ATP CNRS/CNES - TELEDETECTION SPATIALE

INTEGRATION DES DONNEES DE TELEDETECTION DANS UN SYSTEME D'INFORMATION GEOGRAPHIQUE : SUIVI DE LA MORPHOLOGIE ET DE LA DEMOGRAPHIE D'UNE VILLE

RAPPORT INTERMEDIAIRE - AVRIL 1986

ORSTOM
Département D
Unité d'Infographie
Atelier de Télédétection Bondy

ATP CNRS/CNES - TELEDETECTION SPATIALE

INTEGRATION DES DONNEES DE TELEDETECTION DANS UN SYSTEME D'INFORMATION GEOGRAPHIQUE : SUIVI DE LA MORPHOLOGIE ET DE LA DEMOGRAPHIE D'UNE VILLE

RAPPORT INTERMEDIAIRE - AVRIL 1986

ORSTOM
Département D
Unité d'Infographie
Atelier de Télédétection Bondy

AVERTISSEMENT

Ont participé à la réalisation de ce rapport l'ensemble des membres de l'équipe ORSTOM :

- Département D : F. DUREAU, O. BARBARY, A. MICHEL
- Unité d'infographie : M. SOURIS, F. PELLETIER, I.C. ANTIGEOL
- Atelier de télédétection : B. LORTIC, I. RANNOU

* * *

Le présent rapport constitue un bilan intermédiaire des travaux réalisés par l'équipe de recherche ORSTOM sur le thème "Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville". Il s'agit de résultats provisoires à *ne pas diffuser sans l'autorisation des membres de cette équipe.*

R E S U M E

Ce programme de recherche a pour objectif général *d'élaborer un outil permettant le suivi de la morphologie et de la démographie d'une ville, grâce à un système d'information géographique intégrant les données de télédétection* . Les recherches ont débuté en 1985 par des travaux méthodologiques sur Marseille ; le deuxième site d'application sera Quito (EQUATEUR), où le programme constitue un volet particulier de l'Atlas Permanent Informatisé de Quito (APIQ).
L'équipe de recherche ORSTOM est composée de membres du Département D "Urbanisation et socio-systèmes urbains", de l'Unité d'Infographie, et de l'Atelier de Télédétection de Bondy (ATOB) ; en outre, nous collaborons sur Marseille avec le CAMS de l'Ecole des Hautes Etudes en Sciences Sociales, et sur Quito avec les institutions participant à l'APIQ (notamment, Institut Géographique Militaire et CLIRSEN).

Le rapport intermédiaire fait le point à l'issue d'une p*remière année essentiellement consacrée au site de Marseille* La réalisation de l'objectif général du programme nécessite des travaux approfondis dans *trois directions de recherche* (chapitre 1), *qui ont été menées parallèlement après une première phase consacrée à la constitution de la base de données* démographiques et morphologiques : intégration données de télédétection / système d'information géographique, télédétection spatiale en milieu urbain, relations bâti/population et sondages spatiaux en milieu urbain. Ce rapport, réalisé en grande partie à partir des notes de travail et articles rédigés par les différents membres de l'équipe (voir liste p. 8), rend compte de la constitution de la base de données et de l'avancement dans chacune des directions de recherche poursuivies *de façon parallèle* depuis 1985.

La zone d'étude sur Marseille correspond à un rectangle de 512 x 256 pixels de l'image TM 1983 non recalée, centré sur le centre ville et couvrant les différents types de quartiers de Marseille. C'est sur cette zone que nous avons constitué la *base de données* intégrant données morphologiques et démographiques et données de télédétection (chapitre 2); cette opération s'est étendue sur toute l'année 1985, en raison notamment des *difficultés posées par une exploitation précisément localisée des résultats du recensement de la population* (cartographie INSEE inutilisable dans sa forme initiale et délais d'obtention de la bande informatique des résultats par districts de recensement issus de l'exploitation exhaustive).
L'intégration des données démographiques a nécessité plusieurs mois de report des limites de districts sur des plans géographiquement corrects (plans ICOREM 1/2000 et 1/5000, projection Lambert III). Nous disposons maintenant dans la base des résultats par districts de recensement (9 variables décrivant analytiquement les effectifs de population et les nombre de logements), et des fichiers graphiques décrivant les contours des 3531 districts INSEE appartenant à la zone d'étude.
Les données morphologiques sont issues de l'analyse visuelle d'une mosaïque de photographies aériennes non contrôlée au 1/23000. Chacune des 542 zones d'égale apparence déterminées sur la mosaïque sont caractérisées par treize variables décrivant analytiquement la morphologie urbaine , et un indicateur synthétique du type morphologique. Cet indicateur est issu de *traitements d'analyse des données* (Analyse Factorielle des Correspondances et Classification Ascendante Hiérarchique) appliqués au tableau de caractérisation morphologique des 441

zones bâties : outre le fait qu'ils permettent d'établir une *stratification typologique de la zone d'étude marseillaise* (avec une quantification précise des contributions des paramètres morphologiques dans chaque strate, nécessaire pour la partie statistique de cette recherche), ces travaux d'analyse des données ont donné lieu à des *enseignements méthodologiques* utiles aussi bien en télédétection spatiale qu'aérienne.
Grâce aux données morphologiques intégrées dans la base, nous disposons d'une connaissance exhaustive et localisée de certains paramètres de la morphologie urbaine sur l'ensemble de la zone d'étude ; d'autre part, nous pouvons ainsi tirer un échantillon représentatif des types d'occupation du sol présents dans le périmètre d'étude marseillais, et tester l'efficacité d'une stratification morphologique dans l'estimation de la population par sondage (chapitre 4).

La base de données intégrant les données morphologiques et démographiques est gérée par le système TIGER, développé par l'Unité d'Infographie : il s'agit d'un système d'information géographique organisé suivant le schéma de gestion des bases de données relationnelles, enrichi d'opérations algébriques nouvelles pour la manipulation relationnelle de la localisation (chapitre 3).
Si les données graphiques sont structurées et stockées dans la base sous forme vectorielle, le système TIGER utilise pour la réalisation de certaines opérations une structure matricielle sous forme d'une image numérique de pixels associées à un fichier descriptif. Cette *double structure (stockage vectoriel et manipulation matricielle)*, associée au paramétrage de la résolution matricielle, confère au système une souplesse importante : elle permet d'intégrer facilement des données localisées extérieures au système et codées sous forme d'images numériques, telles que les images satellites.
Le seul problème technique pour mettre en relation ces images consiste à faire coïncider les paramètres des deux matrices de pixels : projection géographique, résolution du pixel, localisation dans le plan de projection. Les développements informatiques relatifs à la projection LAMBERT III, et au choix de la fenêtre d'étude par la taille du pixel permettent maintenant de *générer, par une requête sur le système TIGER, des images relatives à la morphologie urbaine avec une résolution donnée du pixel compatible avec la résolution TM* (30 m, 15 m ou 7,5 m par exemple) ; on peut également intégrer dans la base gérée par TIGER les données TM, ce qui nous a permis d'étudier la *concordance géométrique des données TM recalées par l'IGN en projection LAMBERT III et des données morphologiques* (digitalisées sur carte IGN 1/25000 projection LAMBERT III)
Les images de la morphologie générées par TIGER sont alors exploitées en relation avec les six canaux de l'image TM à l'Atelier de Télédétection, qui a développé depuis 1977 des programmes de visualisation, classification et restitution s'articulant autour d'une structure matricielle de fichiers. Nous commençons ainsi à développer de nouvelles méthodes d'analyse de l'image satellite TM : nous débutons par *l'analyse statistique des valeurs radiométriques par variable morphologique et par zone* qui permet de rechercher les diverses combinaisons linéaires de canaux TM qui offrent des taux de corrélation importants avec les différents attributs de la base décrivant la morphologie urbaine. L'analyse systématique de la bibliographie française et anglo-saxonne sur la télédétection urbaine (chapitre 5) montre l'intérêt de l'analyse de structure : l'intégration des données exogènes de la base permet d'envisager une *démarche zonale, permettant une meilleure appréhension de la notion d'agencement*

Si à l'heure actuelle le courant d'information vient plutôt du système TIGER vers le système de l'Atelier de Télédétection, il faut souligner qu'il s'agit d'une *phase transitoire* essentiellement due à des contingences matérielles : à terme, la mise en place définitive se faisant progressivement,le système TIGER réalisera l'analyse radiométrique et structurale, les classifications et utilisera les informations issues de la télédétection.

Outre les travaux réalisés sur Marseille, nous avons préparé la *deuxième phase du programme sur Quito (EQUATEUR)*(partie 2, chapitres 1 à 3), qui débutera en Juillet 1986 avec l'affectation sur place d'une partie de l'équipe ORSTOM, les autres membres de l'équipe fonctionnant avec des missions.
Une mission effectuée en Mai 1985 et les rapports fournis par les membres de l'équipe Atlas de Quito permettent de préciser les *caractéristiques de l'information existante* : couvertures aériennes à très grande échelle (1981 et 1984) couvrant l'ensemble de la ville, recensement de 1982 exploitable au niveau de la "manzana" (pâté de maisons), apportant de nombreuses informations tant sur le logement que sur les caractéristiques socio-économiques des habitants. En ce qui concerne les données de télédétection, nous allons commander l'enregistrement d'une image TM, et 4 scènes SPOT seront enregistrées entre Aout et Novembre 1986 (demande PEPS reprise par le CNES).
D'autre part, M. SOURIS a travaillé durant le second semestre 1985 à l'établissement d'une *proposition d'un système interactif graphique pour l'IGM*, répondant aux deux projets de cet institut membre de l'APIQ : automatisation de la photogrammétrie, et mise en place d'un service d'applications thématiques.

ATP CNRS/CNES - TELEDETECTION SPATIALE

INTEGRATION DES DONNEES DE TELEDETECTION DANS UN SYSTEME D'INFORMATION GEOGRAPHIQUE : SUIVI DE LA MORPHOLOGIE ET DE LA DEMOGRAPHIE D'UNE VILLE

RAPPORT INTERMEDIAIRE - AVRIL 1986

	Page
INTRODUCTION	5
Rappel des objectifs du programme de recherche	5
Le contexte institutionnel et financier	6
La composition de l'équipe ORSTOM	7
Liste des publications, notes de travail et rapports	8
PREMIERE PARTIE : LES TRAVAUX EN COURS SUR MARSEILLE	9
1. LA METHODOLOGIE DEVELOPPEE SUR MARSEILLE	10
1.1. LE CHOIX DE LA ZONE D'ETUDE	11
1.2. PRINCIPES DE L'EXPERIMENTATION	11
2. CONSTITUTION DE LA BASE DE DONNEES SUR MARSEILLE	15
2.1. ETABLISSEMENT D'UNE TYPOLOGIE DE LA MORPHOLOGIE URBAINE	16
2.1.1. Objectifs et données utilisées	16
2.1.2. Analyse visuelle de la mosaïque photographique au 1/23000	19
2.1.3. Analyses statistiques univariées et recodage des descripteurs morphologiques	25
2.1.4. Méthodes d'analyse des données appliquées au tableau de caractérisation morphologique des zones	35
2.1.5. La typologie en huit classes de la morphologie urbaine	63
2.1.6. Enseignements méthodologiques et perspectives pour la suite de l'expérimentation	76
2.2. ANALYSE DE L'INFORMATION A INTEGRER DANS LA BASE	80
2.2.1. Le recensement de population de 1982	80
2.2.2. L'information sur la morphologie urbaine	83
2.2.3. L'image Thematic Mapper	85
2.3. INTEGRATION DES DONNEES DANS UNE BASE GEREE PAR LE SYSTEME TIGER 87	
2.3.1. Préparation des fonds cartographiques à numériser	87
2.3.2. Digitalisation des fonds de carte	91
2.3.3. Intégration de l'information thématique	95
3. DONNEES DE TELEDETECTION ET SYSTEME D'INFORMATION GEOGRAPHIQUE	97
3.1. SYSTEMES D'INFORMATION GEOGRAPHIQUE ET BASES DE DONNEES	97
3.1.1.Généralités sur les systèmes d'information géographique	97

3. 1. 2. Bases de données et information géographique : l'extension du modèle relationnel 108
3. 1. 3. La mise en oeuvre des concepts théoriques 119

3 . 2. INTEGRATION DONNEES DE TELEDETECTION / BASE DE DONNEES GEREE PAR TIGER 129

4. RELATIONS BATI/POPULATION ET SONDAGES SPATIAUX EN MILIEU URBAIN 139
4. 1. INVENTAIRE EXPLORATOIRE DES METHODES MATHEMATIQUES ET STATISTIQUES APPLICABLES 139

4. 2. OBJECTIFS GENERAUX. MARSEILLE. JUIN 85-JUILLET 86 : PHASE DE MISE AU POINT DE LA METHODE 140

4 . 3. ETAPES ET METHODES DU TRAVAIL STATISTIQUE A MARSEILLE 142
4. 3. 1. Analyse des données pour la modélisation des relations entre variables morphologiques et démographiques en milieu urbain 142
4. 3. 2. Plans de sondages aréolaires stratifiés et estimation bayésienne de la population 145

5. LA TELEDETECTION SPATIALE EN MILIEU URBAIN 149
5. 1. LE POINT SUR LA TELEDETECTION URBAINE A TRAVERS LA BIBLIOGRAPHIE FRANCAISE ET ANGLO-SAXONNE SUR LE SUJET 149
5. 1. 1. L'état des connaissances et le bilan des méthodes. 149
5. 1. 2. Les axes de recherche à développer 164

5. 2. DEVELOPPEMENTS INFORMATIQUES 165

5 . 3 . ANALYSE DESCRIPTIVE DES DONNEES SPECTRALES DE L'IMAGE T.M. 167
5. 3. 1. Statistique univariée 167
5. 3. 2. Corrélations entre canaux 171
5. 3. 3. Analyse en composantes principales 175

DEUXIEME PARTIE : LES TRAVAUX PREPARATOIRES SUR QUITO 179

1. ANALYSE DE L'INFORMATION EXISTANTE 181
1. 1. Image satellite 181
1. 2. Photographies aériennes 181
1. 3. Données démographiques 181

2. EQUIPEMENT INFORMATIQUE 186

3 . L'ENQUETE DEMOGRAPHIQUE 188

INTRODUCTION

RAPPEL DES OBJECTIFS DU PROGRAMME DE RECHERCHE

L'objectif général assigné au projet de recherche soumis à l'ATP est d' *élaborer un outil permettant le suivi de la morphologie et de la démographie d'une ville, grâce à un système d'information géographique intégrant les données de télédétection.*
Ce projet s'inscrit dans une dynamique scientifique particulière au sein de l'ORSTOM. En effet, outre les memebres de l'atelier de télédétection il implique des chercheurs travaillant sur deux programmes de recherche : Mise au point d'un système d'information géographique (programme débuté par l'Unité d'Infographie en 1983), et Utilisation de la télédétection pour l'observation des populations urbaines (1) (programme engagé par l'UR 6 du Département D en 1985). Ces deux programmes se rejoignent sur une application commune : l'Atlas Permanent Informatisé de Quito (APIQ) qui comprend, outre la production classique d'un recueil de cartes, un volet de recherches sur la mise au point d'un observatoire urbain informatisé, dont une partie des données serait actualisée à l'aide de la télédétection.

