

2^E MARS 1987

SONDAGES AREOLAIRES SUR IMAGE SATELLITE POUR DES ENQUETES SOCIO-DEMOGRAPHIQUES EN MILIEU URBAIN

MANUEL DE FORMATION

**MUESTREO DE AREAS EN BASE A IMAGENES DE SATELITE PARA ENCUESTAS
SOCIO-DEMOGRAFICAS EN LAS CIUDADES. Manual de capacitación**

**AREA SAMPLING FROM SATELLITE IMAGE FOR SOCIO-DEMOGRAPHIC
SURVEYS IN URBAN ENVIRONMENTS. Training handbook**

F. DUREAU, O. BARBARY, A. MICHEL, B. LORTIC

Éditions de l'ORSTOM

DI>ACTIQUES

RECHERCHE SCIENTIFIQUE POUR LE DÉVELOPPEMENT EN COOPÉRATION

A 30031
872

SONDAGES AREOLAIRES SUR IMAGE SATELLITE POUR DES ENQUETES SOCIO-DEMOGRAPHIQUES EN MILIEU URBAIN

MANUEL DE FORMATION

F. DUREAU, O. BARBARY, A. MICHEL, B. LORTIC

Éditions de l'ORSTOM

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE POUR LE DÉVELOPPEMENT EN COOPÉRATION

Collection  ACTIQUES

PARIS 1989

O.R.S.T.O.M. Fonds Documentaire

N° :

Cote :

A30031

ex 2

21.6.90

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les "copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective" et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, "toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ayants cause, est illicite" (alinéa 1^{er} de l'article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

FICHES DECRIVANT LA MISE EN OEUVRE DE LA METHODE

- 1 La méthode de sondage : description générale, options, coûts
- 2 Introduction à la télédétection spatiale
- 3 Acquisition d'une image satellite SPOT ou THEMATIC MAPPER
- 4 Collecte d'informations sur l'occupation du sol
- 5 Eléments pour la définition du plan de sondage
- 6 Tracé de la limite urbaine
- 7 Classification de l'image satellite
- 8 Stratification de la base de sondage et mesure de superficie des strates
- 9 Calcul des taux de sondage et des tailles des grilles de tirage
- 10 Sélection des îlots de l'échantillon à l'aide des grilles
- 11 Recensement des ménages dans les îlots de l'échantillon
- 12 Mesure de la surface des îlots de l'échantillon
- 13 Sélection des ménages dans les îlots de l'échantillon
- 14 Calcul des estimations et intervalles de confiance. 1 : formulaire
- 15 Calcul des estimations et intervalles de confiance. 2 : mise en oeuvre informatique

A. LE DEVELOPPEMENT DE LA METHODE

Les populations urbaines des pays en développement posent des problèmes spécifiques d'observation du fait des caractéristiques de l'urbanisation dans ces pays : **rythmes de croissance démographique très élevés et modifications rapides du tissu urbain.**

Les recensements exhaustifs font appel à des moyens techniques, financiers et humains considérables, au regard des économies de ces pays, qui interdisent une périodicité inférieure à une dizaine d'années ; de plus, compte tenu de la durée d'exploitation des recensements, les résultats sont souvent caducs dès leur parution.

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. Mais, sans base de sondage complète et à jour, sans cartographie fidèle et exhaustive, les enquêtes par sondage se révèlent difficiles à mettre en oeuvre.

Face à ce constat, il nous a paru nécessaire de développer de nouvelles méthodes de collecte de données démographiques adaptées aux spécificités des villes des pays en développement. A cet effet, la télédétection spatiale (LANDSAT TM et SPOT), qui assure une observation continue et relativement précise de l'occupation du sol, constitue une source de données particulièrement intéressante : comme l'ont montré depuis les années cinquante les travaux réalisés sur photographies aériennes, on peut utiliser avec profit l'information exhaustive sur la morphologie urbaine pour recueillir rapidement, par sondage, des données relatives aux populations urbaines.

Le Département D (Urbanisation et socio-systèmes urbains) de l'ORSTOM a donc initié en 1985 un programme de recherche dont l'objectif central est de développer une méthode de collecte de données démographiques en milieu urbain intégrant l'information apportée par les satellites à haute résolution sur la morphologie urbaine, adaptée aux caractéristiques de l'urbanisation dans les pays en développement et aux contextes locaux d'information, d'équipement, et en moyens financiers et humains. L'idée centrale de la méthode est **d'utiliser l'image satellite comme base de sondage et d'exploiter l'information morphologique sur le milieu urbain que fournissent ces images pour stratifier un plan de sondage aréolaire applicable à une enquête démographique.**

La réalisation de l'objectif général du programme a impliqué des travaux, menés en parallèle par une équipe pluridisciplinaire¹, dans différentes directions de recherche, notamment l'interprétation des images satellite SPOT et LANDSAT TM en milieu urbain et les sondages spatiaux à objectif démographique en milieu urbain. Nous avons également été amenés à travailler sur l'intégration des données de télédétection dans des bases de données géographiques.

A l'issue de deux années de développement méthodologique mené sur les villes de Marseille (FRANCE) et Quito (EQUATEUR), nous avons pu réaliser la **première application en vraie grandeur de notre méthode de collecte** : 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.

La définition du plan de sondage de cette enquête, consacrée aux migrations, repose sur deux phases de la recherche :

- l'expérience menée sur la ville de **Marseille** : cette ville disposait, dès 1985, des informations de référence nécessaires à une première étape de développement méthodologique (image satellite Thematic Mapper, photographies aériennes à grande échelle, et recensement de la population exploitable au niveau de l'îlot) ; nous avons pu non seulement **développer** des méthodes de sondages aréolaires et d'interprétation d'image satellite SPOT et LANDSAT TM, mais encore en **tester la validité** ;
- les recherches effectuées sur le site de **Quito** : l'exploitation de deux enquêtes préliminaires a permis de préciser et de compléter les résultats méthodologiques obtenus sur la ville de Marseille.

Une première collecte d'information (octobre-décembre 1986) a consisté à relever de manière précise l'occupation du sol dans 199 îlots de Quito, afin d'initialiser et tester la fiabilité des traitements des images satellite ; en matière de sondages, l'exploitation des données démographiques recueillies en avril 1987 sur un sous-échantillon de 55 îlots choisis parmi les 199 enquêtés sur le plan morphologique est à l'origine de résultats complémentaires à ceux obtenus sur la ville de Marseille.

Cette enquête a confirmé l'intérêt de la méthode développée et permis **d'évaluer tant son niveau de précision que les coûts** qu'impliquent sa mise en oeuvre.

B. UN EXEMPLE D'APPLICATION DE LA METHODE: L'ENQUETE MIGRATIONS A QUITO

L'enquête Migrations constituant la première application de notre méthode de sondage, elle se devait de répondre à un double objectif, méthodologique et thématique. Elle devait permettre de vérifier que la méthode mise au point possède bien les caractéristiques de rapidité, faible coût et précision que nous lui attribuions au terme de la phase de développement méthodologique ; il importait également de savoir si la technique de sondage employée demeure efficace au delà du domaine thématique restreint des indicateurs démographiques les plus classiques (structures par sexe, âge, activité,...etc). L'enquête avait aussi pour fonction de produire une information fiable sur les deux grands thèmes abordés dans le questionnaire :

- caractéristiques démographiques générales de la population de Quito,
- analyse des systèmes de résidence et pratiques professionnelles des quiténiens.

L'échantillon sélectionné comprend 432 îlots répartis sur l'ensemble de la ville de Quito, y compris dans ses extensions les plus récentes puisque la limite de l'agglomération retenue pour le sondage correspond non pas à des critères administratifs mais à un critère **physique** (continuité de l'espace bâti) issu de l'image satellite.

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 ; 3157 ménages ont été enquêtés.

Cinq des enquêteurs ayant participé aux opérations de collecte ont eu la charge de la codification des données ; ils ont achevé leur tâche en trois mois de travail à temps partiel, de février à avril 1988.

La saisie sur un micro-ordinateur IBM AT 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 3157 logements de l'échantillon. Les corrections sont effectuées à l'aide du logiciel SPSS PC+, principalement le module Data Entry. 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.

¹ L'équipe ORSTOM comprend F. DUREAU (géographe-démographe, responsable du programme de recherche), O. BARBARY (statisticien), A. MICHEL (urbaniste spécialisé en télédétection), B. LORTIC (télédétection). 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).

Puisque nous réalisons à Quito la première application en vraie grandeur de la méthode de sondage à partir d'image satellite, nous avons tenu à utiliser tous les documents disponibles pour valider les différentes étapes du tirage de l'échantillon et les résultats démographiques finaux.

En ce qui concerne le plan de sondage, nous avons procédé aux vérifications suivantes :

- détermination de la limite urbaine sur image SPOT : comparaison avec la limite déterminée par analyse visuelle de la couverture photographique de novembre 1987 à l'échelle du 1/ 40 000 ;
- classification en 6 niveaux de densité du bâti sur image SPOT : comparaison avec la densité observée sur le terrain en novembre 1986 ;
- mesure de surface des 432 îlots : comparaison, sur un échantillon d'îlots, avec les mesures obtenues par planimétrage sur les plans au 1/ 2 000 de l'Institut Géographique Militaire.

Ces vérifications nous conduisent à conclure à la **fiabilité** et à la **validité des techniques adoptées** : elles ont permis une mise en oeuvre correcte du plan de sondage.

Les coefficients de variation des estimateurs, dont le tableau ci-après donne quelques exemples, fournissent une première mesure de la précision des estimations que permet le plan de sondage adopté pour cette enquête. Enfin, les premiers résultats montrent une forte cohérence avec ceux du recensement réalisé en 1982 par l'INEC (Instituto Nacional de Estadística y Censos).

Bien entendu, la précision de la méthode doit être **comparée à celles obtenues par des méthodes de sondage plus traditionnelles**.

Par rapport au sondage équiprobable sur liste, méthode la plus répandue, les comparaisons que nous avons effectuées permettent de conclure que le sondage que nous proposons possède plusieurs avantages :

- sa précision est au moins équivalente à celle du sondage aléatoire simple; elle est meilleure dès que le caractère estimé est bien corrélé à la morphologie ;
- la méthode ne suppose pas l'existence d'une liste d'unités statistiques complète et à jour, de plus elle est directement applicable 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 oeuvre d'un sondage équiprobable "traditionnel" est variable et peut être prohibitif dans bien des villes, le coût lié à l'application du sondage aréolaire sur image satellite possède en revanche une borne supérieure fixe, quelle que soit la taille de la ville où l'on souhaite réaliser une enquête.

Précision de quelques estimations démographiques à partir de l'enquête Migrations (3157 ménages enquêtés, soit 1, 3 % des ménages de Quito)

Indicateur démographique	Estimateur (ensemble de la base)	Estimation du coefficient de variation de l'estimateur (%)	Intervalle de confiance à 95 %
% ménages de 7 personnes et plus	9,84	6,80	+/- 1,34
% plus de 60 ans	5,78	5,40	+/- 0,62
% ménages propriétaires	46,01	2,49	+/- 2,29
Sex ratio	0,957	1,50	+/- 0,029
% nés à Quito	63,17	1,38	+/- 1,74
Taille moyenne du ménage	4,54	1,10	+/- 0,10
Age moyen du chef de ménage	43,96	1,02	+/- 0,90

C. ELEMENTS DE BIBLIOGRAPHIE FONDAMENTALE

1. Statistique et théorie des sondages (ouvrages généraux)

- 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. WILEY, New York.
- GROSBRAS J.M., 1987. - Méthodes statistiques des sondages, ECONOMICA, Paris.
- LEBART L., MORINEAU A., FENELON J.P., 1982. - Traitement des données statistiques, méthodes et programmes. DUNOD, Paris.
- YATES F., 1981. - Sampling methods for censuses and surveys (4^e édition). GRIFFIN, Londres.

2. Collecte des données démographiques

- INED, INSEE, ORSTOM, SEAE, 1973. - Source et analyse des données démographiques. Application à l'Afrique d'expression française et à Madagascar. Première partie : source des données. Paris.
- LOHLE-TART L., CLAIRIN R., FRANCOIS M., GENDREAU F., 1988. - De l'homme au chiffre. Réflexion sur l'observation démographique en Afrique. CEPED - IUESP - IFORD, Paris.
- NATIONS UNIES, 1971. - Methodology of demographic sample surveys. New York.
- TABUTIN D., 1984. - La collecte des données en démographie. ORDINA EDITIONS, Liège.

3. Télédétection spatiale

- American Society of Photogrammetry, 1986. - Manual of remote sensing. SHERIDAN PRESS. Fall Church, 2 vol.
- CNES, SPOT Image, 1986. - Guide des utilisateurs de données SPOT. 2 volumes: 1: Manuel de référence.
2: Guide pratique.



D. BIBLIOGRAPHIE SPECIFIQUE SUR LA METHODE

Tous les textes cités ci-après peuvent être consultés:

- à la mission ORSTOM de Quito : 220 calle Rusia, Quito, EQUATEUR - Tel: 242 820
- au LIA de l'ORSTOM : 70-74 route d'Aulnay, 93140 Bondy, FRANCE - Tel: 48 47 31 95
- au CEDID: ORSTOM, 213 rue Lafayette, 75480 Paris Cedex 10, FRANCE- Tel: 48 03 77 77

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 ATP CNRS/CNES - Août 1987. Paris, 53 p. + 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 CNES - Avril 1988. Paris, 22 p. + 195 p. ann.

BARBARY O., 1988. - 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. Thèse de doctorat, Ecole des Hautes Etudes en Sciences Sociales, 641 p.

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. et GUILLAUME A., 1985. - La population dans l'espace. Télédétection et observation démographique des villes des pays en développement. Communication au XX^e congrès général de l'Union Internationale pour l'Etude Scientifique de la Population, Florence (Italie), 5-12 Juin 1985, 19 p.

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, Traitement des données localisées : l'infographie à l'Orstom, pp.263-286.

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, 6 p.

DUREAU F., BARBARY O., LORTIC B., MICHEL A., 1989. - 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 Dehli (Inde), 20-27 Septembre 1989, 17 p.

DUREAU F., BARBARY O., LORTIC B., MICHEL A., 1989. - L'utilisation de la télédétection pour l'observation démographique en milieu urbain. Convention Ministère de la Coopération - ORSTOM, Rapport final, mai 1989. Paris, 27 p. + ann.

MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F., 1987. - El uso de la teledetección para la observación de las poblaciones urbanas. Investigación en curso en Quito (ECUADOR). Communication au II Simposio Latinoamericano sobre sensores remotos, Colombia, Bogota, 16-20 Novembre 1987, 34 p., (2 versions: français et espagnol).

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, 9 p.

MICHEL A., 1988. - 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 THEMATIC MAPPER en milieu urbain. Thèse de Doctorat, Ecole des Hautes Etudes en Sciences Sociales, 2 vol., 237 p. + 162 p. ann.

MICHEL A., LORTIC B., DUREAU F., BARBARY O., 1988. Une nouvelle méthode de collecte des données démographiques à partir d'images SPOT. Paris, CNES - SPOT IMAGE, SPOT instrument de gestion et de décision, p. 26.

ORSTOM Éditeur
Dépôt légal : 4e trim. 1989

Conception et réalisation Atelier Bordelais de Cartographie.
7, quai de la Grave 33800 BORDEAUX



LA METHODE DE SONDAGE : DESCRIPTION GENERALE, OPTIONS, COUTS

La méthode de sondage développée (fiche 5) consiste en un **sondage aréolaire à 2 degrés** :

- **au 1er degré** : les aires sélectionnées (unités primaires) sont des **îlots**, l'îlot étant défini comme la plus petite unité de l'espace urbain entièrement délimitée par des voies ou des éléments naturels (rivière, ravin, ...). La base de sondage correspond à l'ensemble des îlots inclus dans la zone urbanisée ; une bonne répartition géographique des îlots de l'échantillon est assurée par un tirage systématique à l'aide d'une grille de points. En utilisant une information sur la morphologie des quartiers pour stratifier la base de sondage, on améliore la précision des estimations.

- **au 2ème degré** : les unités enquêtées (unités secondaires) sont des **ménages**. La sélection des ménages à enquêter se fait de façon systématique sur une liste des ménages établie pour chaque îlot de l'échantillon.

