

L'ENQUETE PAR SONDAGE SUR IMAGE SATELLITE :
UNE SOLUTION POUR AMELIORER
L'OBSERVATION DES POPULATIONS CITADINES

par Françoise DUREAU et Olivier BARBARY¹

Rythmes et caractéristiques de l'urbanisation, modes de résidence des habitants et déficience des éléments de connaissance classiques, tels que cartes, plans cadastraux, ou fichiers administratifs, contribuent à rendre particulièrement difficile l'observation des populations citadines des pays en développement. Les instruments traditionnels de l'observation socio-démographique, recensements exhaustifs et enquêtes par sondage sur liste, se révèlent difficiles à mettre en oeuvre et ne satisfont pas correctement les besoins de la recherche ni ceux de la gestion urbaine en matière d'observation *suivie et spatialisée* des populations citadines.

C'est pourquoi, depuis les années cinquante, ont été développées des méthodes de collecte de données démographiques adaptées aux spécificités des villes des pays en développement : celles-ci ont montré que l'on peut utiliser avec profit l'information exhaustive sur la morphologie urbaine pour recueillir rapidement, par sondage, des données relatives aux populations urbaines. Depuis le

¹ Françoise DUREAU et Olivier BARBARY appartiennent à l'ORSTOM, Institut français de recherche scientifique pour le développement en coopération. Ce texte a fait l'objet d'une communication aux Journées d'études "La qualité de l'information dans les enquêtes", Paris, 13-14 juin 1991, organisées par l'Association pour la Statistique et ses Utilisations, le Centre National de la Recherche Scientifique et l'Ecole Nationale Supérieure des Télécommunications.

milieu des années quatre-vingt, la télédétection spatiale², assurant une observation continue et relativement précise de l'occupation du sol, constitue une nouvelle source de données particulièrement intéressante. Dans le cadre d'un programme de recherche initié en 1985, une équipe pluridisciplinaire de l'ORSTOM³ a développé une méthode de collecte de données démographiques en milieu urbain intégrant l'information sur la morphologie urbaine apportée par les satellites à haute résolution. L'idée centrale de la méthode est d'utiliser l'image satellite comme base de sondage et d'exploiter l'information sur la morphologie urbaine fournie par l'image pour stratifier un plan de sondage spatial permettant de sélectionner un échantillon pour une enquête à objectifs démographiques ou socio-économiques.

Après un panorama des travaux relatifs à l'utilisation des photographies aériennes et images satellite pour produire des données socio-démographiques en milieu urbain, la suite de cet article sera consacrée à l'exposé de certains éléments justifiant la méthode développée à l'ORSTOM et à la présentation de sa mise en oeuvre.

I - L'UTILISATION DES PHOTOGRAPHES AERIENNES ET IMAGES SATELLITE POUR PRODUIRE DES DONNEES DEMOGRAPHIQUES EN VILLE

1.1. Problèmes posés par la production d'information sur les populations des villes à croissance rapide des pays en développement

Pour produire des informations sur les populations, le démographe dispose de deux techniques de collecte des données : le recensement, exhaustif, et l'enquête par sondage, où n'est observé qu'un échantillon de la population.

² Pour une présentation des principes de la télédétection spatiale et des méthodes d'interprétation des images satellite en milieu urbain, le lecteur pourra se reporter à la première partie d'un précédent article qui exposait les lignes directrices de ce programme de recherche : DUREAU F. et GUILLAUME A., 1984, "La démographie depuis l'espace : un nouveau système d'observation ?", STATECO n° 38, pp. 5-45.

³ L'équipe de recherche était composée de : BARBARY O. (statisticien), DUREAU F. (géographe-démographe, responsable du programme de recherche), LORTIC B. (télédétection), MICHEL A. (urbaniste spécialisé en télédétection).

Les *recensements* exigent des moyens techniques, financiers et humains considérables qui interdisent une périodicité inférieure à une dizaine d'années et conduisent à une limitation de l'information recueillie au niveau de chaque individu lors de la collecte.

De plus, le temps nécessaire à leur exploitation rend souvent les résultats caducs dès leur parution pour des villes dont les taux de croissance peuvent dépasser 10 % par an. L'absence de cartographie de base dans nombre de villes, principalement dans les zones d'extension récente, et les problèmes de suivi de la masse d'enquêteurs que réclame ce type d'opération ont des répercussions importantes sur la qualité des résultats.

Enfin, l'ampleur des moyens à mettre en oeuvre pour un recensement conduit à une sous-exploitation des données recueillies exhaustivement. Pour alléger l'exploitation de l'information collectée, on ne réalise souvent qu'une exploitation par sondage des questionnaires, ou d'une partie seulement des variables du questionnaire, et ne sont publiés que les résultats globaux sur l'ensemble de la ville : la principale qualité des recensements, l'exhaustivité, qui permet théoriquement d'obtenir des résultats pour tout type de découpage géographique, est remise en cause par une exploitation partielle de l'information recueillie.

Fondées sur l'observation de la seule fraction de la population composant l'échantillon, les *enquêtes par sondage* présentent comme intérêt de réduire les moyens financiers et humains à mobiliser, et d'assurer une meilleure qualité de la collecte, par des enquêteurs moins nombreux et mieux encadrés, qui recueillent des informations plus riches que celles autorisées dans le cadre d'un recensement.

Mais, dans de nombreuses villes des pays en développement, un élément essentiel fait défaut pour procéder à la sélection de l'échantillon à enquêter : une base de sondage complète et à jour. Cette information n'est quasiment jamais disponible dans les villes du Tiers-Monde où les documents cartographiques sont rares et anciens. L'actualisation par de nouveaux relevés de terrain d'une base cartographique vieillie est souvent problématique dans un tissu urbain affecté de modifications rapides. Trop souvent, l'attention portée à

l'actualisation de la cartographie reste très en dessous de celle réclamée par ces travaux et entraîne des biais importants.

D'autre part, les plans de sondage classiques interdisent généralement toute spatialisation des résultats : ils ne permettent pas de connaître les différences internes à la ville en matière de densité de population, ou de composition démographique ou socio-économique. C'est pourtant un élément essentiel de connaissance tant pour les gestionnaires de la ville que pour les chercheurs.

Dans les villes des pays en développement, déficience de l'information cartographique de base et modalités de l'urbanisation se conjuguent pour rendre peu efficaces les techniques classiques de collecte des données démographiques : dans un tel contexte, celles-ci ne permettent pas de réaliser dans de bonnes conditions une observation *suivie* et *spatialisée* des populations citadines. Que proposer pour tenter d'améliorer l'observation des populations citadines du Tiers-Monde ?

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

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

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 de qualité, composée d'îlots précisément identifiés ; d'autre part, une bonne connaissance de la morphologie urbaine permet de stratifier la ville, donc d'améliorer la précision du sondage, et d'obtenir des résultats selon un découpage spatial significatif de la ville enquêtée.

C'est ce type de raisonnement qu'ont mené VERNIERE⁴, puis les différents bureaux d'étude qui ont appliqué les méthodes de production de données démographiques à partir de photographies aériennes : dans le cas des pays africains disposant de couvertures aériennes mais manquant de données démographiques, ces méthodes ont trouvé un champ d'application privilégié.

La mise en évidence des relations entre la morphologie urbaine et les caractéristiques de la population citadine établies par différents travaux a été à l'origine du développement de méthodes d'estimation de population à partir de l'information sur l'occupation du sol apportée par les photographies aériennes.

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.

Une expérience menée au USA permet de se rendre compte du degré de précision de la méthode : il s'agit des travaux de WATKINS (1985), dont la méthodologie et les conclusions sont résumées dans le Tableau 1. 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. La conclusion de WATKINS est celle qui a donné lieu au développement d'un autre type 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 est d'exploiter

⁴ Voir les références bibliographiques en fin d'article.

l'information exhaustive fournie par les couvertures aériennes (exemple : VERNIERE, Tableau 1) :

- 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 allégement de l'échantillon et une spatialisation des résultats. Nombreux sont les bureaux d'étude qui ont appliqué la technique mise au point par VERNIERE pour obtenir des estimations démographiques sur les villes dont ils avaient à réaliser un plan d'aménagement, ou des programmations d'équipement ;

- soit pour actualiser l'effectif de population d'une ville ayant fait l'objet antérieurement d'un recensement de population : 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, banlieue de Dakar, 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 (Tableau 1).

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 lié 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 (Tableau 1), travaillant sur des images LANDSAT MSS, obtiennent de bons résultats sur les grandes villes chinoises en utilisant trois modèles très souvent appliqués, ceux de NORDBECK (Loi de croissance allométrique), TOBLER (1969) et HUXLEY.

