 - T2 + T3) / (NBMANZEQ - 1).  
COMPUTE VARRATST = VARA / ESTTOTX ** 2.  
COMPUTE COEFTOT = 100 * SQRT(VARRATST)/ESTRATST.  
VARIABLE LABEL  
ESTTOTY 'Estimation du total de Y dans la strate' /  
ESTTOTX 'Estimation du total de X dans la strate' /  
ESTRATST 'Estimation du quotient Y/X dans la strate' /  
VARRATST 'Variance est. du quotient' /
```

COEFTOT 'Coeff. de var. est.'
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.

* Sommmation sur les strates 1.

AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = STRDENS /
ZSTRATE NBMANZEQ SZENQ ESTTOTY ESTTOTX SVARA = SUM (ZSTRATE NBMANZEQ SZENQ
ESTTOTY ESTTOTX VARA).
COMPUTE STRATE = 29 + STRDENS.
COMPUTE ESTRATST = ESTTOTY / ESTTOTX.
COMPUTE VARRATST = SVARA / ESTTOTX ** 2.
COMPUTE COEFTOT = 100 * SQRT(VARRATST)/ESTRATST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM'.

* Sommmation sur les strates 2.

GET FILE = 'D:\USR\OLIV\RESULT\TOTST.TEM'.
COMPUTE INDTOT = 1.
AGGREGATE OUTFILE = * /
PRESORTED /
BREAK = INDTOT /
ZSTRATE NBMANZEQ SZENQ ESTTOTY ESTTOTX SVARA = SUM (ZSTRATE NBMANZEQ SZENQ
ESTTOTY ESTTOTX SVARA).
COMPUTE STRATE = 35.
COMPUTE ESTRATST = ESTTOTY / ESTTOTX.
COMPUTE VARRATST = SVARA / ESTTOTX ** 2.
COMPUTE COEFTOT = 100 * SQRT(VARRATST)/ESTRATST.
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.
ADD FILE = 'D:\USR\OLIV\RESULT\STRDLOC.TEM' /
FILE = 'D:\USR\OLIV\RESULT\TOTUN.TEM'.

* Sauvegarde du fichier résultat des strates.

FORMATS MYDZ (F7.6) MXDZ (F7.6) MY2DZ2 (F9.8) MXDZ2 (F9.8) MX2DZ2 (F9.8)
ESTTOTY (F8.0) ESTTOTX (F8.0) ESTRATST (F6.4) VARRATST (F8.6) COEFTOT (F6.4).
SAVE OUTFILE = 'D:\USR\OLIV\RESULT\STRDLOC.SSR'.

* Edition des résultats.

SORT CASES STRATE.
SET PRINTER=ON / WIDTH=130 / LENGTH=500.
REPORT FORMAT=LIST /
VARIABLES = STRATE (LABEL) ZSTRATE NBMANZEQ SZENQ ESTTOTY
ESTTOTX ESTRATST VARRATST COEFTOT /
BREAK = (NOBREAK) (SKIP(3)).

BIBLIOGRAPHIE SELECTIVE

A) Bibliographie générale

- Analyse des données

- BENZECRI J.P. et coll., 1973. L'analyse des données. Tome 1 : La taxinomie, Tome 2 : L'analyse des correspondances. DUNOD, Paris.
- BENZECRI J.P. et coll., 1980. La pratique de l'analyse des données. vol. 2 : abrégé théorique, étude de cas modèles. DUNOD, Paris.
- CHANDON J.L., PINSON S., 1981. Analyse typologique, théories et applications. MASSON, Paris.
- DIDAY E. et coll., 1980 : Optimisation en classification automatique (2 vol.). INRIA, Paris.
- JAMBU M., 1978. Classification automatique pour l'analyse des données, vol. 1 : Méthodes et algorithmes, vol. 2 : Logiciels. DUNOD, Paris.
- LEBART L., MORINEAU A., FENELON J.P., 1982. Traitement des données statistiques, méthodes et programmes. DUNOD, Paris.
- MALINVAUDE E., 1981. Méthodes statistiques de l'économétrie. DUNOD, Paris, 846 p.
- VOLLE M., 1981. L'analyse des données. ECONOMICA, Paris.

- Théorie des sondages et variables régionalisées

- Ouvrage collectif, 1987. Les sondages. Journées d'étude en statistique d'octobre 1986. DROESBEKE, FICHET, TASSI éditeurs.
- ANDRIANASOLO H.H., 1987. Analyse statistique des données de télédétection, statistiques agricoles, application sur Madagascar. Thèse de doctorat en mathématiques appliquées et sciences sociales, EHESS, Paris.
- ASSELIN L.M., 1984. Techniques de sondage avec application à l'Afrique. GAETAN MORIN, Québec.
- COCHRAN W.G., 1977. Sampling techniques (3^e édition). J. WILLEY, New York.
- DE KEERSMAECKER M.L., 1987. Stratégie d'échantillonnage des données de terrain intégrées dans l'analyse des images satellitaires. L'espace géographique, n° 3, pp. 195-205.
- DELINCE J., 1983. Estimations robustes de densité dans le plan, étude comparée de techniques d'échantillonnage, application en forêt. Université Catholique de Louvain, Faculté des Sciences Agronomiques. Louvain.
- DEMING W.E., 1950. Some theory of sampling. J.WILLEY, New York.

- DEROO M., DUSSAIX A.M., 1980. Pratique et analyse des enquêtes par sondage. PUF, Paris.
- DESABIE J., 1966. Théorie et pratique des sondages. DUNOD, Paris.
- DEVILLE J.C., ROY G., 1984. L'échantillon maître fait peu neuve. Courriers des statistiques n° 29, Paris.
- GORDON M., 1979. Analyse partielle des relations interspécifiques entre espèces prises trois à trois. OEcologica Plantarum, tome 14, n° 3.
- GORDON M. et CAMARASA J.M., 1977. Un test hypergéométrique direct pour analyser l'hétérogénéité d'un échantillon "linéaire". OEcologica Plantarum, tome 12, n° 1.
- GOURIEROUX Ch., 1981. Théorie des sondages. ECONOMICA, Paris.
- GROSBRAS J.M. 1987. Méthodes statistiques des sondages, ECONOMICA, Paris
- GUIDICELLI X., LANLY J.P., OUAKAM J.B., PIETRI M., 1972(ISUP). Application de la théorie des processus aléatoires à l'estimation de la précision d'un inventaire forestier par échantillonnage systématique. Annales des Sciences Forestières, Institut National de la Recherche Agronomique. Paris.
