

A PROPOS DU TRAITEMENT INFORMATIQUE DES DONNEES LOCALISEES
Une expérience en cours : télédétection et observation des populations urbaines

Françoise DUREAU, Géographe
 Département D, U.R. 406

Depuis le début de l'année 1985, le Département D a démarré un programme de recherches sur l'utilisation de la télédétection pour l'observation des populations urbaines, en collaboration avec l'Unité d'Infographie de la DIVA et l'Atelier de Télédétection de Bondy (1). L'objectif de ce programme est de mettre au point un système d'observation permanent des populations urbaines utilisant l'information apportée par les satellites à haute résolution.

Le traitement informatique de données géographiques tenant une place très importante dans ces recherches, nous nous proposons d'exposer ici notre méthodologie et les premiers travaux réalisés.

Après avoir rappelé les objectifs généraux du programme de recherches, nous présenterons les travaux méthodologiques en cours de réalisation sur la ville de Marseille, et les traitements de données localisées qu'ils impliquent. Nous nous intéresserons ensuite aux deux premières phases du travail, l'analyse de l'information et l'intégration des données dans une base gérée par le système TIGER. A partir de l'expérience menée sur Marseille seront envisagés les différents problèmes que pose l'intégration de données issues de sources d'information variées. Enfin, nous nous interrogerons sur les répercussions méthodologiques que peut avoir un outil tel que le système d'information géographique TIGER, développé par l'Unité d'Infographie.

1. LES OBJECTIFS DU PROGRAMME

1.1 La nécessité de développer de nouvelles méthodes d'observation des populations urbaines

Les populations des villes des pays en développement posent des problèmes spécifiques d'observation, du fait des caractéristiques de l'urbanisation dans ces pays, des modes particuliers de résidence des habitants (systèmes complexes de pluri-résidences), et de la rareté des éléments de connaissance habituels tels que cartes, plans cadastraux ou fichiers administratifs. Or les rythmes de croissance démographique très élevés et les modifications rapides du tissu urbain nécessitent des observations répétées de ce milieu complexe à observer.

(1) L'équipe est composée de : -Département D : F. DUREAU, O. BARBARY, A. MICHEL, A. CHATELAIN (chercheur associé, EHESS Marseille)
 - Atelier de Télédétection : B. LORTIC, I. RANNOU
 - Unité d'infographie : M. SOURIS, F. PELLETIER, C. ANTIGEOL

Mais les méthodes classiques de collecte des données démographiques (recensements et enquêtes par sondage) ne permettent pas de réaliser une observation suivie et spatialisée des villes des pays en développement.

Les recensements exhaustifs exigent en effet des moyens techniques, financiers et humains considérables qui interdisent une périodicité inférieure à une dizaine d'années ; de plus, le temps nécessaire à leur exploitation (trois ans environ) rend souvent les résultats caducs dès leur parution pour des villes dont les taux de croissance peuvent dépasser 10 % par an.

L'ampleur des moyens à mettre en oeuvre pour un recensement conduit à une sous-exploitation des données recueillies exhaustivement : exploitation par sondage des questionnaires, ou exploitation d'une partie seulement des variables du questionnaire, publication des seuls résultats globaux sur l'ensemble de la ville sont les solutions les plus couramment retenues pour alléger l'exploitation de l'information collectée .

Enfin, notons que l'absence de cartographie de base dans nombre de villes et les problèmes de suivi de la masse d'enquêteurs que réclame ce type d'opération ont des répercussions importantes sur la qualité des résultats.

Quant aux enquêtes par sondage, elles présentent comme intérêt de réduire les moyens financiers et humains à mobiliser, et d'assurer une meilleure qualité de la collecte, par des enquêteurs moins nombreux et mieux encadrés. Mais un facteur essentiel fait généralement défaut pour l'applicabilité de cette méthode dans les villes des pays en développement : *l'absence de base de sondage complète et à jour*, condition sine qua non de tout tirage d'échantillon statistiquement représentatif.

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

L'efficacité de la morphologie urbaine comme vecteur de l'observation démographique

Face à ce constat rapide des limites des méthodes classiques d'observation démographique dans les villes des pays en développement, que proposer pour tenter d'améliorer l'observation des populations urbaines ?

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

Une meilleure connaissance de l'espace intra-urbain nous semble être un élément fondamental d'amélioration des techniques d'enquête par sondage. D'une

part, une bonne connaissance cartographique permet de disposer d'une base de sondage complète et à jour, composée d'îlots précisément identifiés; d'autre part, une bonne connaissance de la morphologie urbaine permet de stratifier la ville en maximisant le gain de précision qu'introduit toute stratification, et d'obtenir des résultats selon un découpage spatial significatif de la ville enquêtée.

L'idée d'utiliser les éléments de connaissance sur la morphologie urbaine à des fins démographiques n'est pas nouvelle : dès 1969, M. VERNIERE (1), géographe à l'ORSTOM, a utilisé des photographies aériennes pour estimer la population de Pikine, banlieue de Dakar. Partant de l'hypothèse qu'il existe des relations entre les caractéristiques morphologiques du milieu urbain et les caractéristiques démographiques et socio-économiques des habitants, le principe de sa méthode est d'exploiter l'information exhaustive fournie par les couvertures aériennes pour recueillir rapidement, par sondage, des données relatives aux populations urbaines.

Les différents travaux menés en appliquant cette méthode ont mis en évidence que cette technique de sondage spatial permet un allègement de l'échantillon d'enquête et une spatialisation des résultats. Nombreux sont les bureaux d'étude qui ont appliqué la méthode de VERNIERE pour obtenir, rapidement, des estimations démographiques sur les villes dont ils avaient à réaliser un plan d'aménagement, ou des programmations d'équipement ; tout récemment, l'IAURIF a testé cette méthode pour l'estimation de caractéristiques socio-économiques des habitants des différents quartiers de Niamey (2).

L'utilisation de l'imagerie satellitaire pour renseigner sur la morphologie urbaine

Toutes les expériences d'observation démographique à partir de photographies aériennes ont clairement montré l'efficacité de la morphologie urbaine comme vecteur pour l'observation des populations résidant en ville. Mais les couvertures aériennes des villes des PED deviennent de plus en plus rares en raison de l'ampleur des moyens techniques, financiers et humains qu'elles réclament : la répétitivité de l'observation aérienne devient une réalité de moins en moins effective. Par contre, elle est offerte par les nouveaux satellites, qui ont maintenant une résolution suffisante (3) pour observer assez finement le tissu urbain.