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

Tableau 1 - Quelques exemples d'estimations de population à partir de photographies aériennes et images satellite

AUTEUR	VILLE	NIVEAU GEO.	SOURCE D'INFO.	REFERENCE	METHODOLOGIE	RESULTATS et CONCLUSIONS
WATKINS (1985)	Boulder (USA)	3 quartiers de 2800 à 6200 Hab en 1970 et 1980	Photo aériennes : Noir et Blanc 1970 : 1/20.000 1980 : 1/6.000 Recensement : Tx de logements vides Nbre moyen de pers. par logement. (Taux calculés pour l'habitat individuel et le collectif, uniques pour les 3 quartiers)	Recensements dém. 1970 et 1980	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., Tx de log. vides, Nb moyen de pers. par logement (pour l'habitat individuel et le collectif)	Erreur relative sur le nombre de logements: entre -4,91% et -0,37% par quartier, et -0,79 et -3,36% pour le total des 3quartiers Erreur relative sur la population : entre -6,64% et +16,57% par quartier, et -1,98 et +8,54% pour le total. des 3 quart. Conclusions * Bon dénombrement des log. sur photo : -Hab. individuel: 1/20.000 à 40.000 -Hab. collectif: 1/8.000 à 10.000 * Pour améliorer la précision, calculer les taux de log. vides et le Nbre moyen de pers. par log., pour chaque quartier de la ville
VERNIERE (1973)	Pikine (SENEGAL)	Zones morpho. homogènes de la banlieue de Dakar (Pop.totale 135.000 Hab.)	Photo. aériennes : Noir et Blanc 1/5.000 1961,1963,1967,1970 Enquête dém. sur échantillon de 8carrés de sondage (1970-71): effectifs de pop.	Recensements OMS, ISEA, administratifs. Enquête stat. nationale	Zonage en espaces homogènes sur photographies de 1970 (morphologie urbaine). Calcul des Coef. de Surface Bâtie Habitable (CSBII) pour 38 carrés de sondage, sur les photo. de 1970. Pour les 8 carrés où la pop. est connue, calcul du coef. Densité de pop./CSBII. En supposant que ce coef. 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 Bilan et perspectives Trois intérêts de la méthode : 1) Spatialiser les résultats d'une enquête démographique globale 2) Actualiser des chiffres de pop. entre 2 recensements 3) Moyen de connaissance autonome en l'absence de toute autre donnée, base pour une enquête par sondage
LO et WELCH (1977)	30 villes chinoises	Villes entières >500.000 Hab.	Images MSS : compositions colorées Cartes et recense. de 150 villes de Taiwan : calcul des équations linéaires et Log entre pop. et surf. des villes (1951-66)	Recensements dém. des 30 villes de Chine	Mesure de la superficie de chaque ville sur l'image MSS au 1/500.000. Estimation de la pop. de chaque ville en appliquant les 2 séries d'équations mises au point sur les villes de Taiwan	Bonne concordance entre résultats calculés par modèles et ceux des recensements dém. Facteurs d'amélioration * Utiliser une image 1/250.000 * N'appliquer le modèle qu'aux villes de surface > 10 Km ² * Utiliser un modèle linéaire pour les villes>2.500.000 Hab. et log. pour les plus petites

des images LANDSAT MSS convienne mieux que les photographies aériennes 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 l'information morphologique sont variés : depuis les méthodes les plus globales, ne retenant de la morphologie urbaine que la surface urbanisée, jusqu'aux méthodes détaillées reposant sur des comptages de logements, 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.

1.3. Des photographies aériennes aux images satellite à haute résolution

Le recours aux photographies aériennes sur les centres urbains des pays en développement devient de plus en plus problématique. En effet, les couvertures aériennes des villes de ces pays sont de plus en plus rares en raison de leur coût et de la priorité souvent donnée au milieu rural : dans ces villes, la répétitivité de l'observation aérienne devient une réalité de moins en moins effective. Il devient donc difficile d'employer les techniques de production de données démographiques fondées sur l'emploi de photographies aériennes.

Cependant, jusqu'au milieu des années quatre-vingt, rares sont les expériences d'utilisation des images satellite pour la production de données démographiques : à partir de notre recherche bibliographique, en 1985 (Tableau 2), nous n'avons eu connaissance que de cinq applications, toutes fondées sur la méthode globale d'estimation de la population totale d'une ville à partir de la superficie urbanisée mesurée sur image satellite, LANDSAT MSS dans la plupart des cas.

**Tableau 2 - Utilisation d'images satellite
pour produire des données démographiques**

Auteur	Année	Ville (PAYS)	Type d'image satellite
SABOL	1968	USA	Image radar
WELLAR	1969	San Antonio et Houston (USA)	Image du satellite GEMINI
REINING	1973	NIGER et BURKINA FASO	Image LANDSAT MSS
MURAI	1974	Tokyo (JAPON)	Image LANDSAT MSS
LO et WELCH	1977	Villes de 500.000 à 2.500.000 hab. (CHINE)	Image LANDSAT MSS
CARDIERI et al.	1989	Sao Paulo (BRESIL)	Image SPOT

Cet état de fait est directement lié à la résolution des images LANDSAT MSS (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 constituants. Avec l'imagerie LANDSAT MSS, 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.

II - LES RECHERCHES MENEES A L'ORSTOM

2.1. Objectifs et hypothèses

La mise en service de nouveaux satellites au milieu des années quatre-vingt change largement la situation de la télédétection spatiale appliquée

au milieu urbain : SPOT et TM ont une résolution suffisante (respectivement 10/20 mètres, et 30 mètres) pour observer assez finement le tissu urbain. Il devient alors possible de faire profiter le milieu urbain des atouts des images satellite, notamment : exhaustivité spatiale, coût raisonnable (environ 10 fois moins que les photographies aériennes), caractère numérique des données, richesse de l'information enregistrée, avec possibilité de vision stéréoscopique (SPOT).

Il nous a donc semblé intéressant de *tester 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'agissait 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 restait la même : utiliser l'information exhaustive sur la morphologie urbaine apportée par les images satellite pour recueillir rapidement, par sondage, les données relatives aux populations urbaines.

L'utilisation d'une image satellite comme *base de sondage* repose sur deux hypothèses :

- il est possible d'extraire d'une image SPOT ou TM les limites d'une agglomération urbaine ;
- on peut sélectionner sur une image satellite un échantillon d'unités spatiales (îlots).

Quant à l'utilisation d'une image en matière de *stratification*, elle se fonde sur les deux idées suivantes :

- il existe certaines relations entre caractéristiques morphologiques du milieu urbain et caractéristiques démographiques et socio-économiques des habitants ;
- on peut identifier sur des images SPOT ou TM des caractéristiques morphologiques pertinentes pour l'observation démographique.

2.2. Les différentes phases de la recherche

Au moment du démarrage de notre programme, en 1985, l'introduction de la télédétection spatiale dans un système de production de données démographiques constituait un champ nouveau de recherches.

Les acquis en télédétection urbaine étaient alors peu nombreux : ils portaient essentiellement sur le traitement multi-dates de l'occupation du sol, l'évaluation de la quantité de végétation et l'observation des franges urbaines. L'analyse précise de l'information apportée par les capteurs à haute résolution en milieu urbain et le développement de méthodes spécifiques d'extraction de cette information restaient à poursuivre.

Quant aux sondages aréolaires, l'expérience demeurait limitée, surtout en démographie où l'habitude est toujours de tirer des échantillons de ménages ou d'individus sur liste. Que ce soit aux Etats-Unis ou en France, l'essentiel de l'expérience dans ce domaine relevait de la statistique agricole : en France et en 1985, seul le SCEES⁵ avait obtenu des résultats intéressants sur l'intégration de la télédétection dans le système de production de statistiques agricoles en développant des méthodes de sondage aréolaires.

Ainsi, la réalisation de l'objectif du programme ne pouvait s'appuyer que sur un acquis méthodologique relativement limité dans chacun des deux grands domaines auxquels il a trait (l'interprétation des images satellite à haute résolution en milieu urbain et les sondages spatiaux en démographie urbaine) et nécessitait des travaux approfondis dans ces deux directions de recherche.

Etant donné l'objectif du projet, proposer une méthode *opérationnelle* pour produire rapidement de l'information démographique en ville, et son caractère tout à fait exploratoire, un double souci a guidé la conception du programme :

⁵ Service Central des Enquêtes et Etudes Statistiques du Ministère de l'Agriculture.