La réalisation de l'objectif général du programme nécessite des travaux approfondis dans trois directions :

- Développement de modules spécifiques aux données de télédétection, qui rejoignent le système d'information géographique mis au point par l'Unité d'Infographie de l'ORSTOM
- Interprétation des images satellites à haute résolution en milieu urbain
- Relations bâti/population et sondages spatiaux en démographie urbaine.

L'étape préalable à ces trois axes de recherche, qui sont développés simultanément depuis le début 1985, est la constitution d'une base de données sur la ville d'expérimentation.

Le présent rapport intermédiaire fait le point des travaux déjà réalisés sur le premier site-test, Marseille, pour la constitution de la base de données et chacune des trois directions de recherche (partie 1) ; dans une seconde partie, nous exposons les travaux préparatoires à la mise en oeuvre des recherches sur le deuxième site du programme, Quito (EQUATEUR) (début des recherches à Quito : Juillet 1986).

---

(1)Le lecteur trouvera un exposé détaillé des objectifs de ce programme de recherche dans les articles de F. DUREAU (voir liste des publications de l'équipe)

LE CONTEXTE INSTITUTIONNEL ET FINANCIER

Trois services de l'ORSTOM participent au programme de recherches :
- L'UR 6 du Département D, "Urbanisation et socio-systèmes urbains"
- L'Unité d'Infographie
- L'Atelier de Télédétection de Bondy
En outre, une collaboration existe sur les travaux concernant Marseille avec l'Ecole des Hautes Etudes en Sciences Sociales (EHESS), en la personne d'A. CHATELAIN.

Quant aux partenaires équatoriens, il s'agit des institutions participant à l'Alas Permanent Informatisé de Quito (APIQ) :
- Institut Géographique Militaire (IGM),
- Municipio de Quito,
- Institut Panaméricain de Géographie et d'Histoire (IPGH),
ainsi que du CLIRSEN, organisme qui réalise les travaux de télédétection en Equateur.

Outre les financements ORSTOM et le financement ATP dont bénéficie le projet, ont été obtenus :

- la fourniture à un tarif préférentiel de 4 scènes SPOT sur la ville de Quito : demande PEPS reprise par le CNES. La ville de Quito se trouvant à cheval sur deux scènes SPOT, le CNES finance les vues verticales en panchromatique et multispectral ; les vues obliques demandées dans le projet PEPS pour travailler en stéréoscopie sont à financer sur le budget du programme ORSTOM.

- un complément de financement par le Comité de Pilotage de la Télédétection en Coopération (Ministère de l'Industrie et de la Recherche et Ministère des Relations Extérieures - Coopération et Développement) : 172.500 F . Ce financement vient en complément de la somme accordée par l'ATP (92.000 F), par rapport à notre demande ATP de 267.500 F. A ce jour, la convention entre l'ORSTOM et le Comité n'est pas encore signée, et nous ne pouvons donc disposer de ces fonds.

LA COMPOSITION DE L'EQUIPE ORSTOM

L'équipe ayant subi des changements importants par rapport à la description qui en était faite dans le projet soumis à l'ATP, nous rappelons ci-dessous la composition effective de l'équipe ORSTOM:

-UR 6 du Département D :

- DUREAU Françoise, chargée de recherche ORSTOM, géographe-démographe.

- BARBARY Olivier, allocataire de recherche MRT, statisticien. Participe au programme depuis Juin 1985.

- MICHEL Alain, allocataire de recherche MRT, urbaniste spécialisé en télédétection. Participe au programme depuis Décembre 1985.

- Unité d'Infographie :

- SOURIS Marc, ingénieur de recherche ORSTOM, responsable de l'Unité d'Infographie, mathématicien -informaticien.

- PELLETIER Françoise, ingénieur de recherche ORSTOM, cartographe-informaticienne.

- ANTIGEOL Immaculée Conception, technicienne ORSTOM, chargée de la digitalisation des cartes.

- Atelier de Télédétection de Bondy (ATOB):

- LORTIC Bernard, responsable de la formation à l'ATOB, spécialiste en télédétection.

- RANNOU Isabelle, technicienne ORSTOM, photographe.

LISTE DES PUBLICATIONS, NOTES DE TRAVAIL ET RAPPORTS

BARBARY O., 1986. -Projet de thèse de mathématiques appliquées aux sciences de l'homme, sous la direction de M. BARBUT, EHESS. Sujet : "Elaboration de plans de sondage pour l'estimation de données spatio-temporelles, en prenant en compte les informations fournies par la télédétection. Application aux données démographiques en milieu urbain des pays en développement". Paris, 20 p.

BARBARY O. et DUREAU F., 1986. - Etablissement d'une typologie de la morphologie urbaine de Marseille à partir de l'analyse visuelle d'une moisaïque photographique au 1/23 000. Paris, ORSTOM, 57p.

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 XXe congrès général de l'Union Internationale pour l'étude scientifique de la population, Florence (Italie), 5-12 Juin 1985.

DUREAU F., 1986. - A propos du traitement informatique de données localisées. Une expérience en cours : télédétection et observation des populations urbaines. A paraître dans la Collection de l'ORSTOM Colloques et séminaires, numéro spécial infographie. Paris, ORSTOM, 25 p.

MICHEL A., 1986. - Projet de thèse de géographie, sous la direction de J.P. RAISON, EHESS. Sujet : " Télédétection et observation suivie de la morphologie et de la démographie des villes des pays en développement. Mise au point des méthodes d'analyse des images satellite à haute résolution et évaluation des informations fournies par ces images " . Paris, 9 p.

MICHEL A., 1986. - Le point sur la télédétection urbaine en 1986 à travers la bibliographie française et anglo-saxonne sur le sujet. Recueil des fiches de dépouillement des articles et ouvrages. Paris, ORSTOM, 50p.

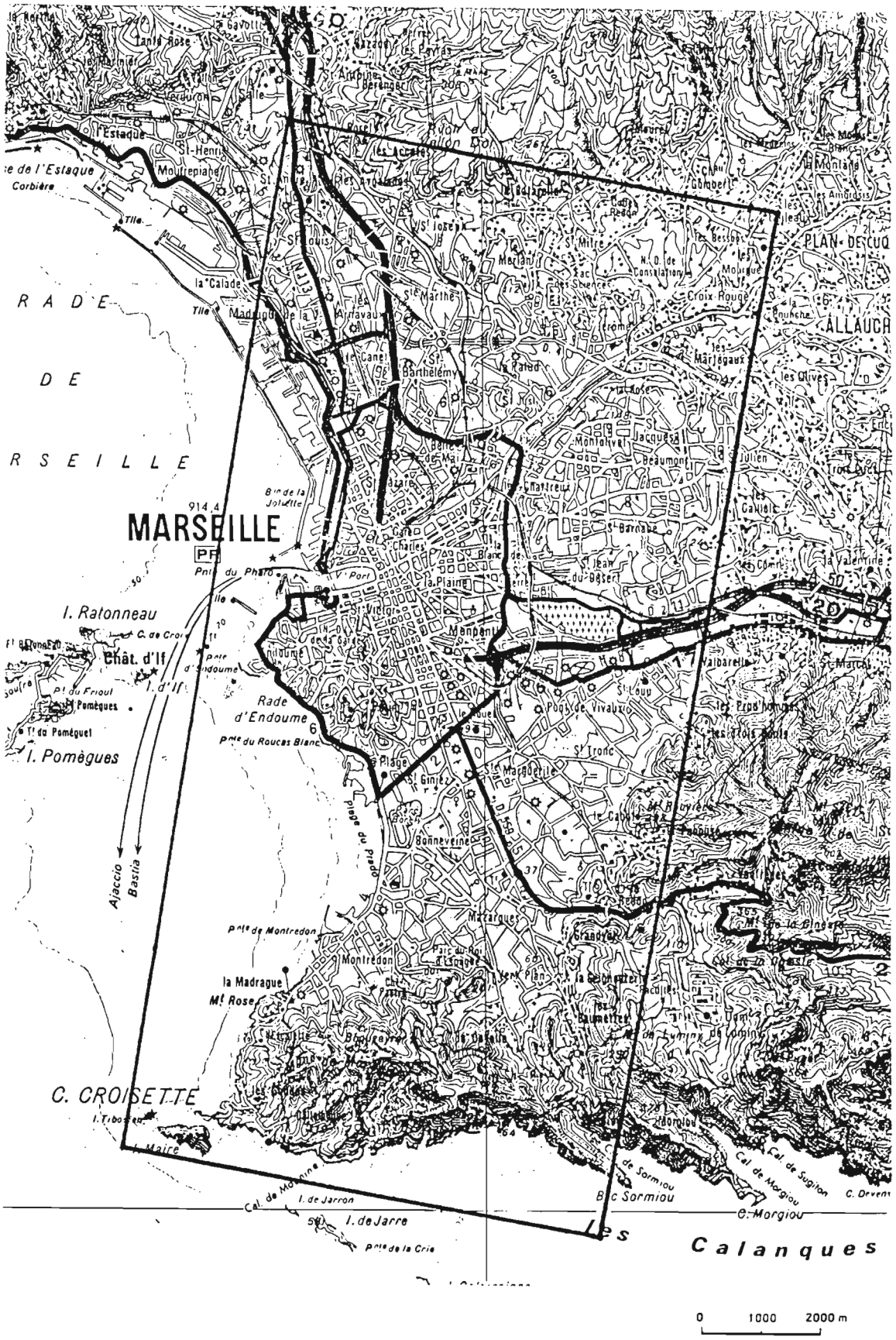
SOURIS M., 1985 -Proposition d'un système interactif graphique pour l'IGM.Quito,9p

SOURIS M. et DUREAU F, 1985. - Rapport de mission en Equateur. 15.05.85 - 29.05.85 (M. SOURIS) et 19.05.85 - 26.05.85 (F. DUREAU). Paris, ORSTOM, 6 p.

SOURIS M., 1986. - Systèmes d'information géographiques et bases de données. A paraître dans la Collection de l'ORSTOM Colloques et séminaires, numéro spécial infographie. Paris, ORSTOM, 65p.

1 . LES TRAVAUX EN COURS SUR MARSEILLE

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



1 . LA METHODOLOGIE DEVELOPPEE SUR MARSEILLE

1. 1. CHOIX DE LA ZONE D'ETUDE

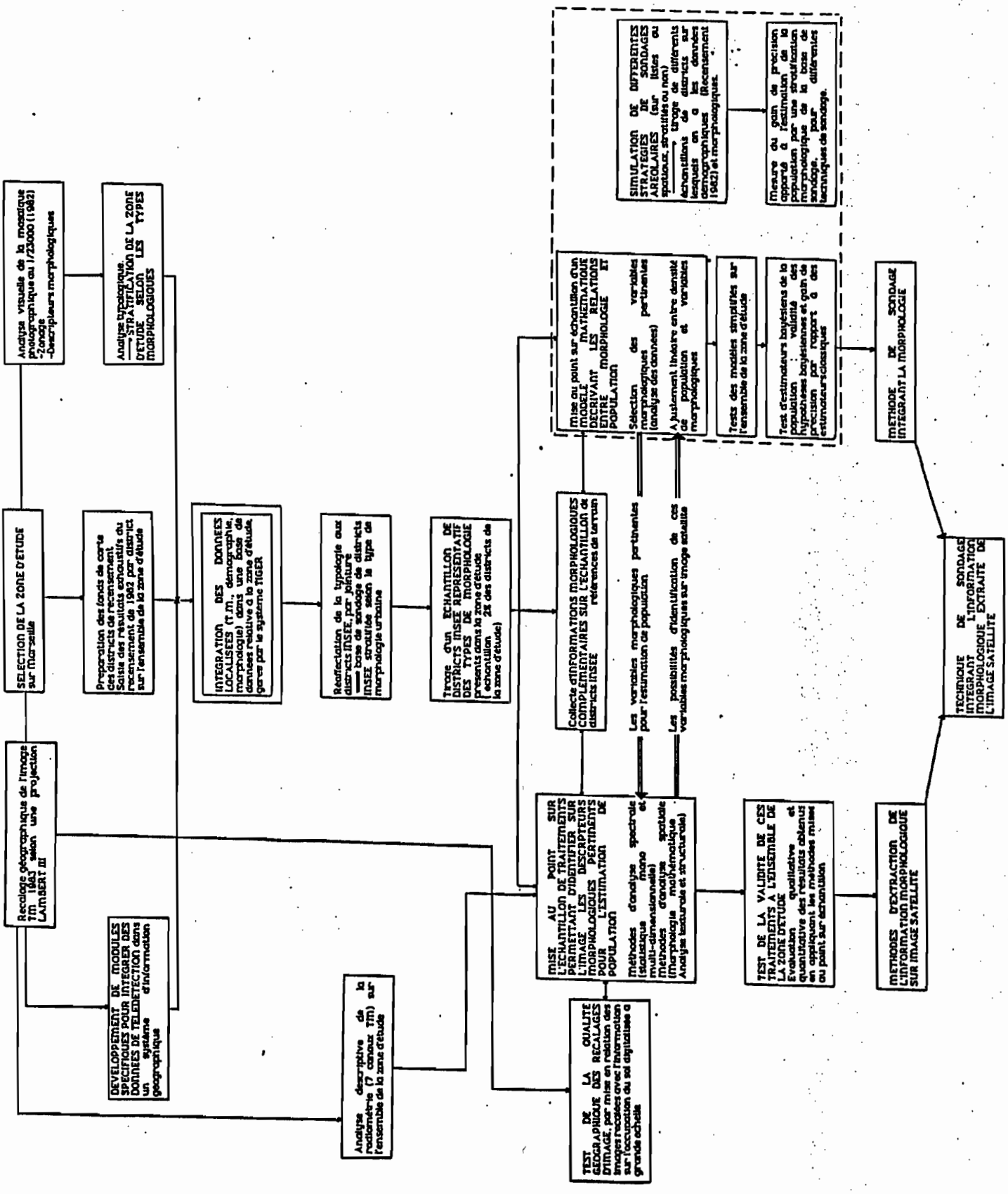
La zone d'étude (voir localisation sur la Figure 1) a été déterminée de manière à *couvrir les différents types de quartiers de Marseille,* soit schématiquement : les quartiers industriels et de grands ensembles ouvriers du nord, les quartiers résidentiels du sud, le vieux centre, et les zones pavillonnaires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

1. 2. PRINCIPES DE L'EXPERIMENTATION

Sur Marseille, les trois directions de recherche évoquées en introduction sont menées en parallèle, leurs résultats se conditionnant mutuellement. Les principes de l'expérimentation menée sur Marseille sont résumés dans la Figure 2 ; on notera que dans cette première phase du programme de recherche, 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 impliquent la mise en relation de données localisées issues de sources d'information variées, et de nombreux traitements de données géographiques . Les exemples qui suivent montrent la nécessité de pouvoir mettre en relation, sans qu'il y ait réduction d'information, les trois types de données intégrées dans la base de Marseille (données démographiques, données morphologiques, et image satellite) : ils précisent l'utilisation que nous

Figure 2 - Schéma d'expérimentation sur Marseille



faisons de la base de données sur Marseille gérée par le système TIGER, et induisent les modules de traitement que le système doit donc comporter (voir chapitre 3).