Les images satellites présentent des qualités indéniables : répétitivité, exhaustivité spatiale, coût raisonnable, caractère numérique des données, richesse de l'information enregistrée (4). Il semble donc nécessaire de tester dès à présent les possibilités de mise en oeuvre de méthodes d'observation démographique intégrant les données morphologiques observables sur image

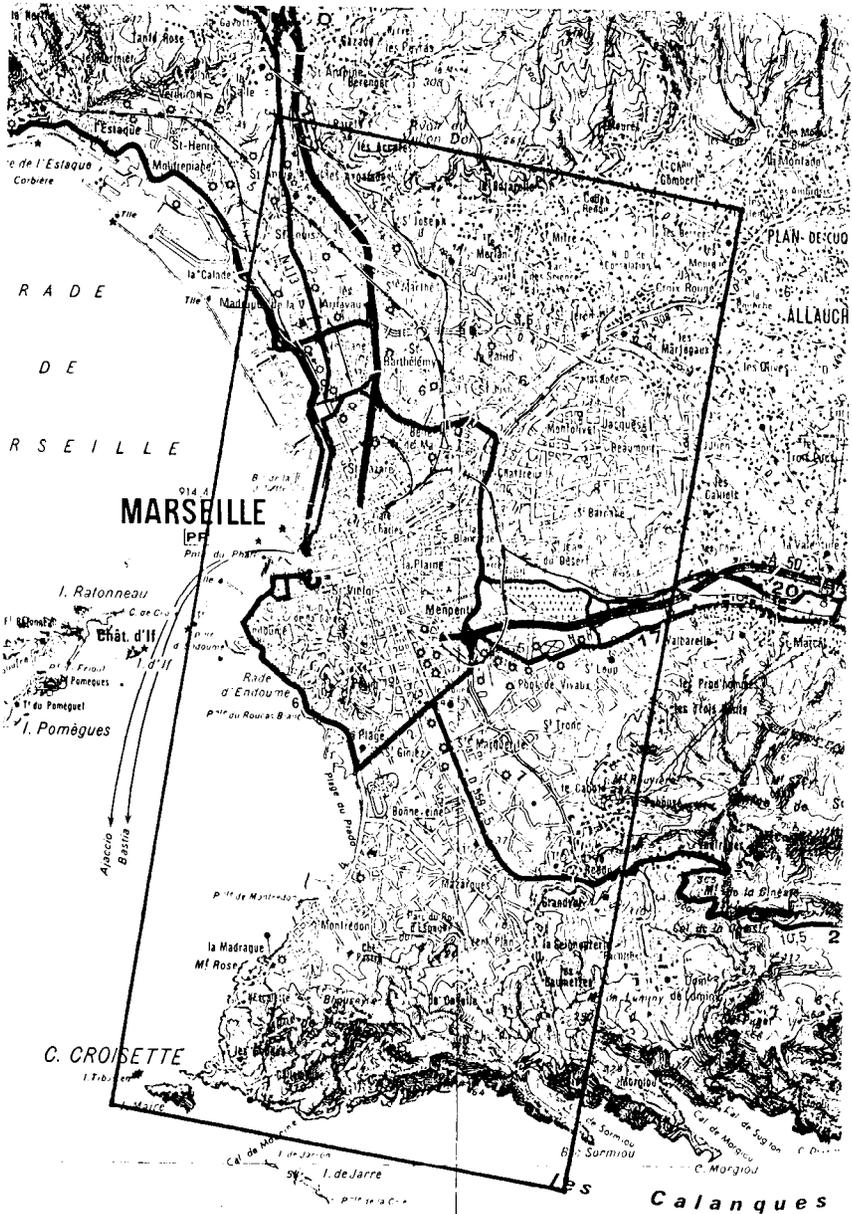
(1) Voir l'exposé de la méthode dans VERNIERE, 1978.

(2) Voir IAURIF, ROC, SIDAU, SEE, 1983.

(3) Résolution de 30 mètres pour Thematic Mapper, et résolution de 20 et 10 mètres pour SPOT.

(4) La vision stéréoscopique deviendra possible avec SPOT.

Figure 1 - Localisation de la zone d'étude sur Marseille



satellite ; mais, si l'on peut s'appuyer sur l'acquis méthodologique constitué par les travaux antérieurs réalisés sur photographies aériennes, il n'en demeure pas moins que l'interprétation des images satellites en milieu urbain est tout à fait nouvelle, et suppose donc une phase préalable de mesure de fiabilité du document et de développement des méthodes d'analyse spécifiques.

L'introduction de la télédétection spatiale dans un système d'acquisition de données démographiques en milieu urbain constitue un champ nouveau de recherches (1), juste balisé par les travaux utilisant les photographies aériennes, et quelques études de télédétection urbaine exploitant des données LANDSAT ou des simulations SPOT.

C'est pourquoi, dans le cadre de ce programme de recherche, plusieurs thèmes doivent faire l'objet de travaux approfondis :

- *L'interprétation des images satellite à haute résolution en milieu urbain* : définir quelles sont les informations relatives à la morphologie urbaine contenues dans ces images et quels sont les traitements permettant d'extraire ces informations.

- *Les relations bâti/population et les sondages spatiaux en démographie urbaine* : rechercher quels sont les critères morphologiques pertinents pour l'observation démographique, et quelle stratégie de sondage aréolaire est la plus efficace pour collecter des données démographiques spatialisées en milieu urbain.

- *L'intégration des données de télédétection dans des bases de données localisées*, ayant pour objectif une observation suivie et spatialisée des villes dans leurs différentes composantes, notamment morphologiques et démographiques.

Ces axes de recherches sont menés depuis Janvier 1985 sur la ville de *Marseille* : étant donné le caractère tout à fait exploratoire du projet, il importe d'avoir des références fiables pour tester la validité des méthodes mises au point au fur et à mesure de l'avancement des travaux.

C'est pourquoi nous avons choisi de travailler d'abord sur une ville française disposant d'une image Thematic Mapper, et d'un recensement récent, puis sur une ville d'un pays en développement disposant également d'un recensement récent, et d'informations localisées sur la morphologie urbaine : il s'agira de *Quito* (EQUATEUR), où nous réaliserons, dans le cadre de l'Atlas de Quito(2), une enquête démographique par sondage spatial, dont le plan sera issu d'une stratification de la ville sur image satellite.

(1) Après une recherche bibliographique, je n'ai trouvé que deux études sur cette question. L'une, réalisée par LO et WELCH, 1977, expose une méthode d'estimation de la population des grandes villes chinoises à partir des données du satellite LANDSAT.

(2) Il s'agit du programme Atlas Permanent Informatisé de Quito, auquel participent plusieurs Unités de Recherches du Département D.