- d'une part, *tester la validité des méthodes mises au point au fur et à mesure de l'avancement des travaux ;*

- d'autre part, *proposer des solutions adaptées aux contextes matériel, financier et humain des villes des pays en développement.*

C'est en fonction de ces deux paramètres qu'ont été déterminés les sites de travail, les différentes phases du programme et les plans d'expérience.

Afin de disposer de références fiables, nous avons choisi de travailler d'abord sur une ville française, Marseille, disposant à une même date (1982) d'une image TM, d'un recensement récent disponible au niveau du district de recensement et d'une couverture aérienne à grande échelle : l'intégration de l'ensemble de ces informations dans une base de données gérée par le système d'information géographique SAVANE⁶ nous a placés dans une situation d'expérimentation quasi-idéale. Toute méthode d'analyse de l'image satellite pouvait être évaluée en comparant ses résultats à l'information morphologique issue de la photo-interprétation et de contrôles terrain. Les travaux en sondages spatiaux bénéficiaient également de ce contexte d'information exhaustive, permettant de mettre en relation les données morphologiques et démographiques et de calculer les variances obtenues selon différents plans de sondage. Ainsi, différentes stratégies de sondage ont pu être évaluées en soumettant les données exhaustives du recensement de population à un programme informatique développé à l'ORSTOM (VARECH) qui, en appliquant les formulaires associés à différents plans de sondage, calculent la variance exacte des estimateurs correspondants.

Après cette première phase de développement méthodologique sur la ville de Marseille, nous avons travaillé sur la ville de Quito (EQUATEUR), qui disposait d'images satellite SPOT et LANDSAT TM. Deux objectifs étaient assignés à cette seconde phase de la recherche, dans une ville d'un pays en développement :

- premier objectif : *compléter et vérifier les résultats obtenus sur Marseille afin d'aboutir à une définition complète de la méthode.*

⁶ Système développé par l'Unité d'Infographie de l'ORSTOM.

A cet effet, nous avons réalisé deux enquêtes préliminaires : l'enquête Morphologie urbaine (relevé précis de l'occupation du sol dans un échantillon de 199 îlots de Quito) et l'enquête Bâti-Population (recueil des informations démographiques sur un sous-échantillon de 55 îlots choisis parmi les 199 enquêtés précédemment). Ces enquêtes étaient nécessaires pour d'une part développer et évaluer des méthodes de traitement d'image SPOT, d'autre part aboutir à une définition complète du plan de sondage le plus efficace. L'application d'outils d'analyse statistique classique (corrélation, régression, analyse de variance) et des programmes de calcul de variance (VARECH et VAR2DEG) aux données recueillies au cours des deux enquêtes préliminaires a permis de vérifier la stabilité des résultats obtenus sur Marseille et d'obtenir des résultats sur deux points : probabilité de sélection des îlots et deuxième degré du sondage.

- deuxième objectif : *réaliser une première application afin de tester son efficacité en termes de rapidité, coût et précision.*

Au cours du dernier trimestre 1987, nous avons effectué une enquête démographique par sondage, sur un échantillon d'îlots tiré sur une image satellite SPOT de Quito. Première application en vraie grandeur de notre méthode de collecte, cette enquête Migrations a confirmé l'intérêt de la méthode développée et permis d'évaluer tant son niveau de précision que les coûts qu'implique sa mise en oeuvre.

A l'issue de cette seconde phase de nos travaux sur Quito, nous avons donc *mis au point et évalué* toutes les étapes d'une méthode de production rapide de données démographiques par sondage aréolaire sur image satellite. Le choix de la stratégie de sondage a reposé sur un certain nombre de résultats statistiques que nous allons présenter maintenant.

III - LES PRINCIPAUX RESULTATS RELATIFS AU PLAN DE SONDAGE

3.1. Un plan de sondage spatial en ville : efficacité de la stratification sur la densité du bâti et des probabilités d'inclusion proportionnelles aux surfaces des îlots

Concernant la technique de sondage, deux questions sont posées au départ de la recherche :

- l'efficacité des informations morphologiques comme critère de stratification de la base de sondage ;
- le mode de sélection des unités primaires, c'est-à-dire des probabilités d'inclusion des îlots dans l'échantillon.

En effet la procédure "naturelle" de tirage d'un échantillon d'unités spatiales dans une base de sondage constituée par un document de type carte, photographie aérienne ou image satellite, est celle du tirage par points. La probabilité d'inclusion d'une unité primaire quelconque est alors proportionnelle à sa surface totale au sol. Mais on peut envisager, même si leur mise en pratique peut s'avérer beaucoup plus difficile, bien d'autres probabilités d'inclusion : équiprobabilité, probabilité proportionnelle à la surface bâtie,... etc. Quelles conséquences ont ces différentes probabilités de sélection sur la précision du sondage ?

Comme nous allons le voir, les effets de la stratification et des probabilités d'inclusion sont étroitement liés. Pour choisir une technique efficace de sélection d'un échantillon d'îlots, il faut en fait répondre à trois questions :

- quel gain de précision sur l'estimation des caractéristiques socio-démographiques de la population apporte la stratification de la base de sondage sur des critères de morphologie urbaine ?
- quelles conséquences auront les probabilités d'inclusion sur la précision des estimations ?
- quelles sont les interactions entre les effets respectifs de la stratification et des probabilités d'inclusion sur la précision du sondage ?

Le Tableau 3 est construit à partir des résultats de l'application du programme VARECH aux bases de données constituées à Marseille et à Quito. Nous prendrons l'exemple de quatre indicateurs démographiques simples pour montrer quels gains de précision sur les estimations permettent d'obtenir, à Marseille et à Quito, différentes stratifications morphologiques de la base sous différents plans de sondage. *L'indicateur du gain de précision est la réduction de la variance de l'estimateur exprimée en pourcentage de la variance du même sondage non stratifié (même base de sondage, mêmes probabilités d'inclusion).* Les estimateurs des totaux sont les estimateurs sans biais classiques d'HORVITZ-THOMPSON associés à chaque jeu de probabilités d'inclusion ; les quotients sont estimés avec un léger biais par l'estimateur du ratio correspondant aux mêmes probabilités.

La première constatation qui s'impose est que *l'introduction d'une stratification morphologique de la base permet un gain de précision très significatif* : la variance est réduite de 20 à 75 % selon les cas. On constate aussi que ces gains sont très variables en fonction des quatre facteurs qui entrent en ligne de compte : la ville (Marseille ou Quito), la variable estimée, la nature de la stratification et enfin les probabilités d'inclusion.

Dès ce stade apparaissent un certain nombre de phénomènes intéressants et on peut réfléchir à des explications d'ordre statistique, géographique (ayant trait aux spécificités des relations entre morphologie et socio-démographie dans les deux villes considérées) ou liées aux techniques de photo-interprétation et de télédétection utilisées pour construire les stratifications. Quelle que soit leur explication, ces faits sont en tout cas d'une grande importance pratique pour construire un plan de sondage efficace ; en voici quelques exemples.

A Marseille la stratification morphologique en neuf classes issue de la photo-interprétation est d'une efficacité comparable sous les deux plans de sondage envisagés : équiprobable ($p=cste$) ou à probabilités proportionnelles aux surfaces des îlots ($p=su$) ; dans les deux cas *la variance est réduite d'au moins 40 %* (lignes 1 et 2).

Sous le plan $p=su$, le critère de densité du bâti, actuellement accessible à partir de l'image satellite, permet à lui seul une diminution de

Tableau 3 : Diminution de la variance des estimateurs due à différentes stratifications morphologiques sous différents plans de sondages (en % de la variance de l'estimateur non stratifié sous le même plan)

Variables à estimer :	Population totale	Population âgée de cinq ans ou plus	Taille moyenne des ménages	Proportion des plus de cinq ans dans la population totale
Sondages :				
1.Marseille, sondages sur la base du recensement INSEE 1982				
Plan équiprobable (p=cste) 9 strates morphologiques d'après photo-interprétation	40 %	39 %	62 %	62 %
Plan à probabilité proportionnelle à la surface de l'ilot (p=su) 9 strates morpho. d'après photo-interprétation	43 %	43 %	40 %	41 %
Plan p=su 6 strates de densité de bâti d'après photo-interprétation	28 %	29 %	25 %	17 %
Plan p=su 6 strates de densité de bâti d'après image satellite TM	26 %	27 %	18 %	11 %
2.Quito, sondages sur la base d'enquêtes préliminaires réalisées en 1986-87				
Plan p=cste 6 strates de densité de bâti d'après enquête Morpho.	59 %	59 %	47 %	44 %
Plan p=su 6 strates de densité de bâti d'après l'enquête Morpho.	77 %	77 %	52 %	41 %
Plan p=su 6 strates de densité de bâti d'après image SPOT	75 %	75 %	51 %	33 %

variance de 20 à 30 % à Marseille. Il fournit donc, dans ce cas précis, plus de la moitié de l'information utile amenée par une stratification morphologique plus fine construite à partir d'une photo-interprétation relativement lourde (lignes 3 et 4 comparées à la ligne 2). Par ailleurs, *ce même critère de stratification est encore beaucoup plus efficace à Quito où il réduit la variance de 35 à 75 % selon les variables estimées (lignes 6 et 7).*

Du point de vue du gain de précision qu'elles permettent, *les stratifications selon la densité du bâti tirées des images satellite sont pratiquement équivalentes aux stratifications "de référence" sur le même critère construites par photo-interprétation à Marseille, ou issues de l'enquête au sol à Quito.*

A Quito la stratification sur la densité du bâti est nettement plus efficace associée à des probabilités d'inclusion proportionnelles aux surfaces des unités primaires que sous un plan équiprobable (lignes 6 et 7 comparées à la ligne 5), ce qui est une des différences majeures existant entre les résultats obtenus sur les deux villes.