. 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 constituants ; 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 de régression,

- 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.

2 . CONSTITUTION DE LA BASE DE DONNEES SUR MARSEILLE

Constituer une base de données sur Marseille intégrant l'information démographique, morphologique et satellitaire a été une opération longue, qui s'est étendue sur toute l'année 1985. Les chapitres qui suivent, relatifs à l'analyse et à l'intégration de l'information à intégrer dans la base mettront en évidence les problèmes rencontrés, les solutions adoptées et les enseignements à tirer de cette expérience.

Nous insisterons ici plus particulièrement sur les données démographiques et morphologiques, le chapitre 3 étant entièrement consacré à l'intégration des données de télédétection (exposé des principes de fonctionnement du système et présentation des modules développés pour l'intégration de l'information satellitaire)

2. 1. ETABLISSEMENT D'UNE TYPOLOGIE DE LA MORPHOLOGIE URBAINE

Comme nous l'avons indiqué au chapitre 1, les caractéristiques morphologiques constituent une des trois informations de base intégrées dans la base de données sur Marseille . Après avoir rappelé l'utilisation qui est faite de ces données dans le programme de recherche, et les objectifs particuliers assignés à la typologie de la morphologie urbaine, nous présenterons les documents qui ont servi à l'élaboration de ces données. La démarche suivie sera exposée chronologiquement, étape par étape, avant de commenter les résultats finaux. En conclusion, nous tirons les enseignements méthodologiques de cette expérience de photo-analyse et traitement des données observées sur photographies aériennes en milieu urbain.

2. 1. 1. Objectifs et données utilisées

. Objectifs

Les objectifs de l'analyse des photographies aériennes sont au nombre de trois :

- *disposer d'une connaissance exhaustive de certains paramètres de la morphologie urbaine sur l'ensemble de la zone d'étude*, afin de pouvoir mettre au point les techniques d'interprétation d'image satellite et tester leur validité : dans ce cas, l'information analytique sur la morphologie urbaine issue de la photo-analyse contrôlée par des vérifications au sol, sert de référence de terrain.

- *tirer un échantillon de travail représentatif des types d'occupation du sol présents dans notre périmètre d'étude de Marseille* : cette opération est nécessaire si l'on veut que la technique de mise au point de méthodes sur échantillon soit efficace (voir chapitre 1. 1.). Il s'agit donc d'établir une typologie de la morphologie urbaine à partir des descripteurs observés sur photographies aériennes : cette typologie sert de base de stratification pour le tirage d'un échantillon de travail représentatif .

- *tester l'efficacité d'une stratification morphologique dans l'estimation de population par sondage* : le principe est de mesurer le gain de précision lié à la stratification morphologique dans le cas de sondages sur listes ou de sondages de points sur une grille (voir chapitre 4. 2.).

Si le premier objectif repose sur les descripteurs analytiques directement issus de la photo-analyse, les deux autres supposent d'établir une typologie de la morphologie urbaine à partir de ces descripteurs : les chapitres 2.1.3, 2.1.4 et 2.1.5 sont consacrés à cette analyse typologique.

. Données utilisées

a) Les photographies aériennes

Pour satisfaire les objectifs assignés à la photo-analyse, les photographies aériennes devaient remplir trois conditions :

- apporter une information exhaustive et homogène sur l'ensemble de la zone d'étude : ce qui suppose une couverture aérienne, à une même échelle et à une date unique (ou des dates très voisines) sur l'ensemble de la zone d'étude (figure 1)
- avoir été enregistrées à une date la plus proche possible des deux autres sources de données : recensement de mars 1982 et image Thematic Mapper de Février 1983
- être à une échelle adéquate pour analyser le tissu urbain et les objets bâtis: 1/15.000 à 1/40.000 environ.

Les photos aériennes réalisées en août 1983 à Marseille dans le cadre de la couverture systématique du littoral français (mission IPLI - Lot 23 - échelle voisine de 1/23000 - panchromatique - noir et blanc) satisfont à ces trois conditions : le décalage temporel est réduit au maximum puisque les photographies aériennes ont été prises à une date intermédiaire entre le recensement et l'image Thematic Mapper, et l'échelle est tout à fait satisfaisante pour l'utilisation prévue ici.
La mission de 1981 (1/15.000 - panchromatique - noir et blanc), que nous avions envisagé d'utiliser, aurait apporté des éléments d'information supplémentaires sur les objets urbains ; malheureusement, elle ne couvre que la moitié nord de notre zone d'étude. Il faut souligner à ce propos que l'agglomération urbaine n'est visiblement pas une notion retenue par l'IGN pour ses missions aériennes : la plupart des couvertures à grande échelle réalisées sur Marseille depuis une dizaine d'années ne couvre qu'une partie de la commune.

Une *mosaïque photographique non contrôlée* a été réalisée à l'ATOB ; c'est sur ce document qu'a été réalisée la photo-analyse.
Une certaine perte d'information existe entre les photographies originales (tirages papier 24 x 24 cm fournis par l'IGN) et la mosaïque. Du fait de la moindre netteté de la mosaïque, le zonage est plus aisé : l'oeil ne perçoit plus nettement le objets urbains bien individualisés, et une certaine intégration, plus proche du niveau d'information recherché pour satisfaire les objectifs de l'analyse visuelle (établissement d'une typologie, et référence de terrain pour la télédétection spatiale), devient ainsi immédiate.

b) Contrôles terrain

Les informations extraites de la photo-analyse ont été contrôlées, par sondage, lors d'une enquête sur le terrain début novembre 1985 .
Le principe du tirage de l'échantillon de zones morphologiques à contrôler au sol a été le suivant : chacun des 17 types morphologiques mis en évidence par les premiers traitements d'analyse des données, AFC (Analyse Factorielle des Correspondances) et CAH (Classification Ascendante Hiérarchique), devaient être représentés par plusieurs zones dans l'échantillon. On a donc procédé à un tirage systématique sur la liste des zones classées par type morphologique, avec un pas variable selon les types :

Nombre de zones par type	Taux de sondage	Nombre de zones dans l'échantillon
226	1 / 15	16
14 à 35	1 / 3	5 à 11
6 à 11	1 / 2	3 à 5
2 à 5	1 / 1	2 à 5

Contrôles terrain - Novembre 1985 - Fiche de relevé

FICHE ILOT PHYSIQUE | N° Classe | N° Zone Morpho

Mesures de l'élément le plus grand dans la zone:
- Longueur :
- Largeur :
- Hauteur :

Mesures de l'élément le plus petit dans la zone:
- Longueur :
- Largeur :
- Hauteur :

Croquis îlot (Noter photo)

	Batiment 1	Batiment 2	Batiment 3	Batiment 4	Batiment 5
Longueur					
Largeur					
Nbre étages					
Nature du toit					
Nature du Non Bati					

Changements depuis 1982:

Au total, l'échantillon tiré selon cette procédure comportait 100 zones morphologiques. La majorité de ces zones occupent une superficie trop importante pour être contrôlées exhaustivement : c'est pourquoi, un deuxième tirage a été effectué au sein de 2/3 des zones les plus grandes de l'échantillon. Il s'est agit de déterminer sur les photographies aériennes, pour chacune de ces grandes zones, un groupe de pâtés de maisons représentatifs.

La fiche de relevé, remplie pour chacune des zones de l'échantillon, rassemble des informations sur :

- le plus grand bâtiment et le plus petit bâtiment de la zone (longueur, largeur, et nombre d'étages). La détermination de ces deux bâtiments a été faite sur photographie aérienne : même si ces bâtiments se trouvaient hors du groupe d'ilôts sélectionnés, nous sommes allés les mesurer.
- les caractéristiques des éléments bâtis majoritaires (longueur, largeur, nombre d'étages, nature du toit), et nature du non bâti. Dans le cas des grandes zones, ces renseignements sont issus de l'observation du groupe d'ilôts sélectionnés. Nous nous sommes efforcés de décrire tous les types de bâtiments présents dans chaque zone morphologique.
- un croquis et des photographies au sol viennent compléter l'information. Près de 150 photographies au sol ont été prises lors des relevés sur le terrain, illustrant tous les types de morphologie urbaine.
- des observations sur les changements intervenus depuis 1982.

Ce contrôle sur le terrain a été à l'origine de certaines modifications dans le codage issu de l'analyse visuelle, comme nous le verrons au chapitre 2.1.4.

2.1.2. Analyse visuelle de la mosaique photographique au 1/23000

La méthodologie retenue pour analyser la mosaïque s'est appuyée sur les enseignements issus des travaux présentés dans la littérature sur le sujet, et l'expérience en photographie de B. LORTIC ; un test sur un petit nombre de zones a permis de cerner la praticabilité des critères retenus a priori. Enfin, la réalisation même de l'analyse a permis de rectifier certaines options prises à l'issue du test.

. Etablissement de la méthodologie d'analyse

Comme le rappellent J.P. GILG et son équipe (1976, Aperçu d'une expérience de télé-analyse et photo-analyse. Informatique et sciences humaines n°30, pp. 21-38), la démarche de l'interprète humain est :
- d'abord globale : délimitation de zones d'égale apparence ou isophènes,
- puis analytique : description des zones, qui valide et module la stratification initiale.

L'efficacité de cette démarche n'est plus à démontrer. Il faut toutefois souligner les problèmes que soulève le zonage manuel dans son application : B. LORTIC (1984, "Une méthode d'analyse visuelle d'une image spatiale. Objectivation du zonage manuel". Mémoire EHESS) évoque *trois difficultés*:

- les variations continues et progressives de l'image : les ruptures existant au sol étant la plupart du temps progressives, il est difficile de localiser précisément les limites des zones homogènes.
- la tendance à privilégier un paramètre visuel au détriment des autres : les variables couleur et valeur sont privilégiées spontanément au détriment des variables texture et structure, qui peuvent pourtant, dans certains domaines, apporter le plus d'information.
- l'existence de différents niveaux d'homogénéité : le seuil de "différence" au dessous duquel on décrète qu'il y a homogénéité est difficile à déterminer et à maintenir constant sur l'ensemble de l'image.

Pour trouver ce seuil, B. LORTIC précise qu'il est nécessaire de procéder par itération :

1) Réaliser un premier zonage sans critère bien défini, en détourant chaque zone semblant homogène et se différenciant de son environnement

2) A partir de ce zonage, déterminer précisément les critères discriminants et les seuils d'homogénéité les plus pertinents

3) Reprendre le premier zonage en décrivant chaque zone à partir de ces critères.

C'est cette démarche itérative que nous avons retenue pour l'analyse visuelle de la mosaïque photographique de Marseille, et que nous exposerons maintenant en détail:

- 1) Zonage en unités d'égale apparence
- 2) Caractérisation de chacune de ces zones à l'aide de descripteurs, avec modifications ponctuelles du zonage initial.

. <u>Zonage en unités d'égale apparence</u>

Le zonage a été établi sur un calque posé sur la mosaïque photographique. La démarche suivie était de détourer les zones d'égale apparence sur le document photographique.

Le découpage spatial ne tient compte que de ce paramètre, et non du découpage en pâtés de maisons ou en districts de recensement : un certain nombre de patés de maisons sont donc divisés, s'ils ne sont pas homogènes. Pour cette étape, on ne prend pas en compte des paramètres bien définis : seule compte la différenciation entre zones, corrolairement l'homogénéité de la zone.

524 zones ont ainsi été délimitées surla zone d'étude présentée dans la figure 1(chapitre 1.1).

Il faut insister sur le fait que la perte d'information de la mosaïque par rapport aux photographies IGN a facilité le zonage, évitant l'écueil majeur en milieu urbain, qui est d'isoler le bâtiment, objet se démarquant généralement de son voisinage immédiat ; nous avons ainsi pu éviter dans la majeure partie des cas d'établir un zonage à un niveau trop fin pour notre application. Toutefois, il existe une certain nombre de zones très petites, dues à la présence d'un objet urbain relativement grand, et très contrasté par rapport à son environnement . Isolées

de ce fait, ces zones ne peuvent pas être décrites de la même façon que les autres zones : le milieu urbain est caractérisé par l'importance de l'effet d'échelle. Lorsque l'on passe du groupe d'ilôts à l'ilôt lui même, les variables descriptives valides ne sont plus les mêmes.

. Test des descripteurs

Cette opération avait pour but de tester le degré de praticabilité d'une première série de descripteurs, et le nombre de modalités qui pouvait être retenues pour chacun d'entre eux.
Ce test a été réalisé sur un petit nombre de zones, dix seulement ; ces zones ont été choisies de manière à couvrir un large éventail des situations rencontrées sur la mosaïque de Marseille. Ont été testés les descripteurs de :

- taille des éléments bâtis,
- forme des éléments bâtis
- agencement du bâti dans l'ilôt
- la valeur (niveau de gris) dominante du bâti
- la variablité de valeur du bâti
- la valeur de la voirie.

Ce premier essai nous a montré que tous ces descripteurs étaient praticables sans problème majeur à l'exception de celui sur la variabilité de la valeur ; d'autre part, nous avons testé le critère de densité du bâti dans la zone, en se fixant des zones étalons.

Le test nous a conduit à étoffer la liste de descripteurs, la première liste choisie s'avérant insuffisante pour rendre compte des diverses situations rencontrées . Quatre axes nous ont guidé dans le choix de ces nouveaux descripteurs :
- distinguer la valeur dominante de la valeur sous-dominante sur la zone, plutôt que de chercher à caractériser directement la variabilité de cette valeur
- séparer la voirie et le non bâti-non voirie, pour la caractérisation de la valeur
- préciser les mesures des éléments bâtis (longueur, largeur, surface et hauteur)
- introduire des descripteurs de l'agencement du réseau de voirie, et de la densité de ce dernier.