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

2 . LES TRAVAUX METHODOLOGIQUES SUR MARSEILLE

La zone d'étude

La zone d'étude (voir localisation sur la Figure 1) a été déterminée de manière à *couvrir les différents types de quartiers de Marseille, soit schématiquement : les quartiers industriels et de grands ensembles ouvriers du nord, les quartiers résidentiels du sud, le vieux centre, et les zones pavillonnaires de l'est de la commune.*

Tous les travaux méthodologiques ne sont pas réalisés sur l'ensemble de cette zone de Marseille, telle qu'elle apparaît sur la Figure 1. En effet, il est nécessaire d'en *extraire un échantillon pour la mise au point de méthodes* telles que celles relatives à l'interprétation d'image satellite : si l'on ne procédait pas ainsi, on ne pourrait pas vérifier la validité des traitements mis au point, c'est à dire tester qualitativement et quantitativement leur efficacité. *Travailler sur un échantillon permet de contrôler en permanence les méthodes développées*, en les testant sur un espace disposant d'une information de référence, les données de population et de morphologie urbaine dans le cas présent.

Pour que cette méthode de travail soit vraiment efficace, il importe de disposer d'un échantillon représentatif des différents types d'occupation du sol présents dans l'ensemble de la zone.

Le plan de recherche

Sur Marseille, les trois directions de recherche évoquées plus haut sont menées en parallèle, leurs résultats se conditionnant mutuellement. Le principe de l'expérimentation menée sur Marseille est résumé dans la Figure 2 ; on notera que dans cette première phase du programme de recherches, purement méthodologique, *les travaux réalisés reposent sur trois types d'information : données démographiques issues du recensement de 1982, caractéristiques morphologiques observées sur photographies aériennes à grande échelle et au sol, et image Thematic Mapper*

Pour cette première étape, il n'est pas réalisé de collecte de données démographiques d'après un sondage aréolaire bâti sur image satellite, comme ce sera le cas à Quito. Le recensement de 1982 est utilisé de deux façons : par sondage, pour modéliser les relations bâti/population et simuler l'enquête démographique sur échantillon aréolaire, et exhaustivement, comme référence pour mesurer le degré de précision des résultats obtenus en mettant en oeuvre différentes stratégies de sondage.

Les différents axes de recherche de ce programme impliquent la mise en relation de données localisées issues de sources d'information variées, et de nombreux traitements de données géographiques : à travers quelques exemples tirés de l'expérience en cours sur Marseille, nous verrons quel peut être l'apport d'un système d'information géographique.

Mise en relation de l'information satellitaire et des données sur la morphologie urbaine

En raison de la résolution insuffisante des images LANDSAT, seules disponibles jusqu'à ces dernières années, la télédétection est restée très peu employée en milieu urbain, caractérisé par une forte hétérogénéité et la faible taille des éléments constitutifs ; l'analyse précise de l'information apportée par les capteurs à haute résolution comme Thematic Mapper et SPOT, et les méthodes d'extraction de cette information n'en sont qu'à leurs débuts.

Dans l'état actuel de la télédétection en milieu urbain, il est primordial de pouvoir confronter spatialement les données satellitaires avec les données sur la morphologie urbaine issues, dans le cas de Marseille, d'observations terrain et de photographies aériennes à grande échelle.

En effet, la mise en relation avec les données localisées gérées par le système d'information géographique s'avère nécessaire à plusieurs niveaux :

- sélection d'un échantillon spatial de l'image représentatif des types morphologiques présents dans la zone d'étude,
- confrontation avec les références de terrain relevées sur cet échantillon, pour développer des méthodes d'interprétation de l'image satellite,
- confrontation avec les données morphologiques connues sur l'ensemble de la zone d'étude afin d'évaluer précisément les résultats obtenus en appliquant les méthodes d'interprétation mises au point sur l'échantillon,
- test de la qualité géographique des recalages d'image satellite, en comparant les images recalées aux données digitalisées à grande échelle.

A l'heure actuelle, étant donné la complexité du milieu étudié et l'état des connaissances en télédétection urbaine, ne pas recourir à une base de données gérée par un système d'information géographique limiterait sérieusement les interprétations de l'image satellite, tant sur le plan thématique que sur le plan spatial.

Mise en relation des données démographiques et des données sur la morphologie urbaine

La mise au point d'un système d'observation démographique intégrant l'information satellitaire suppose de répondre à un certain nombre de questions :

- Quels sont les critères morphologiques pertinents pour établir la stratification spatiale d'une ville en vue d'une enquête démographique par sondage, et quels sont les critères révélant directement des phénomènes démographiques (tels que la dynamique migratoire, par exemple) ?

- Quelle est la stratégie de sondage la plus efficace? Ce qui suppose de définir la taille et le nombre de strates qui maximisent le gain de précision, le type d'unité spatiale à enquêter, le mode de tirage de l'échantillon (aléatoire ou systématique, selon quelle fréquence spatiale).

Pour apporter des éléments de réponse à ces questions, il est nécessaire de mettre en relation et traiter conjointement des données démographiques et morphologiques localisées, c'est à dire de réaliser principalement :

- la jointure entre données démographiques et morphologiques, afin de connaître la répartition de la morphologie selon le découpage géographique des données démographiques, et inversement. Cette information sert, notamment, à tirer un échantillon de districts INSEE représentatif des types morphologiques et à étudier les relations bâti/population par des méthodes d'ajustement,
- des sélections et statistiques spatiales selon les modalités des attributs morphologiques ou démographiques,
- une étude du mode d'implantation spatiale des données afin de déterminer quelles hypothèses peuvent être faites sur leur répartition dans l'espace et donc quelles sont les méthodes d'échantillonnage à tester.
- une évaluation spatialisée et quantifiée des résultats obtenus en simulant différentes stratégies de sondage.

L'utilisation d'un système d'information géographique est fondamentale dans la première phase de mise au point d'un système d'observation démographique utilisant l'information satellitaire sur la morphologie urbaine. Mais il faut souligner également que, de par ses qualités (continuité spatiale et temporelle de l'observation, caractère numérique des données) la télédétection est une source d'information privilégiée pour l'actualisation des données d'un observatoire urbain géré par un système d'information géographique. Il s'agirait non pas d'intégrer l'information satellitaire brute, par pixels, mais d'actualiser les données de la base à partir des informations sur l'occupation du sol extraites des images satellite.

3. L'ANALYSE DES INFORMATIONS A INTEGRER DANS LA BASE DE DONNEES

Comme dans toute étude, il importe d'analyser précisément l'information disponible, afin de déterminer un schéma d'exploitation optimal des données (1) qui soit cohérent avec leur nature et leur degré de fiabilité.

Nous insisterons ici sur l'analyse des caractéristiques thématiques, statistiques et surtout spatiales des données, qui prend une dimension particulière dans le cas du traitement informatisé de données localisées : c'est en effet cette analyse qui permet de préciser le schéma conceptuel des données, donc les bases de leur informatisation et les traitements qui pourront être réalisés.

(1) Puisque déjà médiatisée, toute donnée peut être considérée comme une information.