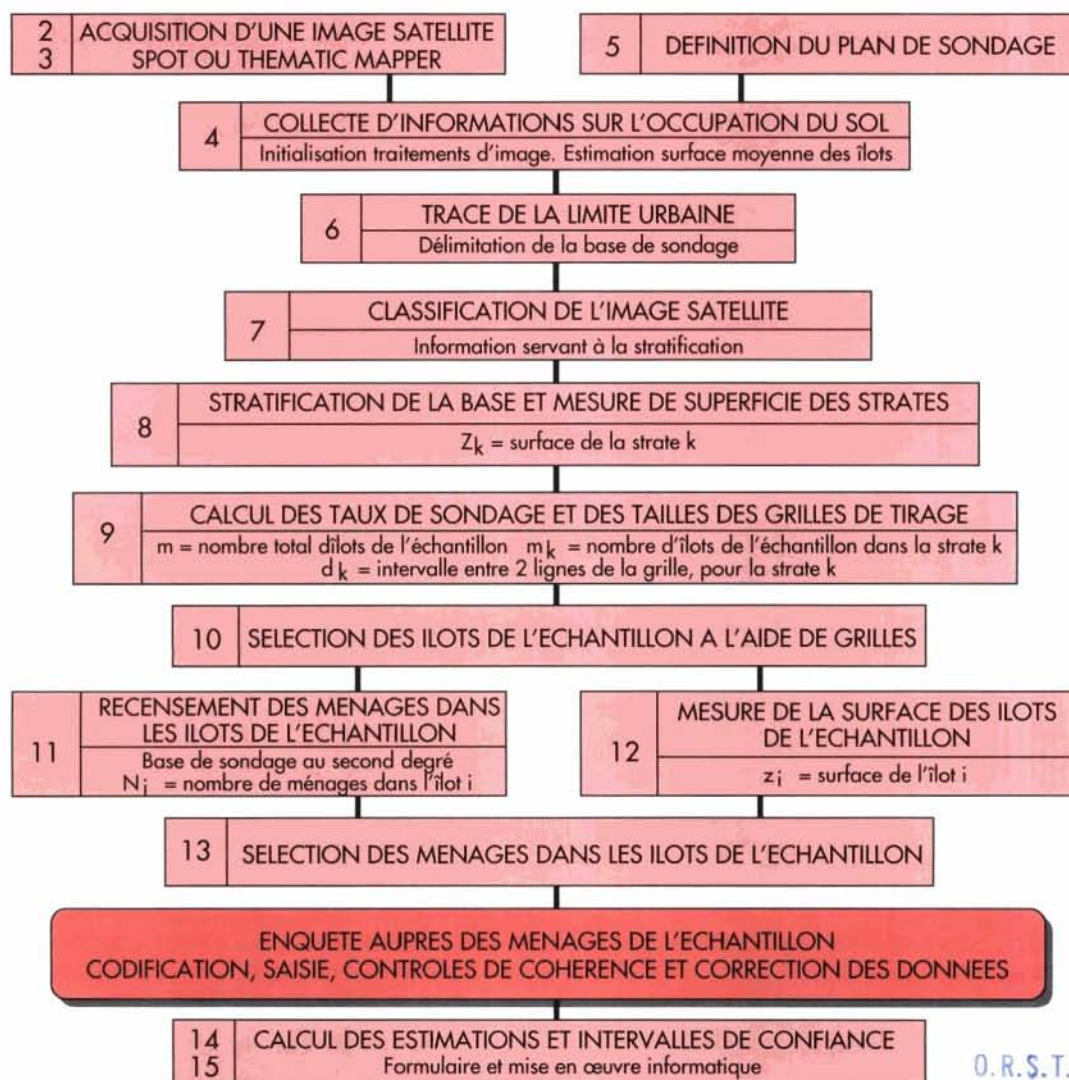
Du fait du caractère marginal de la variance intra-îlots au regard de la variance inter-îlots, il y a intérêt à pratiquer le plus fort taux possible au premier degré, en n'enquêtant qu'une faible fraction des ménages au sein de chaque îlot.

L'**image satellite** sert :

- à la **définition de la base de sondage** : tracé de la limite urbaine ;

- à la **stratification de la base de sondage** : une technique efficace (réduction d'environ 40 % de l'erreur d'échantillonnage) et adaptée au savoir-faire actuel en télédétection urbaine est la stratification sur le critère de la densité du bâti.

L'utilisation d'images satellite permet un échantillonnage rigoureux et efficace pour réaliser une enquête démographique dans une ville qui ne disposerait d'aucune base de sondage classique : dans ce cas, les étapes à suivre sont celles indiquées dans le graphique ci-dessous. Cette démarche ne correspond qu'à une des applications possibles de la méthode (il s'agit là de son application intégrale) ; en effet, le degré d'utilisation des images satellite varie selon le contexte de données disponibles dans la ville à enquêter. Le tableau suivant suggère les cheminements adaptés à d'autres contextes de données.



Les nombres renvoient aux numéros des fiches correspondantes. N° :
Toutes les notations mathématiques utilisées dans le manuel sont répertoriées dans la fiche 14.

GUIDE D'UTILISATION DE LA MÉTHODE

DONNÉES existantes		FPU : Fond de Plan actualisé des Unités primaires (îlots) CTM : Carte Thématique Morphologique actualisée (sans identification des unités primaires, c'est-à-dire des îlots) MPA : Mosaïque de Photographies Aériennes récentes (ou FPU + CTM) IVI : Information sur les Variances Intra-strates des caractères démographiques à estimer							
L I M I T E U R B A I N E	Coût	NÉANT	IVI	F P U	CTM	MPA (ou FPU + CTM)	IVI + FPU	IVI + CTM	IVI + MPA ou IVI + FPU + CTM
	M i n i m u m	Tracé sur tirage du canal panchromatique SPOT à grande échelle (1 / 15 à 30.000)	Tracé sur tirage du canal panchromatique SPOT à grande échelle (1 / 15 à 30.000)	Tracé sur fond de plan à grande échelle (1 / 15 à 30.000)	Tracé sur tirage du canal panchromatique SPOT à grande échelle (1 / 15 à 30.000)	Tracé sur fond de plan ou mosaïque photo à grande échelle (1 / 15 à 30.000)	Tracé sur fond de plan à grande échelle (1 / 15 à 30.000)	Tracé sur tirage du canal panchromatique SPOT à grande échelle (1 / 15 à 30.000)	Tracé sur fond de plan ou mosaïque photo à grande échelle (1 / 15 à 30.000)
	M a x i m u m	Collecte informations sur occupation du sol. Tracé sur tirage du canal panchromatique SPOT d'après image satellite SPOT (XS + Pa.) classée	Collecte informations sur occupation du sol. Tracé sur tirage du canal panchromatique SPOT d'après image satellite SPOT (XS + Pa.) classée	Collecte informations sur occupation du sol. Tracé sur fond de plan d'après image satellite (XS + Pa. de SPOT, ou canaux 3 et 4 TM) classée	Tracé sur tirage du canal panchromatique SPOT d'après image satellite SPOT classée (XS + Pa.)	Tracé sur fond de plan ou mosaïque photo d'après image satellite (XS + Pa. de SPOT, ou TM) classée	Collecte informations sur occupation du sol. Tracé sur fond de plan d'après image satellite (XS + Pa. de SPOT, ou canaux 3 et 4 TM) classée	Tracé sur tirage du canal panchromatique SPOT d'après image satellite SPOT classée	Tracé sur fond de plan ou mosaïque photo d'après image satellite (XS + Pa. de SPOT, ou canaux 3 et 4 TM) classée
S T R A T I F I C A T I O N	M i n i m u m	Stratification d'après connaissance de la ville. Tracé et mesure des strates sur tirage du canal panchromatique SPOT	Stratification d'après connaissance de la ville. Tracé et mesure des strates sur tirage du canal panchromatique SPOT	Stratification d'après connaissance de la ville. Tracé et mesure des strates sur fond de plan à grande échelle	Stratification d'après carte thématique. Tracé et mesure des strates sur tirage du canal panchromatique SPOT	Stratification d'après carte thématique ou interprétation photo. aériennes. Tracé et mesure des strates sur fond de plan ou mosaïque photo.	Stratification d'après connaissance de la ville. Tracé et mesure des strates sur fond de plan à grande échelle	Stratification d'après carte thématique. Tracé et mesure des strates sur tirage du canal panchromatique SPOT	Stratification d'après carte thématique ou interprétation photo. aériennes. Tracé et mesure des strates sur fond de plan ou mosaïque photo.
	M a x i m u m	Collecte informations sur occupation du sol. Stratification d'après image satellite SPOT (XS + Pa.) classée. Mesure des strates sur l'image SPOT Pa.	Collecte informations sur occupation du sol. Stratification d'après image satellite SPOT (XS + Pa.) classée. Mesure des strates sur l'image SPOT Pa.	Collecte informations sur occupation du sol. Stratification d'après image satellite (XS + Pa. SPOT, ou 3,4,5 TM) classée. Report et mesure des strates sur le fond de plan	Stratification d'après image satellite SPOT (XS + Pa.) classée et autres informations sur la morphologie. Report et mesure des strates sur l'image SPOT Pa.	Stratification d'après image satellite (XS + Pa. SPOT, ou TM) classée et autres informations morpho. Report et mesure des strates sur le fond de plan ou la mosaïque photo.	Collecte informations sur occupation du sol. Stratification d'après image satellite (XS + Pa. SPOT, ou 3,4,5 TM) classée. Report et mesure des strates sur le fond de plan	Stratification d'après image satellite SPOT (XS + Pa.) classée et autres informations sur la morphologie. Report et mesure des strates sur l'image SPOT Pa.	Stratification d'après image satellite (XS + Pa. SPOT, ou 3,4,5 TM) classée et autres informations morpho. Report et mesure des strates sur le fond de plan ou la mosaïque photo.
	M i n i	Taux constant dans toutes les strates	Allocation optimale : taux variable selon les strates	Taux constant dans toutes les strates	Taux constant dans toutes les strates	Taux constant dans toutes les strates	Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates
A L L O C A T I O N A U X S T R A T E S	M a x i	Collecte informations démographiques. Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates	Collecte informations démographiques. Allocation optimale : taux variable selon les strates	Collecte informations démographiques. Allocation optimale : taux variable selon les strates	Collecte informations démographiques. Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates	Allocation optimale : taux variable selon les strates
		Image à acquérir pour la solution préconisée	Bande SPOT (XS + Pa.) Film SPOT Pa.	Bande SPOT (XS + Pa.) ou, à défaut, TM	Film SPOT (Pa.)	Pas d'image satellite	Bande SPOT (XS + Pa.) ou, à défaut, TM	Film SPOT (Pa.)	Pas d'image satellite

Le tableau ci-dessus est destiné à guider l'utilisateur dans l'application de la méthode de sondage: il décrit les méthodes adaptées au contexte de données disponibles sur la ville où l'on désire réaliser une enquête. Huit contextes d'information sont individualisés, chacun correspondant à une colonne du tableau.

Pour chacune des ces situations, on s'intéresse aux trois principaux éléments du plan de sondage (définition de la limite urbaine, stratification de la base de sondage, type d'allocation aux strates), en indiquant pour chacun

d'eux quelles sont les solutions minima et maxima; la solution que nous conseillons de choisir est en rouge pâle. Pour chacune des solutions, les numéros des fiches du manuel qu'il faut consulter sont en rouge foncé.

Enfin, au bas de chacune des colonnes du tableau, c'est-à-dire pour chaque contexte de données, figure le type d'image qu'il faut acquérir pour mettre en oeuvre les solutions conseillées.

L'équipement nécessaire à la mise en oeuvre de la méthode

L'application de la méthode réclame l'équipement **minimum** suivant :

- IBM AT (ou compatible), disque dur 20 Mo, imprimante :	15 000 FF ¹
- Logiciel de saisie et traitements statistiques des données d'enquête sur IBM AT. Exemple: SPSS/PC+ (module de base et Data Entry)	12 300 FF
- Logiciel TIMOR de traitement d'image satellite sur IBM AT développé par l'ORSTOM	600 FF
- Boussole, clinomètre et décamètre	2 500 FF
- Planimètre manuel	1 500 FF

Cette configuration minimale peut être améliorée en recourant à des solutions plus coûteuses, mais aussi plus performantes :

- Mesurer les superficies des strates et des îlots avec un planimètre électronique (6000 FF)
- Calculer les surfaces d'îlots à partir des mesures de terrain avec un logiciel de dessin (5 000 FF)
- Réaliser les traitements d'image satellite sur station SUN (3/110C, disque dur 140 Mo, 200 000 FF) avec le logiciel PLANETES développé par l'ORSTOM
- Réaliser l'exploitation de l'enquête démographique avec un logiciel statistique spécialisé dans l'exploitation des enquêtes par sondage.

Les coûts de mise en oeuvre de la méthode

Nous nous plaçons dans le cas de figure suivant : il n'existe ni matériel informatique, ni logiciel, ni image satellite, fond de plan ou carte thématique. L'allocation aux strates est réalisée avec un taux constant dans toutes les strates. Sont inclus tous les salaires du personnel réalisant la collecte, la codification et la saisie de l'information, à l'exception des 6 mois de salaire de la personne qualifiée pour la sélection de l'échantillon et l'encadrement de la collecte, de la codification et de la saisie de l'information.

Postes fixes de dépense (équipement minimum)		Postes de dépense variant selon la taille de l'échantillon et le questionnaire
Equipement informatique	15 000 FF	Mesure surface des îlots : 9,36 FF/îlot Recensement des ménages, collecte démographique sur l'échantillon de ménages, codification et saisie de l'information : - questionnaire court : 11,81 FF/ménage - questionnaire long : 16,65 FF/ménage
Logiciels	12 900 FF	
Equipement de mesure	4 000 FF	
Achat image satellite SPOT (Bandes XS et Pa, film Pa)	36 000 FF	
Collecte d'informations sur l'occupation du sol	6 000 FF	
TOTAL	73 900 FF	

Si l'on considère que l'on enquête, en moyenne, 7,5 ménages par îlot (observation faite à Quito), et m étant le nombre d'îlots sélectionnés, les fonctions de coûts sont donc :

- pour un questionnaire court : $\text{coût} = 73\,900 + 9,36\,m + (7,5 \times 11,81)m = 73\,900 + 97,94\,m$
- pour un questionnaire long : $\text{coût} = 73\,900 + 9,36\,m + (7,5 \times 16,65)m = 73\,900 + 134,24\,m$

Pour l'estimation d'un total sur l'ensemble de la zone d'étude, la fonction de précision caractéristique de la méthode (intervalle de confiance à 95 %) est évaluée par l'expression suivante : $278 / \sqrt{m}$.

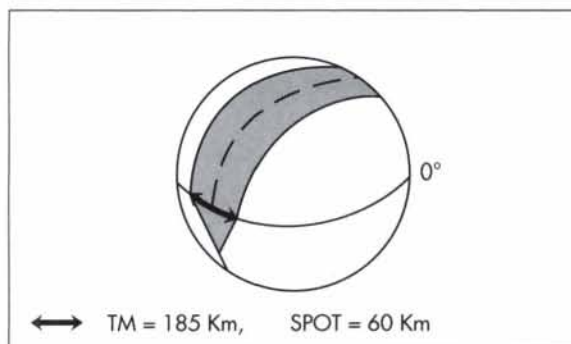
Ces fonctions de coûts et de précision aboutissent donc aux évaluations suivantes :

Nombre d'îlots enquêtés (m)	Nombre de ménages enquêtés	Coût (FF) questionnaire court	Coût (FF) questionnaire long	Précision: intervalle de confiance à 95 % (en % de la pop totale)
100	750	83 694	87 324	+/- 27,8
500	3 750	122 870	141 020	+/- 12,4
1000	7 500	171 840	208 140	+/- 8,8
2000	15 000	269 780	342 380	+/- 6,2

Les frais fixes constituent une part non négligeable du coût d'enquête. De ce fait, si l'on est obligé d'effectuer l'ensemble des dépenses d'achat de données satellitaires, il est peu rentable d'appliquer cette méthode à des enquêtes trop légères (moins de 150 îlots environ). Pour la même raison, lorsqu'un échantillon compte moins de 1000 îlots, le passage d'un questionnaire court à un questionnaire long ne fait pas croître très sensiblement le coût global de l'enquête.

INTRODUCTION A LA TELEDETECTION SPATIALE

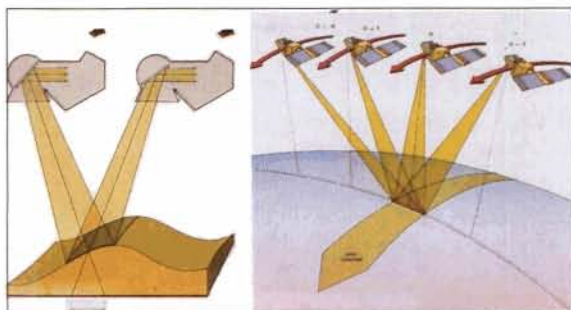
La **télédétection spatiale** peut être définie comme l'ensemble des techniques d'observation de la terre à partir de satellites, qui utilisent les propriétés du **rayonnement électromagnétique**. Dans cette fiche, nous nous limiterons à la présentation des satellites **SPOT** et **LANDSAT Thematic Mapper** puisqu'ils sont à ce jour les plus performants pour l'observation du milieu urbain.