D'autre part, nous avons choisi des descripteurs analytiques. S'ils sont apparemment plus difficiles à maîtriser sur la photographie que des indicateurs synthétiques, il est toujours aisé de construire ensuite un indice à partir de critères observés sur la photographie ; l'on peut ainsi tester différents indices sans avoir à revenir sur le document photographique. C'est ce qui a été fait, par exemple, pour la longueur et la largeur, qui ont permis de construire deux indices : la surface, et le degré d'allongement de l'élément bâti majoritaire.

D'après les enseignements tirés des tests, nous avons établi une liste de descripteurs, dont le nombre de modalités a été fixé en essayant de les appliquer sur des zones les plus diverses possibles de Marseille. La descrition des zones a donc débuté avec un "guide d'analyse" comprenant:
- la liste des descripteurs
-le mode de collecte de chacun de ces descripteurs (mesure directe à la règle, ou

Tableau 1 - Liste des descripteurs

Descripteur	Mode de relevé	Tableau de description
Densité du bâti dans la zone	Estimation visuelle avec zones étalon du rapport S. bâtie/S. totale. 6 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (6 modalités)
Valeur dominante du bâti	Estimation visuelle des niveaux de gris du bâti majoritaire dans la zone. 4 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (4 modalités)
Valeur sous-dominante du bâti	id	id
Valeur dominante de l' espace intersticiel	Estimation visuelle du niveau de gris majoritaire du non-bâti non-voirie. 4 modalités	Conservée dans tableau de description morpho. (4modalités)
Valeur sous-dominante de l'espace intersticiel	id	id
Hauteur du bâti	Estimation visuelle par examen stéréoscopique. 5 modalités dont 1 pour les zones non bâties. (2 niveaux de hauteur x hauteur mixte ou homogène)	Conservée dans tableau de description morpho. (5 modalités)
Surface minimum des éléments bâtis	Mesure de la surface du plus petit élément bâti de la zone, en centièmes de mm^2	Conservée dans tableau de description morpho.
Surface maximum des éléments bâtis	Mesure de la surface du plus grand élément bâti de la zone, en centièmes de mm^2	id
Longueur de l'élément bâti majoritaire	Mesure de la longueur de l'élément bâti majoritaire de la zone, en dixièmes de mm	Conservée dans tableau de description morpho. (non seuillée), utilisée pour variables dérivées *
Largeur de l'élément bâti majoritaire	Mesure de la largeur de l'élément bâti majoritaire de la zone, en dixièmes de mm	id
Forme de l'élément bâti majoritaire	Observation visuelle.4 modalités dont 1 pour les zones non bâties	Conservée dans tableau de description morpho. (4 modalités)

Descripteurs supprimés au cours de l'analyse visuelle :

Agencement du bâti dans l'ilôt	Valeur de la voirie
Agencement du réseau de voiri	Densité du réseau de voirie
Variabilité de la valeur du bâti	Type de non bâti

* Création de 2 variables dérivées à partir des descripteurs de longueur et largeur : surface et rapport longueur/largeur

estimation visuelle) et le nombre de modalités dans le cas de descripteurs estimés visuellement, définies par référence à des zones témoins de la mosaïque.
Le tableau 1 présente la liste des descripteurs et leur mode de recueil sur la mosaïque.

. Description des zones d'égale apparence

La description des zones conduit à *modifier partiellement le premier zonage* en zones d'égale apparence : des limites de zones ont été modifiées lors de la description, de même, certaines zones ont été supprimées, et d'autres ajoutées. C'est d'ailleurs ce dernier cas qui a été le plus courant : nous sommes ainsi passé de 524 zones à 542 zones après caractérisation. Ce phénomène est du au fait que certains voisinages très contrastés tendent à donner une homogénéité apparente à des zones qui, lors de la description, s'avérent ne pas pouvoir être décrites de façon homogène ; d'où des cas de subdivision lors de la caractérisation.

Notons que nous avons choisi de procéder à la *description des zones critère par critère, ou par couple de critères similaires* (exemples : longueur et largeur de l'élément bâti majoritaire, valeur dominante et sous-dominante du bâti), et non zone par zone.
Etant donné la surface à analyser et le nombre important de zones à décrire, il nous a en effet semblé préférable de procéder ainsi pour garder une certaine constance dans l'appréhension des critères qualitatifs : il nous semble en effet que le risque est déjà grand de ne pas rester constant pour des critères qui réclament plusieurs jours de travail pour être établis sur l'ensemble de la zone d'étude. En procédant par zone, celà aurait signifié un écart qui se chiffre en semaines entre le début et la fin du travail de caractérisation : il est totalement utopique de penser que l'on puisse rester constant dans ces jugements sur une période de temps aussi longue. Procéder par critère ou type de critères permet, à notre avis, de limiter le degré de variabilité de la description visuelle, qui ne peut jamais être totalement éliminée pour des critères tels que la valeur, dont la perception par l'oeil est influencée par des paramètres tels que l'ensoleillement de la pièce où est réalisée l'analyse photographique.

Pour une partie des descripteurs, outre la hauteur qui nécessitait la stéréoscopie , nous avons du retourner au document d'origine, c'est à dire les photographies aériennes de l'IGN, la mosaïque se révélant d'une précision insuffisante. Ce gain de précision ne remet pas en cause la méthode sur mosaïque "lissée" pour établir le zonage.

Sur les 17 variables retenues initialement, 6 ont été éliminées dès la phase de description des zones L'examen de la mosaïque pour les 11 premiers critères ont en effet mis en évidence :

- soit que certains de ces 6 critères étaient redondants avec ceux déjà établis : ainsi, l'agencement du bâti dans l'ilôt a été abandonné après application à la moitié des zones, car ce critère n'apportait aucune information nouvelle par rapport aux autres descripteurs.

- soit qu'ils n'étaient pas praticables pour un trop grand nombre de zones,

trop petites pour qu'ils puissent être appliqués : c'est le cas des caractéristiques du réseau de voirie (agencement et densité), qui supposent que la zone contienne plusieurs pâtés de maisons pour qu'ils aient un sens.

. Le tableau de description des zones

A l'issue de la photo-analyse, nous avons donc obtenu un tableau croisé, décrivant les 542 zones par 11 variables analytiques de la morphologie urbaine (voir Tableau 1). Ce tableau a été saisi sur support informatique, afin de servir à l'établissement de la typologie de la morphologie urbaine, à l'étude statistique des relations bâti/population, et à l'interprétation des images satellites Le zonage, reporté sur carte topographique IGN au 1/25000 a été digitalisé par l'Unité d'Infographie (voir chapitres 2. 2et 2. 3 pour l'intégration dans la base de données sur Marseille).

Avant tout traitement, il importe de bien préciser les caractéristiques de ce tableau initial de données sur la morphoplogie urbaine :
- *les variables sont de nature différente, au sens mathématique* : on trouve des variables métriques (mesures de longueur et largeur) aussi bien que des variables qualitatives (valeurs, densité du bâti, par exemple)
- étant donné leur mode de production, *toutes les variables n'ont pas exactement le même degré de fiabilité*
- *les différentes modalités prises par une variable ordonnée n'ont pas un écart constant entre elles, mais font référence à des situations types* : il existe des phénomènes de "saut" qui font que l'on ne peut rencontrer tous les cas de figures pour une variable. On ne peut donc assimiler une échelle de modalités ordonnées à une variable quantitative continue qui aurait été seuillée.
- *la caractérisation visuelle doit être complétée par une enquête terrain*, pour confirmer certaines opérations, volontairement limitées mais néanmoins effectives, d'interprétation : c'est le cas principalement de l'interprétation "élément bâti". Cette enquête est également nécessaire pour l'utilisation qui sera faite de cette analyse visuelle de photographies aériennes : il s'agit de passer de l'analyse à l'interprétation, c'est à dire faire correspondre aux objets présents sur la photographie les objets présents sur le terrain, afin de pouvoir utiliser la photographie aérienne comme élément de référence terrain lors de mise au point et validation des méthodes d'extraction de l'information sur image satellite.

*2.1.3. ANALYSES STATISTIQUES UNIVARIEES ET RECODAGE DES DESCRIPTEURS MORPHOLOGIQUES.

A : LES DISTRIBUTIONS DES DESCRIPTEURS MORPHOLOGIQUES QUALITATIFS.

Remarques préliminaires :

1. La variable qualitative "FORM", qui décrit la forme de l'élément bâti majoritaire en deux modalités : parallèlépipèdes et autres formes, apparait dés l'examen de sa distribution, sans grand intérêt. 417 zones sur les 441 existantes ont été codées avec la première modalité. En fait, quelque soit la forme des constructions, on a toujours pu donner un sens aux mesures de longueur et de largeur sur le bâtiment typique de la zone. On préfèrera donc remplacer la variable qualitative "FORM" par la mesure calculée du rapport : longueur/largeur caractérisant l'élément bâti majoritaire. La distribution et le recodage de cette variable seront décrits au paragraphe concernant les variables quantitatives du tableau.

Description des distributions (fig. 2.1.3 1) :

1 : Densité du bâti (5 modalités)
La majorité des zones (380 sur 441) a une densité de bâti faible ou moyenne:
modalité 1 : densité très faible 99 zones
modalité 2 : densité faible 120 zones
modalité 3 : densité moyenne 159 zones
Les zones denses et très denses sont relativement peu nombreuses :
modalité 4 : densité forte 53 zones
modalité 5 : densité très forte 10 zones

2 : Valeur dominante dans l'espace bâti (3 modalités, fig.2.1.3 2)
Les modalités dominantes correspondent aux valeurs moyennes (220 zones) et claires (159 zones).Seules 62 zones ont une valeur majoritaire foncée de leur bâti. Ce sont, soit des zones au bâti ancien, à toit de tuiles canal sombres, soit des zones ou dominent de vieux bâtiments industriels, hangars à toit de tuiles mécaniques .

3 : Valeur sous-dominante du bâti (3 modalités, fig 2.1.3 3)
On observe en gros la même répartition que pour la valeur dominante:
modalité 1 (claire) 192 zones
modalité 2 (moyenne) 206 zones
modalité 3 (foncée) 43 zones