Le Tableau 4 permet de compléter l'analyse de la dépendance qui existe entre les effets respectifs des stratifications et des probabilités d'inclusion. Il est lui aussi construit à partir des résultats de l'application du programme VARECH et donne pour les mêmes caractères démographiques que le Tableau 3, d'une part les coefficients de variation des différentes estimations et d'autre part en gras, l'écart de précision existant entre le plan de sondage à probabilités d'inclusion inégales considéré (probabilités proportionnelles aux surfaces des îlots ou à leurs surfaces bâties dans le cas de Quito) et le même plan équiprobable (même base et même stratification). *Cet indicateur est donné en pourcentage de la variance du sondage équiprobable, affecté du signe de la variation*⁷.

A Marseille il se confirme que le sondage, stratifié ou non, est peu sensible au changement des probabilités d'inclusion, du moins en ce qui concerne l'estimation des totaux. Pour les ratios, le tirage p=su améliore la précision en l'absence de stratification alors qu'il la dégrade au contraire dans le cas stratifié (lignes 1 et 2, 3 et 4).

⁷ Il s'agit en fait du concept de "Design Effect" (effet du plan de sondage) développé par L. KISH.

Tableau 4 : Effet des probabilités d'inclusion sur la précision du sondage, coefficients de variation des estimateurs et écarts de variance en % de la variance sous plan équiprobable

Sondages :	Variables à estimer :	Population totale	Population âgée de cinq ans ou plus	Taille moyenne des ménages	Proportion des plus de cinq ans dans la pop. totale
1.Marseille (recensement INSEE 1982) sondages non stratifiés					
Plan équiprobable (p=cste)					
	<i>Coefficient de variation :</i>	9.83	9.66	3.72	0.47
	<i>Ecart de variance :</i>	0	0	0	0
Plan à probabilité proportionnelle à la surface de l'ilot (p=su)					
	<i>Coefficient de variation :</i>	10.13	10.07	3.23	0.43
	<i>Ecart de variance :</i>	+ 6%	+ 9%	-25%	-16%
2.Marseille (INSEE 1982) sondages stratifiés en 9 strates morphologiques					
Plan équiprobable (p=cste)					
	<i>Coefficient de variation :</i>	7.62	7.56	2.28	0.29
	<i>Ecart de variance :</i>	0	0	0	0
Plan à probabilité proportionnelle à la surface de l'ilot (p=su)					
	<i>Coefficient de variation :</i>	7.67	7.59	2.50	0.33
	<i>Ecart de variance :</i>	+1%	+1%	+20%	+30%
3.Quito (enquêtes préliminaires 1987) sondages non stratifiés					
Plan équiprobable (p=cste)					
	<i>Coefficient de variation :</i>	24.91	25.34	3.71	1.02
	<i>Ecart de variance :</i>	0	0	0	0
Plan à probabilité proportionnelle à la surface de l'ilot (p=su)					
	<i>Coefficient de variation :</i>	25.42	26.45	4.07	1.12
	<i>Ecart de variance :</i>	+4%	+9%	+20%	+21%
Plan à probabilité proportionnelle à la surface bâtie de l'ilot (p=sb)					
	<i>Coefficient de variation :</i>	15.97	16.64	3.08	0.99
	<i>Ecart de variance :</i>	-59%	-7%	-31%	-6%
4.Quito (1987) sondages stratifiés en 6 classes de densité de bâti					
Plan équiprobable (p=cste)					
	<i>Coefficient de variation :</i>	15.92	16.14	2.68	0.83
	<i>Ecart de variance :</i>	0	0	0	0
Plan à probabilité proportionnelle à la surface de l'ilot (p=su)					
	<i>Coefficient de variation :</i>	12.30	12.77	2.83	0.85
	<i>Ecart de variance :</i>	-40%	-37%	+12%	+5%
Plan à probabilité proportionnelle à la surface bâtie de l'ilot (p=sb)					
	<i>Coefficient de variation :</i>	12.23	12.85	2.81	0.88
	<i>Ecart de variance :</i>	-41%	-37%	+10%	+12%

Le phénomène le plus intéressant est celui que révèlent les résultats obtenus à Quito, très différents de ceux de Marseille. A Quito, en l'absence de stratification, le sondage $p=su$ est peu efficace (lignes 5 et 6) alors qu'un sondage avec probabilités proportionnelles aux surfaces bâties ($p=sb$) serait nettement meilleur que l'équiprobable (lignes 5 et 7). Mais lorsqu'on stratifie la base sur un critère de densité du bâti, la conclusion s'inverse : *le sondage $p=su$ réduit sensiblement la variance des estimations de totaux* et ne dégrade qu'assez peu celle des ratios (lignes 8 et 9), qui de toute façon profitent beaucoup plus de la stratification qu'il ne pâtissent de ces probabilités (voir Tableau 4). *Quant au sondage $p=sb$, il a sous cette stratification, des performances totalement équivalentes au sondage $p=su$* (lignes 9 et 10).

La synthèse des observations faites à partir des deux tableaux permet d'affirmer qu'*une fois stratifiée la base de sondage selon la densité du bâti, le sondage à probabilité proportionnelle à la surface totale des unités est la solution qui s'impose*. D'une part il est *meilleur que le sondage équiprobable stratifié et a fortiori non stratifié*, d'autre part sous ce plan *le gain induit par la stratification est important (réduction de la variance de 35 à 75 %)*. Enfin, une fois retenu le principe de la stratification sur la densité du bâti, *le tirage $p=sb$ n'améliore pas l'estimation par rapport au tirage $p=su$* . Il est donc inutile de rechercher plus avant un mode de tirage qui garantisse une probabilité approximativement proportionnelle aux surfaces bâties ; cette recherche se heurterait d'ailleurs à de nombreuses difficultés pratiques. Ce dernier résultat trouve une explication intuitive simple : à l'intérieur d'une même classe de densité de bâti, surface totale et surface bâtie des unités primaires sont *grosso modo* proportionnelles et présentent donc les mêmes corrélations avec les caractères démographiques.

3.2. L'effet du tirage systématique à l'aide d'une grille de points

Pour mettre en oeuvre, à partir de l'image satellite, un sondage stratifié sur le critère de la densité du bâti à probabilités d'inclusion proportionnelles aux surfaces des unités primaires, on utilise une technique de sélection de l'échantillon par tirage de points dans chaque strate matérialisée sur l'image. On a alors le choix entre une sélection des points par tirage de coordonnées aléatoires et une sélection systématique au moyen d'une grille de points.

Les caractères démographiques que nous cherchons à estimer n'étant pas indépendants de la localisation des ménages dans l'agglomération urbaine, le tirage systématique est certainement avantageux. Il est malheureusement impossible de mesurer rigoureusement le gain sur la précision des estimations parce qu'il n'existe pas, dans le strict cadre de la théorie des sondages probabilistes, d'expression de la variance des estimateurs sous un plan de sondage systématique.