L'image Thematic Mapper

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

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

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

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

Le recensement de population de 1982

Le recensement, réalisé en Mars 1982 sur l'ensemble du territoire français, a donné lieu à différents types de résultats, en fonction des étapes du schéma d'exploitation des questionnaires :

- Le sondage au 1/20 : première exploitation destinée à une publication rapide de résultats, selon un découpage géographique grossier (département, avec distinction urbain et rural). Le sondage est réalisé en tirant une feuille de logement sur quatre (échantillon au 1/4), puis une feuille de logement sur cinq au sein de cet échantillon.
- Le sondage au 1/4 : correspond à l'exploitation de toutes les feuilles de ménage de l'échantillon au 1/4 tiré précédemment. Les résultats de cette exploitation au 1/4 sont statistiquement valables au niveau du quartier. Une ville comme Marseille étant divisée en 111 quartiers, dont la délimitation ne repose pas sur des critères de type de tissu urbain, ou de densité, ce découpage ne peut être pertinent pour notre application.

Tableau 1 - Données individuelles saisies pour les résultats exhaustifs du recensement de 1982 (INSEE)

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

-L'Exhaustif : les résultats exhaustifs sont issus de la fusion d'un extrait des variables du fichier du 1/4, et du fichier provenant de la saisie des 3/4 des bulletins collectés restants. La saisie des 3/4 n'est que partielle : ne sont saisies que quelques unes des réponses portées sur les questionnaires.

Des deux pages du questionnaire que chacun de nous est censé avoir rempli pour le recensement, ne sont saisies sur les 3/4 des questionnaires, et donc exploitées pour l'exhaustif, que les informations mentionnées dans le Tableau 1 ; on mesure donc l'importante réduction d'information qui s'opère maintenant dans les recensements de population, comme nous le signalions en première partie de cet article. Ainsi, à ce stade de l'exploitation, on perd toute information sur une notion aussi primordiale que la catégorie socio professionnelle ; mais, cette exploitation exhaustive étant la seule qui permette de travailler à un niveau géographique suffisamment fin, on est dans l'obligation d'y recourir pour toute étude démographique précise en milieu urbain.

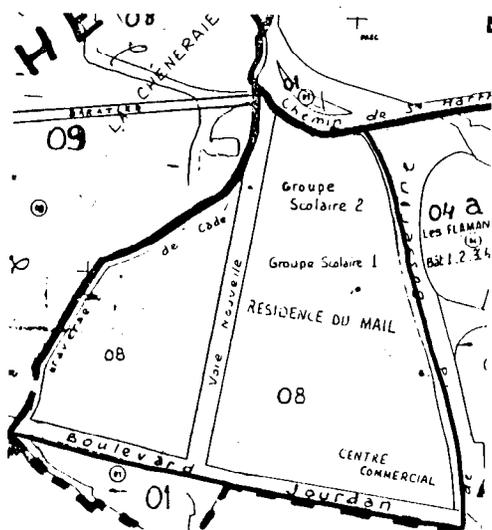
C'est donc sur les résultats exhaustifs que nous travaillons sur Marseille. Nous disposons de la bande magnétique des résultats exhaustifs sur la commune, qui comprend, outre l'effectif de population, les répartitions de cette population selon les variables citées dans le Tableau 1, prises indépendamment ou en les croisant.

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

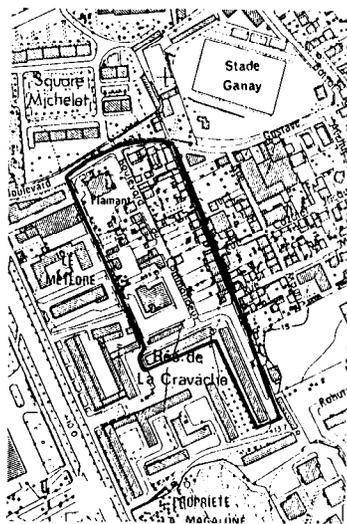
- *Un critère administratif*, lié à la fonction première du recensement qui doit fournir des chiffres de population légale pour les différentes unités territoriales de France : ainsi, les districts de recensement doivent respecter les limites de sections cadastrales, de cantons et de communes.

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

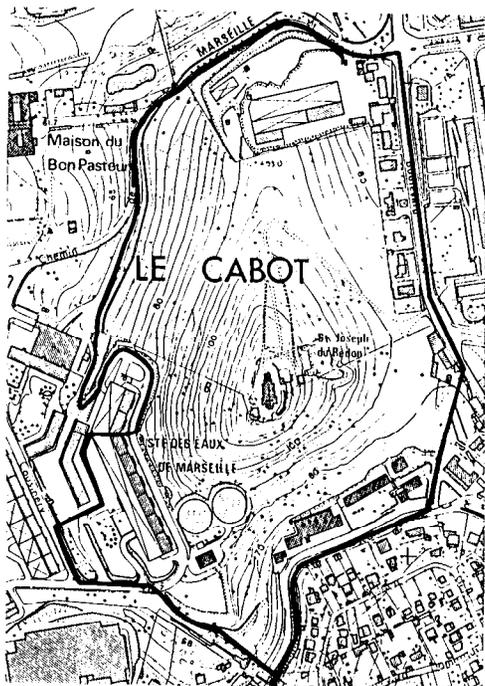
Figure 3 - Marseille . Découpage INSEE en districts de recensement



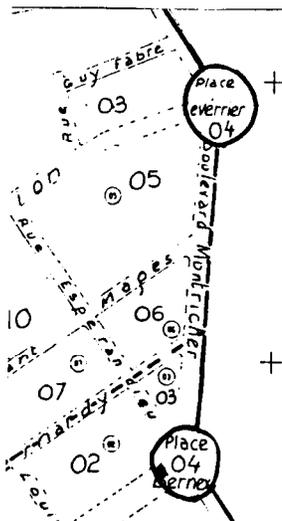
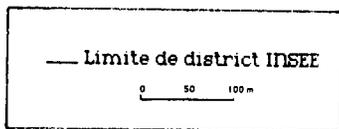
Exemple 1



Exemple 2



Exemple 3



Exemple 4

personnes.

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

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

La figure 3 illustre quelques unes de ces divergences:

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

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

. La morphologie urbaine

La connaissance de la morphologie urbaine sur l'ensemble de la zone d'étude repose sur *l'analyse visuelle d'une mosaïque de photographies aériennes au 1/23000* (mission IGN d'Août 1982), avec contrôle terrain . Le choix de cette source d'information a été dicté par deux contraintes : avoir une information homogène sur l'ensemble de la zone d'étude, et la plus proche de la date des deux autres sources de données, recensement et image Thematic Mapper. Ces deux conditions ne sont réunies que dans le cadre de la couverture systématique du littoral effectuée par l'IGN en 1982 ; l'agglomération urbaine n'étant généralement pas une notion retenue par l'IGN pour ses missions aériennes, les autres couvertures aériennes à grande échelle réalisées sur Marseille depuis une dizaine d'années ne couvrent qu'une partie de la commune.