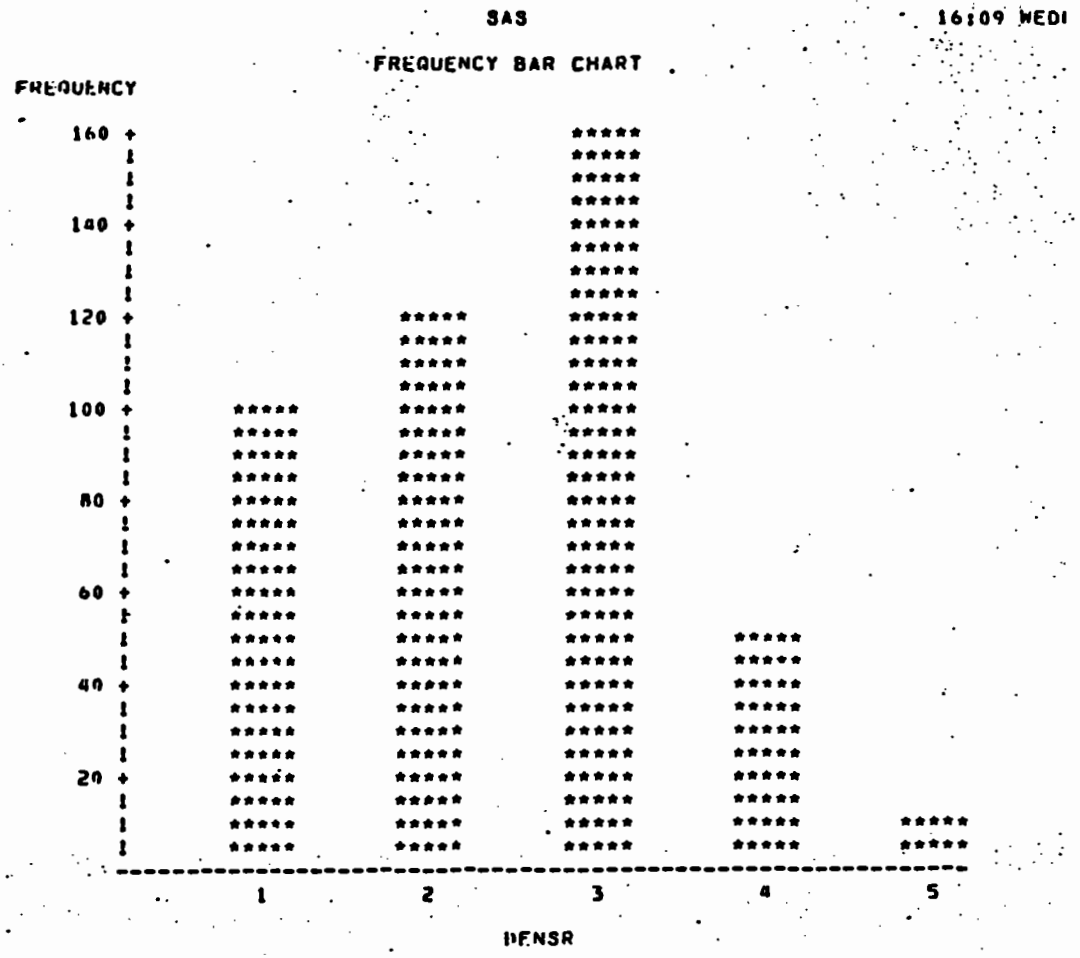


Fig. 2.1.3. 1

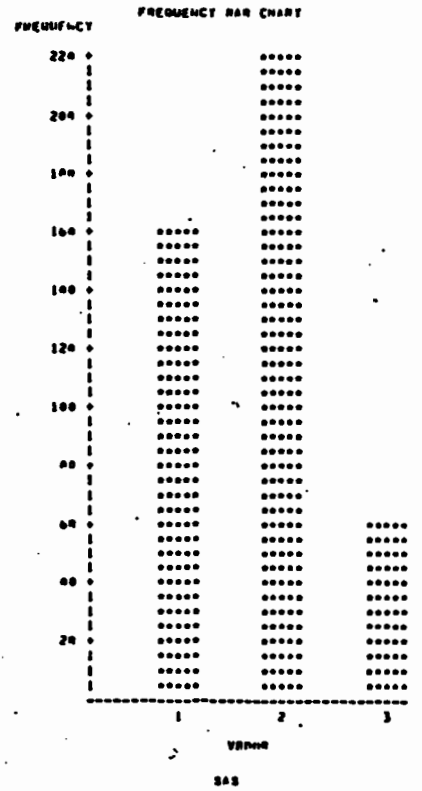


Fig. 2.1.3. 2

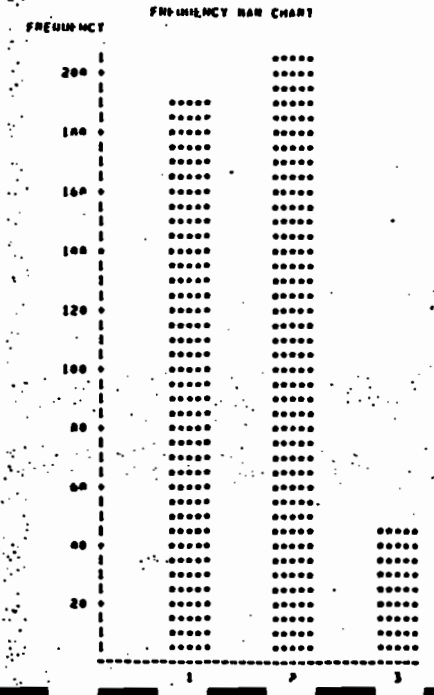


Fig. 2.1.3. 3

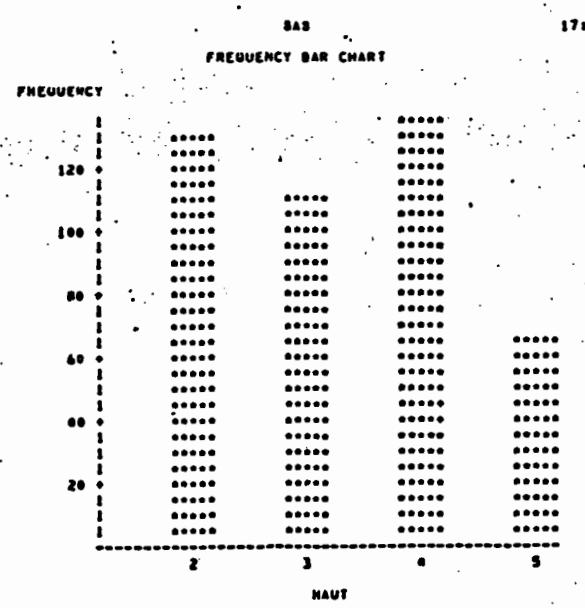


Fig 2.1.3. 6

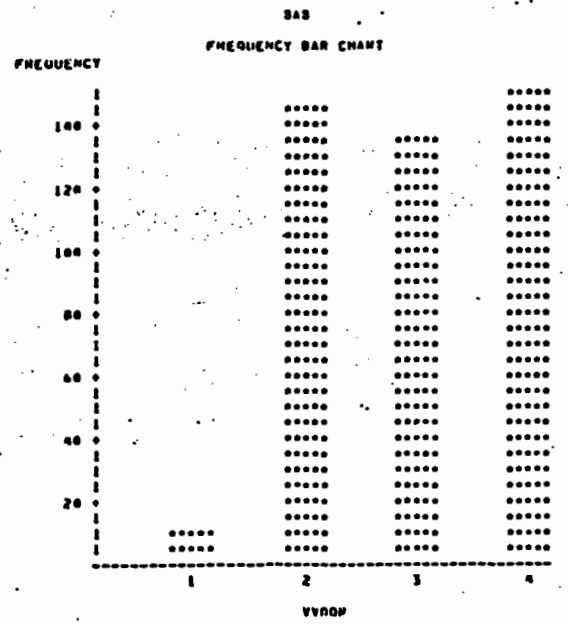


Fig 2.1.3. 4

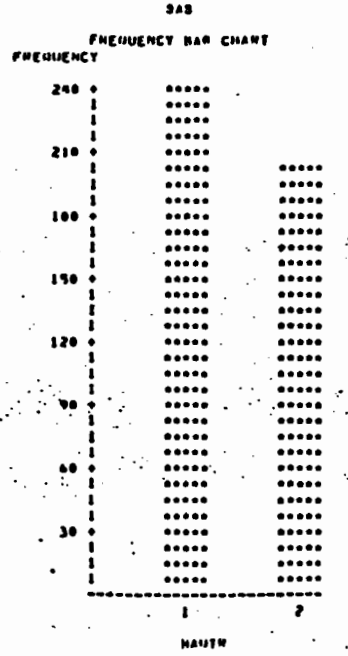


Fig. 2.1.3. 7

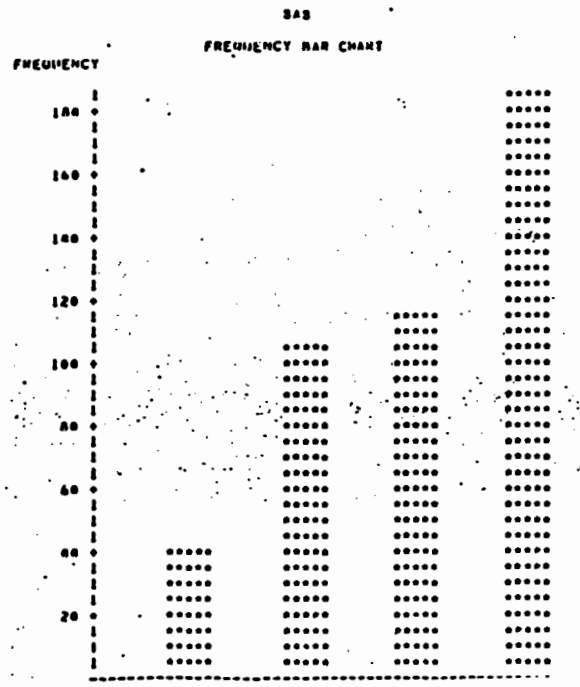
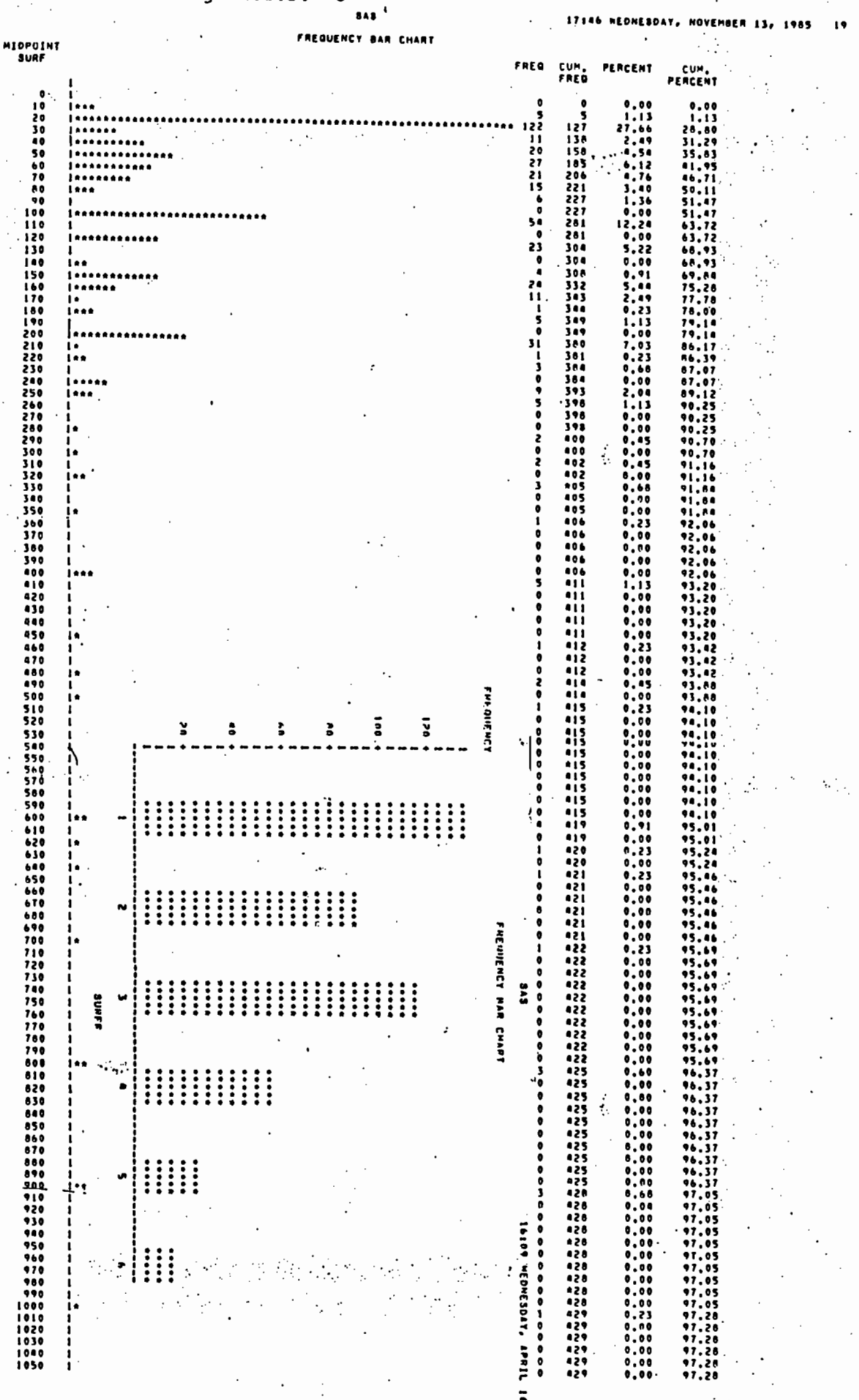


Fig. 2.1.3. 5

Fig. 2.1.3. 8



SAS 1 17:46 WEDNESDAY, NOVEMBER 13, 1985 19

FREQUENCY BAR CHART

MIDPOINT SURF	FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
0	0	0	0.00	0.00
10	5	5	1.13	1.13
20	122	127	27.66	28.80
30	11	138	2.49	31.29
40	20	158	4.54	35.83
50	27	185	6.12	41.95
60	21	206	4.76	46.71
70	15	221	3.40	50.11
80	6	227	1.36	51.47
90	0	227	0.00	51.47
100	54	281	12.24	63.72
110	0	281	0.00	63.72
120	23	304	5.22	68.93
130	0	304	0.00	68.93
140	4	308	0.91	69.84
150	24	332	5.44	75.28
160	11	343	2.49	77.78
170	1	344	0.23	78.00
180	5	349	1.13	79.14
190	0	349	0.00	79.14
200	31	380	7.03	86.17
210	1	381	0.23	86.39
220	3	384	0.68	87.07
230	0	384	0.00	87.07
240	9	393	2.04	89.12
250	5	398	1.13	90.25
260	0	398	0.00	90.25
270	0	398	0.00	90.25
280	2	400	0.45	90.70
290	0	400	0.00	90.70
300	2	402	0.45	91.16
310	0	402	0.00	91.16
320	3	405	0.68	91.84
330	0	405	0.00	91.84
340	0	405	0.00	91.84
350	1	406	0.23	92.06
360	0	406	0.00	92.06
370	0	406	0.00	92.06
380	0	406	0.00	92.06
390	0	406	0.00	92.06
400	5	411	1.13	93.20
410	0	411	0.00	93.20
420	0	411	0.00	93.20
430	0	411	0.00	93.20
440	0	411	0.00	93.20
450	1	412	0.23	93.42
460	0	412	0.00	93.42
470	0	412	0.00	93.42
480	2	414	0.45	93.88
490	0	414	0.00	93.88
500	1	415	0.23	94.10
510	0	415	0.00	94.10
520	0	415	0.00	94.10
530	0	415	0.00	94.10
540	0	415	0.00	94.10
550	0	415	0.00	94.10
560	0	415	0.00	94.10
570	0	415	0.00	94.10
580	0	415	0.00	94.10
590	0	415	0.00	94.10
600	4	419	0.91	95.01
610	0	419	0.00	95.01
620	1	420	0.23	95.24
630	0	420	0.00	95.24
640	1	421	0.23	95.46
650	0	421	0.00	95.46
660	0	421	0.00	95.46
670	0	421	0.00	95.46
680	0	421	0.00	95.46
690	0	421	0.00	95.46
700	1	422	0.23	95.69
710	0	422	0.00	95.69
720	0	422	0.00	95.69
730	0	422	0.00	95.69
740	0	422	0.00	95.69
750	0	422	0.00	95.69
760	0	422	0.00	95.69
770	0	422	0.00	95.69
780	0	422	0.00	95.69
790	0	422	0.00	95.69
800	3	425	0.68	96.37
810	0	425	0.00	96.37
820	0	425	0.00	96.37
830	0	425	0.00	96.37
840	0	425	0.00	96.37
850	0	425	0.00	96.37
860	0	425	0.00	96.37
870	0	425	0.00	96.37
880	0	425	0.00	96.37
890	0	425	0.00	96.37
900	3	428	0.68	97.05
910	0	428	0.00	97.05
920	0	428	0.00	97.05
930	0	428	0.00	97.05
940	0	428	0.00	97.05
950	0	428	0.00	97.05
960	0	428	0.00	97.05
970	0	428	0.00	97.05
980	0	428	0.00	97.05
990	0	428	0.00	97.05
1000	1	429	0.23	97.28
1010	0	429	0.00	97.28
1020	0	429	0.00	97.28
1030	0	429	0.00	97.28
1040	0	429	0.00	97.28
1050	0	429	0.00	97.28

4 : Valeur dominante de l'environnement du bâti, voirie exclue (4 modalités, fig 2.1.3. 4)
La distribution est concentrée sur les trois modalités de rangs 2, 3 et 4 et il y a aproximativement equi-répartition entre elles :
modalité 2 (claire) 143 zones
modalité 3 (gris moyen) 138 zones
modalité 4 (gris foncé) 149 zones
La modalité 4 correspond à une dominante boisée de l'environnement, la modalité 3 aux pelouses, prairies, terrains agricoles ou parterres et la modalité 2, dejà assez claire, est le plus souvent liée aux espaces bitumés : parking, entrée d'immeuble etc... ou aux terrains vagues. La modalité 1, qui est marginale (11 zones), n'apparait que dans les endroits à très forte réflectance : sols nus, roches et carrières, etc...