- HOUSEMAN E., 1975. Area frame sampling in agriculture. Statistical Reporting Service, United States Department of Agriculture.
- MATHERON G., 1965. Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature. MASSON, Paris.
- MEYER ROUX J., 1981. Cartographie et statistique en télédétection. SCEES.
- RAJ D., 1968. Sampling Theory. MC GRAW HILL, New York.
- RAJ D., 1972. The design of sample surveys. MC GRAW HILL, New York.
- RAO J.N.K., HARTLEY H.O., COCHRAN W.G., 1962. On a simple procedure of unequal probability sampling without replacement. Journal of Royal Statistical Society B24, pp. 482 - 491.
- SUKHATME P.V., SUKHATME B.V., 1970. Sampling theory of surveys with applications. FAO, Rome.
- SAEBO H.V. Land use and environmental statistics obtained by point sampling. Meddelanden fran Staten Skogforskningsinstitut, n° 1, pp. 317-340.
- THIONET P., 1958. La théorie des sondages. INSEE, Paris.
- YATES F., 1981. Sampling methods for censuses and surveys (4^e édition). GRIFFIN, Londres.

_ Collecte de données démographiques

ANTOINE P., 1984. Recensement : une efficacité relative. Actuel développement, n° 59.

NATIONS UNIES et CEA, 1974. Manuel des enquêtes par sondage en Afrique.

TABUTIN D., 1984. La collecte des données en démographie. ORDINA EDITIONS, Liège.

- Utilisation des données de télédétection aérienne et spatiale pour des estimations démographiques

ADENIYI P. O., 1983. An aerial photographic method for estimating urban population. Photogrammetric engineering and remote sensing, vol. 49, n° 4, pp 545-560.

BALLUT A., GAUTHIER M., 1983. Une méthode pour la production de données socio-économiques spatialisées dans le cadre d'un schéma directeur d'aménagement et d'urbanisme. IAURIF. Paris.

GREEN N.E., 1956. Aerial photographic analysis of residential neighbourhoods : an evaluation of data accuracy. Social forces, n° 35, pp. 142-247.

HOLZ R.K., HUFF D.L., MAYFIELD R.C., 1969. Urban spatial structure based on remote sensing imagery. Sixth International Symposium Remote Sensing Environment, vol. 2, pp. 819 - 830.

HSU S.Y., 1971. Population estimation. Photogrammetric engineering and remote sensing, vol. 37, n° 5, pp. 819 - 830.

IAURIF, ROC, SIDAU, SEE, 1983. Les données urbaines. Système permanent de production rapide et économique. Recherche et synthèse. Paris.

KRAUS S.P., SENGER L.W., RYERSON J.M. 1974. Estimating population from photographically determined residential land use types. Remote sensing of environment, n° 3, pp. 35 - 42.

LO C.P., WELCH R., 1977. Chinese urban population estimates. Annals of the Association of American Geographers, vol. 67, n° 2, pp 246 - 250.

PORTER P.W., 1956. Population distribution and land use in Liberia. London School of Economics and Political Sciences. Thèse de Ph. D.

VERNIERE M., 1978. Méthode d'analyse quantitative de la croissance urbaine dans l'espace et dans le temps. Exemple d'une banlieue de Dakar (Sénégal). Photo interprétation, n° 1, pp. 34 - 55.

WATKINS J.F., 1985. Small area production estimates using aerial photography. Photogrammetric engineering and remote sensing, vol. 51, n° 2, pp. 1933 - 1935.

- Télédétection spatiale en milieu urbain

BALLUT A., 1980. L'étude de l'occupation du sol par satellite. Cahiers de l'IAURIF, vol. 61. Paris.

BOQUET E., 1980. Réactualisation du mode d'occupation du sol par l'utilisation des simulations SPOT (81 et 83). Mémoire de maîtrise de géographie. Université Paris VII. Paris.

CNRS, 1982. Débats des journées télédétection en milieu urbain (6 - 7 mai 1982). CERG, Paris.

IGN, 1984. Recherches méthodologiques sur l'utilisation de la télédétection spatiale en milieu urbain. IGN, Paris.

THIBAUT C. BIANCALE M. DELAVIGNER., 1983. La télédétection par satellite appliquée au milieu urbain : test de validité de l'indice de végétation dans la proche couronne. IAURIF, Paris.

WILMET J., SOYER J., 1982. Lumbumbashi et le sud-est du haut Shaba : interprétation de données LANDSAT. Bulletin de la Société Belge d'Etude Géographiques. pp 87 - 100.

- Informatique, programmation

COHEN R., 1986. Programmer en dbase III et dbase III plus. éditions du PSI, Paris.

- Logiciels utilisés

DBASE III PLUS (ASTHON TATE) : Gestion de base de données.

SPSS PC + (SPSS inc.) : Traitements statistiques.

STAT ITCF (Institut Technique des Céréales et Fourages) : Traitements statistiques.

ADAD MICRO (ADAD) : Analyse des données.

WORD (MICROSOFT) : Traitement de texte.

B) Publications et notes de travail de l'équipe ORSTOM
(peuvent être consultées au Département SDU de l'ORSTOM)

Articles

- DUREAU F. et GUILLAUME A., 1984. - La démographie depuis l'espace : un nouveau système d'observation?. Paris, STATECO n° 38, pp. 5-46.
- DUREAU F., 1986. - A propos du traitement informatique de données localisées. Une expérience en cours : télédétection et observation des populations urbaines. Paris, ORSTOM, collection Colloques et séminaires, Numéro sur le traitement des données localisées, pp.263-286.
- DUREAU F. et BARBARY O., 1987. - Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale. A paraître dans la collection Informatique et sciences humaines, EHESS, 42p.
- MICHEL A., DUREAU F., LORTIC B., SOURIS M., 1987. - Mise au point des méthodes d'analyse des images satellite à haute résolution et évaluation des informations fournies par ces images. 1 : Etude statistique du descripteur "densité du bâti". Présentation des classifications 2. Mise en évidence des réseaux routiers sur une image SPOT panchromatique. Etude de faisabilité sur Quito (Equateur). Paris, Service Technique de l'Urbanisme, Bilan sur la télédétection urbaine en France, Octobre 1987.