Tournant autour de la terre à une distance de plus de 800 Km, LANDSAT TM comme SPOT, sont deux satellites d'observation de la terre à défilement héliosynchrones (c'est-à-dire qu'une région quelconque du globe est toujours survolée à la même heure solaire locale).

LANDSAT TM observe une bande de **185 Km** de large et couvre l'ensemble du globe en **18 jours**.

SPOT repasse à la verticale du même point tous les **26 jours** ; ce même point peut également être observé par SPOT en visée oblique, avec une fréquence moyenne de 3 jours. Le satellite enregistre sur une bande de **60 Km** (en visée verticale) à **80 Km** (en visée de 27°/verticale).

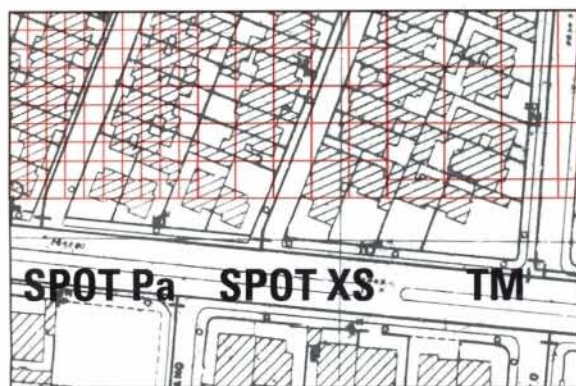


Grâce aux possibilités de visée latérale de SPOT (+ ou -27° par rapport à la verticale), il est possible d'observer des zones situées dans une bande de 950 Km de large à la verticale de la trace. Deux propriétés en découlent :

- pendant une période de 26 jours, la possibilité d'observer une même zone 7 fois (à l'équateur) ou 11 fois (à une latitude de 45°),
- l'obtention de couples stéréoscopiques dont le traitement permet l'appréhension du relief.

Les capteurs embarqués sur les satellites mesurent une **grandeur physique**, appelée **luminance spectrale**, qui après avoir été codée entre 0 et 255, est transmise au sol. Les valeurs reçues par les stations de réception sont stockées sur support magnétique sous forme d'image matricielle ; elles peuvent être aisément reproduites sur film photographique.

La mesure de luminance est globale au sein d'une surface au sol de taille fixe : à cette unité de surface correspond une unité élémentaire de l'image appelée **pixel**.



La **résolution spatiale** est définie par la taille de la surface élémentaire au sol (pixel) :

SPOT :

panchromatique : 10 x 10 m soit 100 m²

multispectral : 20 x 20 m soit 400 m²

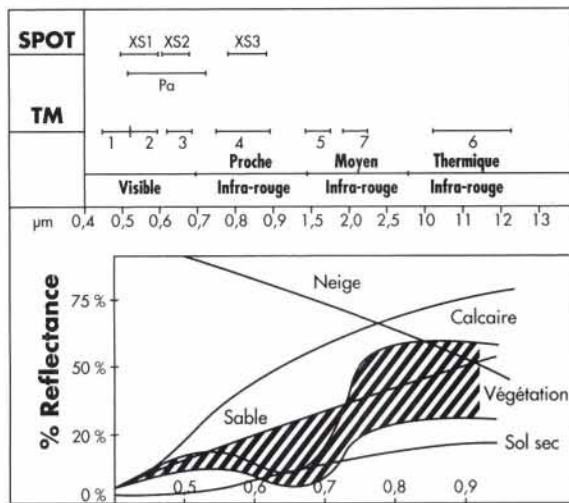
Thematic Mapper :

- infra-rouge thermique : 120 x 120 m soit 1440 m²

- autres canaux : 30 x 30 m soit 900 m²

La **luminance** est mesurée à l'intérieur de plusieurs **bandes du spectre visible ou infra-rouge**, dont les bornes varient selon le capteur. Ces bornes sont soigneusement choisies pour **minimiser l'absorption due à l'atmosphère terrestre**, et **éviter la redondance de l'information**.

La **résolution spectrale** d'un capteur est définie par le nombre et les caractéristiques des fenêtres à l'intérieur desquelles les mesures sont effectuées ; pour chaque pixel, il y a donc autant de mesures que de bandes. O.R.S.T.O.M. Fonds Documentaire



SPOT

Deux modes de fonctionnement :

- mode panchromatique : 1 bande
- mode multispectral : 3 bandes (XS1, XS2, XS3)

LANDSAT TM

- mode multispectral : 7 bandes (TM1 à TM7).

L'interprétation d'une image satellite est actuellement fondée sur l'hypothèse qu'il est possible de reconnaître un certain nombre d'éléments présents à la surface de la terre à partir des valeurs de luminance de ces objets dans les différentes fenêtres du spectre électromagnétique, c'est-à-dire leur **signature spectrale**.

Pré-traitements standards

Les images reçues par les stations de réception ne sont pas directement utilisables ; avant d'être diffusées, elles doivent faire l'objet de corrections géométriques et radiométriques.

Les corrections **géométriques** sont destinées à corriger les effets du mouvement secondaire du satellite dans les trois dimensions (lacet, roulis, tangage)

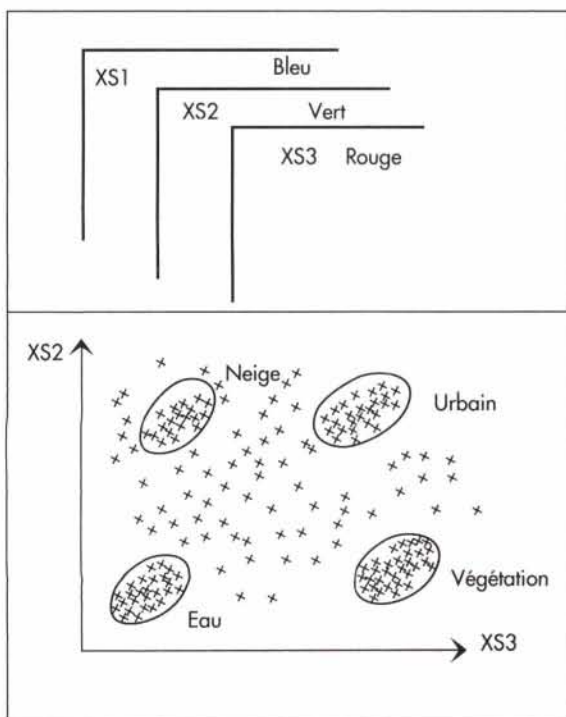
Les corrections **radiométriques** sont fondées sur le réétalonnage du radiomètre, qui permet une égalisation relative des mesures de ses détecteurs.

Les images SPOT (niveau de correction géométrique 1B) et TM, nécessaires à l'application de la méthode décrite dans ce manuel, font systématiquement l'objet des corrections citées ci-dessus. A ce niveau de correction géométrique, en cas de relief important, les images ne sont pas superposables à une carte ; ce résultat ne peut être obtenu que par un traitement particulier nécessitant une information sur le relief de la zone (carte topographique, modèle numérique de terrain).

Exploitation thématique de l'image

Pour extraire de l'information thématique à partir d'une image, deux types de traitements peuvent être mis en oeuvre :

- traitements photographiques, à partir des films. Exemples : amélioration de contraste, composition colorée ;
- traitements numériques à partir des bandes magnétiques. Exemples : calcul d'indices, classifications.



Compositions colorées .

Une couleur primaire est affectée à chacun des trois canaux (par exemple pour SPOT, XS1 : bleu, XS2 : vert, XS3 : rouge) ; puis, les trois canaux sont superposés afin d'obtenir une représentation en couleur qui allie ainsi les propriétés discriminantes de chacun d'eux. Pour SPOT, un tel traitement fait apparaître la végétation en rouge, couleur affectée au canal infra-rouge, dans lequel la végétation possède une valeur de luminance forte.

Classification automatique

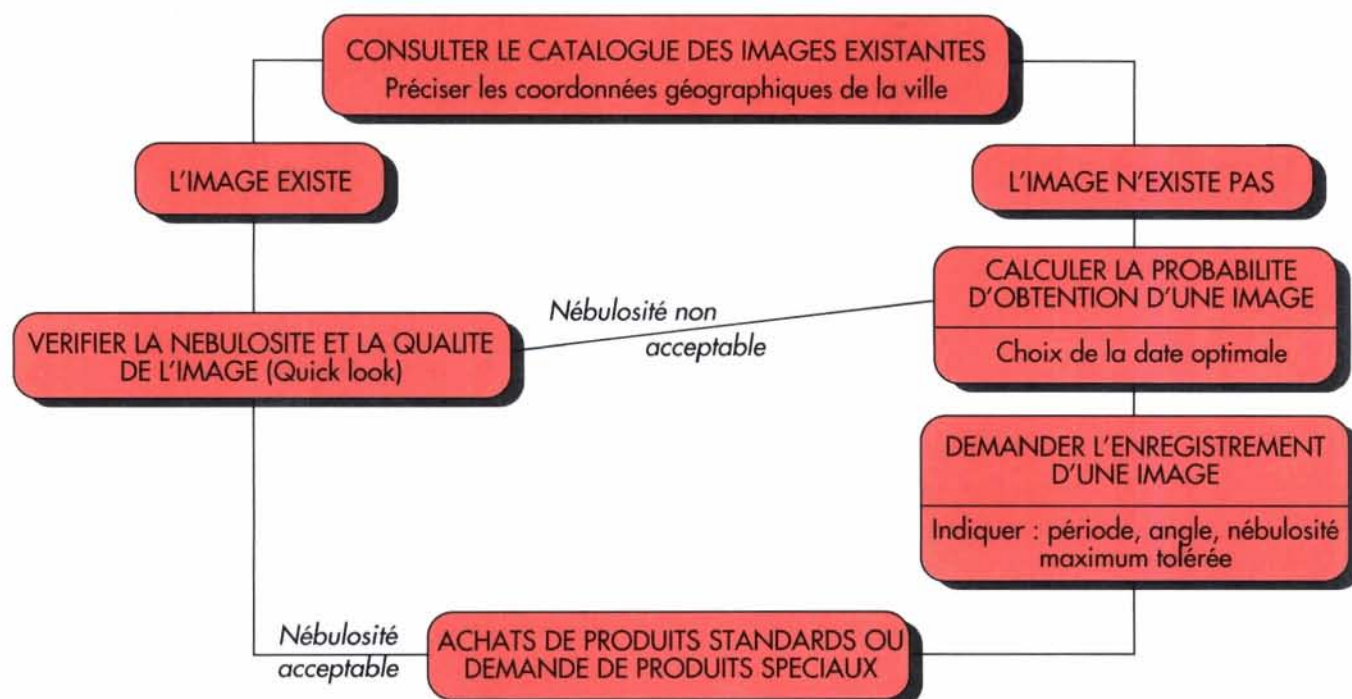
L'étude statistique des signatures spectrales permet la discrimination des grands types d'occupation du sol en faisant appel à des techniques classiques de classification.

Par exemple, l'utilisation conjointe des canaux XS2 et XS3 conduit à une première différenciation en 4 classes d'occupation du sol.

ACQUISITION D'UNE IMAGE SATELLITE SPOT OU THEMATIC MAPPER

Quelle est la démarche à suivre?

Les satellites SPOT et LANDSAT TM n'enregistrent pas de façon continue : l'enregistrement est programmé par les sociétés exploitant les satellites en fonction des commandes et de leurs propres impératifs. Il existe donc, pour SPOT comme pour TM, un catalogue des images enregistrées, avec leurs caractéristiques (date d'enregistrement, pourcentage de nuages sur l'image, qualité de l'image, angle de visée pour SPOT).



A qui s'adresser pour connaître les images existantes, ou commander une image?

Images SPOT: distributeur local, ou
 SPOT IMAGE - Service clientèle
 16 bis avenue Edouard Belin - BP 4359
 31030 - Toulouse Cedex - FRANCE
 Te: (33) 61 53 99 76
 Telex: 532 079 F SPOTIM
 Telefax: (33) 61 27 46 05

Images TM: distributeur local ou
 EOSAT
 4300 Forbes Blvd - Lanham
 Maryland 20706 - USA
 Te: 1800 367 2801
 Telex: 277 685 LSAT UR
 Telefax: 605 594 6589

SPOT ou TM?

La réponse à cette question est fonction du contexte local d'informations : la fiche 1 précise, pour quelques situations types, quel doit être le type d'image à acquérir. L'élément décisif est la résolution différente de TM (30 mètres) et de SPOT (10 et 20 mètres) (fiche 2).

La stratification de l'espace urbain sur un critère comme la densité du bâti est possible aussi bien à partir de TM que de SPOT (le traitement décrit fiche 7 est toutefois plus facile à réaliser avec SPOT) ; par contre, seule l'image **SPOT** permet de repérer le réseau de voirie, et donc de **localiser et délimiter les îlots** de l'échantillon d'enquête. S'il n'existe pas de cartographie à jour, la seule solution est donc d'utiliser une image SPOT pour sélectionner un échantillon d'îlots ; s'il existe une cartographie, on a le choix entre acquérir une image SPOT, ou à défaut une image TM, qui sert alors **uniquement pour délimiter et stratifier la base de sondage**.

Quelle est la probabilité d'obtention d'une image sur la ville à enquêter?

Pour statuer sur le degré de faisabilité d'un plan de sondage aréolaire sur image satellite dans une ville donnée, et définir les mois de l'année les plus favorables pour obtenir une image, il est nécessaire de calculer la probabilité d'obtention d'une image satellite SPOT ou TM sur cette ville.

Cette probabilité dépend : de la nébulosité, du rythme de passage du satellite, de l'angle de visée souhaité (pour SPOT). Connaissant, pour chaque mois de l'année, le nombre de jours ayant, à l'heure de passage du satellite, une nébulosité maxima donnée, on calcule la probabilité d'obtenir une image pour un mois donné, une nébulosité maximum donnée et un angle de visée maximum fixé en appliquant la formule suivante :

$$1 - (1 - f)^P$$

f = Fréquence relative moyenne (sur la période de référence des statistiques disponibles, 10 ans par exemple) des jours du mois vérifiant la condition de nébulosité fixée.

P = Nombre moyen de passages du satellite pendant un mois

Valeurs de P à l'équateur :

TM : P= 2;

SPOT : P=1,25 pour une image verticale, P=3,58 pour un angle inférieur à 10°, P= 8,17 pour un angle quelconque.

Exemple de calcul :

En janvier, durant les dernières années, on a observé les statistiques de nébulosité suivantes, à 9h30, heure de passage du satellite TM au dessus d'une ville située à l'équateur : 9% des jours sans nuage et 48 % des jours avec une nébulosité < 50 % .

La probabilité d'obtenir une image TM sans nuage est de : $1 - (1 - 0,09)^2 = 1 - 0,91^2 = 17 \%$

La probabilité d'obtenir une image TM avec une nébulosité < 50 % est de : $1 - (1 - 0,48)^2 = 1 - 0,52^2 = 73 \%$

Quel est le prix des images, en 1988 ?

Type de produit	SPOT (niveau 1B) scène : 60 Km x 60 Km			THEMATIC MAPPER scène : 185 Km x 185 Km		
FILM NEGATIF						
Scène entière	11 000 FF	panchro	1/400.000	976 FF	par canal	1/1.000.000
	8 000 FF	3 can. XS	1/400.000			
1/4 scène (au choix)	14 000 FF	panchro	1/200.000	Pas de film pour 1/4 de scène		
	11 000 FF	3 can. XS	1/200.000			
BANDE MAGNETIQUE						
Scène entière	14 000 FF	panchro		20 130 FF	7 canaux	
	11 000 FF	3 canaux XS				
1/4 scène (au choix)	Pas de bande 1/4 de scène			10 065 FF	7 canaux	
Demande programmation	3 000 FF (en sus prix scène)			9 760 FF (en sus prix scène)		
Contrainte de nébulosité				1 678 FF (en sus prix scène)		
Redécoupage le long de la trace	3 000 FF (en sus prix scène)			Pas de possibilité de redécoupage		

Pour de nombreuses villes, un quart de scène (soit 30 x 30 Km pour SPOT, et 90 x 90 Km pour TM) suffit à couvrir l'ensemble de l'agglomération.

Depuis la fin de l'année 1988, il est possible d'acquérir, pour une somme de 1 500 FF, une disquette correspondant à un extrait d'image SPOT de 10 x 10 Km, à condition que l'image existe déjà sur bande magnétique.