5 : Valeur sous-dominante de l'environnement (4 modalités, fig 2.1.3 5)
Les fréquences vont en croissant avec le rang des modalités :
modalité 1 (très claire) 38 zones
modalité 2 (claire) 105 zones
modalité 3 (gris moyen) 114 zones
modalité 4 (gris foncé) 184 zones

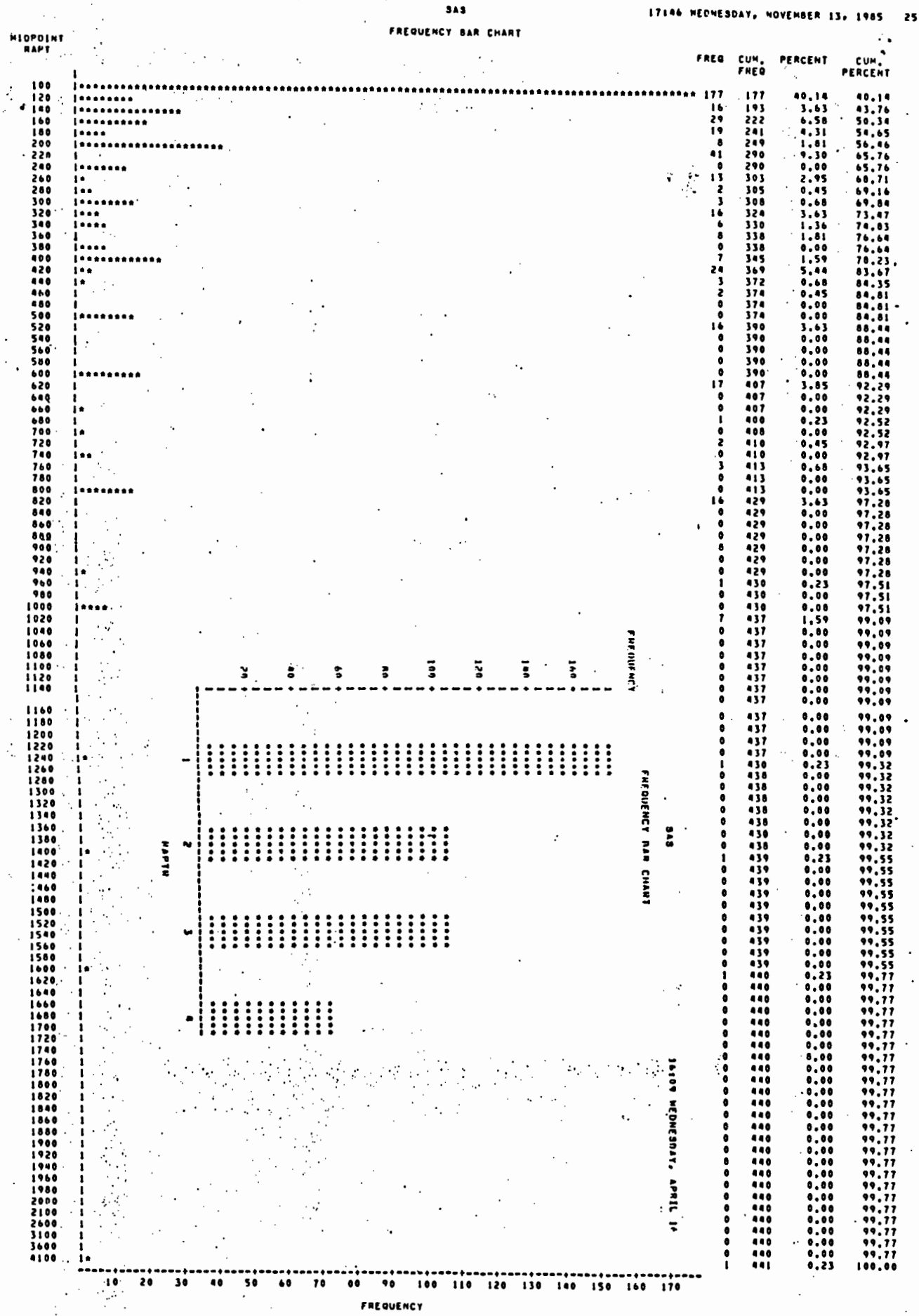
6 : Elévation du bât, fig 2.1.3 6 et 7)
A l'origine, la description de l'élévation du bâti a été faite en quatre modalités. On observait les fréquences suivantes :
modalité 1 (bâti uniformément bas) 131 zones
modalité 2 (élévation mixte à dominante basse) 105 zones
modalité 3 (élévation mixte à dominante haute) 114 zones
modalité 4 (bâti uniformément haut) 184 zones
Par la suite, considérant d'une part que la distinction entre les modalités 1, 2 et 3, 4 pose parfois problème lorsqu'on se satisfait d'une évaluation rapide (sans stéréoscope) et que, d'autre part, dans l'optique typologique qui nous intéresse pour l'instant, c'est le seuil entre dominante haute et dominante basse qui importe le plus, on a pris le parti de regrouper les modalités 1 et 2 d'une part, 3 et 4 d'autre part. Il en résulte une description dichotomique de l'élévation du bâti, ou les zones à dominante basse sont légèrement plus nombreuses que celles à dominante haute (241 pour 200).
On verra plus tard que, pour palier à la déscription trés sommaire de ce caractère morphologique important du point de vue thématique, on sera amené à augmenter son poids dans l'analyse par pondération de ses deux modalités.

B : DISTRIBUTION DES CARACTERES QUANTITATIFS, SEUILLAGE ET RECODAGE EN QUALITATIF

Remarque préliminaire :

Les quatre variables quantitatives s'attachent à la description des éléments bâtis de chaque zone. Leur surface maximum et leur surface minimum ont été mesurées directement sur l'image grâce à des rectangles étalons. On a également mesuré la longueur et la largeur de l'élément majoritaire dans chaque zone. Plutôt que de retenir ces deux descripteurs tels quels, on a préféré, pour faciliter l'interprétation des résultats d'analyse, calculer deux indicateurs synthétiques : surface de l'élément bâti majoritaire et rapport de forme (longueur/largeur) du bâtiment en question.

Fig. 2.1.3.‘ 9


SAS
17:46 WEDNESDAY, NOVEMBER 13, 1985 25
FREQUENCY BAR CHART
MIDPOINT RAPT
FREQ
CUM. FREQ
PERCENT
CUM. PERCENT
FREQUENCY
FREQUENCY BAR CHART
SAS
17:09 WEDNESDAY, APRIL 1
RAPTN
FREQUENCY

Les distributions des descripteurs et leur recodage.

Il s'agit ici, en étudiant les histogrammes de ces quatre variables, d'identifier les modes principaux de chaque distribution pour fixer ensuite les bornes des intervalles qui résumeront au mieux ces mesures en autant de variables qualitatives; chaque modalité représentera alors un mode principal. Pour mieux évaluer ce qui est conservé et ce qui est perdu dans l'information qu'apporte chaque descripteur, on pourra comparer les histogrammes originaux aux histogrammes des variables qualitatives obtenues, qui sont juxtaposés sur les figures 8 à 11. Dans le commentaire, on se contente de donner les intervalles de chaque modalité et d'indiquer les fréquences et les principales valeurs "centrales" .

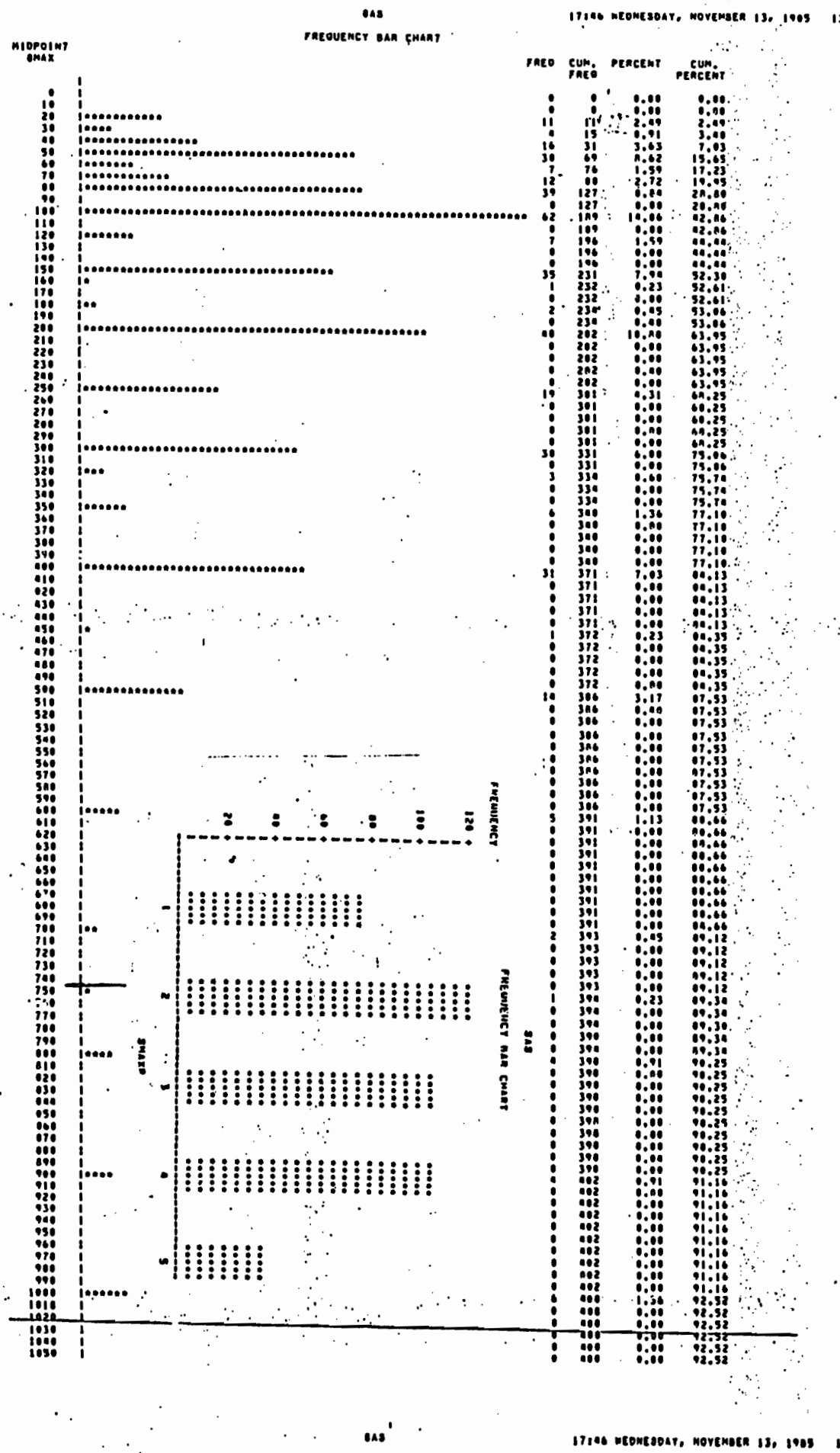
1. Surface de l'élément bâti majoritaire (fig 2.1.3 8) :
Cinq modes principaux sont observables dans la série qui donneront six modalités de la variable recodée (la partie terminale de l'histogramme a été coupée en deux modalités)
- modalité 1 (0 < SURF < 30) : 136 zones offrant les plus petites tailles de bâtiments, centrées autour de la valeur 25 (112 zones) qui représentent 31% du total.
- modalité 2 (30 < SURF < 90) : 92 zones autour de la valeur 50 soit 21% des zones.
- modalité 3 (90 < SURF < 190) : 117 zones (27%) qui prennent préférentiellement les valeurs 100, 120 et 150.
- modalité 4 (191 < SURF < 360) : 57 zones (13%) dont 31 ont la valeur 200 et 14 les valeurs 240 et 250.
- modalité 5 (360 < SURF < 950) : 25 zones (6%).
- modalité 6 (950 < SURF < 3000) : 14 zones (3%).

2 . Rapport longueur/largeur de l'élément bâti majoritaire (fig.2.1.3 9)
Les quatre modes principaux de la distribution fournissent quatre modalités.
- modalité 1 (RAPT = 1) : 177 zones ou la plupart des bâtiments sont aproximativement carrés, soit 40% du total des zones.
- modalité 2 (1 < RAPT < 2) : Les formes peu allongées dominent dans ces 107 zones (24%).
- modalité 3 (2 < RAPT < 5) : 103 zones (23%).
- modalité 4 (5 < RAPT < 40) : 54 zones ou les formes sont trés allongées (12%).
La plupart des zones à grands bâtiments d'habitat collectif ou à bâtiments industriels se trouvent dans les modalités 3 et 4.

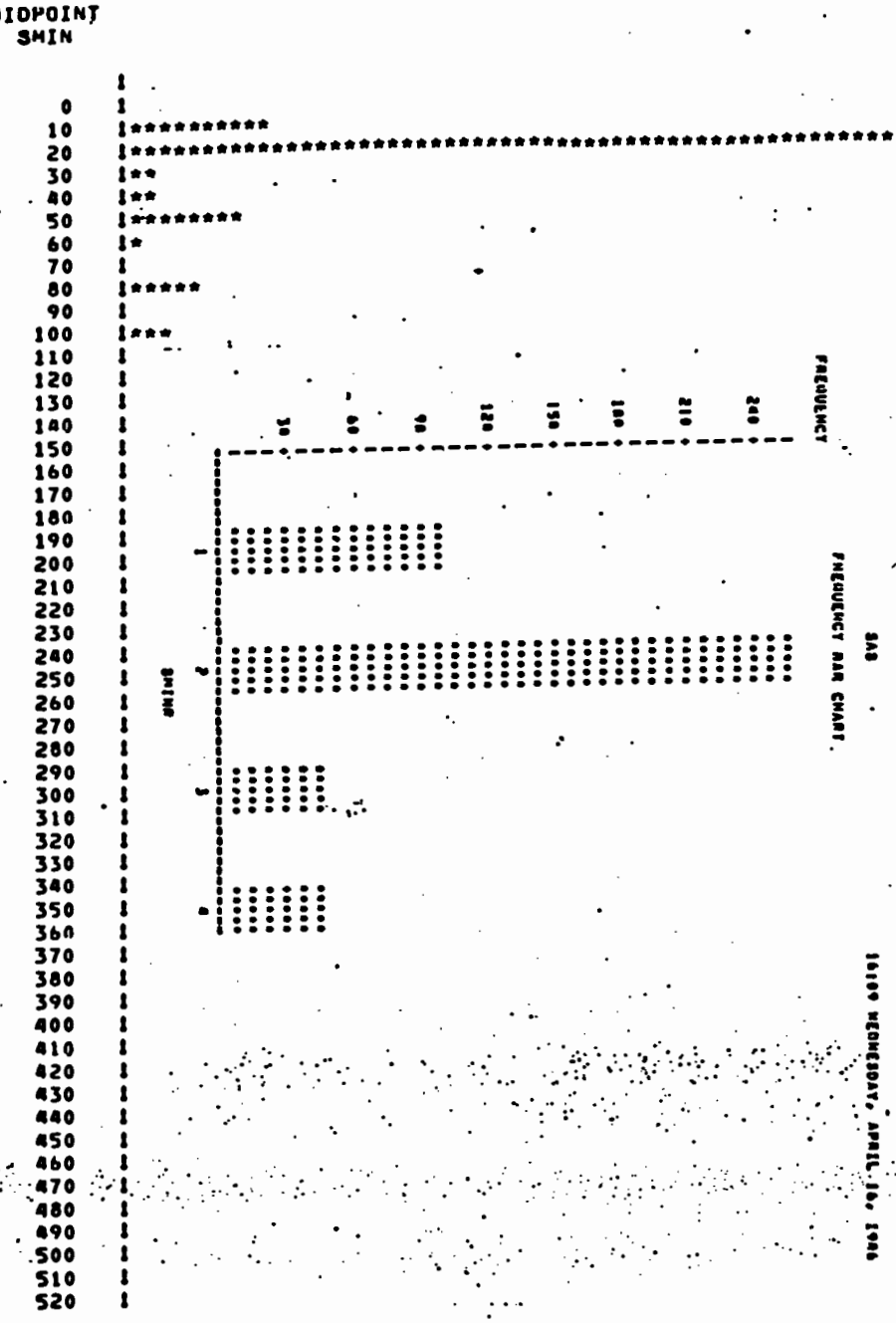
3 . Surface maximum des éléments bâtis (fig 2.1.3 10).
Cinq modes principaux donnent cinq modalités;
- modalité 1 (0 < SMAX < 60) : 75 zones (17%) centrées autour de la valeur 50 (38 zones).
- modalité 2 (60 < SMAX < 149) : 122 zones (28%) centrées autour des valeurs 80 et 100 (respectivement 39 et 69 zones).
- modalité 3 (149 < SMAX < 299) : 105 zones centrées autour de la valeur 200 (48 zones), soit 24% du total.
- modalité 4 (299 < SMAX < 1100) : 103 zones (23%) ou les valeurs majoritaires sont 300 : 30 zones, 400 : 31 zones et 500 : 14 zones.
- modalité 5 (SMAX > 1100) : 36 zones (8%) dont les valeurs s'étendent jusqu'à 7500.