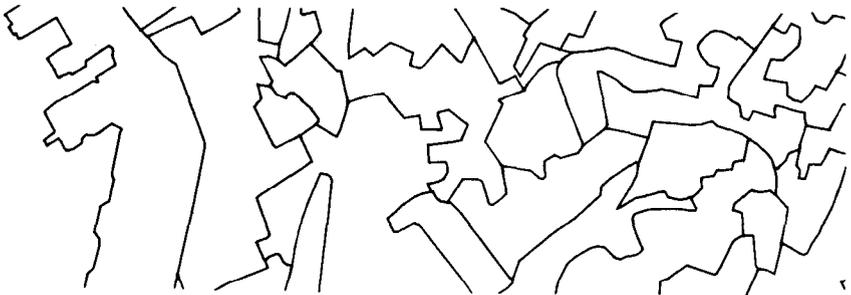
L'analyse de la mosaïque photographique a été réalisée, classiquement, en deux étapes : zonage en unités homogènes, puis description des 542 zones ainsi dégagées à l'aide de onze caractères morphologiques.

Figure 4 - L'implantation géographique des trois types de données à intégrer dans la base de données sur Marseille



Découpage INSEE en districts de recensement

0 250 500 m



Zonage morphologique

0 250 500 m

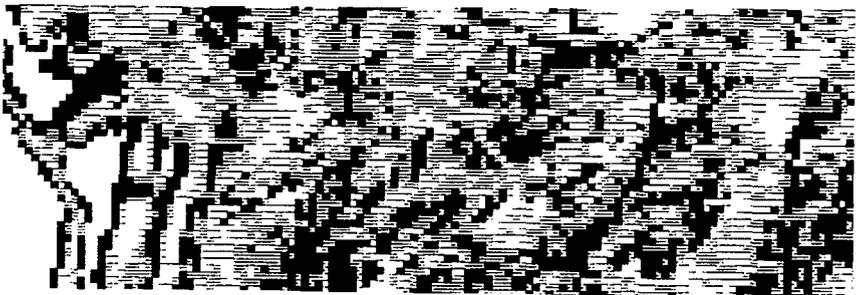


Image Thematic Mapper

0 250 500 m

Les traitements d'analyse des données sur le fichier descriptif des zones et les vérifications sur le terrain ont permis de définir une *typologie morphologique en huit postes*. Après différents essais de traitements (analyse des correspondances, classification ascendante hiérarchique, nuées dynamiques), il s'est avéré qu'une méthode de classification efficace par rapport à notre objectif consiste à retenir dix des variables observées sur photographies aériennes et à appliquer à ce fichier la chaîne de traitements suivante : un tri éliminant toutes les zones non bâties, suivi d'une première classification sur l'ensemble des zones bâties, puis nouvelles classifications indépendantes sur chacune des deux grandes classes définies par le premier passage, en faisant varier les coefficients de pondération des différentes variables.

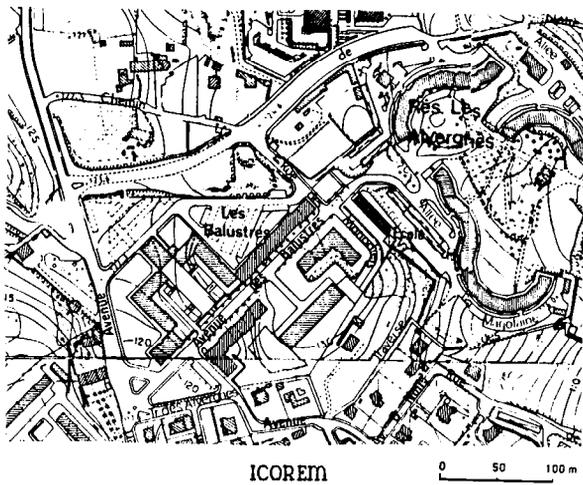
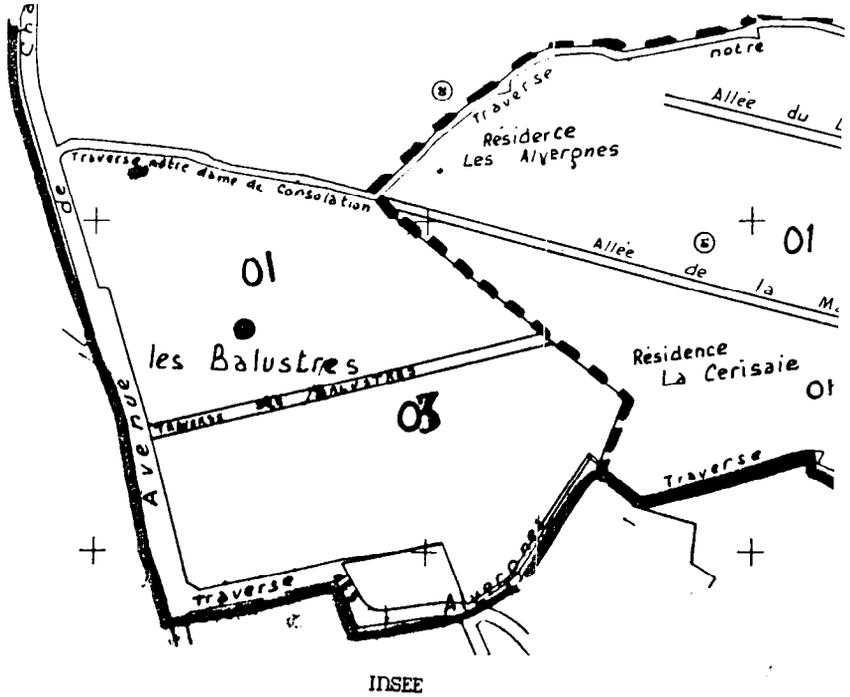
A l'issue de l'analyse visuelle et des traitements d'analyse des données, nous disposons donc d'une information homogène sur l'ensemble de la zone d'étude, décrivant la morphologie urbaine en 1982, de façon synthétique (typologie en 8 postes), et analytique (onze descripteurs pour chacune des zones)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

Comme le montre la Figure 4, *les trois types de données localisées à intégrer dans la base s'inscrivent dans des découpages géographiques différents*, ce qui justifie le recours au système TIGER pour les mettre en relation et les gérer. En effet, un logiciel de cartographie automatique classique ne pourrait effectuer des traitements sans avoir ramené toutes les données à un découpage géographique unique, ce qui ferait perdre tout intérêt au croisement de données localisées tel que nous le concevons dans ce programme.