COLLECTE D'INFORMATIONS SUR L'OCCUPATION DU SOL

La collecte d'informations sur l'occupation du sol dans un certain nombre d'îlots de la ville est indispensable pour interpréter l'image satellite devant servir de base de sondage à l'enquête démographique. La réalisation de cette opération est décrite dans le tableau suivant :

Date de la collecte sur le terrain	Concomitance maximale avec la date d'enregistrement de l'image satellite
Nombre d'îlots	Environ 100 îlots pour une ville d'une superficie de 15 à 20 000 hectares.
Sélection des îlots	Tirage systématique par grille superposée à l'image satellite, ou à un plan de la ville (si ancien, ajouter des îlots localisés dans les extensions récentes non cartographiées) Introduction éventuelle d'une stratification basée sur des informations exogènes à l'image.
Données à collecter sur le terrain	Doivent correspondre aux critères de stratification à extraire de l'image satellite. Par exemple, pour une stratification sur la densité du bâti, on doit recueillir pour chaque îlot : - la surface occupée par les bâtiments - la surface totale de l'îlot.

Au dos de cette fiche 4, figure un exemple de fiche d'enquête à utiliser pour des relevés de terrain relatifs à la densité du bâti. Le principe de remplissage de la fiche est le suivant.

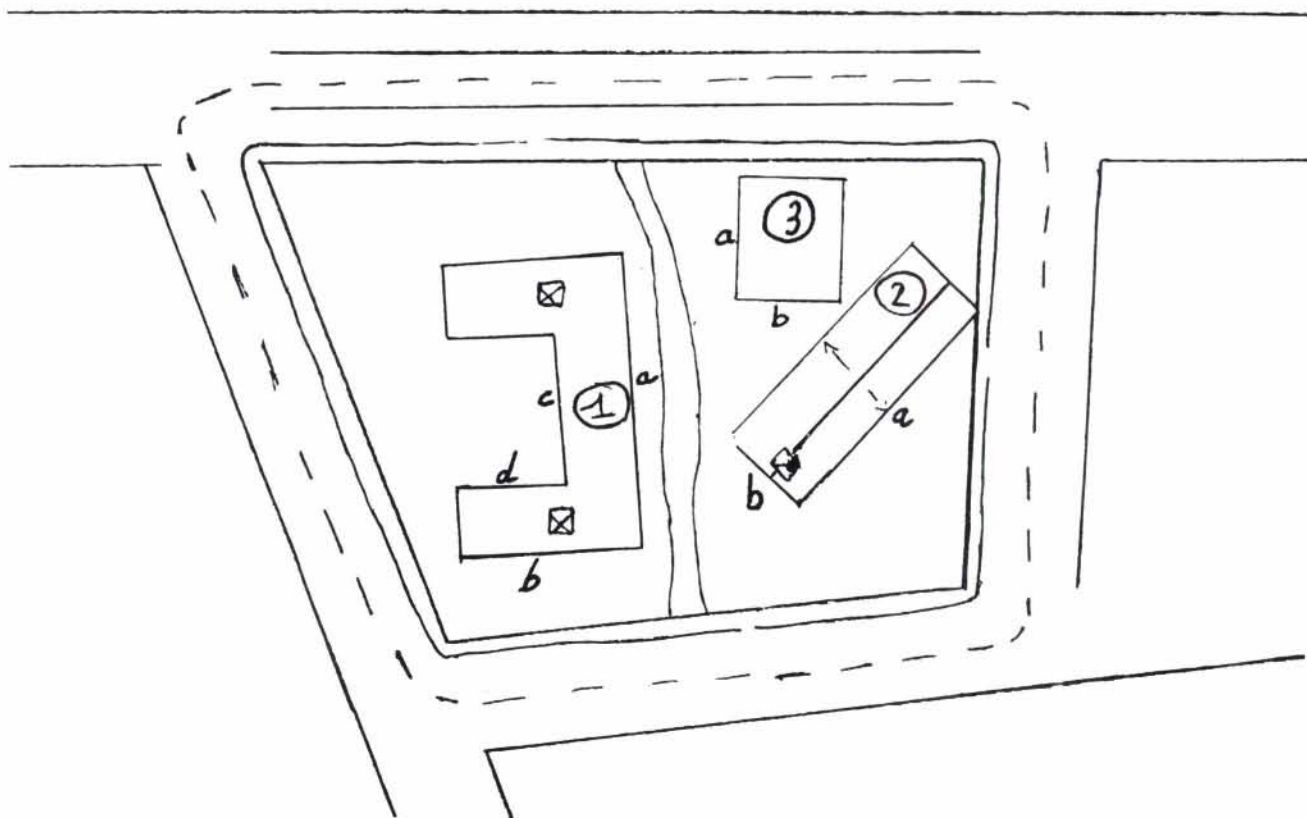
L'enquêteur recherche un point surplombant l'îlot (terrasse d'immeuble, relief) afin d'en avoir une vision d'ensemble ; il dresse un schéma des bâtiments de l'îlot et les numérote.

L'enquêteur se rend ensuite sur l'îlot lui-même et procède à la mesure des bâtiments et de l'îlot. Pour le calcul de la densité du bâti, la précision d'une mesure en nombre de pas est suffisante, à condition de connaître la taille moyenne du pas de l'enquêteur (on répètera et enregistrera les mesures au pas et au décimètre durant les deux premiers jours d'enquête).

La densité du bâti dans l'îlot est calculée au bureau : somme des surfaces des bâtiments / surface de l'îlot (l'élévation des bâtiments n'est pas prise en compte : c'est-à-dire que l'on ne cumule pas les surfaces correspondant aux différents étages d'un même bâtiment).

Numéro d'îlot : 48		DESCRIPTION DES BATIMENTS				
Date : 1/10/88 Heure : 15 ~						
Numéro du bâtiment correspondant au repère porté sur le plan		1	2	3		
D I M E N S I O N S	Batiment	☒	☒	☐	☐	☐
	Terminé	☐	☐	☒	☐	☐
	En construction					
	Dimensions au sol:					
	Côté a	15 m	12 m	8 m	---	---
	Côté b	5 m	6 m	6 m	---	---
	(Indiquer ces côtés sur le plan)	c = 7 m d = 3,5 m				

SCHEMA DES BATIMENTS DE L'ÎLOT



Durée de la collecte sur le terrain (mesure de l'îlot et des bâtiments) : 4 îlots/ enquêteur/jour

ELEMENTS POUR LA DEFINITION DU PLAN DE SONDAGE

Dans la fiche 1, la méthode de sondage est décrite dans ses grandes lignes. Ce type de sondage est, comme nous l'avons dit, un des résultats de notre recherche et a été testé lors de la première application de la méthode à Quito. C'est lui qui sert de base à ce manuel mais **sa définition n'est pas pour autant totalement figée ; l'utilisateur de la méthode peut être amené à y apporter certaines modifications**, en prenant en compte les spécificités du site et les informations qu'il possède.

Comme toujours, s'agissant de sondage, il importe de rechercher un **bon compromis entre les contraintes liées au site, aux données existantes et au budget disponible d'une part, et l'amélioration de la précision des estimations d'autre part**. Pour faciliter cette recherche, cette fiche examine les deux étapes de la définition du plan de sondage pour lesquelles des adaptations sont possibles sans qu'il soit nécessaire d'effectuer d'autres recherches méthodologiques.

1. Définitions de la base et de la structure du sondage

Ces définitions résultent principalement de l'utilisation d'une base de sondage spatiale (image satellite SPOT, photographies aériennes ou plan) et de l'objet de l'enquête. Lorsque l'on s'intéresse aux caractéristiques d'unités statistiques du type "logement", "ménage", "individu", etc..., le plus souvent il n'existe pas de liste complète et à jour pouvant servir de base à la sélection directe d'un échantillon d'unités statistiques. En revanche, divers documents sur lesquels on peut identifier sans ambiguïté les îlots de la ville peuvent constituer des bases de sondage pour le tirage d'un échantillon d'unités spatiales : image satellite, mosaïque aérienne récente ou plan détaillé actualisé du site urbain. La situation que nous venons de décrire débouche donc nécessairement sur un plan de sondage comprenant au moins deux degrés, c'est-à-dire que **la sélection de l'échantillon s'effectuera en deux temps : sélection des îlots, puis sélection des ménages dans les îlots**.

La définition des unités statistiques (logements, ménages, individus, etc...) dépend bien sûr de l'objet fixé à l'enquête. On peut avoir intérêt, pour des raisons pratiques ou pour améliorer la précision des estimations, à **introduire des degrés intermédiaires dans le plan de sondage** (par exemple, sélectionner dans chaque îlot un échantillon d'immeubles, puis un échantillon de ménages ou d'individus dans chaque immeuble sélectionné).

Enfin, pour définir complètement la structure du sondage, il reste à **fixer les taux aux différents degrés**, c'est-à-dire, dans le cas simple à deux degrés, le nombre d'îlots à sélectionner et le nombre d'unités statistiques à enquêter dans chaque îlot (fiche 9). C'est là un des principaux paramètres sur lequel on peut agir pour améliorer la précision d'un sondage à deux degrés, le nombre total d'unités statistiques enquêtées étant fixé. En effet, dans ce cas, l'erreur d'échantillonnage comprend deux termes : l'un est fonction de la variabilité du caractère à estimer **entre** les unités primaires (c'est à dire entre îlots distincts) et l'autre est fonction de la variabilité de ce même caractère **au sein** des unités primaires (c'est à dire entre les ménages appartenant à un même îlot).

L'un des résultats importants des recherches menées pour développer la méthode est que le premier terme de l'erreur (variance inter-unités primaires) détermine presque entièrement l'erreur d'échantillonnage. C'est donc ce terme qu'il faut minimiser : pour cela, **on cherchera à enquêter un maximum d'unités primaires (îlots) en réduisant en conséquence le nombre d'unités statistiques échantillon (ménages) enquêtées dans chacune d'elles**.

2. Stratification

Pour l'exposé du principe de la stratification et des modalités de sa réalisation en utilisant les différentes sources d'information disponibles, le lecteur se reportera à la fiche 8. Il faut seulement souligner ici que le gain de précision permis par cette opération dépend de **deux facteurs** :

- d'une part du niveau d'homogénéité que présentent les îlots appartenant à une même strate relativement aux variables à estimer et de son corollaire, l'hétérogénéité d'une strate à l'autre ;
- d'autre part de la répartition de l'échantillon entre les strates (ce qu'on appelle l'allocation aux strates).

Concernant la **définition proprement dite des strates**, les variables à estimer étant inconnues *a priori*, la règle générale est d'utiliser conjointement **toutes les sources d'information disponibles** concernant des variables que l'on suppose bien corrélées au thème de l'enquête, afin d'aboutir aux strates les plus homogènes possible (fiche 8).

Une fois définie la stratification, deux types d'allocation aux strates sont possibles :

- **l'allocation proportionnelle** (au nombre d'îlots) consiste à répartir l'échantillon en adoptant le même taux dans toutes les strates de la base de sondage. Cette solution, dont la mise en œuvre est très simple (fiche 9), sera retenue en l'absence d'informations chiffrées sur les variances des caractères à estimer au sein des strates. Elle permet un gain de précision appréciable si les critères de stratification sont pertinents ; et même si ce n'est pas le cas, cette solution ne risque pas de donner un moins bon résultat qu'un sondage non stratifié.

- **les allocations "optimisées"** reposent sur le principe qu'il faut augmenter le taux de sondage dans les strates "hétérogènes". Plus précisément, le taux de sondage doit être modulé selon la contribution de chaque strate à la variance de l'estimateur sur l'ensemble de la base. Etant donné le plan de sondage pratiqué, la probabilité d'inclusion des unités primaires est proportionnelle à leur surface. C'est la quantité S_{ky} suivante qui mesure la contribution de la strate k à l'erreur d'échantillonnage sur l'estimation du caractère y :

$$S_{ky} = \left(Z_k \sum_{i=1}^{M_k} \left(\frac{y_i^2}{z_i} \right) - Y_k^2 \right)^{1/2}$$

où M_k est le nombre d'unités primaires (îlots) total dans la strate k . Alors, si h est le nombre de strates dans l'univers et m le nombre d'unités primaires à sélectionner dans l'échantillon complet (toutes strates confondues), le nombre optimal d'unités primaires dans la strate k (m_k) est donné par la formule suivante :

$$m_k = m \cdot \frac{S_{ky}}{\sum_{k=1}^h S_{ky}}$$

La définition, à partir de cette formule, d'une allocation "optimisée" pose donc **deux problèmes** :

1. comment disposer, pour chaque strate, d'une information chiffrée concernant S_{ky} ?
2. l'allocation optimale fournie par la formule étant valable pour un caractère y donné, comment concilier plusieurs allocations différentes (chacune optimale pour une variable importante du questionnaire) pour définir une allocation "optimisée" qui améliorera la précision d'un ensemble d'estimations ?

Les réponses à ces questions dépendent bien entendu du **contexte de données** disponibles dans lequel on se trouve, et suivant ce contexte on pourra juger opportun ou non de tenter d'optimiser l'allocation (une allocation mal optimisée pouvant donner de moins bons résultats que l'allocation proportionnelle et même, en cas d'erreur grave sur les m_k , qu'un sondage non stratifié).

Deux observations peuvent aider l'utilisateur à prendre cette décision :

- L'information concernant les **S_{ky}** au sein des différentes strates peut provenir de sources existantes (enquêtes ou recensements récents concernant les mêmes thèmes) ou d'une opération préliminaire légère de collecte démographique. Il faut prendre garde que pour estimer avec une fiabilité acceptable la variance intra-strate de certains caractères démographiques liés à des événements rares (comme des taux de mortalité ou de fécondité), l'échantillon correspondant à l'enquête préliminaire devra être assez important.
- L'allocation ne peut être optimale à la fois pour toutes les variables du questionnaire ; cependant, on constatera très souvent, à cause des corrélations existant entre variables ayant trait à un même thème, qu'une tendance commune se dégage lorsque l'on compare les allocations optimales propres à une série de variables distinctes. La définition d'une allocation "optimisée" est alors possible.

TRACE DE LA LIMITE URBAINE

Cette étape est particulièrement importante, puisqu'elle détermine les limites de la base de sondage.

Rappelons que l'on ne peut extraire d'une image satellite ou d'une photographie aérienne qu'une limite traduisant un **phénomène physique**. Une limite d'ordre administratif ou politique ne peut en aucun cas être détectée sur ce type de document.

Il convient impérativement de préciser clairement le concept d'agglomération urbaine retenu pour délimiter la base de sondage. Une solution praticable, décrite dans cette fiche, est de définir la ville comme une **zone continue d'espace bâti**.

Quelle que soit la définition retenue et le document servant de support au tracé (plan, mosaïque de photographies aériennes, tirage du canal panchromatique SPOT), quelques règles minimales doivent être suivies :



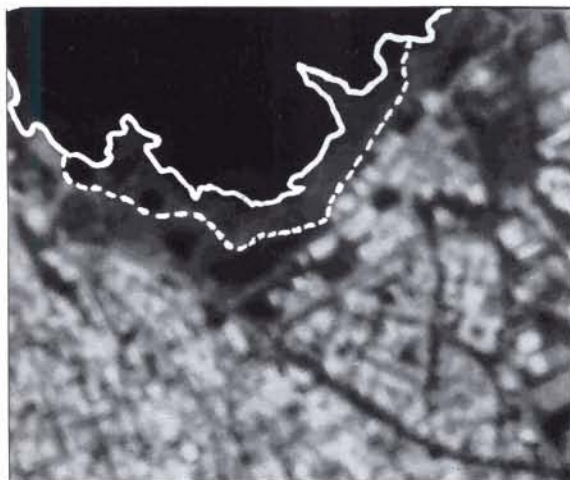
S'efforcer d'emprunter une **limite** naturelle (rivière, ravin, ...) ou artificielle (voirie,...) **repérable** sans ambiguïté à la fois sur l'image satellite et sur le terrain.

Exclure de la base les zones non bâties occupant une surface importante (au moins la taille d'un îlot moyen) à l'**intérieur** de la ville. Ces exclusions permettront une meilleure précision des estimations.

Il faut **contrôler sur le terrain** la limite urbaine extraite de l'image, surtout s'il existe un intervalle de temps important entre la date d'enregistrement de l'image et celle de l'enquête démographique.

Si l'on dispose seulement d'un film du canal panchromatique SPOT.

Globalement, la végétation apparaît sur le tirage photographique d'autant plus foncée que l'activité chlorophyllienne est importante. L'espace urbain (minéral) apparaît en plus clair, à l'exception des centres-ville très denses recouverts de toitures très foncées (type tuile canal) ou des espaces bitumés. De plus, la prise en compte de critères autres que les seules teintes de gris, comme l'agencement général, la structure ou la texture apportent une information intéressante.