Communications à des colloques

- BARBARY O., 1988. - Développement d'une méthode de sondage démographique en milieu urbain intégrant l'information satellitaire. Présentation des principaux résultats obtenus à Marseille et à Quito. Communication au Colloque sur les sondages aréolaires dans les pays andins, Quito (Equateur), 2-7 mai 1988, 7 p.
- DUREAU F. et GUILLAUME A., 1985. - La population dans l'espace. Télédétection et observation démographique des villes des pays en développement. Communication au XXe congrès général de l'Union Internationale pour l'Etude Scientifique de la Population, Florence (Italie), 5-12 Juin 1985, 19 p.
- DUREAU F., LORTIC B., MICHEL A., SOURIS M., 1987. - Télédétection et système d'information géographique. Communication au Forum International de l'Instrumentation et de l'Information géographiques, Lyon (France), 10-13 Juin 1987, 15 p.
- DUREAU F., LORTIC B., MICHEL A., SOURIS M., 1987. - Informatique, télédétection et observation des populations urbaines. Une recherche en cours à Quito (EQUATEUR). Communication à la Primera Conferencia Latinoamericana sobre informatica en geografia, San José, Costa Rica, 5-9 Octobre 1987, 29 p.
- DUREAU F., BARBARY O., LORTIC B., MICHEL A., 1988. - Utiliser la télédétection spatiale pour produire rapidement des données démographiques en ville. L'exemple de Quito (Equateur). Communication au colloque Gestion urbaine et développement ISTED, Lyon (FRANCE), 20-22 septembre 1988, 6p.

- DUREAU F., 1988. - A propose de l'analyse des systèmes résidentiels. Présentation de l'enquête Migrations réalisée à Quito (EQUATEUR) en décembre 1987. Communication présentée au colloque Migrations, changements sociaux et développement, ORSTOM, Paris, 20-22 septembre 1988, 17 p.
- MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F., 1987. - El uso de la teledeteccion para la observacion de las poblaciones urbanas. Investigacion en curso en Quito (ECUADOR). Communication au II Simposio Latinoamericano sobre sensores remotos, Columbia, Bogota, 16-20 Novembre 1987, 34 p.
- MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F., 1987. - L'utilisation de la télédétection pour l'observation des populations urbaines. Une recherche en cours à Quito, (EQUATEUR). Communication au colloque SPOT, Paris, 23-27 Novembre 1987, 9p.
- MICHEL A., LORTIC B., BARBARY O., DUREAU F., SOURIS M., 1988. - Segmentation et classification d'une image satellite SPOT en milieu urbain. Application à la ville de Quito (Equateur). Communication au colloque PIXIM 88, Paris, 24-28 Octobre 1988, 17 p.

Notes de travail et rapports

- Equipe ORSTOM (UR 406, ATOB, Unité d'infographie), 1986. - Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville. Rapport intermédiaire - Avril 1986. Paris, 190 p.
- Equipe ORSTOM (UR 406, ATOB, Unité d'infographie), 1987. - Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville. Rapport final - Août 1987. Paris, 53p.+538 p.
- Equipe ORSTOM (UR 15 du département SDU, Unités de télédétection et d'infographie du LIA), 1988. - L'utilisation de l'imagerie SPOT pour l'observation démographique en milieu urbain. Rapport final - Avril 1988. Paris, 22 p. + 195 p. ann.
- BARBARY O., 1988. - Mise au point de plans de sondage pour l'estimation de caractères démographiques à Marseille. Evaluation des précisions permises par l'emploi de diverses techniques de sélection, de stratification et d'estimation. Quito, 184 p.
- BARBARY O., 1988. - Définition et mise en oeuvre d'un plan de sondage aréolaire sur image SPOT pour une enquête démographique à Quito (Equateur). Quito, 204 p.
- BARBARY O., DUREAU F., MICHEL A., 1988. - Définition et mise en oeuvre d'un plan de sondage aréolaire sur image SPOT pour une enquête démographique à Quito (EQUATEUR). Quito, 204 p.
- DUREAU F. et MICHEL A., 1986. - Quito : enquête morphologie urbaine. Instructions aux enquêteurs. (2 versions : français et espagnol). Quito, 19 p.
- DUREAU F., 1987. - Quito : enquête bâti/population. Instructions aux enquêteurs. (2 versions : français et espagnol). Quito, 27 p.

- DUREAU F., 1987. - Quito. Enquête démographie/migrations. Instructions aux enquêteurs. (2 versions : français et espagnol). Quito, 33 p.
- DUREAU F., 1988. - Eléments de calcul des coûts d'enquête à Quito. Quito, 13 p.
- EBERHARD J.M., 1987 - Télédétection urbaine : éléments de caractérisation géographique des quartiers de Quito à l'aide d'une image SPOT. Mémoire de DEA, EHESS, 98p.
- MICHEL A., 1986 et version actualisée 1987. - Le point sur la télédétection urbaine en 1986 à travers la bibliographie française et anglo-saxonne sur le sujet. Recueil des fiches de dépouillement des articles et ouvrages. ORSTOM, 49p.
- MICHEL A., 1987. - Mise au point des méthodes d'analyse des images satellite à haute résolution : étude statistique du descripteur "densité du bâti", présentation des classifications. (Image TM Marseille 1983). Paris, 99p.
- MICHEL A., 1987. - Influence des facteurs météorologiques sur la probabilité d'obtention d'une image satellite SPOT ou Thematic Mapper. Etude sur le site de Quito, Quito 14 p.