Pour tourner cette difficulté, nous avons choisi une méthode d'approximation qui s'inspire d'une proposition faite par DESABIE pour le tirage systématique sur liste⁸ et que nous adaptons au tirage systématique spatial. L'idée est la suivante : la sélection systématique se fait dans une liste d'unités d'échantillonnage triée selon un certain ordre ; *si cet ordre est informatif par rapport aux variables qu'on cherche à estimer, alors on peut considérer que le sondage bénéficie d'une sorte de stratification implicite selon le ou les critères qui définissent cet ordre.* On est alors fondé à utiliser un formulaire d'inférence pour les estimateurs et leurs variances correspondant à un sondage stratifié à partir de l'ordre de la liste. Dans le cas du sondage systématique qui nous intéresse, l'ordre qui structure la base de sondage et que l'on juge informatif est la localisation géographique des îlots. *L'effet de stratification implicite peut donc être approché en considérant la base comme stratifiée géographiquement, par exemple selon un maillage carré de l'image.*

Mais bien sûr il n'y a pas non plus indépendance entre localisation et caractéristiques morphologiques des îlots. Par conséquent, cette stratification géographique implicite doit être croisée avec la vraie stratification morphologique afin de ne pas rendre compte du seul effet isolé du tirage systématique mais des effets cumulés de ce mode de sélection et de la stratification sur la densité du bâti.

Une fois définies ces stratifications croisées, c'est à nouveau le programme VARECH qui est notre outil d'évaluation. Le nombre des strates géographiques doit logiquement être maximum pour obtenir une bonne approximation des variances réelles du sondage systématique. C'est donc à la base de données exhaustive constituée à Marseille, qui comprend un nombre suffisant

⁸ J. DESABIE, "Théorie et pratique des sondages" (DUNOD 1966) : Le tirage systématique envisagé comme un sondage stratifié, Ch. 7.7.7., p. 171.

d'unités primaires (3534 îlots), que la méthode a été appliquée. De ce fait les temps de calcul sur micro-ordinateur correspondant à des stratifications géographiques très fines sont vite prohibitifs. Cependant, comme nous avons constaté une convergence assez rapide de la suite des variances obtenue avec un nombre de strates augmentant jusqu'à 15, nous avons retenu ce nombre comme un bon compromis entre la précision de l'approximation et des temps de calcul raisonnables.

Le Tableau 5 donne, pour trois caractères démographiques, les coefficients de variation des estimations et les gains de variance correspondant à deux stratifications morphologiques de la base (en neuf strates morphologiques issues de la photo-interprétation et en six classes de densité du bâti), croisées avec la stratification géographique en quinze classes. Le gain de variance est là encore mesuré en pourcentage de la variance du même sondage non stratifié : ici le plan de sondage considéré dans tout le tableau est celui à probabilités d'inclusion proportionnelles aux surfaces des îlots.

On remarque d'abord, à l'examen du tableau, que *les gains permis par la bonne répartition spatiale de l'échantillon peuvent être égaux et même supérieurs à ceux dus à la stratification morphologique* : ils varient suivant le caractère démographique auquel on s'intéresse, entre 25 et 55 % de la variance de l'estimateur déjà stratifié sur le critère morphologique. Au total l'effet cumulé de la stratification et du tirage systématique permet une diminution de la variance comprise entre 45 et 80 % alors que le sondage aléatoire avec les mêmes stratifications ne la réduit que de 20 à 50 %.

**Tableau 5 : Coefficients de variation⁹ 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
(calculés sur MARSEILLE)**

Caractères :	POPSDC (*)		$\frac{POP+5}{POPTOTAL}$ (**)		$\frac{POPFRANC}{POPTOTAL}$ (**)	
	(1)	(2)	(1)	(2)	(1)	(2)
Sondages :						
Non stratifié	10.13	0	0,43	0	3,67	0
1) Marseille : croisements avec la stratification morphologique en neuf classes d'après photo-interprétation						
Strat. morph.(9 cl.)	7.67	43%	0.33	41%	2.55	52%
morph.(9 cl.) x loc.(15 cl.)	6.83	55%	0.27	61%	1.72	78%
2) Marseille : croisements avec la stratification en six classes de densité du bâti d'après photo-interprétation						
Strat. dens.(6 cl.)	8.59	28%	0.39	18%	2.87	39%
Dens.(6 cl.) x loc.(15 cl.)	7.48	45%	0.31	48%	1.91	73%

(*) : Population totale (sans doubles comptes), estimateur sans biais

(**) : respectivement, proportion de la population totale âgée de plus de cinq ans et proportion de la population totale de nationalité française, estimateurs par le ratio légèrement biaisés

(1) : Coefficient de variation de l'estimation

(2) : Gain de précision, en % de la variance du sondage non stratifié

Autre constatation intéressante : le gain qu'apporte la "stratification spatiale" est équivalent, pour un caractère donné, sous les deux stratifications

⁹ Plan à probabilités proportionnelles aux surfaces des districts, estimation sans biais des totaux et estimation légèrement biaisée des quotients.

morphologiques de départ ; nous avons d'ailleurs pu vérifier cette stabilité pour d'autres stratifications. Il semble donc que *les deux techniques, stratification et tirage systématique, permettent de prendre en compte des niveaux géographiques distincts de variation des caractères démographiques*. Le découpage de l'espace urbain en strates morphologiques homogènes rend compte d'une partie des variations du caractère (variance "macro" ou inter-strates) tandis qu'à l'intérieur des strates, le tirage systématique permet d'appréhender des variations plus locales (variance "micro" ou intra-strate). Une conséquence pratique de cet état de fait est qu'on peut penser que *même sous des stratifications morphologiques plus fines et donc plus efficaces, le sondage systématique conserve son intérêt*.

Enfin, au plan méthodologique, soulignons que la méthode d'approximation de la variance d'un sondage systématique spatial employée ici, si elle a l'avantage d'être intuitive et simple, est en contrepartie assez grossière : *elle peut engendrer une surestimation importante des variances*. Ce défaut fut peu gênant dans la phase de développement de la méthode où le but poursuivi était d'évaluer l'intérêt de la technique, mais il l'est beaucoup plus lors de l'exploitation d'une enquête menée à l'aide de la méthode que nous proposons, où il importe de mesurer le plus précisément possible l'erreur d'échantillonnage. D'autres méthodes existent dont certaines ont déjà fait leurs preuves comme les méthodes dites de ré-échantillonnage : BRR, JACKKNIFE, BOOTSTRAP. Dans la mesure où elles sont disponibles dans les logiciels d'exploitation d'enquête, nous conseillons leur utilisation.

3.3. L'effet de grappe au second degré du sondage

Jusqu'à présent, les résultats que nous avons présentés ont montré que sur l'image satellite stratifiée selon la densité du bâti un sondage systématique à l'aide d'une grille de points, qui entraîne une probabilité d'inclusion des îlots proportionnelle à leur surface, était une technique efficace de sélection des unités primaires. Pour accéder à l'information sur les ménages, unités d'observation de l'enquête, il faut ajouter un second degré au sondage où l'on sélectionne, dans chaque unité primaire de l'échantillon, un échantillon de ménages. *Quelle stratégie adopter pour ce second degré ?*

Dans un sondage à deux degrés, l'expression de la variance comporte deux termes correspondant aux échantillonnages à chacun des degrés. Le premier terme représente l'erreur aléatoire 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, l'erreur due au second degré du sondage, vient de la variabilité du caractère dans l'ensemble des ménages d'une même unité primaire. Dans chacun des deux termes interviennent bien sûr les taux au premier et second degré du sondage. *Evidemment, 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 en n'enquêtant en revanche qu'une fraction des ménages dans chaque îlot de l'échantillon.*

Le programme VAR2DEG a permis de vérifier cette hypothèse à partir des données collectées dans l'enquête préliminaire réalisée à Quito. Il réalise la stricte application du formulaire de la variance d'estimation d'un total sous un plan de sondage à deux degrés en isolant ses deux termes : A, variance due au premier degré ou variance inter-îlots, B variance due au second degré ou somme des variances intra-îlots. Les résultats que nous présentons correspondent au plan de sondage qui nous intéresse : au premier degré les probabilités d'inclusion sont proportionnelles à la surface des unités primaires, le second degré est équiprobable sans remise.

On constate dans le Tableau 6 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 dans un grand nombre de pâtés de maisons est bien préférable à un sondage en grappe où les pâtés de maisons sont enquêtés exhaustivement.* On doit donc choisir un taux au second degré le plus faible possible pour optimiser la précision du sondage, mais apparaissent alors deux contraintes. D'une part, il faut conserver un nombre suffisant de ménages dans chaque unité primaire pour pouvoir estimer précisément la variance des estimateurs. D'autre part et surtout, l'augmentation du taux au premier degré entraîne un surcoût lié aux frais de transport et aux opérations de collecte strictement attachées aux unités primaires : dénombrement des ménages et mesure de surfaces.