Ainsi que l'analyse des informations à intégrer dans la base de données sur Marseille l'a mis en évidence, *ces différents découpages résultent bien sur du mode d'implantation spatiale propre à un phénomène, mais aussi de la méthode de collecte et d'exploitation de l'information décrivant ce phénomène*. Ce deuxième facteur déforme nécessairement la réalité des faits que l'on cherche à observer : les données à intégrer dans la base correspondent à un certain niveau de médiatisation, donc à une certaine modélisation du réel, que nous maîtrisons plus ou moins selon la source d'information dont elles sont issues. Cette modélisation est complétée par les choix qui sont faits au niveau des modalités d'intégration des données dans la base, notamment les choix de l'échelle de saisie, du découpage géographique, des relations et des attributs (1).

(1) Terminologie employée dans les bases de données relationnelles (voir article de M. SOURIS, p. 66).

Figure 5 - Comparaison de la cartographie INSEE des districts de recensement et d'un plan ICOREM



4 . L'INTEGRATION DES DONNEES DANS UNE BASE GERE PAR LE SYSTEME TIGER

L'intégration des données dans la base comporte trois phases : la préparation des fonds cartographiques avant numérisation, la numérisation de ces fonds, et l'intégration de l'information thématique s'y rapportant. Nous ne traiterons en détail que du premier point à propos de l'information démographique et morphologique. La phase de numérisation est décrite dans l'article de M. SOURIS (cf p.) ; elle ne soulève pas de problèmes méthodologiques pour l'utilisateur du système TIGER, tout comme l'intégration de l'image satellite puisque le système TIGER accepte des données maillées.

. PREPARATION DES FONDS CARTOGRAPHIQUES A NUMERISER

L'intégration des fonds de cartes dans la base gérée par le système TIGER se fait par numérisation sur table à numériser BENSON : le principe est de suivre avec un pointeur le tracé des limites de zones, et d'enregistrer à l'aide du pointeur les coordonnées des points situés le long des limites que l'on souhaite conserver dans la base. Il est bien évident que le tracé enregistré correspond au tracé présent sur le document soumis à numérisation : il importe donc de soumettre à la numérisation des fonds de cartes corrects, et remplissant les conditions nécessaires à l'exploitation qui en sera faite.

Dans le cas de Marseille, la précision des fonds de carte est d'autant plus importante que l'on veut croiser à grande échelle les différentes informations, et que les limites de zones sont significatives et objets d'étude : par exemple, nous utiliserons le tracé de la voirie, issu de la carte des limites d'îlots, pour tester la qualité géométrique du recalage des images satellite.

Trois paramètres doivent être considérés : la précision du tracé sur le document à numériser, l'échelle de numérisation, et le système de projection géographique de la carte à numériser. A travers l'expérience de Marseille, nous évoquerons successivement ces trois points et les questions qu'il soulèvent.

. La précision du tracé sur le document à numériser

La figure 5 montre un exemple de la *cartographie INSEE* : il est clair que l'on ne peut numériser directement ce type de document, au tracé très approximatif. La confrontation d'un extrait de carte INSEE et d'un extrait d'un plan correct (représentant la même zone de Marseille) met en évidence l'ampleur des déformations présentes sur les cartes INSEE. Inexploitables directement, les fonds de carte établis par l'INSEE doivent donc être redessinés avant d'être numérisés.

La commune de Marseille est presque entièrement couverte par une cartographie à grande échelle (1/5000 et 1/2000), réalisée par l'ICOREM à partir de la fin des années soixante-dix : l'ICOREM ayant développé un système informatisé de gestion de données cartographiques et urbaines, cette société assure une mise à jour constante de ses fichiers géographiques en fonction des relevés de terrain faits

par ses nombreux enquêteurs. Les cartes éditées par l'ICOREM renseignent très précisément sur l'occupation du sol, comme le montre l'exemple de la Figure 5 : les méthodes de travail utilisées pour établir ces cartes nous assurent de leur très bonne qualité.

C'est donc sur les cartes ICOREM que nous avons reporté le tracé des 3354 districts INSEE composant notre zone d'étude marseillaise, ce qui nous a donné une série de fonds de cartes géographiquement corrects, prêts pour la numérisation.

Le tracé des limites INSEE sur les plans ICOREM a nécessité plusieurs mois de travail particulièrement fastidieux . Toute personne devant exploiter cartographiquement le recensement se trouve confrontée au même problème : on voit que cela limite sérieusement les analyses du recensement et l'on conçoit aisément que, dans ces conditions, même les agences d'urbanisme s'en tiennent à une exploitation du recensement par arrondissements ou, au mieux, par quartiers. Outre cet effet des déficiences de la documentation cartographique relative aux recensements , il faut insister sur l'intérêt que l'INSEE pourrait trouver à utiliser ces documents ICOREM lors de la phase de cartographie qui précède tout recensement ; comme le dit très justement R. CLAIRIN, "*On ne le répètera jamais assez, si on ne dispose pas d'une cartographie adéquate, la collecte démographique se fait dans le brouillard*" (1). L'INSEE dispose à Marseille d'une arme pour lever ce brouillard, on comprendrait difficilement que cet institut persiste à ne pas l'employer.

Quant au *zonage morphologique* dont on a vu qu'il a été réalisé sur une mosaïque de photographies aériennes non contrôlée, il comporte par conséquent de nombreuses déformations géométriques, et ne correspond à aucune projection cartographique . Le zonage ne pouvait donc pas être digitalisé sous cette forme. C'est pourquoi, avant numérisation, il a été nécessaire de reporter ces zones sur un fond de carte géographiquement correct : nous avons choisi la carte IGN au 1/25 000, dont l'échelle est proche des documents photographiques utilisés pour ce zonage morphologique, et qui comporte les principaux éléments du réseau de voirie.

Enfin, notons que *l'image Thematic Mapper* a été recalée géographiquement par l'IGN , selon une projection LAMBERT III : ce recalage s'appuie sur une dizaine d'amers (points caractéristiques de la surface terrestre bien identifiés sur l'image satellite et localisés sur une carte topographique à grande échelle). L'opération de recalage est un préalable nécessaire avant toute intégration de données issues de la télédétection dans un système d'information géographique. C'est pourquoi nous nous proposons de tester sur Marseille la qualité de différentes procédures de recalage d'image satellite.

(1) R. CLAIRIN, 1981. *Quelques réflexions sur la collecte démographique*. STATECO n°25.

Cité par G. THEODORE, 1985.

L'échelle des documents à numériser

L'échelle de numérisation est un paramètre important à considérer : même si, techniquement, le système TIGER permet de créer des images à n'importe quelle échelle à partir d'un fond numérisé à une échelle donnée, *l'échelle de numérisation est un facteur limitant qu'il faut prendre en considération, pour deux raisons :*

- *l'une est d'ordre purement cartographique* : le degré de précision d'un tracé cartographique est directement lié à son échelle. Plus on réduit l'échelle, plus le tracé est volontairement simplifié : les cartographes procèdent ainsi, manuellement ou automatiquement maintenant, à une "généralisation" du tracé, en ne retenant que les éléments majeurs de celui-ci. Agrandir une carte dans des proportions importantes par rapport à son échelle initiale comporte donc le risque d'obtenir une restitution très simplifiée, et même fausse si le rapport d'agrandissement est exagéré. On peut toujours réduire l'échelle d'une carte, à l'aide d'une procédure de généralisation automatique ou non ; par contre, l'ordinateur ne peut inventer des détails de tracé à partir d'un tracé simplifié.