4 Surface minimum des éléments batis (fig 2.1.3 11)

Fig. 2.1.3. 10


SAS
FREQUENCY BAR CHART
17:46 WEDNESDAY, NOVEMBER 13, 1985 13
MIDPOINT
OMAX
FREQ
CUM. FREQ
PERCENT
CUM. PERCENT
FREQUENCY
STAND
SAS
FREQUENCY BAR CHART
SAS
17:46 WEDNESDAY, NOVEMBER 13, 1985 15

FREQUENCY BAR CHART



FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
0	0	0.00	0.00
52	52	11.79	11.79
281	333	63.72	75.51
12	345	2.72	78.23
8	353	1.81	80.05
39	392	8.84	88.89
3	395	0.68	89.57
1	396	0.23	89.80
26	422	5.90	95.69
0	422	0.00	95.69
15	437	3.40	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
0	437	0.00	99.09
1	438	0.23	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
0	438	0.00	99.32
1	439	0.23	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55
0	439	0.00	99.55

Fig. 2.1.3. 11

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

Ici quatre modes principaux fournissent quatre modalités.
- modalité 1 (0 < SMIN < 24) : 97 zones (22%) ou les valeurs les plus fréquentes sont 10 :52 zones et 20 : 45 zones.
- modalité 2 (24 < SMIN < 35) : 252 zones centrées sur la valeur 25 (236 zones).
- modalité 3 (35 < SMIN < 65) : 48 zones (11%) centrées autour de la valeur 50 (39 zones).
- modalité 4 (65 < SMIN < 4000) : 44 zones (10%) ou les valeurs majoritaires sont 80 : 26 zones et 100 : 15 zones.

On aura remarqué, dans l'indication des valeurs les plus fréquemment prises, une concentration autour des comptes ronds liée à l'emploi d'étalons.
Ce biais montre le degré d'approximation des mesures effectuées et relativise la perte d'information qui vient de l'agglomération des valeurs autour de la valeur centrale de chaque mode (quand elle est unique). On pourrait donc relever ces valeurs comme notes affectées aux zones mais pour ramener tous les descripteurs à une nature qualitative, il faut passer au codage disjonctif et, ce faisant, on perd l'information relative aux écarts entre valeurs centrales des modes.

2.1.4. METHODES D'ANALYSE DES DONNEES APPLIQUEES AU TABLEAU DE CARACTERISATION MORPHOLOGIQUE DES ZONES.

-A : ANALYSE DU TABLEAU GLOBAL SANS PONDERATION

Les données

Dans le but d'établir une typologie des quartiers marseillais, on exploitera, grâce à des outils maintenant classiques en analyse des données (AFC et CAH), l'information morphologique tirée de la photo-interprétation, concernant 441 zones de l'espace urbain de Marseille qui contiennent des bâtiments. A l'issue de l'analyse monodimensionnelle des descripteurs morphologiques qui a permis le codage des variables en un ensemble de descripteurs qualitatifs, ainsi qu'une première sélection, et le calcul des descripteurs pertinents, on dispose d'un tableau disjonctif complet croisant 441 zones bâties d'égale apparence avec les 40 modalités possibles des 10 descripteurs morphologiques retenus.(voir fig. 2.1.4 1). Ce tableau sera donc soumis successivement à l'analyse factorielle des correspondances et à une classification ascendante hiérarchique maximisant le moment d'ordre 2 de la partition associé à la distance du X2 (programmes ANCOR et CAH2CO de l'ADDAD exécutés sous S.A.S. au CIRCE).

Analyse factorielle des correspondances.

- Examen des valeurs propres.

Comme toujours lorsqu'il s'agit du tableau disjonctif complet associé à une correspondance multiple, les valeurs propres exprimées en pourcentages de la trace de la matrice diagonalisée (égale à la somme des valeurs propres), reflètent mal la part d'inertie totale expliquée par chacun des axes. Dans l'ensemble des analyses factorielles qui vont suivre, on prendra donc, comme mesure de cette quantité, les valeurs propres issues de l'A.F.C. du tableau de Burt associé, également

Fig. 2.1.4. 2.

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE: MARSEILLE : A . F . C . DU TABLEAU COMPLET , SANS PONDERATION .25 / 3 / 86

NOMBRE DE POINTS : 40

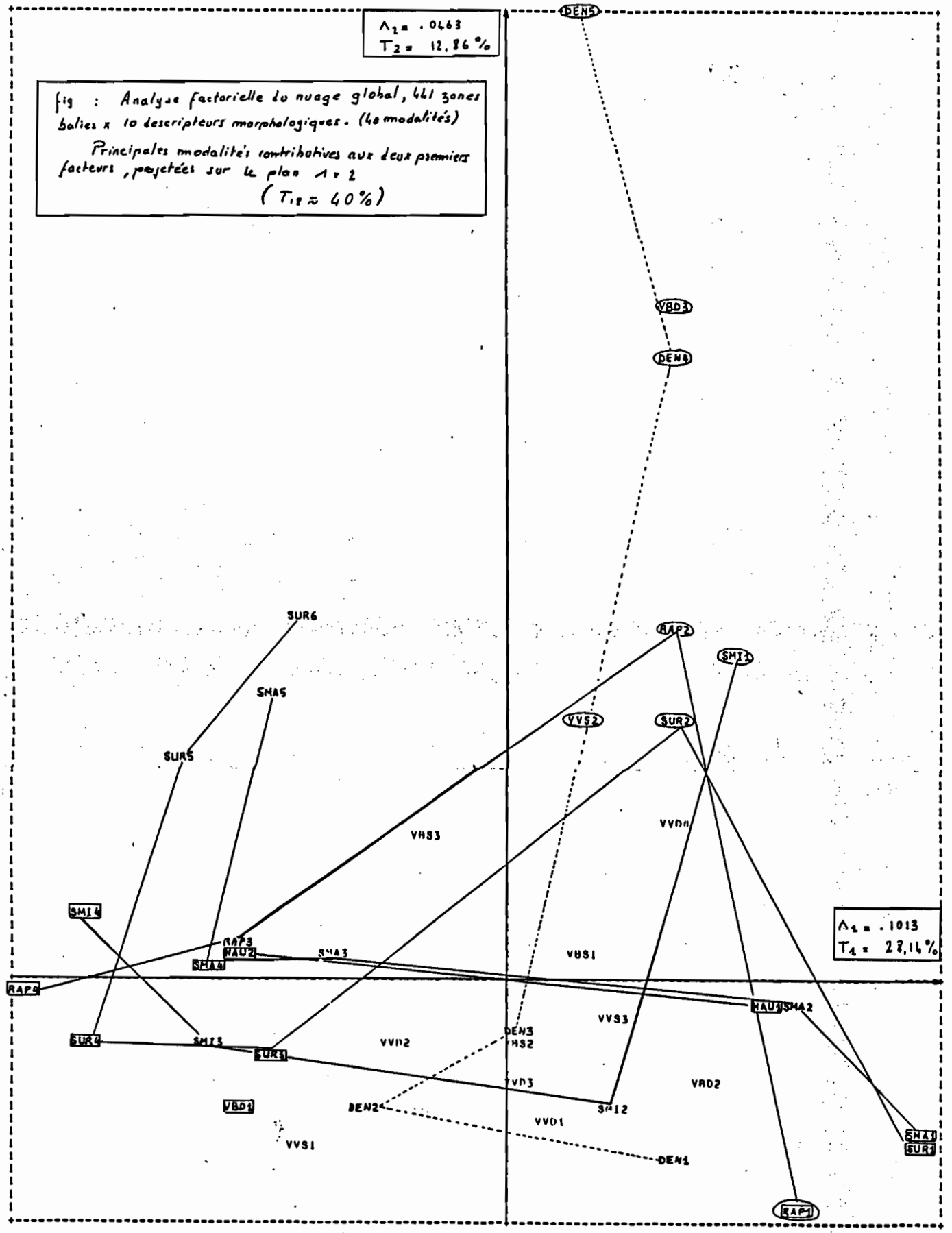


fig : Analyse factorielle du nuage global, 441 zones balies x 10 descripteurs morphologiques. (40 modalités)

Principales modalités contributives aux deux premiers facteurs, projetées sur le plan 1 x 2 ($T_{12} \approx 40\%$)

NOMBRE DE POINTS SUPERPOSES : 2

RAP3(MAU2) VVS4(VVD3)

rapportées à leur somme. Ces valeurs propres appelées Λ_i ne sont en fait que les carrés des λ_i issus du tableau disjonctif. Le tableau ci-dessous fourni la correspondance entre les λ_i et les Λ_i pour les 10 premiers facteurs. La valeur propre est exclue puisqu'elle vaut toujours 1.

rangs	λ_i	$\Lambda_i = \lambda_i^2$	$\left(\Lambda_i / \sum \Lambda_i\right)$%	cumul	histogramme
2	.3182	.1013	28.14	28.14	**********
3	.2151	.0463	12.86	41.00	*********
4	.1879	.0353	9.8	50.8	*******
5	.1622	.0263	7.3	58.1	*****
6	.1490	.0222	6.17	64.27	****
7	.1434	.0206	5.72	69.99	****
8	.1374	.0189	5.25	75.24	****
9	.1251	.0156	4.33	79.57	***
10	.1168	.0136	3.78	83.35	***
11	.1121	.0126	3.5	86;85	**

Les 7 premiers facteurs pourraient être considérés comme intéressants : ils ont une contribution à l'inertie du nuage supérieure à deux fois la contribution moyenne, mais on ne rendra compte ici que des 4 premiers qui permettent d'interpréter la partition en 8 classes fournie par la classification.

-Examen des facteurs et des plans factoriels

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

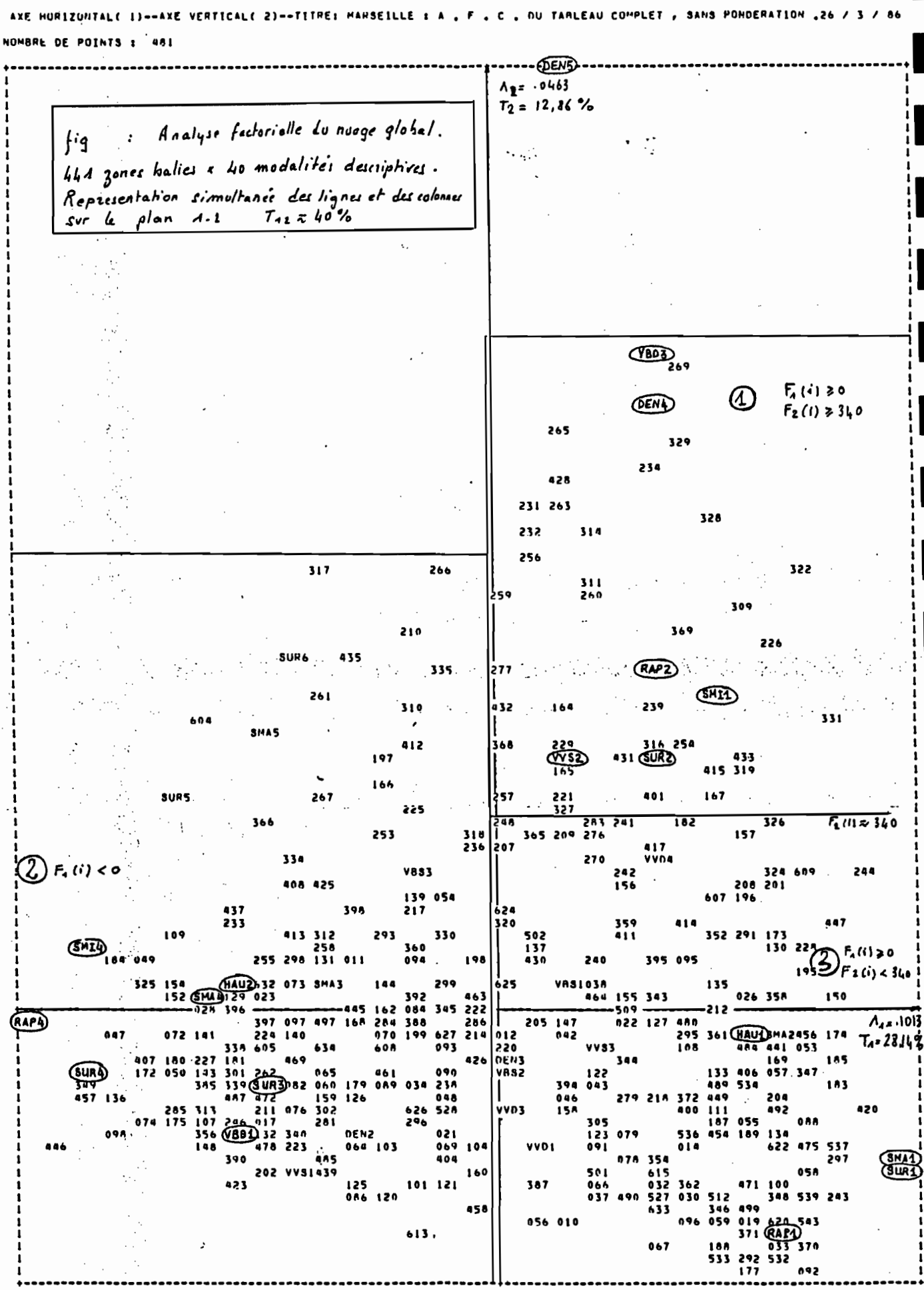
rangs	modl.	fact.	CTR%	cumul	poids
1	SUR1	1065	110	110	140
2	HAU2	-750	82	192	200
3	VBD1	-753	65	257	160
4	SMA1	1073	63	320	70
5	RAP4	-1434	58	378	50
6	RAP1	650	54	432	180
7	HAU1	550	52	484	240
8	SMA4	-784	48	532	100
9	SUR4	-1061	46	578	60
10	SUR3	-713	45	623	120
11	SMI4	-1199	44	667	40

Le premier facteur, qui explique environ 28% de l'inertie du nuage, apparait comme classant grossièrement les zones selon la taille des éléments bâtis, leur forme et leur hauteur : l'ensemble des modalités caractérisant les petites tailles de bâti, les formes carrées ou peu allongées et les faibles hauteurs, sont groupées à droite de l'axe (coordonnées positives) et inversement, les grands bâtiments allongés et dont la hauteur dépasse majoritairement deux niveaux se situent à droite de l'axe (voir figures 2.1.4 2 et 3).
Cette opposition explique 82% de l'inertie de l'axe si l'on prend en compte les modalités des cinq variables : SURF, SMIN, SMAX, RAPT, HAUT,

Fig. 2.1.4. 3

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE: MARSEILLE : A . F . C . DU TABLEAU COMPLET , SANS PONDERATION .26 / 3 / 86

NOMBRE DE POINTS : 481



37.5% étant du aux modalités de rangs 1 et 2 qui caracterisent le demi-plan droit et 44% aux modalités de rangs superieurs.