Une solution acceptable du problème est *d'enquêter un nombre constant de ménages dans chaque îlot*. En effet, si l'on décompose à nouveau le calcul de la variance pour évaluer la contribution de groupes de pâtés de maisons contenant des nombres de ménages à peu près constants, 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 premier degré le plus fort*, tandis qu'on peut enquêter exhaustivement les pâtés de maisons les moins peuplés. Les probabilités d'inclusion finale des ménages dans l'échantillon vont donc dépendre d'une part de la surface de l'îlot dans lequel ils se trouvent, d'autre part du nombre de ménages qui y résident.

Enfin le programme permet de mesurer les conséquences sur la précision du sondage de cet effet de grappe dans les îlots : par rapport à un sondage en grappe sous le même plan où l'on enquêterait tous les ménages de chaque unité primaire, la variance est divisée par cinq.

Tableau 6 - Comparaison des variances dues aux premier et second degrés du sondage (Programme VAR2DEG)

Variances : Caractères :	A (x 10 ⁻⁶)	B (x 10 ⁻⁶)	$\frac{A}{A+B}$
Population totale	40,75	0,25	0,99
Population des plus de 5 ans	31,47	0,23	0,99
Population des non migrants	13,16	0,32	0,98

A = variance due au premier degré ou variance inter-îlots

B = variance due au second degré ou somme des variances intra-îlots

Les résultats des expériences menées sur les données du recensement de 1982 à Marseille et à partir des enquêtes préliminaires réalisées à Quito, que

nous avons présentés ici, nous ont apporté tous les éléments pour élaborer un plan de sondage complet. Le point suivant décrit comment ce plan doit être mis en oeuvre pour réaliser une enquête démographique par sondage à partir d'une image satellite.

IV - LA METHODE D'ENQUETE DEMOGRAPHIQUE PAR SONDEGE AREOLAIRE SUR IMAGE SATELLITE

4.1. Description générale

Prenant en considération les contraintes d'application de la méthode dans des villes des pays en développement, les conséquences de l'utilisation d'une image satellite comme base de sondage et les résultats des expérimentations préliminaires sur les villes de Marseille et Quito, nous avons abouti à la définition d'un plan de sondage aréolaire sur image satellite adapté à notre objectif : la production rapide de données démographiques en ville. Le Tableau 7 en précise les principales caractéristiques.

La méthode de sondage développée consiste en un sondage aréolaire à deux degrés. Au 1er degré, les aires sélectionnées sont des *îlots* (pâtés de maisons) ; au 2ème degré, les unités enquêtées sont des *ménages*.

L'image satellite sert :

- à la *définition* de la base de sondage : la limite urbaine est tracée sur l'image, à partir de l'information fournie par celle-ci ;
- à la *stratification* de la base de sondage : une technique efficace et adaptée au savoir-faire actuel en télédétection urbaine est la stratification sur le critère de densité du bâti ;
- à la *sélection d'un échantillon d'îlots* géographiquement répartis sur l'ensemble de la ville.

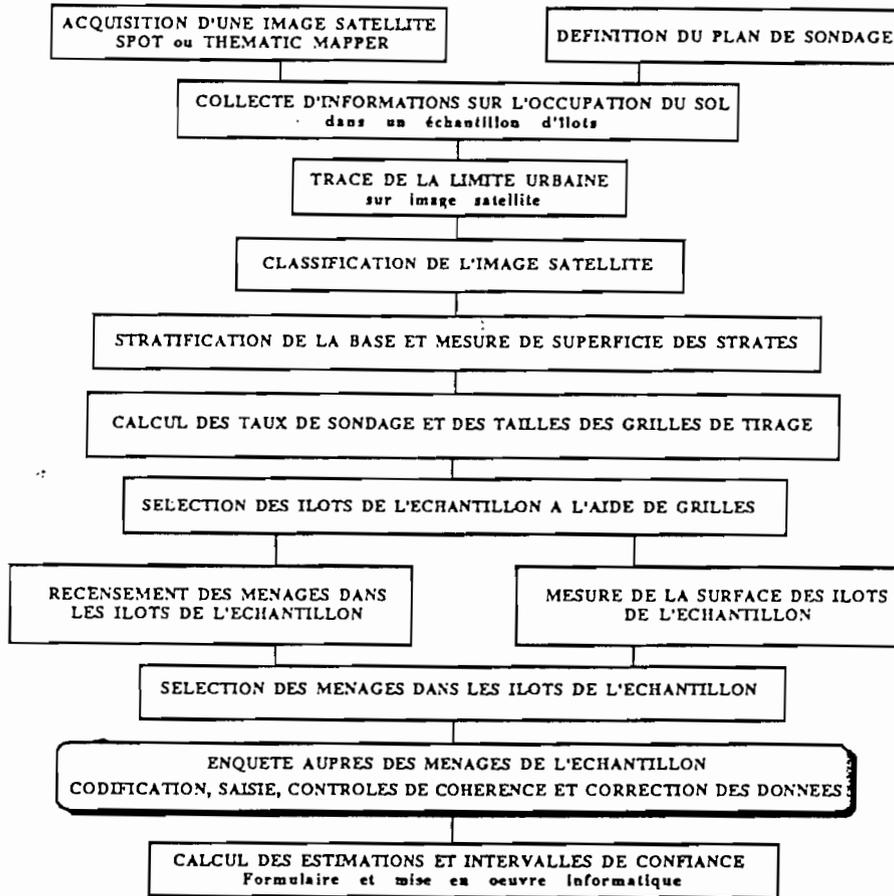
Tableau 7 - Caractéristiques du plan de sondage

Structure générale	Sondage aréolaire stratifié à deux degrés Unités primaires = îlots Unités secondaires = ménages
Base de sondage	Ensemble de l'aire urbaine, dont sont exclues les zones non bâties
Stratification	Stratification à partir de la classification de l'image satellite en fonction de la densité du bâti
Sélection des îlots (premier degré)	Sondage spatial systématique à l'aide d'une grille de points superposée à l'image satellite, le placement initial de la grille sur l'image étant aléatoire. Taux respectant la règle d'allocation aux strates choisie : proportionnelle ou optimale. Probabilité de sélection proportionnelle à la surface des îlots
Sélection des ménages (deuxième degré)	Tirage équiprobable systématique sur liste des ménages de chaque îlot de l'échantillon. Taux variable selon le nombre de ménages de l'îlot
Estimateurs	Total : estimateur sans biais sous plan à probabilités proportionnelles aux surfaces des îlots Quotient : estimateur du ratio sous plan à probabilités proportionnelles aux surfaces des îlots

4.2. Les étapes de sa mise en oeuvre

Pour mettre en oeuvre le plan de sondage décrit ci-dessus, un certain nombre d'opérations doivent être réalisées ; la figure 1 résume l'enchaînement de ces différentes étapes. Les paragraphes suivants sont consacrés à une brève description de chacune des principales étapes.

Figure 1 - Les différentes étapes de la mise en oeuvre de la méthode



* La délimitation de la base de sondage sur image satellite

Délimiter la base de sondage consiste à tracer la limite urbaine sur l'image satellite. Etant donné le type d'information fournie par les images satellite, il est bien évident qu'on ne peut extraire de celles-ci que des limites traduisant un phénomène physique ; une solution intéressante est de définir la ville comme une zone continue d'espace bâti.

Si l'on retient cette définition, deux techniques différentes peuvent être mises en oeuvre, selon le type d'image dont on dispose :

- si l'on ne dispose que du film de l'image panchromatique SPOT : le tracé est réalisé sur un tirage photographique de cette image à une échelle comprise entre le 1/30 000 et le 1/15 000, en utilisant les teintes de gris de l'image, l'agencement général, la structure et la texture qui renseignent sur le

degré de végétation ou de minéralisation (construction) et le type de tissu urbain ; la limite est tracée là où se termine une zone continue d'espace bâti, en s'efforçant de suivre une limite (naturelle ou artificielle) visible sur l'image et sur le terrain ;

- si l'on dispose, en plus, d'une bande magnétique SPOT XS ou TM : il est possible de réaliser un traitement numérique simple pour aider à la détermination de la limite urbaine, qui sera établie, comme précédemment, sur un tirage à grande échelle du canal panchromatique. L'objet de ce traitement est de classer l'image selon l'indice de végétation et le degré d'hétérogénéité des valeurs radiométriques ; faisant l'hypothèse que la zone urbaine correspond à un espace à dominante minérale et fortement hétérogène, on peut ainsi isoler, numériquement, la zone urbanisée.

Dans le cas où la base de sondage doit nécessairement correspondre à des limites administratives, le travail de délimitation consiste uniquement à reporter le tracé de ces limites, connues par un document cartographique quelconque, sur l'image satellite.