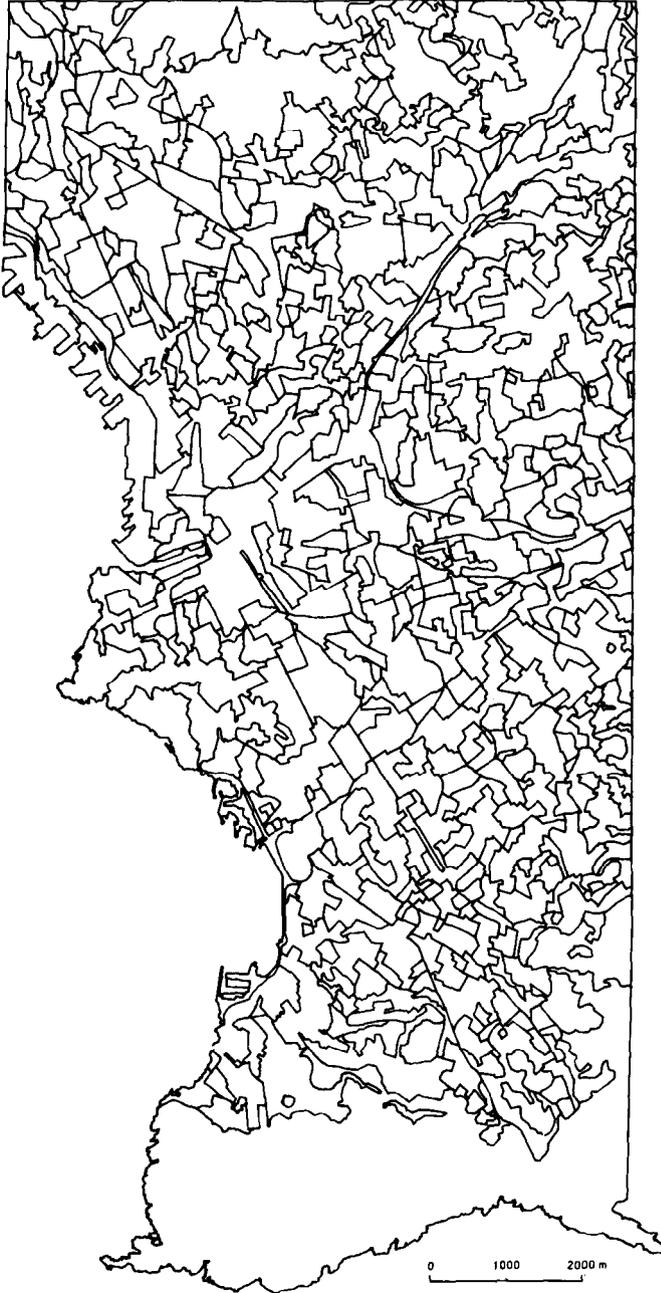
- *l'autre est d'ordre conceptuel* : toute information a un domaine de validité propre, que l'on ne doit pas dépasser. Ainsi, une typologie de l'occupation du sol est directement liée à l'échelle à laquelle elle a été établie et ne sera pas la même si on la bâtit au 1/1000 ou au 1/40000 : un changement d'échelle n'est pas une simple opération géométrique, mais influe directement sur la nature des phénomènes observés. En milieu urbain où un faible changement d'échelle modifie largement l'objet d'observation et la perception de l'organisation spatiale, il importe d'être particulièrement attentif à l'échelle de collecte, ou d'élaboration, de l'information.

Sur Marseille, les données démographiques issues du recensement correspondent à une cartographie numérisée au 1/5000 (et 1/2000 sur une petite partie de la zone d'étude, non couverte par les cartes COREM au 1/5000) ; quant aux données relatives à la morphologie urbaine, leur fond de carte a été saisi au 1/25.000.

L'écart entre les échelles de saisie des deux informations n'est pas négligeable (rapport de 1 à 5 dans la majeure partie de la zone d'étude, et de 1 à 12 parfois) : dans l'opération de jointure entre zones INSEE et zones morphologiques, il est nécessaire de prendre en considération cette différence d'échelle de saisie, qui se traduit par une imprécision relative plus importante dans le tracé des zones morphologiques que dans celui des zones INSEE.

D'autre part, comme nous le disions plus haut, il existe une *limite conceptuelle, qui nous interdit d'utiliser sans précaution une donnée à une échelle trop différente de son échelle d'établissement* : il faut bien être conscient qu'avoir réalisé le zonage morphologique sur une mosaïque photographique au 1/23000 induit directement un certain type de zonage et un ensemble de descripteurs, différents de ceux que l'on aurait pu retenir en travaillant à une autre échelle (au 1/500, par exemple, on aurait perdu la notion du tissu urbain pour ne plus observer que le bâtiment en tant qu'objet bin individualisé à cette échelle). Si l'on travaille à une très grande échelle, au niveau de l'îlot, les descripteurs et types de morphologie issus de l'analyse visuelle d'une mosaïque au 1/23000 sont, pour l'îlot, un attribut

Figure 6 - Marseille. Zonage morphologique d'après analyse visuelle d'une mosaïque photographique au 1/23000 de 1982



— Limite de zone morphologiquement homogène

d'appartenance à un type de tissu urbain, et non un attribut décrivant précisément la morphologie au sein de l'ilôt.

Le système de projection géographique des documents à intégrer dans la base

Nous évoquons ce deuxième point pour mémoire : dans la majeure partie des cas, cette question ne pose pas problème.

Le système TIGER conserve l'information géographique en coordonnées géographiques : après la numérisation, les coordonnées table sont transformées en coordonnées géographiques à partir des coordonnées géographiques connues pour les trois ou quatre points de calage saisis en même temps que le fond cartographique. Cette transformation suppose bien évidemment que l'on connaisse le système de projection du fond de carte numérisé.

Ce principe étant admis, il est clair que *l'on peut intégrer dans une base gérée par TIGER des informations géographiques correspondant à différents systèmes de projection.*

Dans le cas de Marseille, les districts INSEE sont cartographiés en projection LAMBERT III, puisque le tracé a été repris sur les plan ICOREM, qui sont tous en LAMBERT III. De même, le zonage morphologique est en LAMBERT III, puisque c'est la projection utilisée par l'IGN pour sa carte au 1/25 000. Quant à l'image satellite, l'IGN a effectué le recalage géographique selon une projection de même type, LAMBERT III.