Deux photo-interprètes tracent de façon indépendante les limites de la ville sur un tirage de ce canal (échelle 1/15 000 à 1/30 000).

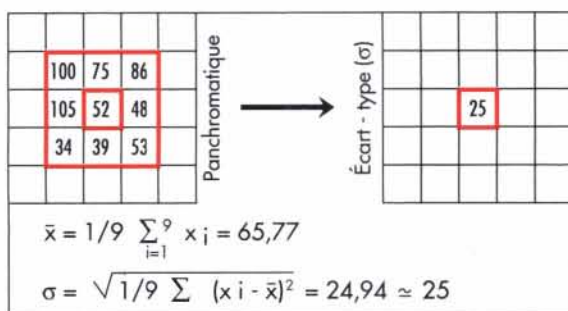
Les limites obtenues sont ensuite confrontées afin de retenir une délimitation unique, les endroits où les tracés divergent fortement devant faire l'objet d'un soin tout particulier.

Si l'on dispose d'une bande magnétique SPOT XS et Pa, ou d'une bande magnétique TM

Les traitements numériques à effectuer sont simples mais en revanche, ils ne constituent qu'une **aide** à la détermination de la limite urbaine, qui sera établie sur un tirage du canal panchromatique ou un plan (échelle 1/15 000 à 1/30 000). L'hypothèse de base est que la zone urbaine correspond à un **espace à dominante minérale et fortement hétérogène** (compte tenu de la résolution spatiale du panchromatique).

SPOT : $128 \times (1 + (XS3 - XS2) / (XS3 + XS2))$

TM : $128 \times (1 + (TM4 - TM3) / (TM4 + TM3))$

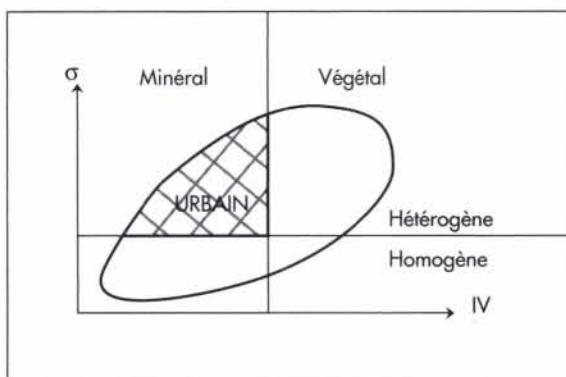


Calcul de l'indice de végétation normalisé (I.V.).

Calcul de l'écart-type local sur le canal panchromatique de SPOT ou sur le canal 4 de TM.

Recherche des seuils pertinents de l'indice de végétation et de l'écart-type local sur des zones test bien connues.

Réalisation d'un histogramme bidimensionnel sur des fenêtres échantillon. Quatre classes sont définies par croisement de l'indice de végétation et de l'écart-type local seuillés.



Classification de l'ensemble de l'image par segmentation selon les seuils retenus.

Tracé manuel de la limite sur le tirage du canal panchromatique ou le plan, en utilisant la classification, et en suivant les règles citées au début de cette fiche.



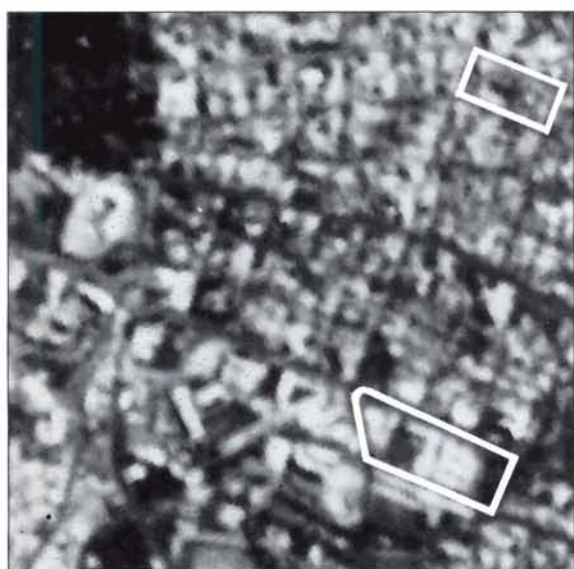
Durée du tracé de la limite urbaine : environ 2 jours pour une ville de 15 à 20 000 Hectares

CLASSIFICATION DE L'IMAGE SATELLITE

La classification supervisée de l'image satellite va servir de fondement pour stratifier la base de sondage. La méthode de classification choisie, selon **la densité de l'espace bâti** (définie fiche 4) est d'ores et déjà opérationnelle ; **aisément modélisable** à partir des images satellite, la densité du bâti est d'autre part **fortement liée** aux informations de type démographique ou socio-économique que l'on désire collecter.

Il est nécessaire de posséder une information précise sur la densité du bâti, dans un certain nombre de zones test pour **initialiser** la classification. Cette connaissance *a priori* provient d'une collecte d'information sur un échantillon d'îlots (fiche 4).

Pour réaliser une classification selon la densité du bâti à partir d'une bande magnétique SPOT XS et Pa ou d'une bande magnétique Thematic Mapper, il faut effectuer les opérations suivantes :



Localisation des îlots appartenant à l'échantillon sur l'image du canal panchromatique SPOT ou du canal 4 TM. Si l'on utilise une image TM, cette opération est très délicate à réaliser.

Ilot	SPOT I.V.	TM3	TM4	TM5
1	154,1	22,1	30,2	36,2
2	167,9	18,7	34,2	40,3
3	133,8	25,4	30,39	37,38
4	123,5	26,7	29,5	38,3

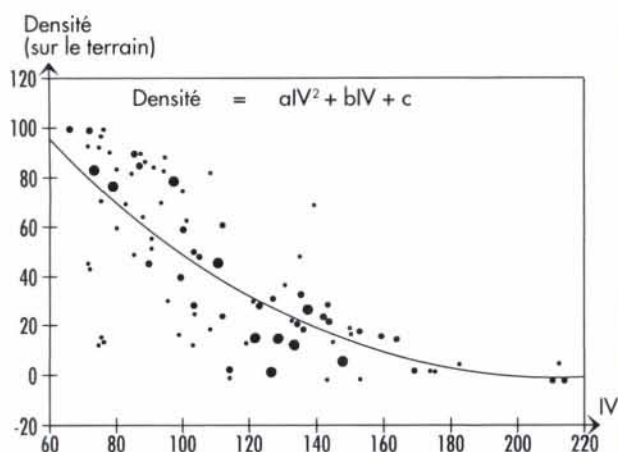
Calcul de la moyenne radiométrique :

- de l'indice de végétation (SPOT)

ou

- de chacun des canaux 3, 4 et 5 (TM)

à l'intérieur de chacun des îlots appartenant à l'échantillon enquêté.

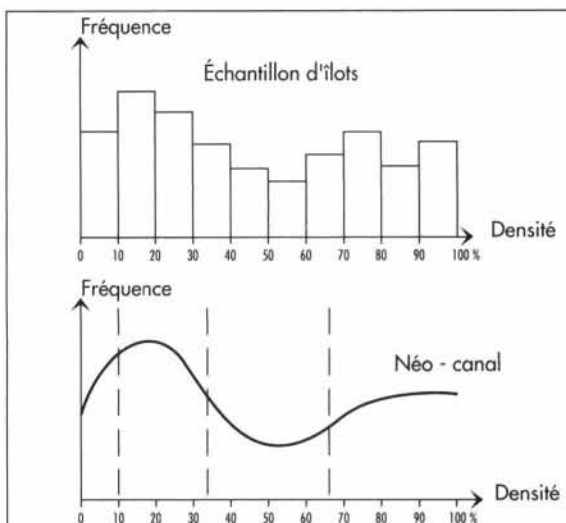


Modélisation de la relation densité du bâti/ image satellite : recherche des coefficients de la meilleure régression polynomiale simple (SPOT) ou multiple (TM) entre la densité du bâti calculée sur le terrain et la(les) moyenne(s) radiométrique(s). Dans un cas comme dans l'autre, le coefficient de corrélation prévisible se situe aux alentours de **0,75**.

0	18	22	20	34	74	90	100	100	100
15	20	25	42	53	61	70	85	95	98
7	25	27	38	48	51	55	66	80	100
8	26	31	40	60	58	65	74	87	55
10	10	30	18	30	49	59	75	80	80

Image du néo - canal "densité du bâti"

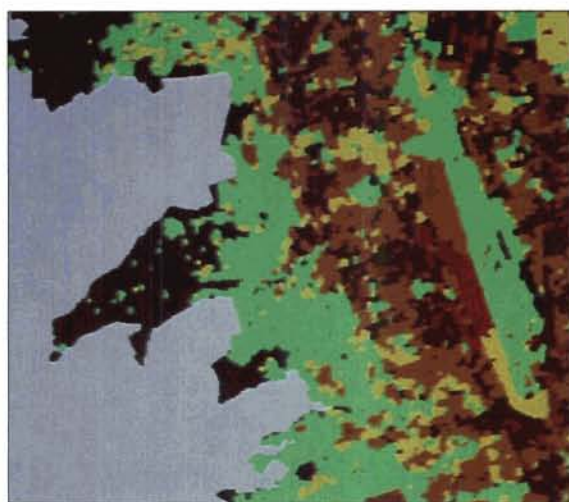
Création du néo-canal "densité du bâti" à partir des coefficients de la régression sur l'ensemble de la base. L'image ainsi obtenue possède des valeurs variant entre 0 et 100 et constitue une approximation en pourcentage de la densité du bâti par pixel ¹.



Seuillage de l'histogramme des fréquences du néo-canal "densité du bâti" en fonction de trois éléments :

- sa forme (nombre de pics et de creux) ;
- l'histogramme des fréquences de la densité du bâti calculée d'après l'information recueillie au sol sur un échantillon d'îlots;
- l'information exogène liant densités du bâti et types de bâti.

Pratiquement, on retient entre 3 et 5 classes.



Segmentation de ce néo-canal en fonction de ces seuils.

On obtient alors une classification de l'espace urbanisé en h classes de densité du bâti.

Il serait évidemment possible de stratifier l'espace urbanisé en fonction d'autres critères éventuellement plus liés avec les données socio-démographiques que l'on désire collecter, par exemple la couleur des toits ou la hauteur des bâtiments. Dans ce cas, il serait impératif de recueillir l'information adéquate lors de l'enquête (fiche 4) puis de modéliser la relation entre ce critère et la (ou les) variable(s) radiométrique(s) correspondante(s).

1 : Si l'on dispose d'un logiciel de traitement d'image satellite complet, on peut envisager les étapes suivantes : extraction de la voirie sur image panchromatique à partir d'opérateurs de morphologie mathématique, segmentation et "labellisation" des zones obtenues à partir de la voirie, classification des zones. La seule différence notable réside dans l'individu à classer, soit le pixel, soit une zone se rapprochant de l'îlot.

Durée de la classification de l'image : environ 8 jours pour une ville de 15 à 20 000 Hectares

STRATIFICATION DE LA BASE DE SONDAGE ET MESURE DE SUPERFICIE DES STRATES

Qu'est-ce qu'un sondage stratifié, et quel intérêt a-t-on à stratifier ?

Un sondage stratifié est un sondage où les opérations de sélection de l'échantillon et d'extrapolation à partir de cet échantillon se font, **séparément et indépendamment, dans plusieurs sous-ensembles appelés "strates"** distinguées à l'avance dans l'univers.

La stratification peut poursuivre deux types d'objectifs :

- d'une part, puisqu'elle permet de contrôler la taille de l'échantillon dans chaque strate, elle est souvent utilisée pour produire des **estimations séparées dans des sous-populations** particulières ;
- d'autre part, si les sous-ensembles définis présentent des particularités importantes par rapport à l'enquête, et sont chacun relativement homogène, le fait de stratifier le sondage permettra, pour une même taille de l'échantillon, une **réduction importante de l'erreur d'échantillonnage** au niveau global (ensemble des strates).

Quelles sources de données utiliser pour réaliser une stratification ?

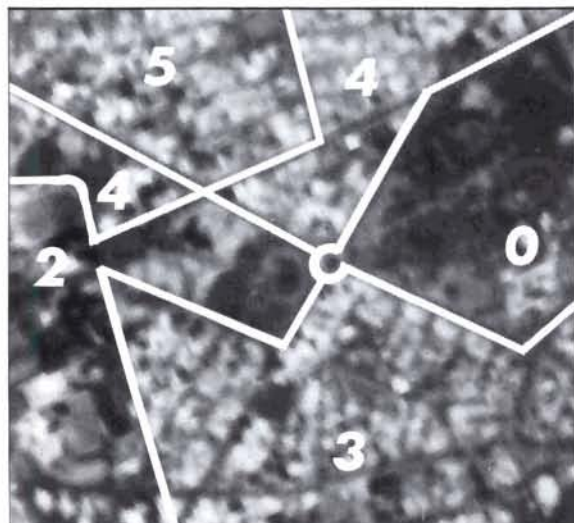
La stratification de la base de sondage peut venir :

- de l'interprétation d'une classification de l'image satellite : par exemple, la classification selon le niveau de densité du bâti, et/ou
- d'un zonage de la ville fondé sur des informations exogènes : par exemple, une typologie du bâti issue d'une photo-interprétation, d'une enquête sur le terrain ou d'une autre source statistique existante.

Quelles sont les règles de base à respecter pour réaliser une stratification à partir d'une image satellite classée ?

L'élaboration de la stratification consiste à établir un zonage, c'est-à-dire délimiter des zones le plus homogènes possibles au regard du critère de stratification. Selon le document dont on dispose, le zonage est tracé sur un plan, ou sur un tirage à grande échelle du canal panchromatique SPOT.

Quelques règles importantes doivent être suivies :



Les limites des zones doivent emprunter des frontières naturelles (rivières, par exemple) ou artificielles (rues) de façon à ne pas couper les îlots, qui constitueront l'unité de base de l'échantillon : il s'agit d'affecter les **îlots** aux différentes strates.

Les plages homogènes ne doivent être individualisées que si elles contiennent un nombre entier d'îlots.

Deux interprètes réalisent, indépendamment, le zonage ; les résultats sont ensuite confrontés pour établir une synthèse, qui servira de stratification. Ce procédé permet de réduire la subjectivité du zonage.

Il est utile de confronter, et éventuellement modifier, la stratification ainsi obtenue avec d'autres informations (résultats d'études spécifiques, connaissance personnelle) pouvant permettre de l'améliorer.

Pourquoi mesurer la superficie des strates?

Le calcul des estimateurs à partir des informations collectées par enquête sur échantillon fait appel à la superficie des strates dans la base de sondage (fiches 14 et 15).

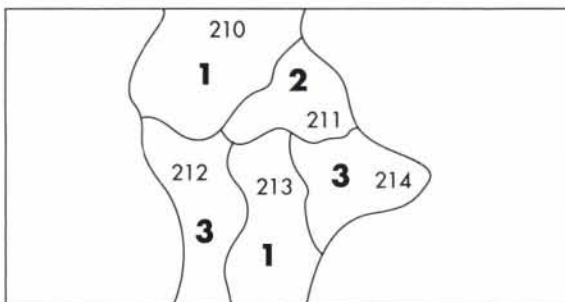
Comment réaliser la mesure de superficie des strates ?

Outre la solution informatique (comptage du nombre de pixels inclus dans chaque strate), trois techniques de mesure manuelle peuvent être pratiquées :

- mesure au planimètre (manuel ou électronique) ;
- mesure avec un polyester millimétré ;
- découpage de chaque zone et pesée sur une balance précise.

La mesure au planimètre semble être à la fois la plus rapide et la plus fiable.

Quelle que soit la technique de mesure, la démarche à suivre est la suivante:



Numéroter les zones et les strates auxquelles elles appartiennent sur le document où a été dessiné le zonage.

Zones	Strate 1	Strate 2	Strate 3
210	250 mm ²		
211		125 mm ²	
212			175 mm ²
213	200 mm ²		
214			150 mm ²

Mesurer la superficie de chaque zone.

Zones	Strate 1	Strate 2	Strate 3
360		550 mm ²	
361	280 mm ²		
Total	1 250 040 mm ²	2 784 610 mm ²	4 575 640 mm ²

Sommer les superficies des zones appartenant à une même strate, pour obtenir la superficie de chacune des h strates de la base.

TOTAL des strates	8 610 290 mm ²
-------------------	---------------------------

Sommer les superficies des strates (Z_k), pour obtenir la superficie totale de la base de sondage (Z).

La superficie des strates doit être cohérente avec celle des îlots appartenant à l'échantillon. Dans le cas où les îlots sont mesurés sur le terrain, il ne faut donc pas oublier de convertir les mesures de strates sur plan, en mesures de superficie sur le terrain, en utilisant l'échelle du document sur lequel ont été mesurées les strates.