* La stratification de la base d'après l'information fournie par l'image satellite

La stratification de la base de sondage utilise une classification supervisée de l'image satellite, c'est-à-dire initialisée sur une information précise collectée, au sol, dans un certain nombre d'îlots.

La méthode de classification que nous préconisons est une classification selon la densité de l'espace bâti : aisément modélisable à partir des images satellite, la densité du bâti est, d'autre part, un critère de stratification efficace pour une enquête socio-démographique.

Cette classification est fondée sur la recherche de la meilleure régression entre la densité du bâti observée sur le terrain dans les îlots test et l'indice de végétation (dans le cas d'une image SPOT) ou les moyennes radiométriques des canaux 3, 4 et 5 d'une image TM ; les coefficients de la régression sont ensuite appliqués à l'ensemble de l'image. Après avoir partitionné l'image résultante en n classes, on dispose donc d'une classification de la zone urbanisée en fonction de n niveaux de densité du bâti.

Une fois en possession de cette classification, il s'agit d'élaborer la stratification proprement dite, c'est-à-dire de délimiter des zones les plus homogènes possibles au regard de ce critère de densité du bâti. Cette opération est réalisée sur un tirage à grande échelle de l'image satellite, en s'efforçant de suivre des frontières, naturelles ou artificielles, afin de ne pas couper les pâtés de maisons, qui constitueront l'unité de base de l'échantillon au premier degré.

Enfin, la mesure de superficie des strates étant une information nécessaire pour le calcul des estimateurs, on mesure les zones ainsi délimitées.

* La sélection de l'échantillon d'îlots

Le premier travail consiste à déterminer le nombre total de pâtés de maisons à enquêter et leur répartition dans les strates. Le tirage de l'échantillon de pâtés de maisons s'effectue ensuite indépendamment dans chaque strate à l'aide d'une grille placée sur l'image satellite, la maille de la grille étant calculée pour respecter le taux prévu ; pour obtenir la surface de la maille de la grille de tirage dans chaque strate, il suffit de diviser la surface totale de la strate mesurée sur l'image par le nombre d'îlots à sélectionner.

Une fois que l'on s'est assuré que l'on trouve dans chaque strate le nombre de points de sondage voulus, l'étape suivante consiste à identifier sur l'image les îlots correspondant aux points de la grille. Lorsque le type de quartier, et donc le degré de lisibilité de la voirie sur l'image rendent impossible une identification directe des îlots sur l'image, ce travail est réalisé, avec le superviseur chargé de la zone, directement sur le terrain.

* Le recensement des ménages et la mesure des surfaces des îlots de l'échantillon

Le recensement des ménages qui occupent les pâtés de maisons sert à sélectionner l'échantillon de ménages à enquêter et à estimer les totaux, moyennes et quotients au sein des unités primaires, c'est-à-dire les pâtés de maisons.

Ce travail de recensement des ménages dans chaque îlot de l'échantillon est confié aux superviseurs qui établissent, sur une fiche prévue à cet effet, un schéma simplifié du pâté de maisons sur lequel figurent : les

renseignements nécessaires à l'identification du pâté de maisons pour l'enquêteur (noms des rues bordant l'îlot, éléments naturels servant de limites...etc), le dessin de l'ensemble des bâtiments de l'îlot, le nombre de ménages dans chacun des bâtiments, la numérotation des ménages de 1 à N, par ordre croissant en suivant le tour de l'îlot.

Après le recensement des ménages, le superviseur effectue la mesure de la surface de l'îlot : cette mesure est indispensable pour pratiquer les estimations des totaux, moyennes et quotients au niveau de chacune des strates, de la base entière ou d'un autre éventuel découpage de la ville.

A l'aide d'un décamètre, d'un clinomètre et d'une boussole, le superviseur établit 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. La surface est ensuite calculée, informatiquement ou manuellement, à partir du schéma et des mesures effectuées par le superviseur.

* La sélection des ménages à enquêter

Comme on l'a vu au point III, la stratégie retenue pour le deuxième degré du sondage est d'enquêter, autant que possible, le même nombre de ménages dans tous les îlots. En fait, on adopte la stratégie suivante de sélection des ménages : on enquête un petit nombre constant de ménages (n_1 , de l'ordre de 5 ou 10) dans tous les îlots contenant au moins n_1 ménages, et on enquête tous les ménages dans les îlots contenant moins de n_1 ménages.

Si l'îlot compte plus de n_1 ménages, le superviseur procède à un tirage systématique sur la liste des ménages de l'îlot numérotés de 1 à N ; il établit ainsi la liste des n_1 ménages à enquêter. L'enquêteur doit enquêter tous les ménages de la liste.

* La production des résultats : estimations et intervalles de confiance

La dernière étape consiste à estimer les résultats démographiques au niveau de la population entière à partir des informations collectées sur l'échantillon de ménages. Le calcul des estimateurs est simple à condition, comme nous l'avons vu, de ne pas tenir compte du tirage systématique de l'échantillon d'îlots. En assimilant le tirage au premier degré à un sondage aléatoire avec

Tableau 8 - Notations et formulaire pour l'estimation d'un total sous le plan préconisé

a. Notations

Notations	Significations
1) Base de sondage et échantillon N_i n_i m_k	Nombre de ménages dans l'îlot i Nombre de ménages de l'îlot i dans l'échantillon Nombre d'îlots de la strate k dans l'échantillon
2) Variable à estimer y_{ij} \hat{y}_i \hat{Y}_k	Valeur du caractère pour le ménage j de l'îlot i Estimateur du total de y dans l'îlot i Estimateur du total de y dans la strate k
3) Variable définissant la probabilité d'inclusion dans l'échantillon des unités primaires z_i Z_k	Surface totale de l'îlot i Surface totale de la strate k
4) Sommations sur les échantillons n_i $S_{j=1}$ m_k $S_{i=1}$	Somme sur l'échantillon de ménages dans l'îlot i Somme sur l'échantillon d'îlots de la strate k
5) Estimation des erreurs d'échantillonnage $\hat{V}(\hat{Y}_k)$	Estimations de la variance de l'estimateur \hat{Y}_k

b. Formulaire

Quantité à estimer	Estimateur dans l'îlot i	Estimateur dans la strate k	Estimateur de la variance de l'estimateur dans la strate k
Y_k	$\hat{y}_i = N_i \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i}$	$\hat{Y}_k = \frac{Z_k}{m_k} \sum_{i=1}^{m_k} \frac{\hat{y}_i}{z_i}$	$\hat{V}(\hat{Y}_k) = \frac{1}{m_k(m_k-1)} \sum_{i=1}^{m_k} \left(\frac{Z_k \hat{y}_i}{z_i} - \hat{Y}_k \right)^2$

remise d'unités spatiales ayant des probabilités d'inclusion proportionnelles à leurs surfaces, le formulaire d'estimation découle de la théorie des estimateurs sans biais avec probabilités d'appartenance inégales. Au second degré, le plan est équiprobable sans remise et l'estimateur sans biais est bien connu. Le Tableau 8 donne l'exemple du formulaire d'estimation d'un total que nous avons appliqué pour l'exploitation de l'enquête Migrations.

V - CONCLUSIONS ET PERSPECTIVES

Les travaux effectués depuis 1985 sur Marseille et Quito ont vérifié l'hypothèse de base du programme de recherche : la prise en compte de la morphologie urbaine telle qu'elle peut être extraite d'une image satellite SPOT ou TM permet de définir un plan de sondage rigoureux et performant. Ces recherches sont à l'origine de résultats de portée générale en sondage aréolaire et télédétection urbaine, résultats utilisés pour établir la chaîne complète d'une méthode de production de données démographiques intégrant la télédétection spatiale.

La première application de cette méthode à Quito, en décembre 1987, a confirmé son intérêt et permis d'évaluer les coûts de sa mise en oeuvre et la précision des résultats : dans une ville ne disposant d'aucune base de sondage classique, nous avons pu sélectionner de façon rigoureuse à partir d'une image satellite SPOT, un échantillon pour une enquête démographique.

La méthodologie que nous avons développée peut apporter des éléments de réponse aux problèmes posés par l'observation des populations des villes des pays en développement : en suivant les étapes décrites dans le point 4 de cet article, cette technique de production de données démographiques pourrait être appliquée dès maintenant dans d'autres villes. Avec cette méthode, il est possible de mettre en oeuvre rapidement une enquête par sondage dans une ville ne disposant d'aucune base de sondage classique. *Rapidité et spatialisation* des résultats en fonction de découpages divers de la ville sont des caractéristiques de la méthode qui devraient séduire les différents acteurs de la gestion urbaine. Sélectionner l'échantillon d'une enquête socio-démographique sur image privilégie aussi la mise en évidence des *différences internes* à la ville et l'analyse de la

dynamique des sous-populations citadines. L'intérêt de ces approches de l'urbanisation est souvent souligné dans les recherches sur les villes des pays en développement. L'allègement de l'échantillon enquêté grâce au gain de précision apporté par le plan de sondage autorise, pour un même coût global, une *observation plus fouillée* que par les méthodes traditionnelles de sondage.