Toutes nos informations géographiques concernant Marseille relèvent donc d'un même système de projection, LAMBERT III ; mais il s'agit d'un concours de circonstances, lié au fait que nous utilisons des documents de l'IGN ou se recalant sur les documents IGN, et non d'une obligation du système TIGER.

LA NUMERISATION DES FONDS DE CARTE

La numérisation des cartes des districts de recensement et du découpage morphologique a été réalisée par l'Unité d'Infographie, selon la procédure exposée par M. SOURIS dans son article. Etant donné la surface occupée par la cartographie des districts de recensement et le nombre de zones qu'elle comporte, la zone d'étude a été digitalisée par parties, en 26 feuilles différentes.

La Figure 6 présente le zonage morphologique numérisé au 1/25000, tel qu'il est restitué sur table traçante : sur cet exemple, ne figurent que les arcs (contours de zones).

L'INTEGRATION DE L'INFORMATION THEMATIQUE

L'information thématique relative aux fonds de carte numérisés constitue les "attributs", pour reprendre le vocabulaire des bases de données relationnelles :

dans le cas de Marseille, on peut distinguer *deux relations (morphologie et démographie), avec deux séries d'attributs, l'une correspondant au zonage morphologique, l'autre aux districts du recensement.*

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

- *morphologie urbaine* : vingt cinq variables décrivant analytiquement la morphologie, et un indicateur synthétique du type morphologique. Ces attributs renseignent sur la densité du bâti, la taille, l'élévation, la valeur et la forme des éléments bâtis, et la valeur du non bâti : tous les descripteurs sont issus de la photo-analyse.
- *démographie* : pour chaque district INSEE, neuf variables décrivant les nombres de logements et les effectifs de population selon différentes définitions liées à des critères de résidence.

EN GUISE DE CONCLUSION : QUELQUES REPERCUSSIONS METHODOLOGIQUES D'UN SYSTEME DE GESTION DE DONNEES LOCALISEES

A travers la présentation d'une expérience en cours sur Marseille, nous avons tenté de montrer quelques exemples d'utilisation d'un système d'information géographique, en l'occurrence le système TIGER développé par l'Unité d'Infographie. Mais l'on ne peut s'en tenir à cet aspect pragmatique de la question : un tel système ne sert pas qu'à automatiser des tâches manuelles fastidieuses de cartographie et statistiques spatiales.

Outre la multiplication des possibilités de traitement permise par l'informatique, un aspect essentiel fait que cet outil peut avoir des répercussions méthodologiques importantes : la localisation est au centre du système. Ce qui signifie qu'il ne s'agit pas de traiter statistiquement des unités géographiques par un logiciel statistique enrichi de possibilités graphiques, mais de *mettre en relation et traiter des informations localisées conservées chacune dans leur propre mode d'implantation spatiale*. Cette possibilité est essentielle : lever l'obstacle du découpage géographique unique pour croiser différentes données permet de ne plus abandonner ou même réduire une des composantes essentielles de l'information localisée, sa dimension géographique.

Cette possibilité a des effets directs sur les traitements qui peuvent être réalisés, et les raisonnements qu'ils induisent ; mais elle engendre aussi *une réflexion sur l'information utilisée, son niveau de médiatisation par rapport au phénomène que l'on étudie*. On est également conduit à analyser précisément tout le travail de production de données localisées de synthèse, visualisées classiquement sous forme de cartes : favorisant une décomposition de la démarche, un système d'information géographique peut donc avoir des répercussions importantes sur les méthodes de disciplines scientifiques productrices de données localisées de synthèse comme la géographie, la pédologie ...etc.

Enfin, ces systèmes peuvent dynamiser certains champs de recherche : citons notamment, dans le cas des études urbaines, l'étude des répartitions spatiales (régularités ou discontinuités) des citadins, et les recherches sur les effets même de la localisation, qui peut devenir une variable explicative pour comprendre les implantations des groupes sociaux, leurs modes d'appropriation de l'espace intra-urbain.

Les spécialistes de l'étude des phénomènes localisés, de leur répartition et de leurs relations que sont les géographes ne peuvent qu'être intéressés par un outil qui donne à l'espace une fonction de première importance : gageons que ceux-ci n'y verront pas qu'un outil sophistiqué, mais intégreront dans leur pratique scientifique les implications méthodologiques qu'il engendre, et en feront un facteur de renouvellement et d'avancée pour leur discipline.

ELEMENTS DE BIBLIOGRAPHIE

- BALLUT A. et GAUTHIER M., 1983. - Une méthode pour la production de données socio-économiques spatialisées dans le cadre d'un schéma directeur d'aménagement et d'urbanisme. IAURIF, Paris, 55 p..
- BRUGIONI D.A., 1983. - The Census : It can be done more accurately with space-age technology. Photogrammetric Engineering and Remote Sensing, Vol. 49, n°9, pp.1337-1339.
- CNRS, 1982. - Débats des journées télédétection en milieu urbain (6-7 mai 1982). CERG, Paris, 141 p..
- DEVAUGES R., 1982. - Fichiers informatisés, géomatique et recherche en sciences humaines. Bulletin de la Société Languedocienne de Géographie, Tome 16, Fasc. 1-2, pp. 187-196.
- DEVAUGES R. et GENDREAU F., 1979. - Les banques de données pour l'aménagement. Réflexions sur quelques expériences. La recherche urbaine à l'ORSTOM, Tome 1, Orientations et projets, pp. 123-130.
- DUREAU F. et GUILLAUME A., 1984. - La démographie depuis l'espace : un nouveau système d'observation ? STATECO, n° 38, pp. 5-46.
- HAKIM C., 1979. - Le recensement de la population et ses sous-produits : des bases de données pour la recherche. Revue Internationale des Sciences Sociales, Vol. 31, n° 2, pp.373-381.
- IAURIF, ROC, SIDAU, SEE, 1983. - Les données urbaines. Système permanent de production rapide et économique. Recherche et synthèse. Paris, 82 p.
- IGN, 1984. - Recherches méthodologiques sur l'utilisation de la télédétection spatiale en aménagement urbain. IGN, Paris, 19 p..
- INSEE, 1982. - Recensement général de la population de 1982. Guide d'utilisation. Tome 1 : Présentation générale, Organisation, Dénombrement, Exploitation statistique, Publications. INSEE, Paris, 95 p..
- LO C.P. et WELCH R., 1977. - Chinese urban population estimates. Annals of the Association for American Geographers, Vol. 67, n° 2, pp. 246-250.
- THEODORE G., 1985. - Similitudes et différences dans la méthodologie des recensements de population dans les pays industrialisés et les pays en développement. Paris, STATECO, n° 42, pp. 66-86
- VERNIERE M., 1978. - Méthode d'analyse quantitative de la croissance urbaine dans l'espace et dans le temps. Exemple d'une banlieue de Dakar (Sénégal). Photo interprétation, n° 1, pp. 34-55.