Exemple: strate de 16 000 mm² mesurés sur un document au 1/ 30 000

Surface de la strate sur le terrain = 16 000 x (30 000)² mm² = 14 400 000 000 000 mm² = 14 400 000 m², ou 1440 Hectares.

**Durée de la stratification de la base et de la mesure de superficie des strates :
environ 4 jours pour une ville de 15 à 20 000 Hectares**

CALCUL DES TAUX DE SONDAGE ET DES TAILLES DES GRILLES DE TIRAGE

1. Calcul du nombre total de ménages à enquêter

Si le budget disponible pour l'enquête proprement-dite (c'est-à-dire n'incluant pas les postes fixes de dépense répertoriés au verso de la fiche 1) est fixé à l'avance, le montant de ce budget (B) et le coût unitaire d'enquête par ménage (c) conduisent au nombre total de ménages que l'on peut inclure dans l'échantillon (n) :

$$n = B/c$$

Lorsque l'on cherche, au contraire, à atteindre une précision (s: amplitude du demi intervalle de confiance à 95 %) fixée à l'avance pour l'estimation d'un ratio R, on peut utiliser comme approximation de n la formule :

$$n = 4R (1-R) / s^2$$

En toute rigueur, cette dernière formule correspond à un sondage aléatoire simple; étant donné le plan de sondage décrit dans ce manuel, la précision obtenue avec un échantillon de n ménages sera meilleure que celle souhaitée (s).

2. Calcul du nombre total d'îlots à sélectionner

La **détermination du nombre de ménages à enquêter dans chaque îlot** (donc du nombre d'îlots à sélectionner) est le résultat d'un compromis entre deux objectifs contradictoires :

- assurer, par un petit nombre d'enquêtes dans chaque îlot, la dispersion géographique maximale de l'échantillon au premier degré, ce qui assure une meilleure précision des estimations globales ;
- conserver dans chaque îlot un nombre suffisant de ménages enquêtés pour produire des estimations fiables dans chaque unité primaire.

Le premier objectif est généralement largement prioritaire, ce qui conduit à adopter la stratégie suivante de sélection des ménages.

On enquête un petit nombre constant de ménages (q, de l'ordre de 5 à 10) dans tous les îlots contenant au moins q ménages et on enquête tous les ménages des îlots contenant moins de q ménages.

n et q étant fixés, **pour calculer le nombre total d'îlots de l'échantillon (m)**, il faut connaître : la proportion (p) d'îlots de l'échantillon qui contiendront moins de (q) ménages et le nombre moyen de ménages (q') que l'on devra enquêter dans ces îlots. La formule permettant de calculer **m** est alors :

$$m = \frac{n}{pq' + (1-p)q}$$

Les quantités **p** et **q'** sont bien entendu inconnues *a priori*, il faut par conséquent les **estimer**, soit à partir d'informations exogènes disponibles (recensement ou enquête antérieure par exemple), soit à partir d'une collecte démographique préliminaire (fiche 5).

En l'absence de toute information supplémentaire on appliquera la formule $m=n/q$ qui fournit un nombre **m** d'îlots légèrement insuffisant, tant que q reste petit, pour enquêter le nombre **n** de ménages fixé par le budget disponible ou la précision désirée.

3. Calcul des taux de sondage

Pour gérer correctement le taux de sondage au premier degré dans chaque strate (il y a h strates), il faut disposer d'une **estimation du nombre d'îlots existant dans chaque strate (M_k , k variant de 1 à h)**. On l'obtient en divisant la surface totale de chaque strate (Z_k , connue d'après les mesures de surface faites sur l'image (fiche 8)) par la surface moyenne des îlots de cette strate $\bar{z}_k$, dont on a une estimation d'après la collecte d'information sur l'occupation du sol (fiche 4) ou d'après les mesures planimétrées sur une base cartographique éventuellement disponible (dans ce cas, on pourra procéder par comptage des îlots pour estimer directement M_k) :

$$M_k = \frac{Z_k}{\bar{z}_k}$$

On calculera donc M_k (estimation du nombre d'îlots total) pour chaque strate k.

Le **taux global au premier degré (t)** pour l'ensemble de la base de sondage est alors :

$$t = \frac{m}{h \sum_{k=1}^h M_k}$$

La **répartition de l'échantillon entre les strates** dépend bien sûr du type d'allocation retenu :

- si l'allocation est **proportionnelle** (taux constant dans toutes les strates), l'application du taux global (t) au nombre total d'îlots estimé de chaque strate (M_k) donne le nombre d'îlots à enquêter dans chaque strate (m_k) ;
- si l'étude des possibilités d'**optimisation de l'allocation** aux strates a pu être menée à bien, on a déterminé, en fonction de m, des valeurs estimées des M_k et de l'information disponible sur les écarts type des caractères à estimer au sein des strates (S_{ky}), les nombres m_k d'îlots à sélectionner dans chaque strate (fiche 5).

4. Calcul des tailles des grilles de tirage dans chaque strate

Pour obtenir dans chaque strate la surface de la maille de la grille de tirage, il suffit de diviser la surface de la strate (Z_k , exprimée en mm^2 mesurée sur image ou plan) par le nombre d'îlots à sélectionner (m_k) ; la longueur du côté de la maille (d_k) s'obtient en prenant la racine carrée de la surface de la maille.

5. Exemple d'application

Pour une enquête donnée, le budget alloué permet d'enquêter environ $n = 4\,000$ ménages et on choisit d'enquêter dix ménages dans chaque îlot sélectionné contenant au moins dix ménages et tous les ménages des autres îlots. La proportion d'îlots de moins de dix ménages attendue dans l'échantillon est de $p = 22,7\%$ et ces îlots contiendront en moyenne à peu près $q' = 6,9$ ménages (estimations faites d'après les sources d'information disponibles). On a donc :

$$m = \frac{4000}{0.227 \times 6.9 + 0.773 \times 10} = 430$$

La base de sondage étant, dans cet exemple d'enquête, stratifiée en $h = 5$ strates, les calculs effectués pour déterminer la taille de l'échantillon dans chacune des cinq strates et les tailles des grilles de sélection correspondantes sont résumés dans le tableau ci dessous :

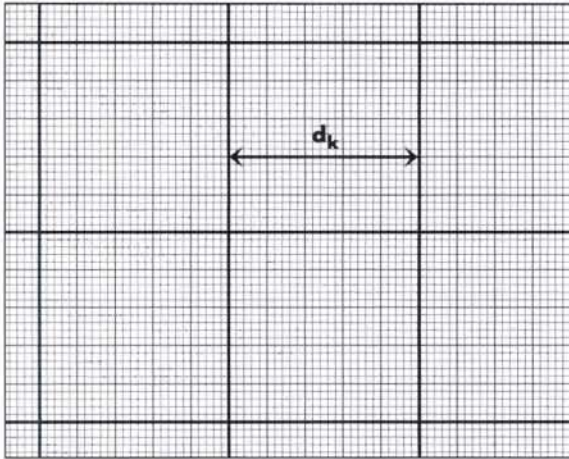
Strates	1	2	3	4	5	Total
Z_k (mm^2)	161 963	151 408	152 757	149 033	48 186	663 347
$\bar{z}_k$ (mm^2)	146,3	120,3	94,7	57,8	56,3	89,5
M_k	1 107	1 259	1 613	2 577	856	7.412
m_k (alloc. proportionnelle)	64	73	94	149	50	430
m_k (alloc. optimale)	28	27	108	229	38	430
d_k (mm) (alloc. optimale)	76,1	74,9	37,6	25,5	35,6	

Durée des calculs et du tracé des grilles : 1 demi-journée

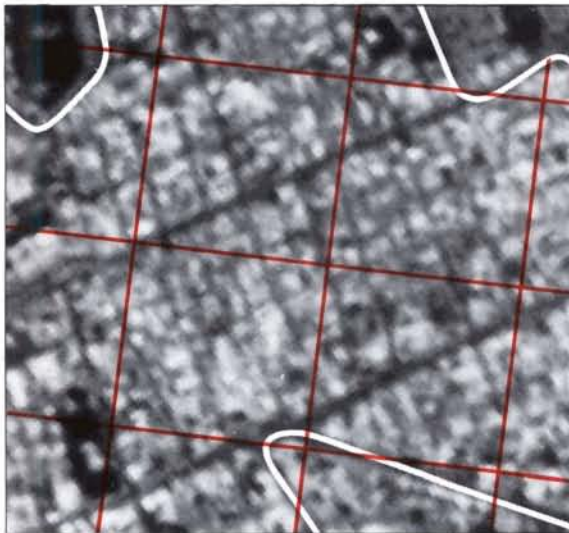
SELECTION DES ILOTS DE L'ÉCHANTILLON A L'AIDE DES GRILLES

La sélection des îlots s'effectue indépendamment dans chaque strate à l'aide d'une grille dont la longueur du côté de la maille (d_k) a été calculée pour respecter le taux prévu dans la strate (fiche 9).

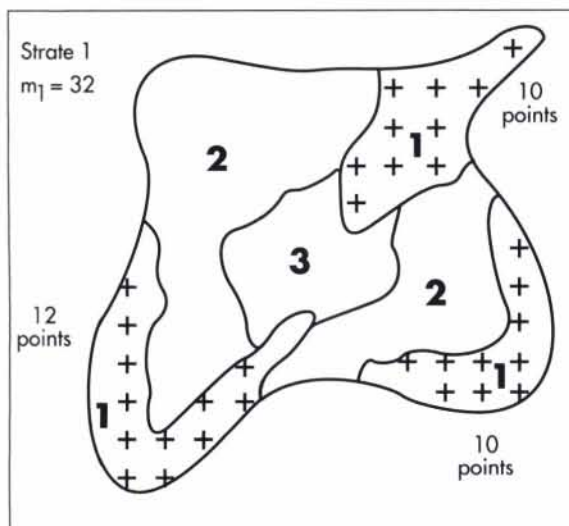
Pour chaque strate, les **étapes du tirage** sont les suivantes :



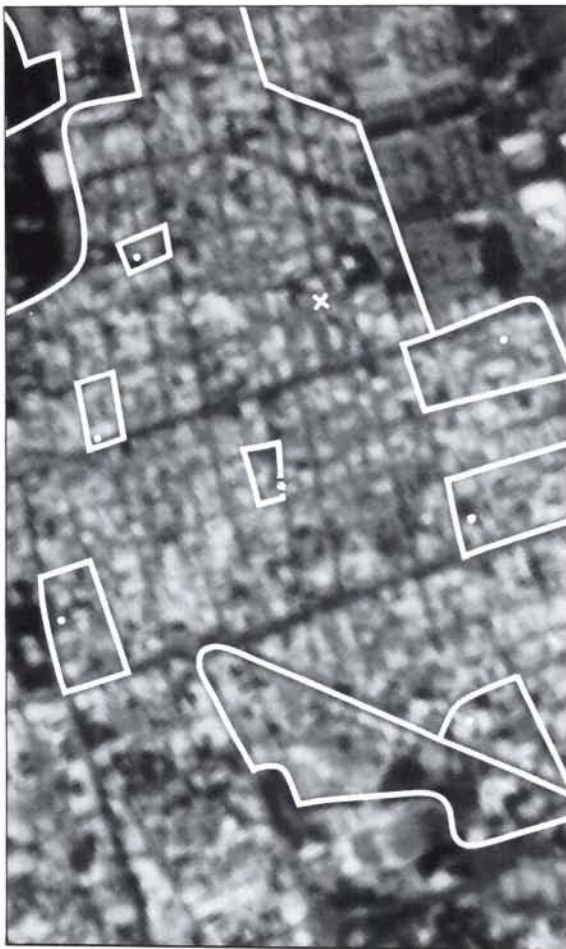
Construire la grille sur un support transparent millimétré (d_k = intervalle entre 2 lignes de la grille, pour la strate k).



Superposer la grille à l'image satellite (tirage papier à une échelle voisine du 1/15 000) ou au plan, en choisissant le point de calage de la grille aléatoirement et en évitant d'orienter les axes principaux de la grille (horizontal, vertical ou diagonal) parallèlement à un axe principal du réseau de voirie.



Compter le nombre de points (intersections) de la grille inclus dans la strate. Si ce nombre ne correspond pas au nombre d'îlots souhaité (m_k), répéter l'opération en positionnant différemment la grille, jusqu'à obtenir le nombre d'îlots défini dans le plan de sondage (fiche 9 pour le calcul du nombre d'îlots de l'échantillon par strate).



Identifier les îlots correspondant aux points de la grille :

- si le point est situé à l'intérieur d'un îlot, c'est l'îlot en question qui est sélectionné ;
- si le point est situé sur la voirie, entre des îlots, appliquer une règle systématique pour la sélection de l'îlot : par exemple, l'îlot situé immédiatement au sud du point ;
- si deux points de la grille appartiennent à un seul et même îlot, cet îlot est sélectionné, ainsi que son voisin, en appliquant une règle systématique pour rechercher l'îlot voisin : par exemple, l'îlot situé immédiatement au sud de celui déjà sélectionné.

Selon le type de quartier et donc le degré de lisibilité de la voirie sur l'image satellite, **les limites de chacun des îlots** de l'échantillon sont déterminées



directement sur l'image satellite,

ou sur le terrain, avec le superviseur.

Pour les parties de la ville où existe une cartographie, il est intéressant de l'utiliser pour reporter les îlots de l'échantillon;



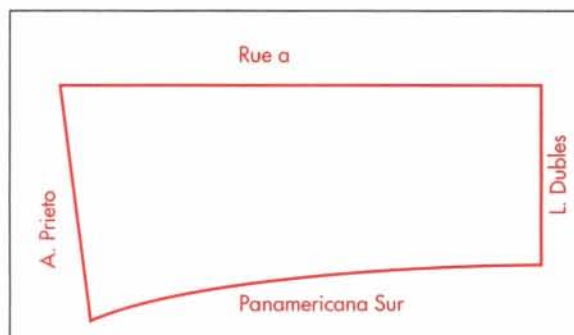
ainsi, les enquêteurs disposent d'un **plan de situation des îlots** à enquêter.

Dans les zones non cartographiées, ou pour les îlots aux limites complexes, le superviseur se rendra sur le terrain avec l'enquêteur pour lui définir précisément les îlots d'enquête.