Si la méthode développée peut d'ores et déjà être utilisée en appliquant le schéma proposé actuellement, elle est aussi, bien évidemment, perfectible.

D'une part, la multiplication des applications dans des contextes urbains différant tant sur un plan physique que socio-économique contribuera nécessairement à un perfectionnement de la méthode de sondage aréolaire sur image satellite que nous avons présentée dans cet article.

D'autre part, il est évident que la seule information sur la densité du bâti ne rend compte que de manière très fruste de la morphologie d'une ville et ne constitue qu'une infime partie du "potentiel" d'informations que constitue une image satellite SPOT ou TM : le développement de méthodes pour extraire de l'image des informations sur la hauteur et l'agencement des bâtiments et l'intégration de celles-ci dans la stratification et/ou dans les estimateurs permettraient d'augmenter la précision des résultats de l'enquête démographique.

Accumuler des expériences d'application de notre méthode, intégrer les innovations que connaîtra la télédétection urbaine contribueront à améliorer la solution que nous proposons aujourd'hui.

INDICATIONS BIBLIOGRAPHIQUES

1. TELEDETECTION AERIENNE OU SPATIALE ET DEMOGRAPHIE

ADENIYI P.O. : An aerial photographic method of estimating urban population. *Photogrammetric engineering and remote sensing*, Vol. 49, n° 4, pp. 545-560, 1983.

BALLUT A. et GAUTHIER M. : Une méthode pour la production de données socio-économiques spatialisées dans le cadre d'un schéma directeur d'aménagement et d'urbanisme. IAURIF, Paris, 55 p., 1983.

CARDIERI L. et NASCIMENTO I. : Evaluation de l'occupation et de l'utilisation du sol de la région métropolitaine de Sao Paulo par télédétection : méthode d'estimation et projection des populations à partir d'une image SPOT. Paris, IAURIF, EMPLASA, MS21, 1989.

COLLINS W.G. et EL BEIK A.H. : Population census with the aid of aerial photographs : an experiment in the city of Leeds. *Photogrammetric record* 7, n° 37, pp. 16-26, 1971.

GREEN N.E. : Aerial photographic analysis of residential neighborhoods : an evaluation of data accuracy. *Social Forces*, n° 35, pp. 142-1247, 1956.

HOLZ R.K., HUFF D.L., MAYFIELD R.C. : Urban spatial structure based on remote sensing imagery. *Sixth International Symposium Remote Sensing Environment*, Vol. II, pp. 819-830, 1969.

HSU S.Y. : Population estimation. *Photogrammetric engineering and remote sensing*, Vol. 37, n° 5, pp. 449-454, 1971.

IAURIF, ROC, SIDAU, SEE : Les données urbaines. Système permanent de production rapide et économique. Recherche et synthèse. Paris, 82 p., 1983.

- IAURIF, ICEA, ENSTP : Expérimentation d'une méthode simplifiée de recueil de données urbaines à Bouaké. Côte d'Ivoire. Première phase et deuxième phase. Paris, IAURIF, 78 p., 1987 et 1989.
- KRAUS S.P., SENGER L.W., RYERSON J.M. : Estimating population from photographically determined residential land use types. *Remote sensing of environment*, n° 3, pp. 35-42, 1974.
- LO C.P. : Accuracy of population estimation from mediuscale aerial photography. *Photogrammetric engineering and remote sensing*, Vol. 52, n° 12, pp. 1859-1869, 1986.
- LO C.P. et WELCH R. : Chinese urban population estimates. *Annals of the Association for American Geographers*, Vol. 67, pp. 246-250, 1977.
- OGROSKY C.E. : Population estimates from satellite imagerie. *Photogrammetric engineering and remote sensing*, Vol. XLI, ° 6, pp. 707-712, 1986.
- POLLE V.F.L. : Population estimation from aerial photos for non-homogeneous urban presidential areas. Cases studies from Teheran and Colombo. *ITC Journal*, n° 2 pp. 116-122, 1984.
- PORTER P.W. : Population distribution and land use in Liberia. London School of Economics and Political Science, Thèse de Ph. D, 1956.
- TOBLER W.R. : Satellite confirmation of settlement size coefficients. *Area*, Vol. 1, n° 3, pp. 30-34, 1969.
- VERNIERE M. : 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, 1978.
- WATKINS J.F. : Small area population estimates using aerial photography. *Photogrammetric engineering and remote sensing*, Vol. 51, n° 2, pp. 1933-1935, 1985.

2. THEORIE DES SONDAGES

- ASSELIN L.M. : Techniques de sondage avec application à l'Afrique, GAETAN MORIN, Québec, 697 p., 1984.

COCHRAN W.G. : Sampling techniques (3^e édition), J. WILEY, New York, 428 p., 1977.

DESABIE J. : Théorie et pratique des sondages, DUNOD, Paris, 483 p., 1966.

DROESBEKE, FICHET, TASSI éditeurs (Ouvrage collectif) : Les sondages, Journées d'étude en statistique d'octobre 1986, ECONOMICA, Paris, 310 p., 1987.

GOURIEROUX Ch. : Théorie des sondages, ECONOMICA, Paris, 272 p., 1981.

GROSBRAS J.M. : Méthodes statistiques des sondages, ECONOMICA, Paris, 342 p., 1987.

3. BIBLIOGRAPHIE SELECTIVE DE L'EQUIPE ORSTOM

BARBARY O. : 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. Paris, EHESS, Thèse de doctorat, 641 p., 1988.

BARBARY O., DUREAU F., MICHEL A. : Définition et mise en oeuvre d'un plan de sondage aréolaire sur image SPOT pour une enquête démographique à Quito (Equateur). Quito, ORSTOM, 204 p., 1988.

DUREAU F. et GUILLAUME A. : La population depuis l'espace : un nouveau système d'observation ? *STATECO*, n° 38. Paris, INSEE, pp. 5-46, 1985.

DUREAU F., BARBARY O., LORTIC B., MICHEL A. : Une nouvelle méthode de collecte d'information sur les populations urbaines : l'enquête démographique par sondage aréolaire sur image satellite. Communication au XXI^e congrès général de l'Union Internationale pour l'Etude Scientifique de la Population, New Delhi (Inde), 20-27 septembre 1989, 17 p., 1989.

DUREAU F., BARBARY O., LORTIC B., MICHEL A. : L'utilisation de la télédétection pour l'observation démographique en milieu urbain. Convention Ministère de la Coopération-ORSTOM. Rapport final. Paris, ORSTOM, 31 p., 1989.

DUREAU F., BARBARY O., MICHEL A., LORTIC B. : Sondages aréolaires sur image satellite pour des enquêtes socio-démographiques en milieu urbain. Manuel de formation. (3 versions : français, anglais, espagnol). Paris, ORSTOM, *Collection Didactiques*, 40 p., 1989.

Equipe ORSTOM (UR 406, ATOB, Unité d'infographie) : Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville. ATP CNRS/CNES. Rapport intermédiaire, avril 1986. Rapport final, août 1987. Paris, ORSTOM, 190 p. et 591 p., 1986 et 1987.

Equipe ORSTOM (UR 15 du département SDU) : L'utilisation de l'imagerie SPOT pour l'observation démographique en milieu urbain. Rapport final CNES, avril 1988. Paris, ORSTOM, 217 p., 1988.

MICHEL A. : Stratification de l'espace urbain à partir d'images satellite pour réaliser un sondage à objectif démographique. Mise au point et évaluation des méthodes d'analyse des images SPOT et LANDSAT TM en milieu urbain. Paris, EHESS, Thèse de doctorat, 369 p., 1988.

MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F. : L'utilisation de la télédétection pour l'observation des populations urbaines. Une recherche en cours à Quito (Equateur). in : SPOT 1. *Utilisation des images, bilan, résultats*. Paris, novembre 1987. Paris, CNES, pp. 505-514, 1987.

Dureau Françoise, Barbary Olivier. (1990).

L'enquête par sondage sur image satellite : une solution pour améliorer l'observation des populations citadines.

Statéco, (67), 63-100.

ISSN 0224-098X