LISTE DES TABLEAUX, FIGURES ET PHOTOGRAPHIES

	Pages
Tableau 2.1 - Synthèse de sept ouvrages de référence consacrés aux sondages.....	41
Tableau 2.2 - Utilisation de données de télédétection aérienne et spatiale pour des estimations démographiques.....	49
Tableau 2.3 - Quelques exemples d'estimations de population à partir de photographies aériennes et images satellite.....	51
 Tableau 3.1 : Les étapes de la recherche et les choix méthodologiques effectués (synthèse).....	 61
Figure 4.1 - Localisation de la zone d'étude sur Marseille.....	70
Figure 4.2 - Marseille . Découpage INSEE en districts de recensement.....	73
Figure 4.3 - L'implantation géographique des trois types de données à intégrer dans la base de données sur Marseille.....	75
Figure 4.4 - Marseille. Zonage morphologique d'après l'analyse visuelle d'une mosaïque photographique au 1/23 000 de 1982.....	78
Figure 4.5 - Marseille. Découpage INSEE en districts de recensement, restitution au 1/5 000.....	79
Tableau 4.6 - Relation morphologie. Liste des attributs.....	80
Tableau 4.7 - Relation population. Liste des attributs.....	80
Figure 4.8 - Plan d'expérience sur Marseille.....	82
Figure 4.9 - Schéma d'établissement de la typologie de la morphologie urbaine de Marseille.....	85
Figure 4.10 - Schéma de l'étape de mise au point et d'évaluation de plans de sondages stratifiés à Marseille.....	87
 Tableau 5.1 - Extraits du tableau de données morphologiques, variables originales et descripteurs qualitatifs recodés.....	 90
Figure 5.2 - Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2) Principales modalités contribuant aux deux premiers facteurs, projetées sur le plan 1 x 2.....	 93
Figure 5.3 - Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2) Représentation simultanée des lignes et des colonnes, projetées sur le plan 1 x 2.....	 94
Figure 5.4 - Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3) Principales modalités contribuant aux facteurs 1 et 3, projetées sur le plan 1 x 3.....	 96
Figure 5.5 - Analyse factorielle du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3) Représentation simultanée des lignes et des colonnes, projetées sur le plan 1 x 3.....	 97
Figure 5.6 - Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie Histogramme des indices de niveau de la hiérarchie des partitions.....	 99
Figure 5.7 - Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie Projection des centres de classes de la partition en 8 classes sur le plan factoriel 1 x 2.....	 100

	Pages
Figure 5.8 - Classification ascendante hiérarchique du nuage global : 441 zones bâties x 40 modalités descriptives de la morphologie Projection des centres de classes de la partition en 8 classes sur le plan factoriel 1 x 3.....	101
Tableau 5.9 - Analyse factorielle du sous fichier des zones à grands bâtiments ; 198 zones x 40 modalités descriptives de la morphologie Contribution des modalités des descripteurs morphologiques aux 7 premiers facteurs.....	105
Figure 5.10 - Analyse factorielle du sous-fichier des zones à grands bâtiments : 198 zones x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2) Principales modalités contribuant aux facteurs 1 et 2, projetées sur le plan 1x2.....	107
Figure 5.11 - Analyse factorielle du sous-fichier des zones à grands bâtiments : 198 zones x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2) Représentation des zones projetées sur le plan 1x2.....	108
Figure 5.12 - Classification ascendante hiérarchique du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie Histogramme des indices de niveau de la hiérarchie de partitions.....	109
Figure 5.13 - Classification ascendante hiérarchique du sous-fichier des zones à grands bâtiments : 198 zones bâties x 40 modalités descriptives de la morphologie Projection des centres de classes de la partition en 5 classes sur le plan factoriel 1 x 2.....	110
Tableau 5.14 - Analyse factorielle du sous-fichier des zones à petits bâtiments ; 191 zones x 40 modalités descriptives de la morphologie. Contribution des modalités des descripteurs morphologiques aux 7 premiers facteurs.....	112
Figure 5.15 - Analyse factorielle du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 2) Principales modalités contribuant aux facteurs 1 et 2, projetées sur le plan 1x2.....	113
Figure 5.16 - Analyse factorielle du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie (plan factoriel 1 X 3) Principales modalités contribuant aux facteurs 1 et 3, projetées sur le plan 1x3.....	115
Figure 5.17 - Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie Histogramme des indices de niveau de la hiérarchie de partitions.....	116
Figure 5.18 - Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie Projection des centres de classes de la partition en 7 classes sur le plan factoriel 1 x 2.....	117
Figure 5.19 - Classification ascendante hiérarchique du sous-fichier des zones à petits bâtiments : 191 zones bâties x 40 modalités descriptives de la morphologie Projection des centres de classes de la partition en 7 classes sur le plan factoriel 1 x 3.....	118
Figure 5.20 - Fréquence des différentes modalités des descripteurs morphologiques selon les strates	122

	Pages
Tableau 5.21 - Surfaces occupées par les strates (récapitulation d'une interrogation du système SAVANE)	123
Figure 6.1 - $f(n/N)=(N-n)/n$ c'est-à-dire $g(t)=(1/t) - 1$	135
Tableau 6.2 - Précision de l'estimation du total de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.	148
Tableau 6.3 : Précision de l'estimation de la moyenne de densité de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.....	150
Figure 6.4 - Précision des estimations partielles du total de population sans doubles comptes au taux de 5 %	143
Figure 6.5 - Précision des estimations partielles de la moyenne de la densité de population sans doubles comptes au taux de 5 %	143
Figure 6.6 - L'intégration et les croisements de données sous dbase III	165
Tableau 7.1.a - Résultats du test mené avec le programme VARECH. Estimation du total de population sans doubles comptes : deux plans de sondage, trois estimateurs, neuf strates morphologiques synthétiques, taux de 6,6% appliqué a une base de 2619 districts (R.G.P. Marseille 1982).	170