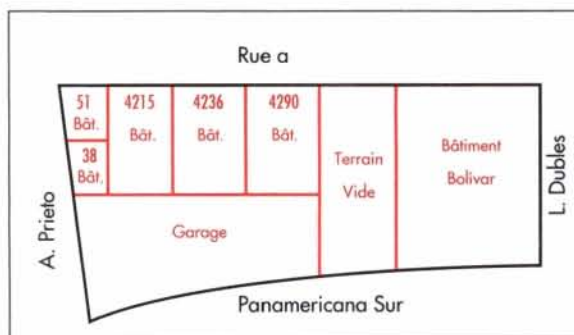
Durée de la sélection des 80 à 100 îlots d'une strate (inclus le report sur plan) : 3 jours

RECENSEMENT DES MENAGES DANS LES ILOTS DE L'ECHANTILLON

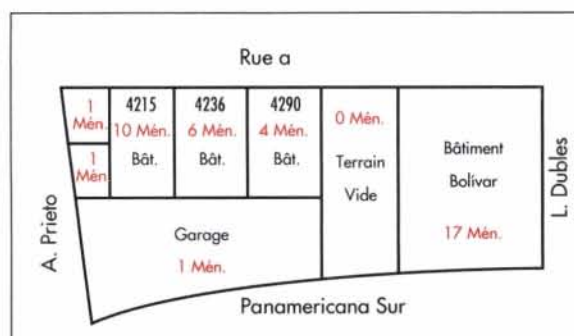
Pour chaque îlot i de l'échantillon, les superviseurs établissent, sur une fiche du type de celle figurant au verso, un schéma simplifié de l'îlot sur lequel figurent :



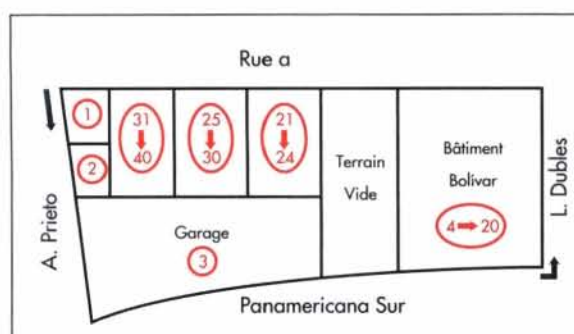
- les renseignements nécessaires à l'identification de l'îlot par l'enquêteur : noms des rues bordant l'îlot, éléments naturels servant de limites (ruisseau, ravin,...etc) ;



- le dessin de l'ensemble des bâtiments de l'îlot ;



- le nombre de ménages dans chacun des bâtiments : pour obtenir cette information, les superviseurs interrogent les habitants de chacun des bâtiments ;

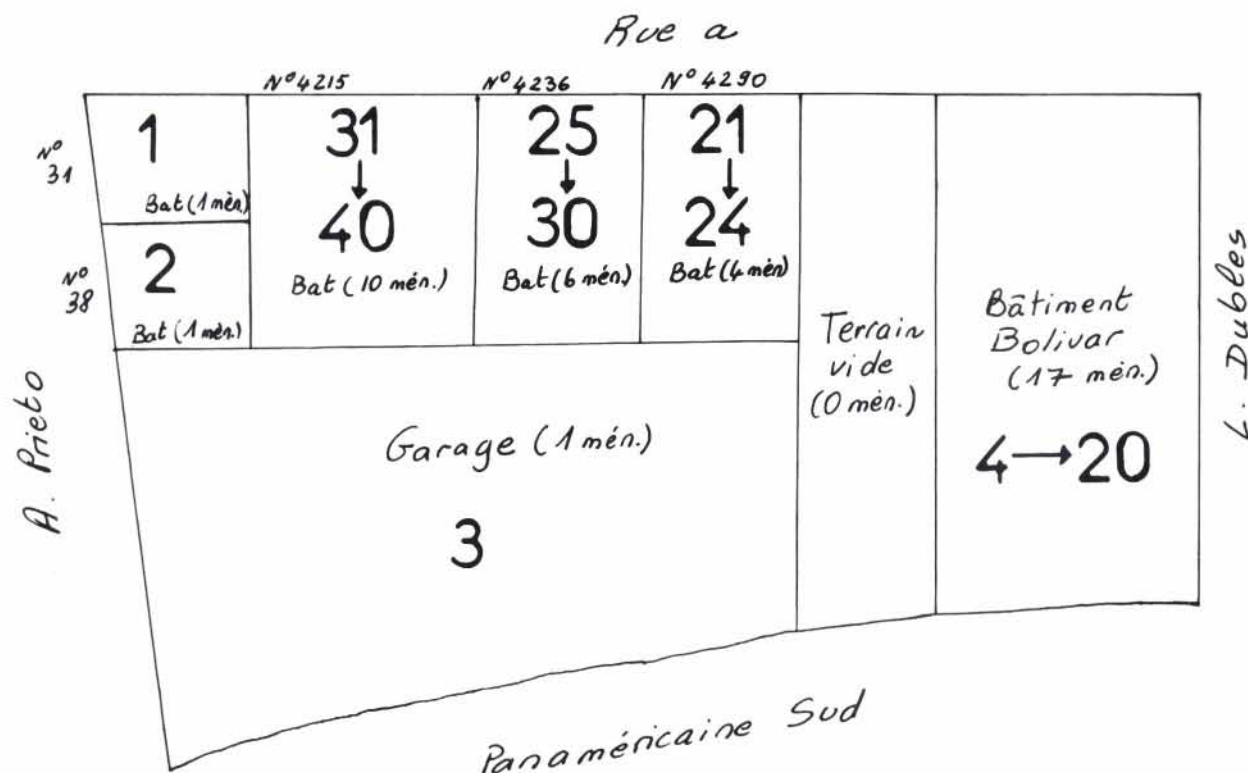


- la numérotation des ménages dans l'îlot, de 1 à N_i , par ordre croissant en suivant le tour de l'îlot.

FICHE ILOT

Nom SUPERVISEUR <i>Henry ALBAN</i>	N° Ilot <i>253</i>
Nom ENQUETEUR <i>Juan PUENTE</i>	Nombre de ménages : TOTAL <i>40</i>

SCHEMA DE L'ILOT



ECHANTILLON DE MENAGES

Liste des ménages

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Collecte / Codification

Enquête ☐

Vérif. enquête (1) ☐

Vérif. enquête (2) ☐

Corrections question. ☐

Codification ☐

Vérif. codification ☐

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Durée du recensement (îlot moyen de 35 logements) : 5 îlots/jour/superviseur

ORSTOM

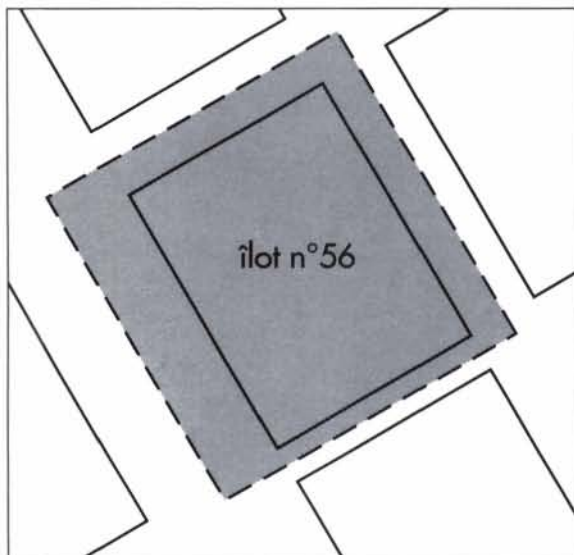
MESURE DE LA SURFACE DES ILOTS DE L'ECHANTILLON

La mesure de surface des îlots, information nécessaire au calcul des estimateurs (fiche 14), peut être réalisée selon trois techniques :

- avec un support polyester millimétré superposé à un tirage papier de l'image satellite (canal panchromatique de SPOT) à une échelle voisine du 1/15 000 ;
- par programme informatique, calculant le nombre de pixels de l'image satellite (canal panchromatique de SPOT, pixel=10 m x 10 m, soit 100 m²) inclus dans chaque îlot ;
- sur le terrain.

Les tests de précision effectués ont démontré que cette dernière solution est la plus précise (+/- 8%, pour une personne réalisant pour la première fois ce type de mesure) ; **c'est donc cette méthode**, expliquée dans cette fiche, **qu'il est conseillé d'utiliser**.

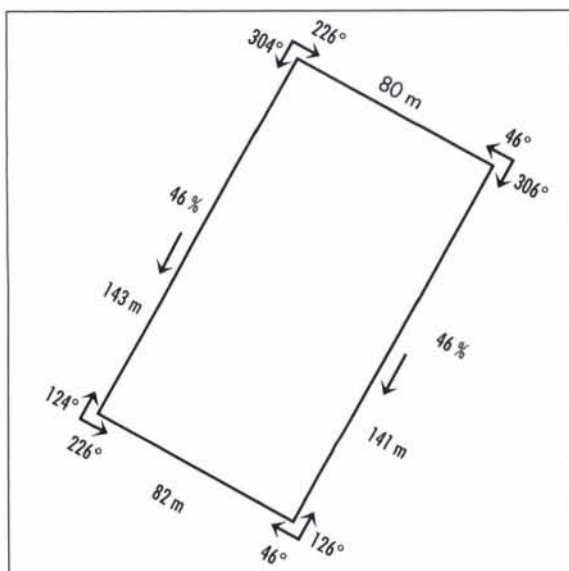
Pour éviter toute erreur d'identification des îlots, le travail de mesure doit être confié au même superviseur que celui qui effectue le recensement des ménages (fiche 11) ; l'organisation optimale du travail des superviseurs conduit ceux-ci à réaliser, lors d'une seule et même visite sur un îlot, le recensement des logements et la mesure de la surface de l'îlot. Ainsi, on s'assure que les deux informations correspondent au même îlot.



Afin d'être cohérent avec la mesure de surface des strates, effectuée sur document projeté (image ou plan) et incluant la voirie, la "surface de l'îlot" **inclut la demi-voirie** bordant l'îlot, et correspond à la surface **projetée sur un plan horizontal**.

Matériel nécessaire au superviseur : un décimètre, une boussole, un clinomètre.

Sur le terrain, au verso de la fiche servant au recensement des logements, le superviseur fait un schéma de l'îlot comportant les informations suivantes :



Longueurs des côtés, décomposés, si nécessaire, en segments rectilignes et de même pente :

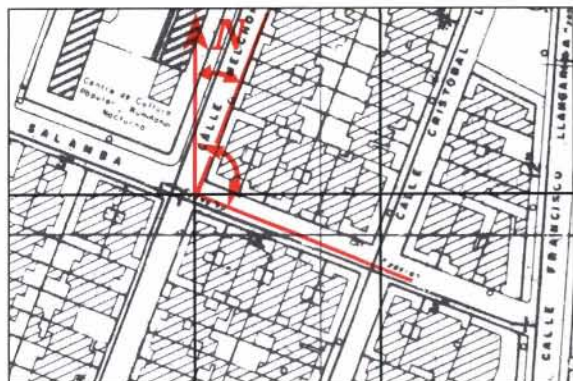
- sur terrain en pente, les longueurs sont mesurées au décimètre ;
- sur terrain plat, les longueurs sont mesurées en nombre de pas ; la longueur du pas est connue par la mesure simultanée en pas et au décimètre durant les deux premiers jours de travail du superviseur.

Pente de chaque segment : en cas de pente forte (> 30 %), la pente de chaque segment est mesurée au clinomètre, et reportée sur le schéma de l'îlot.

O.R.S.T.O.M. Fonds Documentaire

N° :

Cote :



Angle entre les segments : à chaque extrémité d'un segment, le superviseur mesure, à la boussole, l'angle entre la direction du segment et le nord.

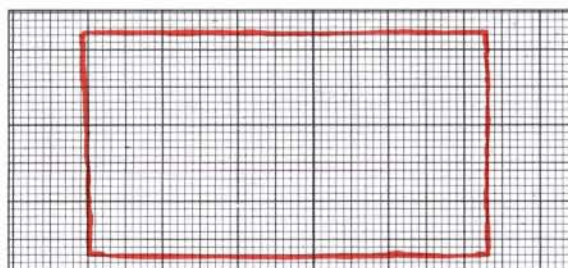
Au bureau, un opérateur réalise le calcul de la surface de l'îlot (z_i) à partir du schéma complet établi par le superviseur :



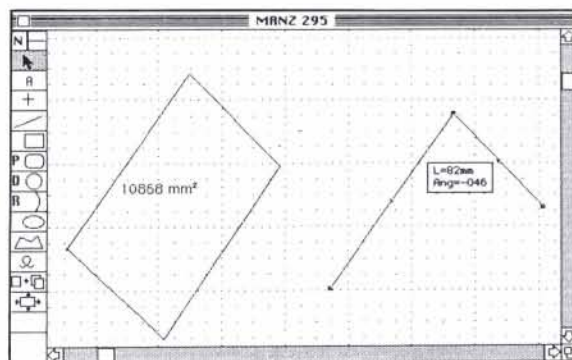
Transformation en mètres des longueurs mesurées en pas ;

	A	B	C
1	LONGUEUR	PENTE	LONG. PROJ.
2	51	0,31	48,71
3	64	0,35	60,41
4	102	0,41	94,38
5	118	0,37	110,67
6	141	0,46	128,09
7	143	0,46	129,91

Transformation des longueurs mesurées sur un terrain pentu **en longueurs projetées** sur un plan horizontal ;



Réalisation manuelle du calcul de surface : dessin à l'échelle sur papier millimétré, planimétrage ou comptage du nombre de mm^2 inclus dans le polygone ;



ou **réalisation informatique** : avec le programme Mac Draft sur Macintosh, ou tout autre programme de dessin équivalent sur micro-ordinateur compatible IBM.

Avec MacDraft (exemple ci-contre), l'opérateur dessine le polygone composé des segments en respectant les longueurs et les angles prescrits par le schéma du superviseur ; l'ordinateur affiche la surface du polygone ainsi défini.

Durée de la mesure des îlots sur le terrain : 8 îlots/jour/superviseur
Durée du calcul de surface (inclus conversion des longueurs) : 40 îlots/jour/opérateur

SELECTION DES MENAGES DANS LES ILOTS DE L'ECHANTILLON

Les îlots de l'échantillon comptant **moins de q ménages** (détermination de ce seuil q, décrite fiche 9) font l'objet d'une enquête **exhaustive** : tous les ménages de l'îlot doivent être enquêtés.

Dans un îlot i comptant **q ménages ou plus**, le superviseur doit établir une liste des q ménages à enquêter. Il établit cette liste en appliquant **la procédure de tirage systématique suivante** :

$$\left. \begin{array}{l} N_i = 40 \\ q = 10 \end{array} \right\} \longrightarrow k = 4$$

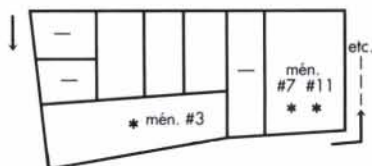
Soit : N_i = nombre total de ménages de l'îlot i

et : q = nombre de ménages à enquêter

$$k = \text{pas du tirage systématique} = N_i / q$$

Si $Y = 3 \longrightarrow$ Ménages sélectionnés
 # 3
 # 7 (3 + 4)
 # 11 (7 + 4)
 ...
 etc.

Tirage d'un nombre Y compris entre 1 et k, au hasard : le ménage portant ce numéro Y est sélectionné, ainsi que les ménages portant le numéro $Y+k$, $Y+2k$,etc



ECHANTILLON DE MENAGES

Liste des ménages ... 3 ... 7

11 15 19 23

27 31 35 39

Ménages de remplacement.....

Dans la partie inférieure de la fiche îlot, le superviseur indique la liste des numéros des ménages sélectionnés.

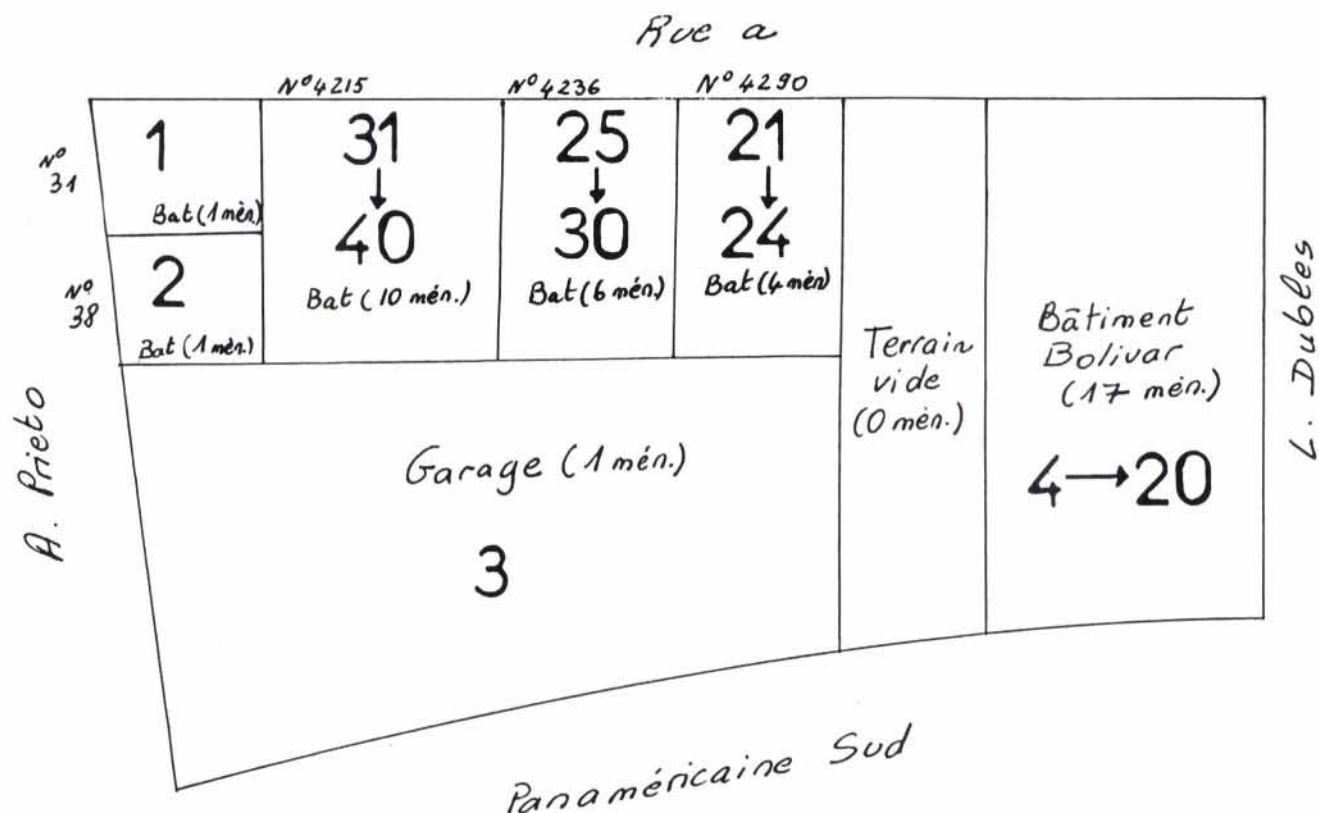
L'enquêteur doit enquêter tous les ménages de la liste. Si, en dépit de plusieurs visites, l'enquêteur ne peut réaliser l'enquête dans un des ménages de la liste, il procède à l'enquête dans le ménage ayant le numéro immédiatement supérieur à celui qui n'a pas pu être enquêté, afin de respecter le nombre total d'enquêtes à réaliser dans l'îlot.