Graphique 7.1.b - Résultats du test mené avec le programme VARECH Estimation du total de population sans doubles comptes. Comparaison graphique des coefficients de variation pour les trois techniques de sondages (allocation optimale)	171
Tableau 7.2 - Liste des caractères démographiques ayant fait l'objet du test	187
Tableau 7.3.a - Résultats du test mené avec le programme VARECH Coefficients de variation des estimations stratifiées et non stratifiées de 15 caractères démographiques (effectifs). Mesure du gain dû à la stratification	174
Graphique 7.3.b - Résultats du test mené avec le programme VARECH Comparaison, pour trois sondages, des coefficients de variation des estimateurs stratifiés de 15 mesures d'effectifs. Allocation optimale aux strates	176
Tableau 7.4 - Résultats du test mené avec le programme VARECH Coefficients de variation des estimations stratifiées et non stratifiées de 13 caractères démographiques (ratios). Mesure du gain dû à la stratification	179
Tableau 7.5 - Résultats du test mené avec le programme VARECH. Coefficients de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques (première partie, estimation sans biais d'effectifs).....	182
Tableau 7.5 - Résultats du test mené avec le programme VARECH. Coefficients de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (seconde partie, estimation de ratios)	183
Tableau 7.6 - Résultats du test mené avec le programme VARECH. Quantités à estimer et intervalles de confiance à 95% par strate pour trois caractères démographiques.....	186
Tableau 7.7 - Liste des caractères démographiques considérés dans le chapitre.....	198
Tableau 7.8.a - Coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.....	191
Graphique 7.8.b - Comparaison des gains de précision apportés par trois stratifications morphologiques à l'estimation de 11 caractères démographiques (en % du coefficient de variation non stratifiée)	192

	Pages
Tableau 7.9 a. - Coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.....	196
Graphique 7.9.b - Comparaison des gains de précision apportés par trois stratifications morphologiques à l'estimation de 11 caractères démographiques (en % du coefficient de variation non stratifiée)	197
Figure 8.1 - Schéma du nombre de districts par strate dans les deux stratifications spatiales.....	203
Tableau 8.2. - Croisement des stratifications spatiales et morphologiques, nombre de classes des stratifications résultantes.....	204
Tableau 8.3 - Coefficients de variation et gain de précision obtenus avec les stratifications morphologiques et spatiales pour trois caractères démographiques.....	205
Tableau 8.4 - Coefficients de variation et gains de précision obtenus sur l'estimation de trois caractères démographiques à partir de stratifications croisant un critère morphologique et un critère de localisation.....	207
Figure 9.1 - Plan d'expérience sur Quito.....	217
Figure 9.2 - Schéma de la démarche de classification.....	223
Figure 9.3 - Structure du fichier descriptif des bâtiments.....	224
Figure 9.4 - Structure du fichier descriptif des îlots.....	225
Tableau 9.5 - Typologie des bâtiments.....	226
Tableau 9.6 - Typologie des îlots.....	227
Tableau 10.1 - Liste des champs du fichier "pâté de maisons".....	236
Tableau 10.2 - Matrice des corrélations entre variables démographiques (logiciel SPSS).....	240
Tableau 10.3 - Coefficients de corrélation entre variables démographiques et mesures de taille des unités primaires du sondage. (logiciel SPSS).....	242
Tableau 10.4 - Résultats résumés des régressions et analyses de la variance entre caractères démographiques (effectifs ou densité) et surface bâtie ou densité du bâti. (Logiciel SPSS).....	243
Tableau 10.5 - Coefficients de variation des estimations des variables démographiques avec différentes définitions des probabilités d'appartenance à l'échantillon, cas non stratifié et cas stratifié (Programme VARECH).....	245
Tableau 10.6 - Coefficients de variation des estimations des variables démographiques avec différentes stratifications.(Programme VARECH).	248
Tableau 10.7 - Comparaisons des allocations optimales obtenues pour l'ensemble des caractères démographiques considérés.(Programme VARECH).....	250
Figure 11.1 - Regroupement des 38 types morphologiques de bâtiments en 14 nouveaux types, rappel de la typologie des bâtiments en 38 types... ..	257
Figure 11.2 - Regroupement des 38 types morphologiques de bâtiments en 14 nouveaux types, regroupements effectués pour parvenir aux 14 types	258
Tableau 11.3 - Résultats résumés des analyses de la variance à un critère (Programme "ONEWAY", SPSS).....	261
Tableau 11.4 - Regroupements effectués d'après les analyses de la variance à un critère pour obtenir des catégories de revenus.....	264
Tableau 11.5 - Pourcentage de variance expliquée dans les analyses de variance à un, deux et trois critères.....	265

	Pages
Tableau 11.6 - Pourcentage de variance expliquée dans les analyses de variance à un, deux et trois critères, après regroupement de certaines modalités des variables explicatives et prise en compte des interactions d'ordre 1	266
Tableau 11.7 - Comparaison des variances dues aux premier et second degrés du sondage (Programme VAR2DEG).....	273
Tableau 11.8 - Comparaison des variances dues aux premier et second degrés du sondage selon la taille des pâtés de maisons (Programme VAR2DEG).....	275
Tableau 11.9 - Variances des estimateurs, décomposées selon la taille des pâtés de maisons, pour deux stratégies de sondage au second degré (Programme VAR2DEG).....	276
Tableau 12.1 - Notations adoptées dans le formulaire.....	292
Tableau 12.2 - Formulaire d'estimation de totaux, moyennes et quotients, pour une unité primaire i (deuxième degré).....	295
Tableau 12.3 - Formulaire d'estimation de totaux, moyennes et quotients, pour une strate donnée k, négligeant la variance intra-unité primaire, (premier degré).....	296
Tableau 12.4 - Résumé des calculs effectués pour la détermination du nombre de pâtés de maisons à enquêter dans chaque strate	302
Tableau 12.5 - Calcul des mailles de grilles dans chaque strate	303
Figure 12.6 - Illustration du processus de sélection dans les différentes strates sur une partie de la base de sondage	305
Figure 12.7 - Schéma du pâté de maisons établi à l'issue du recensement des ménages, servant pour la sélection des ménages à enquêter	312
Figure 12.8 - Croquis d'ilot pour les mesures de surface sur le terrain	313
Figure 12.9 - Exemple de calcul de surface effectué à l'aide du logiciel MacDraft™ sur Macintosh.....	314
Graphique 12.10 - Mesures de surfaces.....	315
Tableau 13.1 - Croisement des stratifications sur les critères de densité du bâti et de localisation	324
Tableau 13.2 - La stratification résultante.....	324
Tableau 13.3 - Liste et définitions des caractères démographiques estimés	328
Tableau 13.4 - Estimations des caractères démographiques à Quito et leur précision	329
Figure 13.5 - Intervalles de confiance des estimations par strate	335
Tableau 13.6 - Estimation du total de population dans 50 ilots de l'échantillon.....	337
Figure 15.1 - Précision en fonction de la taille de l'échantillon.....	371
Figure 15.2 - Coûts nécessaires à une précision donnée.....	371
Tableau 15.3 - Taille des échantillons, coûts et précisions pour l'enquête par sondage pratiquée à Quito.....	372

PLAN DETAILLE

	Pages
Sommaire	3
Avertissement	5
INTRODUCTION	6
1. Petite chronique d'une croissance urbaine	6
2. La recherche et la gestion urbaine ont besoin de données actualisées, localisées et diversifiées.....	8
3. Les méthodes traditionnelles de collecte ne suffisent pas à la demande d'information sur les populations urbaines.....	9
3.1 Le recensement.....	9
3.2. Les enquêtes par sondage.....	11
4. Utiliser en ville les images satellite à haute résolution comme base de sondage, un projet de recherche à l'ORSTOM	12
5. Les objectifs principaux de l'analyse statistique et de la recherche sur les sondages spatiaux	15
6. Pour une recherche appliquée : justification du principe empirique de la méthode.....	17

PREMIERE PARTIE : PROBLEMATIQUE, CONTEXTE THEORIQUE, CHOIX DES METHODES ET ETAPES DU TRAVAIL

1 : LE PROBLEME GENERAL : UN SONDAGE AREOLAIRE STRATIFIE, A DEUX DEGRES	21
Introduction : Quel sondage pour une nouvelle méthode de collecte démographique en ville?	21
1.1 : Détermination de l'objet de la recherche : possibilités et contraintes.....	22
1.1.1 : Détermination de l'objet de l'enquête	22
1.1.2 : Définition de la base de sondage	22
1.1.3 : Définition des domaines d'intérêts, stratification	23
1.1.4 : Techniques de sélection des unités aréolaires	25
1.2 : Comment évaluer les performances de la technique envisagée?..	26
Conclusion : l'idéal et le praticable	27
2 : SONDAGES AREOLAIRES ET TELEDETECTION, ETAT DE LA RECHERCHE	29
Introduction	29
2.1 : Synthèse de sept ouvrages de théorie des sondages.....	30
2.1.1 : Généralités et bases de sondage.....	30
2.1.2 : Le sondage équiprobable.....	31
2.1.3 : Les sondages aréolaires	32
2.1.4 : La stratification.....	32
2.1.5 : Les sondages à plusieurs degrés.....	34
2.1.6 : Les sondages à probabilités inégales	35

2.1.7 : Utilisation d'informations exogènes dans l'estimateur	36
2.1.8 : Les sondages systématiques	37
2.1.9 : Conclusion	39
2.2 : Les sondages dans le domaine des statistiques agricoles, forestières et minières	44
2.2.1 : Les sondages aréolaires en agriculture	44
2.2.2 : Les recherches récentes sur l'échantillonnage par points	45
2.2.3 : L'utilisation de la télédétection spatiale dans le domaine des statistiques d'utilisation du sol	46
2.3 : Bilan des expériences dans le domaine de l'estimation de données urbaines à partir d'informations morphologiques	48
2.3.1 : Utiliser la morphologie urbaine pour estimer les populations citadines : une méthode déjà ancienne pratiquée sur photographies aériennes	48
2.3.2 : Des photographies aériennes aux images satellite à haute résolution	52
2.3.2.1 : Quelques rares expériences d'estimation de la population totale des villes à partir de la tache urbaine mesurée sur image satellite	52
2.3.2.2 : Raréfaction des couvertures aériennes et progrès de la télédétection spatiale, système d'observation continue de la morphologie urbaine	52

3 : LES QUESTIONS ET LES OUTILS STATISTIQUES DANS LES DIFFÉRENTES PHASES DE LA RECHERCHE..... 54

Introduction	54
3.1 : Récapitulation des questions	54
3.1.1 : Questions liées à l'utilisation de l'image satellite comme base de sondage	54
3.1.2 : Questions liées à l'utilisation de l'information morphologique extraite de l'image pour améliorer le sondage	56
3.1.3 : Questions indépendantes de l'utilisation de l'image satellite...	57
3.2 : Inventaire des méthodes mathématiques applicables.....	58
3.3 : Les étapes de la recherche et les choix méthodologiques effectués	60
3.3.1 : Marseille, juin 85 - juillet 86: phase de mise au point de la méthode.....	62
3.3.2. : Quito, juillet 86 - septembre 88 : première application.....	63

DEUXIEME PARTIE : MISE AU POINT DE LA METHODE A MARSEILLE, STRATIFICATIONS ET APPROCHE DU PREMIER DEGRE DU SONDAGE

Introduction : objectifs généraux et principe de la méthode d'évaluation des plans de sondage	66
---	----

4 : BASE DE DONNEES, OBJECTIFS ET PLAN D'EXPERIENCE A MARSEILLE. 69

Introduction	69
4.1 : La zone d'étude, les données intégrées et le système de gestion de ces données	69
4.1.1 : La zone d'étude.....	69
4.1.2 : Les données intégrées.....	71

4.1.3 : Le système de gestion de ces données et les étapes de constitution de la base.....	76
4.2 : Le plan général de l'expérience.....	81
4.3 : L'étape d'analyse typologique des données morphologiques.....	84
4.4 : L'étape de mise au point et d'évaluation de plans de sondages stratifiés.....	86
5 : ANALYSE TYPOLOGIQUE POUR LA CONSTRUCTION D'UNE STRATIFICATION MORPHOLOGIQUE DE L'AGGLOMERATION MARSEILLAISE.....	89
Introduction.....	89
5.1 : Analyse du tableau global.....	91
5.1.1 : Analyse factorielle des correspondances.....	91
5.1.2 : Classification ascendante hiérarchique.....	98