FICHE ILOT

Nom SUPERVISEUR... *Heary ALBAN*...
 Nom ENQUETEUR... *Juan PUENTE*...

N° Ilot *253*
 Nombre de ménages : TOTAL *40*

SCHEMA DE L'ILOT



ECHANTILLON DE MENAGES

Liste des ménages *3 7*
11 15 19 23
27 31 35 39

Ménages de remplacement.....

Collecte / Codification

Enquête ☐
 Vérif. enquête (1) ☐
 Vérif. enquête (2) ☐
 Corrections question. ☐
 Codification ☐
 Vérif. codification ☐

Durée de la sélection (îlot moyen de 35 ménages) : 30 îlots/superviseur/jour

ORSTOM

CALCUL DES ESTIMATIONS ET INTERVALLES DE CONFIANCE

(1. FORMULAIRE)

Notations adoptées dans le formulaire

Objets	Notations	Significations
1) Base de sondage	N_i	Nombre total de ménages recensés dans l'îlot i
2) Echantillon	n_i m_k	Nombre de ménages de l'îlot i appartenant à l'échantillon Nombre d'îlots de la strate k appartenant à l'échantillon
3) Variables dont on doit estimer le total ou la moyenne (y) ou intervenant au dénominateur du quotient à estimer (x, N)	y_{ij}, x_{ij} $\hat{y}_i, \hat{x}_i$ $\hat{Y}_k, \hat{X}_k$ $\hat{N}_k$ $\hat{\bar{y}}_k, \hat{\bar{x}}_k$	Valeurs des caractères pour le ménage j de l'îlot i Estimateurs ¹ des totaux dans l'îlot i Estimateurs des totaux dans la strate k Estimateur du nombre de ménages dans la strate k Estimateurs des moyennes dans la strate k
4) Variable définissant la probabilité d'inclusion dans l'échantillon des unités primaires (z)	z_i Z_k	Surface totale de l'îlot i Surface totale de la strate k
5) Quotients à estimer (r)	$\hat{r}_i$ R_k $\hat{R}_k$	Estimateur du quotient dans l'îlot i ($\hat{r}_i = \hat{y}_i / \hat{x}_i$) Valeur du quotient dans la strate k ($R_k = Y_k / X_k$) Estimateur du quotient dans la strate k ($\hat{R}_k = \hat{Y}_k / \hat{X}_k$)
6) Sommations	n_i $\sum_{j=1}$ m_k $\sum_{i=1}$	Somme sur les ménages de l'îlot i appartenant à l'échantillon Somme sur les îlots de la strate k appartenant à l'échantillon
7) Erreurs d'échantillonnage et estimation des erreurs	$\hat{V}(\hat{Y}_k)$ et $\hat{V}(\hat{Y}_k)$ $\hat{E}(\hat{R}_k - R_k)^2$	Estimations des variances des estimateurs $\hat{Y}_k$ et $\hat{\bar{y}}_k$ dans la strate k (premier degré du sondage) Estimation de l'écart quadratique moyen de l'estimateur $\hat{R}_k$ dans la strate k (premier degré du sondage)

1 : Le symbole : $\hat{}$ placé au dessus de la lettre désignant une quantité donnée, signifie que l'on parle d'un estimateur de cette quantité. Par exemple $\hat{y}_i$ est l'estimateur d'un total de population dans l'îlot i et $\hat{V}(\hat{Y}_k)$ est l'estimateur de la variance de l'estimateur $\hat{Y}_k$

Sondages, estimateurs, erreurs d'échantillonnage et intervalles de confiance

Le principe du **sondage** consiste, lorsqu'on veut recueillir une information sur une population considérée dans son ensemble, à n'observer qu'une fraction de cette population appelée **échantillon**. De cette information partielle, il faut ensuite déduire des indications sur les caractéristiques de la population entière.

Les formules mathématiques données dans les deux premières colonnes du tableau ci-dessous combinent les informations obtenues sur l'échantillon pour fournir des approximations des statistiques recherchées ; ce sont les **estimateurs** de ces statistiques. Bien sûr, la valeur fournie par l'estimateur n'est généralement pas exactement égale à la vraie valeur de la statistique sur l'ensemble de la population.

L'écart entre ces deux valeurs a deux composantes :

- la première, introduite lors de la mise en œuvre de l'enquête du fait des non réponses ou réponses fausses, etc..., s'appelle **l'erreur de mesure** ; elle existe dans toute procédure de collecte d'information, qu'il s'agisse de sondages ou de recensements et c'est le soin apporté à l'organisation de la collecte sur le terrain qui permettra de la réduire ;
- la seconde composante, appelée **erreur d'échantillonnage**, est celle qui résulte de l'observation partielle de la population. Elle est donc spécifique aux procédures de collecte par sondage et dépend du **plan de sondage**, c'est-à-dire de l'ensemble des règles adoptées pour sélectionner l'échantillon et du **choix des estimateurs**. La dernière colonne du tableau ci-dessous fournit une estimation de cette erreur pour le plan de sondage que nous préconisons ici.

Cette indication doit être interprétée de la façon suivante.

Elle permet de construire pour chaque estimation **l'intervalle de confiance** pour un **seuil de probabilité** donné. Par exemple, au seuil de probabilité de 95%, l'intervalle de confiance (appelé intervalle de confiance à 95%) est l'intervalle qui contiendra 95% des résultats d'estimation si l'on répète un grand nombre de fois l'expérience d'échantillonnage. Pour ce même seuil de probabilité (95%), l'intervalle de confiance, centré sur l'estimation obtenue, a pour amplitude totale quatre fois la racine carrée de l'estimateur de la variance donnée dans la dernière colonne du tableau formulaire.

Exemple : Dans l'enquête réalisée à Quito, la taille moyenne du ménage (hors ménages collectifs) pour l'ensemble de l'agglomération a été estimée à 4,535 personnes et la variance de cette estimation à 0,0024885. L'intervalle de confiance associé à cette estimation (lui aussi estimé) vaut donc : $4,535 \pm 2 \times \sqrt{0,0024885}$, c'est à dire :

$I = [4,435 ; 4,635]$ ou, exprimé en % de la quantité à estimer, $4,535 \pm 2,2 \%$.

Formulaire

Types de variable	Estimateur dans l'ilot i	Estimateur dans la strate k	Estimateur de la variance de l'estimateur dans la strate k
Totaux Y_k	$\hat{y}_i = N_i \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i}$	$\hat{Y}_k = \frac{Z_k}{m_k} \sum_{i=1}^{m_k} \frac{\hat{y}_i}{z_i}$	$\hat{V}(\hat{Y}_k) = \frac{1}{m_k (m_k - 1)} \sum_{i=1}^{m_k} \left(\frac{Z_k \hat{y}_i}{z_i} - \hat{Y}_k \right)^2$
Moyennes $\bar{Y}_k = \frac{Y_k}{N_k}$	$\hat{\bar{y}}_i = \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i}$	$\hat{\bar{Y}}_k = \hat{Y}_k / \hat{N}_k$ où, $\hat{N}_k = \frac{Z_k}{m_k} \sum_{i=1}^{m_k} \frac{N_i}{z_i}$	$\hat{V}(\hat{\bar{Y}}_k) = \frac{1}{m_k (m_k - 1)} \sum_{i=1}^{m_k} \frac{Z_k^2}{\hat{N}_k^2} \left(\frac{\hat{y}_i - \hat{\bar{Y}}_k}{z_i} \right)^2$
Ratios $R_k = \frac{Y_k}{X_k}$	$\hat{r}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{\sum_{j=1}^{n_i} x_{ij}}$	$\hat{R}_k = \frac{\sum_{i=1}^{m_k} \frac{\hat{y}_i}{z_i}}{\sum_{i=1}^{m_k} \frac{\hat{x}_i}{z_i}}$	$\hat{E}(\hat{R}_k - R_k)^2 = \frac{1}{m_k (m_k - 1)} \sum_{i=1}^{m_k} \frac{Z_k^2}{\hat{X}_k^2} \left(\frac{\hat{y}_i - \hat{R}_k \hat{x}_i}{z_i} \right)^2$

CALCUL DES ESTIMATIONS ET INTERVALLES DE CONFIANCE (2. MISE EN ŒUVRE INFORMATIQUE)

Etant donné le grand nombre de calculs d'estimations différentes auxquelles peut conduire l'exploitation d'une enquête par sondage, il n'est en général pas possible d'appliquer complètement le formulaire d'estimation fourni dans la fiche **14** à l'ensemble des variables. C'est pourquoi nous allons proposer maintenant deux solutions de mise en œuvre informatique de ce formulaire pour exploiter les données recueillies lors de l'enquête. Elles sont praticables à l'aide de logiciels statistiques généraux (voir ci-dessous) ; mais l'utilisation d'un logiciel spécialisé d'exploitation d'enquête par sondage simplifiera énormément le travail de programmation.

1. Fichiers de données et programme informatique pour la production d'estimations et des intervalles de confiance associés

Il s'agit ici de programmer, généralement à l'aide des commandes d'un logiciel approprié (par exemple SPSS/PC, SAS/PC, voire Dbase III), l'ensemble du formulaire exposé dans la fiche **14**. Cette solution permettra les calculs d'estimation des statistiques recherchées et **l'évaluation de l'erreur d'échantillonnage associée à chacune d'elles**.

Sur un micro-ordinateur, quel que soit l'environnement logiciel disponible, ces opérations exigeront des temps de calcul assez longs. Très souvent, elles ne peuvent donc pas être envisagées pour l'ensemble de l'exploitation de l'enquête mais nous ne saurions trop insister sur le fait que l'application d'une procédure de ce type à un nombre suffisant de variables peut seule fournir une **évaluation de la fiabilité** des informations recueillies.

Pour éviter les redondances d'information lors de la constitution des fichiers contenant les données issues de l'enquête et faciliter la programmation du formulaire, on constituera, à partir des fichiers issus de l'étape de saisie (fichiers ménages, individus, migrations, etc...), trois fichiers correspondant aux contenus suivants :

Fichiers	Contenu		Provenance des informations (fiches décrivant la collecte)
	Enregistrements	Variables	
Unités secondaires (j)	Un enregistrement par ménage	1- Identification de l'îlot d'appartenance (i) 2- Variables à estimer (y_{ij})	Résultat de l'enquête après agrégation de l'information au niveau du ménage
Unités primaires (i)	Un enregistrement par îlot	1- Identification de l'îlot (i) 2- Nombre de ménages présents dans l'îlot (N_i) 3- Surface de l'îlot (z_i) 4- Identification de la strate d'appartenance (k)	Recensement des ménages (fiche 11) Mesures sur image, sur plan ou sur le terrain (fiche 12)
Strates (k)	Un enregistrement par strate	1. Identification de la strate (k) 2. Surface de la strate (Z_k)	Mesures sur image ou plan (fiche 8)

Les étapes du calcul informatique **des estimateurs et de leurs variances** sont alors les suivantes :

- Sommation sur le fichier d'unités secondaires (ménages), calcul des $\hat{y}_j$ (ou $\hat{r}_j$ dans le cas de l'estimation d'un ratio) par application des formules (première colonne du tableau de la fiche **14**).
- Fusion horizontale des résultats de la première étape avec le fichier d'unités primaires.
- Sommation sur le fichier d'unités primaires par strate, calcul des $\hat{Y}_k$ ($\hat{Y}_k$ ou $\hat{R}_k$ dans les cas d'estimations de moyennes ou de ratios) par application de la seconde colonne du formulaire (fiche **14**) et des $\hat{V}(\hat{Y}_k)$ ($\hat{V}(\hat{Y}_k)$ ou $\hat{E}(\hat{R}_k - R_k)^2$ dans les cas d'estimation de moyennes ou de ratios) par application de la troisième colonne du formulaire (fiche **14**).
- Fusion horizontale des résultats de la troisième étape avec le fichier des strates.
- Sommation sur les strates pour l'obtention des estimations et des variances sur l'univers entier et/ou les domaines d'étude. Les sommations sont directes dans le cas des totaux (estimation totale = somme des estimations par strate, variance totale = somme des variances par strate). Dans le cas des moyennes et ratios, les sommations doivent être effectuées indépendamment pour les quantités figurant au numérateur et au dénominateur des estimateurs, comme indiqué dans le tableau ci dessous, avec les notations de la fiche **14**.

Types de variable	Estimateur	Estimateur de la variance
Moyennes $\bar{Y} = \frac{Y}{N}$	$\hat{\bar{Y}} = \frac{\sum_{k=1}^h \hat{Y}_k}{\sum_{k=1}^h \hat{N}_k}$	$\hat{V}(\hat{\bar{Y}}) = \frac{\sum_{k=1}^h \frac{Z_k^2}{m_k(m_k-1)} \sum_{i=1}^{m_k} \left(\frac{\hat{y}_i - N_i \hat{Y}_k}{z_i} \right)^2}{\left(\sum_{k=1}^h \hat{N}_k \right)^2}$
Ratios $R = \frac{Y}{X}$	$\hat{R} = \frac{\sum_{k=1}^h \hat{Y}_k}{\sum_{k=1}^h \hat{X}_k}$	$\hat{E}(\hat{R} - R)^2 = \frac{\sum_{k=1}^h \frac{Z_k^2}{m_k(m_k-1)} \sum_{i=1}^{m_k} \left(\frac{\hat{y}_i - \hat{R} \hat{x}_i}{z_i} \right)^2}{\left(\sum_{k=1}^h \hat{X}_k \right)^2}$

2. Calcul d'estimations par pondération des enregistrements des fichiers de données

Comme on le constate en examinant le formulaire de la fiche **14**, l'estimateur du total d'une variable y sur l'ensemble de la population est une **combinaison linéaire des valeurs prises par y sur chacune des unités statistiques observée dans l'échantillon** : ménages si y est une caractéristique attachée aux ménages, individus si y décrit des individus, etc.

Si on laisse de côté l'estimation de la variance des estimateurs, il y a donc une façon simple et rapide de calculer un grand nombre d'estimations de totaux (et par conséquent d'estimer toutes les statistiques simples qui se déduisent de ces totaux : moyennes, quotients, etc.). En effet, si l'on **pondère chaque enregistrement du fichier original de données** (fichier ménages, individus, chefs de ménage, etc.) par le coefficient qu'aurait cette unité dans l'estimateur (coefficient que nous appellerons désormais "**poids**" de l'unité statistique), l'estimation d'une statistique donnée sur l'ensemble de la population est alors égale à la valeur de cette même statistique sur le fichier "échantillon" pondéré.

Etant donné le plan de sondage pratiqué (lequel détermine la formule de l'estimateur), le poids d'une unité statistique quelconque de l'échantillon **ne dépend que de l'îlot et de la strate** auxquels elle appartient (ce poids sera donc le même quelle que soit l'unité considérée, ménage, individu, etc...). Le poids doit donc être calculé une fois pour toutes pour l'ensemble des unités primaires (îlots) de l'échantillon, et conservé dans un fichier "îlot". Il suffira ensuite d'attribuer, dans le fichier de données, le **même poids à toutes les unités appartenant à un même îlot**.

La formule donnant le poids p_i d'un îlot donné de l'échantillon (îlot i , appartenant à la strate k) est la suivante (avec les notations de la fiche **14**) :

$$p_i = \frac{Z_k N_i}{n_i m_k z_i}$$

3. Estimations dans divers domaines

Quelle que soit la solution informatique retenue, les estimations (et leur précision dans le premier cas) peuvent être obtenues pour l'ensemble de l'univers étudié, par strate constituée lors du tirage de l'échantillon, ou pour un quelconque autre domaine d'étude, c'est à dire un sous-ensemble géographique de l'univers de départ, pourvu qu'il contienne un nombre entier d'îlots, bien sûr.

Le principe des calculs reste inchangé; ils s'appliqueront simplement à des fichiers de données où **toutes les unités exclues du domaine d'étude** prennent pour la variable à estimer la valeur **zéro**.



DI > ACTIQUES

**Qui vise à instruire.
Qui appartient à la langue des sciences et des techniques.**

Ouvrage publié avec la participation du CNES

ISSN : 1142-2580
ISBN : 2-7099-0971-5
Editions de l'ORSTOM
70, route d'Aulnay - 93143 BONDY Cedex

© SPOT IMAGE - Image