5.1.3 : Vérifications sur les photographies aériennes et sur le terrain : critique thématique de la partition en huit classes.....	102
5.2 : Analyse des sous-fichiers pondérés issus de la première classification.....	104
5.2.1 : Le sous-fichier des zones à grands bâtiments.....	104
5.2.2 : Le sous-fichier des zones à petits bâtiments.....	111
5.3 : La typologie en huit classes de la morphologie urbaine.....	119
5.3.1 : Caractérisation morphologique des strates.....	119
5.3.2 : Répartition spatiale des strates.....	121
5.4 : Enseignements méthodologiques et perspectives pour la suite de l'expérimentation, conclusions.....	127
6 : MISE AU POINT DE LA METHODE D'EVALUATION DES PLANS DE SONDAJE, PREMIERE EXPERIENCE, FORMULAIRE ET PROGRAMME INFORMATIQUE.....	131
Introduction.....	131
6.1 : Mise au point de la séquence du test sur un cas simple : plan équiprobable sans remise, estimation sans biais d'effectif et de densité de population, stratification morphologique synthétique.....	132
6.1.1 : Base de sondage, caractéristiques des échantillons.....	133
6.1.2 : Taux de sondage.....	134
6.1.3 : Formulaire du plan équiprobable sans remise.....	136
6.1.3.1 : Sans stratification.....	136
6.1.3.2 : Avec stratification.....	138
6.1.3.3 : Allocation optimale de NEYMAN.....	139
6.1.4 : Application et résultats, commentaires.....	139
6.1.5 : Conclusions, perspectives.....	145
6.2 : Généralisation du test : plans équiprobables, plan à probabilités inégales, estimateurs sans biais, estimateurs par le ratio.....	152
6.2.1 : Position du problème.....	152
6.2.2 : Formulaire général.....	153
6.2.2.1 : Préliminaire.....	153
6.2.2.2 : Conventions et notations.....	153
6.2.2.3 : Plan équiprobable, estimateur sans biais.....	154
6.2.2.4 : Plan équiprobable, estimateur par le ratio.....	155
6.2.2.5 : Plan à probabilités inégales, estimateur sans biais.....	157
6.2.2.6 : Plan à probabilités inégales, estimation d'un ratio ou d'un total par le ratio.....	158
6.3 : Le programme VARECH associé au logiciel Dbase III.....	163
6.3.1 : Matériel, système gestionnaire de fichiers et base de données	

Dbase III sur Marseille.....	163
6.3.2 : Le programme VARECH (VARIance d'EChantillonnage), architecture générale et fonctionnement.....	164
Conclusion	
7 : EVALUATION DE PLANS DE SONDAGE STRATIFIES POUR L'ESTIMATION DE CARACTERES DEMOGRAPHIQUES (RESULTATS DES TESTS MENES GRACE AU PROGRAMME VARECH SUR LA BASE DE DONNEES CONSTITUEE A MARSEILLE).....	167
Introduction	167
7.1 : Comparaison de quatre techniques d'estimation pour une même variable	168
7.2 : Extension du test à d'autres caractères démographiques.....	172
7.2.1 : Résultats sur l'ensemble de la base de sondage, estimateurs et caractères démographiques.....	172
7.2.2 : Résultats par strate : variabilité des caractères démographiques et précision des estimations partielles	181
7.2.3 : Premières conclusions.....	184
7.3 : Stratifications construites sur le critère de la densité du bâti issu de la photo-interprétation, puis de l'image satellite.....	188
7.3.1 : Préliminaires.....	188
7.3.2 : Provenance et élaboration des stratifications	188
7.3.3 : Evaluation des stratifications issues de la photo-interprétation	189
7.3.3.1 : Comparaison des stratifications selon la densité du bâti en trois et six classes	193
7.3.3.2 : Contribution du caractère morphologique "densité du bâti" à une stratification morphologique synthétique	193
7.3.4 : Evaluation de la stratification issue de l'image satellite.....	194
7.3.4.1 : Comparaison à la référence "photographies aériennes".	194
7.3.4.2 : Gain permis par le critère de densité du bâti et comparaison à la stratification morphologique synthétique	195
7.3.5 : Perspectives pour la suite de la recherche.....	195
Conclusion.....	199
8 : APPORT DU TIRAGE ARÉOLAIRE SYSTÉMATIQUE.....	201
Introduction	201
8.1 : La méthode : approximation du tirage systématique aréolaire par stratification de la base de sondage sur un critère de localisation	202
8.2 : Les données nécessaires au test : élaboration des stratifications et remarques sur l'interprétation des résultats.....	203
8.3 : Résultats	204
Conclusion.....	208
CONCLUSION.....	209
1. Bilan de l'expérience menée à Marseille : objectifs, déroulement, résultats.....	209
2. Analyse des spécificités du plan défini	211
3. Questions en suspens, perspectives de travail à Quito	212

**TROISIEME PARTIE : DEFINITION ET MISE EN ŒUVRE D'UN PLAN DE
SONDAGE AREOLAIRE SUR IMAGE SPOT POUR UNE ENQUETE DEMOGRAPHIQUE
A QUITO**

Introduction..... 215

**9 : OBJECTIFS GENERAUX, PLAN D'EXPERIENCE, REALISATION
DES ENQUETES PRELIMINAIRES..... 216**

9.1 : Objectifs généraux, plan d'expérience.....	216
9.2 : Réalisation de l'enquête "morphologie".....	218
9.2.1 : Sélection de l'échantillon.....	218
9.2.2 : Elaboration des documents d'enquête.....	219
9.2.3 : Collecte des données.....	220
9.2.4 : Codification, saisie, contrôles de cohérence et recodage.....	220
9.2.5 : Exploitation de l'enquête morphologie en vue du tirage d'un échantillon représentatif.....	221
9.3 : Réalisation de l'enquête "relations bâti/population".....	228
9.3.1 : Sélection de l'échantillon.....	228
9.3.2 : Elaboration des documents d'enquête.....	229
9.3.3 : Collecte des données.....	229
9.3.4 : Codification et saisie.....	231

**10 : ENQUETE PRELIMINAIRE "BATI / POPULATION",
EXPLOITATION POUR LA MISE AU POINT DU PREMIER DEGRE DU
SONDAGE : ECHANTILLONNAGE DES PATES DE MAISONS..... 232**

10.1 : Introduction : le contexte et les questions posées.....	232
10.2 : Les données et les outils statistiques utilisés.....	234
10.2.1 : Les données.....	234
10.2.1.1 : Données démographiques.....	235
10.2.1.2 : Taille des unités primaires.....	238
10.2.1.3 : Stratifications.....	238
10.2.2 : Les outils statistiques utilisés.....	239
10.3 : Résultats.....	240
10.3.1 : Phase exploratoire.....	240
10.3.2 : Phase conclusive.....	244
10.3.2.1 : Probabilités d'appartenance des unités primaires à l'échantillon.....	244
10.3.2.2 : Stratifications.....	247
10.3.2.3 : Allocations optimales.....	249
10.4 : Conclusions.....	252

**11 : ENQUETE PRELIMINAIRE "BATI / POPULATION",
EXPLOITATION POUR LA MISE AU POINT DU SECOND DEGRE DU
SONDAGE : ECHANTILLONNAGE DES MENAGES DANS LES PATES
DE MAISONS..... 255**

11.1 : Introduction, objectifs principaux et la démarche.....	255
11.2 : Analyses préliminaires : les possibilités "d'explication" des caractéristiques démographiques des ménages par la morphologie du bâtiment, la morphologie et la localisation de l'ilot.....	256
11.2.1 : Les questions posées, les données et les outils statistiques	

utilisés	256
11.2.1.1 : Les questions posées.....	256
11.2.1.2 : Les données.....	256
11.2.1.3 : Les outils statistiques utilisés	259
11.2.2 : Résultats.....	260
11.2.2.1 : Analyse de la variance à un caractère : pouvoir discriminant des facteurs "explicatifs" sur les différents caractères démographiques étudiés.....	260
11.2.2.2 : Analyse de la variance multi-critères : redondance/ indépendance de l'information apportée par les trois facteurs "explicatifs"	264
11.2.3 : Conclusions	267
11.3 : Détermination du second degré du sondage : échantillonnage de ménages dans un pâté de maisons.....	268
11.3.1 : Position du problème.....	269
11.3.2 : Formulaire, programme informatique	272
11.3.2.1 : Formulaire	269
11.3.2.2 : Programme informatique	272
11.3.3 : Résultats.....	273
11.3.4 : Conclusions	277

12 : LE PLAN DE SONDAGE ET SA MISE EN ŒUVRE POUR L'ENQUETE DEMOGRAPHIE/MIGRATIONS..... 279

12.1 : Introduction : le double objectif méthodologique et thématique de l'enquête, présentation du questionnaire	279
12.2 : Description du plan de sondage, étapes de sélection et d'inférence	286
12.2.1 : Résumé des résultats obtenus lors des phases d'expérience ...	286
12.2.2 : Description du plan de sondage : étapes de sélection aux premier et second degrés	287
12.2.3 : Formulaire d'estimation pour l'étape d'inférence.....	288
12.2.3.1 : Préliminaire, les problèmes posés par l'inférence dans le cas du sondage systématique	288
12.2.3.2 : Formulaire	290
12.3 : Mise en œuvre du plan de sondage : sélection de l'échantillon et étapes du travail de collecte.....	297
12.3.1 : Définition de la base de sondage sur l'image: la limite urbaine	297
12.3.1.1 : Les documents utilisés	297
12.3.1.2 : Mise en oeuvre et règles de décision	297
12.3.1.3 : Validation	298
12.3.2 : Stratification de la base de sondage sur l'image : les critères retenus et la réalisation pratique, mesures de surface de chaque strate.....	298
12.3.2.1 : Utilisation de la classification des pixels de l'image SPOT selon la densité du bâti	298
12.3.2.2 : Evaluation de la classification	298
12.3.2.3 : Elaboration de la stratification	299
12.3.3 : Sélection de l'échantillon de pâtés de maisons.....	300
12.3.3.1 : Détermination du nombre total de pâtés de maisons de l'échantillon et de la répartition entre les strates	300
12.3.3.2 : Définition des grilles	302
12.3.3.3 : Sélection	303
12.3.4 : Recensement des ménages dans les pâtés de maisons et sélection des ménages à enquêter	310
12.3.4.1 : Recensement des ménages.....	310
12.3.4.2 : Sélection des ménages à enquêter.....	310
12.3.5 : Mesures de surface des unités primaires	311

12.3.5.1 : Les méthodes de mesure employées	311
12.3.5.2 : Evaluation de la précision des techniques de mesure....	315
12.4 : Conclusion.....	316

QUATRIEME PARTIE : EVALUATION DES PERFORMANCES ET DES CONDITIONS D'APPLICATION DE LA METHODE.

Introduction	319
13 : PRECISION DE LA METHODE : LES PREMIERS RESULTATS DE L'ENQUETE REALISEE A QUITO.....	320
Introduction	320
13.1 : Démarche d'inférence : les programmes d'estimation	321
13.1.1 : Estimation séparée des deux termes de la variance (premier et second degrés).....	321
13.1.2 : Approximation de la variance du sondage systématique	323
13.1.3 : Les programmes d'estimation	323
13.2 : Estimation des caractères démographiques.....	327
13.2.1 : Précision des estimations globales	327
13.2.2 : Estimations par strate et variation des caractères au sein de la base de sondage.....	334
13.2.3 : Estimation séparée des deux termes de la variance	336
13.2.4 : Précision des estimations dans les unités primaires de l'échantillon	336
13.3 : Cohérences des résultats de l'enquête par rapport aux chiffres du recensement de 1982 et aux tendances d'évolution connues.....	339
Conclusion.....	340
14 : DOMAINES ET CONDITIONS D'APPLICATION DE LA METHODE...	342
Introduction	342
14.1 : Sites et thématiques possibles pour l'application de la méthode..	342
14.1.1 : Les sites urbains d'application	343
14.1.2 : Les thématiques que l'on peut aborder.....	344
14.2 : Les conditions d'application liées à l'utilisation de l'image satellite, à la définition du plan de sondage et à sa mise en oeuvre.....	345
14.2.1 : Conditions liées à l'utilisation de l'image	345
1) Conditions d'obtention d'une image exploitable comme base de sondage	345
2) Conditions nécessaires au traitement de l'image en vue de la stratification de la base	347
14.2.2 : Conditions liées à la définition du plan de sondage et à sa mise en oeuvre	348
1) Environnement informatique nécessaire	349
2) Etapes de la mise au point du plan : les enquêtes préliminaires	350
3) Conditions de la collecte.....	352
Conclusion.....	352
15 : COUTS ET DELAIS DE LA MISE EN ŒUVRE.....	354
Introduction	354

15.1 : Eléments pour le calcul des coûts et des délais de mise en œuvre de la méthode.....	355
15.1.1 : Equipement.....	355
15.1.2 : Images satellite.....	356
15.1.3 : Sélection de l'échantillon.....	357
15.1.4 : Collecte des données sur le terrain.....	358
15.1.5 : Codification des données.....	361
15.1.6 : Saisie des données.....	363
15.1.7 : Synthèse des coûts.....	364
15.2 : Coûts nécessaires à l'obtention d'une précision fixée à l'avance..	367
15.2.1 : Les fonctions de coût attachées aux types d'enquêtes réalisées à Quito.....	367
15.2.2 : La fonction précision.....	369
15.2.3 : Budget nécessaire pour atteindre une précision fixée.....	370
Conclusion.....	373
CONCLUSION : SYNTHESE DES PRINCIPAUX RESULTATS, PERSPECTIVES POUR LES RECHERCHES FUTURES.....	374
Annexes.....	383
Bibliographie.....	622
Liste des tableaux, figures et photographies.....	629