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

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

Le second facteur (13% de l'inertie du nuage) classe les zones suivant la densité du bâti, la forme et la valeur dominante dans l'espace bâti. Le nuage s'étire du côté positif de l'axe où sont regroupées les zones denses. Le bâti est ici moyennement allongé et de valeur sombre : on est autour des modalités DEN4, DEN5, RAP2, VBD3, VVS2, SMI1, SUR2. A l'inverse les valeurs négatives du deuxième facteur sont le fait des zones peu denses ou à densité moyenne (modalités DEN1, DEN2, DEN3) où les bâtiments sont soit carrés soit de formes très allongées (modalités RAP1 ou RAP4) et où la valeur du bâti est claire (VBD1 ou 2), (voir figures 2.1.4 2 et 3).
Cette opposition explique 73% de l'inertie de l'axe dont 58% dûe aux modalités de coordonnées positives : DEN4, DEN5, RAP2, VBD3, SMI1, VVS2.

Plan factoriel 1-2 :
En examinant, sur la figure 2.1.4 2, la projection simultanée des zones et des modalités, on distingue principalement trois régions du plan factoriel.
1 : La région supérieure droite du plan, caractérisée par des valeurs fortement positives du second facteur (F2(i) > 340) et positives du premier. S'y trouvent regroupées les zones à forte densité de bâti (DEN4 et DEN5), la surface des constructions y est petite (SMI1, SUR2), leur valeur foncée (VBD3) et leur formes peu allongées (RAP2).
2 : La région gauche du plan (premier facteur négatif) est caractérisée par les modalités de rang élevé des variables : "surface de l'élément majoritaire" (SUR3, SUR4, SUR5, SUR6), "surface maximum de l'élément bâti" (SMA3, SMA4, SMA5) et "rapport longueur/largeur de l'élément majoritaire" (RAP3, RAP4). Les bâtiments y sont majoritairement élevés (HAU2) et de valeur claire (VBD1, VBD2).
3 : Dans la région droite du plan (premier facteur positif, deuxième facteur négatif ou faiblement positif : -815 < F2(i) < 560) sont regroupées les zones à petits bâtiments (SUR1, SMA1) de formes carrées (RAP1) majoritairement bas (HAU1) ainsi que les zones peu denses (DEN1).

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

Fig. 2.1.4. 4.

AXE HORIZONTAL(1)--AXE VERTICAL(3)--TITRE: MARSEILLE : A . F . C . DU TABLEAU COMPLET , SANS PONDERATION .26 / 3 / 86

NOMBRE DE POINTS : 40

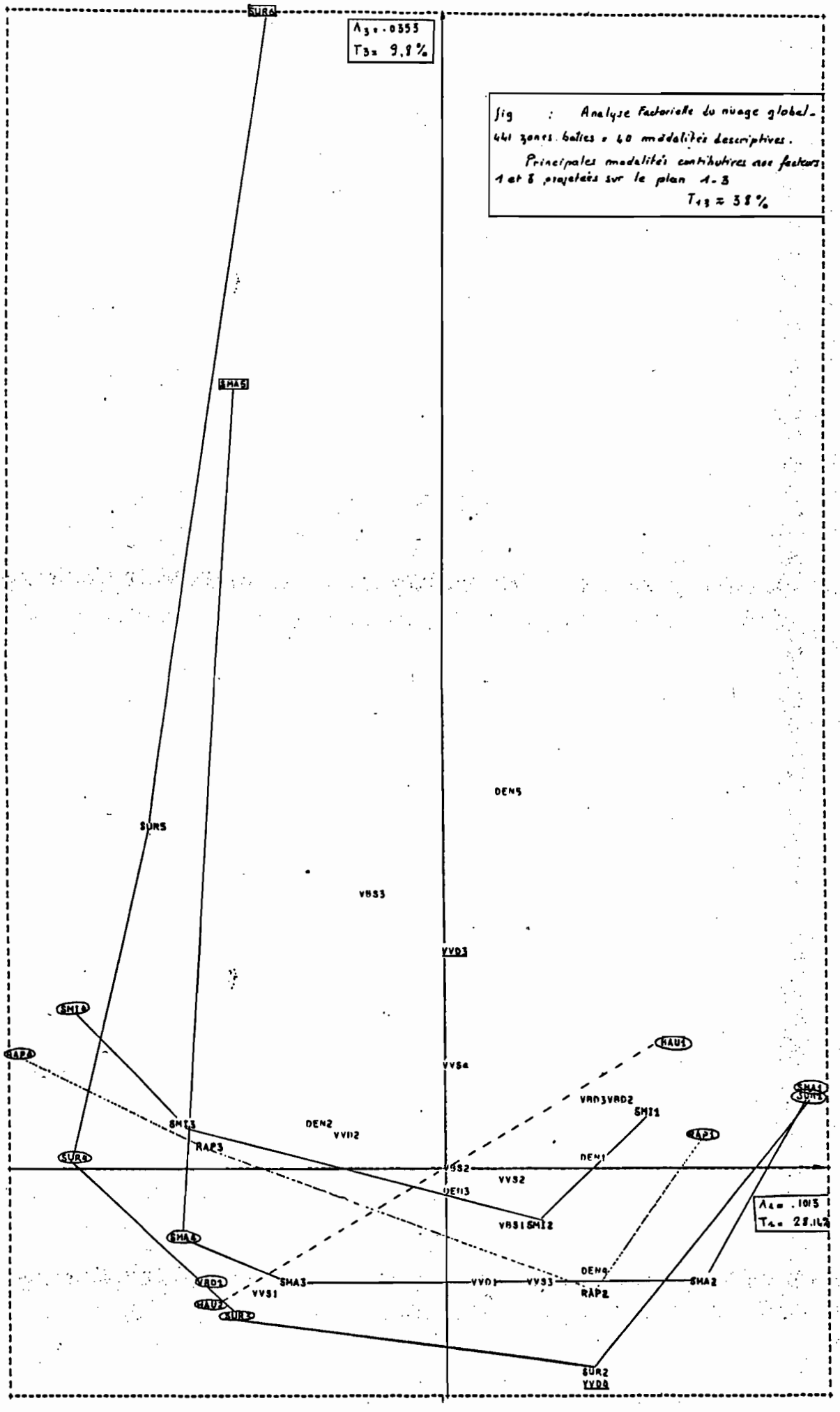


Fig : Analyse factorielle du nuage global - 441 zones balles - 40 modalités descriptives.

Principales modalités contributrices aux facteurs 1 et 3 projetées sur le plan 1-3

$T_{1\,3} \simeq 38\%$

Fig. 2.1.4.3

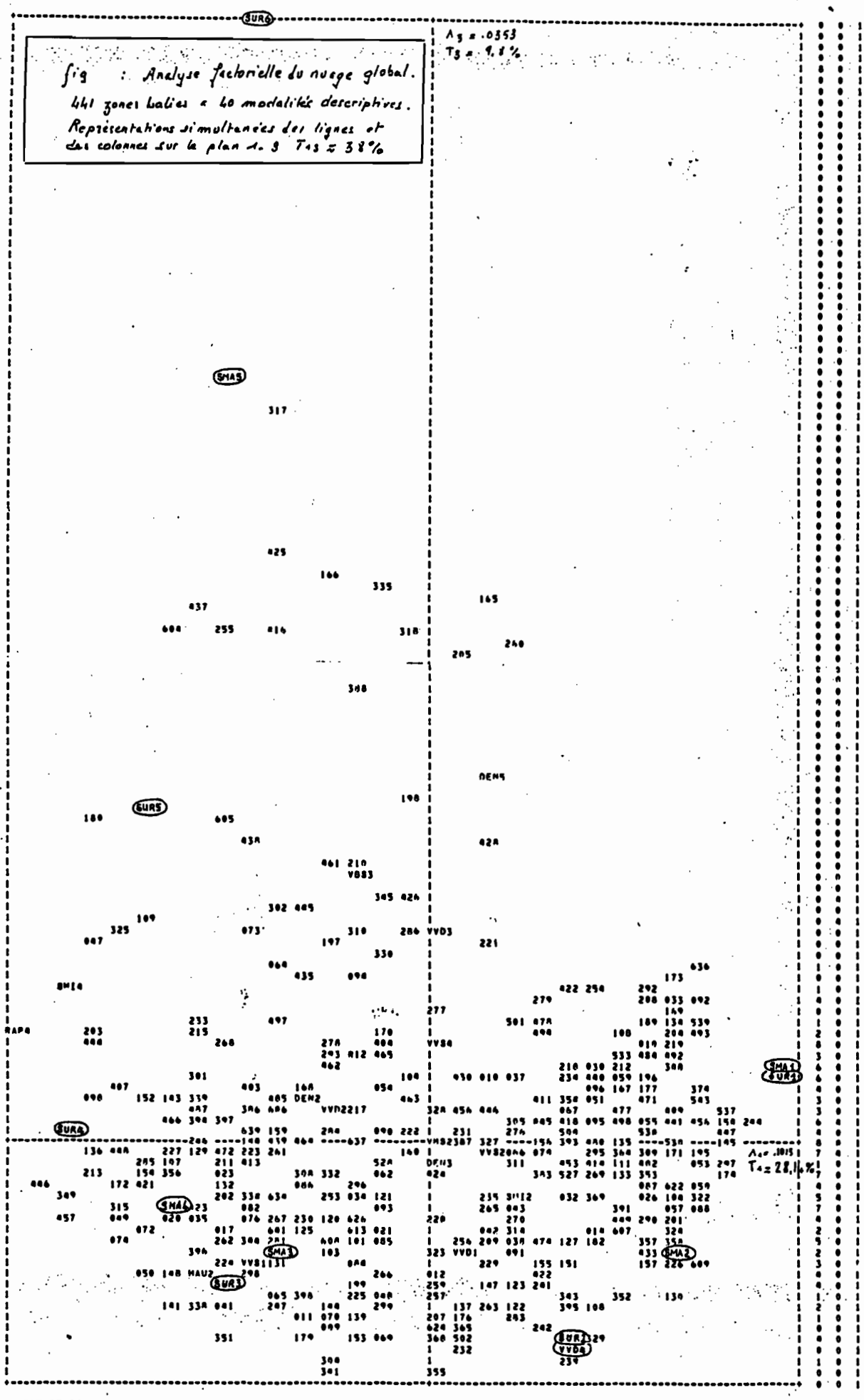

AXE HORIZONTAL(1)--AXE VERTICAL(3)--TITRE: MARSEILLE : A . F . C . DU TABLEAU COMPLET , SANS PONDERATION .24 / 3 / 86
NOMBRE DE POINTS : 481
fig : Analyse factorielle du nuage global.
441 zones balies à 40 modalités descriptives.
Représentations simultanées des lignes et des colonnes sur le plan 1. 3 T13 ≈ 38%
A3 = .0353
T3 = 9.8 %
SUR6
SMA5
SUR5
SUR4
SMA1
SUR1
SMA4
SMA3
SMA2
SUR3
SUR2
VVD4
DEN5
VVD3
VVD1
DEN3
VVB4
VVB1
MAU2
AP6
BMI4
A4 = .0015
T4 = 28.1 %
NOMBRE DE POINTS SUPERPOSES : 124

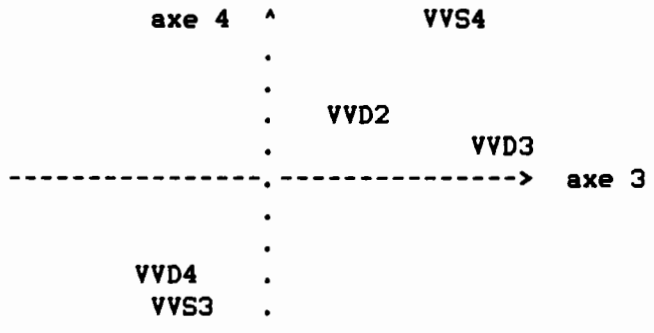
Fig. 2.1.4. 6

SOMME DES INDICES DE NIVEAU 0.25000E+01

J	I(J)	A(J)	B(J)	T(J)	T(Q)	HISTOGRAMME DES INDICES DE NIVEAU DE LA HIERARCHIE
395	236	393	394	94	94	**********
394	118	391	392	47	142	**********
393	104	388	390	42	183	**********
392	90	389	377	36	219	**********
391	71	361	333	29	248	**********
390	68	384	373	27	275	**********
389	67	380	387	27	302	**********
388	66	364	386	26	328	**********
387	65	358	379	26	354	**********
386	51	378	385	20	375	**********
385	42	372	383	17	391	**********
384	40	354	362	16	407	********
383	39	381	382	16	423	********
382	37	360	376	15	438	********
381	34	352	375	14	451	*******
380	34	367	374	14	465	*******
379	31	356	368	12	477	*******
378	28	340	292	11	488	******
377	27	357	314	11	499	******
376	26	359	370	10	509	*****
375	22	350	363	9	518	*****
374	21	369	371	9	527	*****
373	21	290	366	8	535	****
372	21	296	365	8	543	****
371	21	329	298	8	552	****
370	20	322	346	8	560	****
369	20	345	291	8	567	****
368	19	306	299	7	575	****
367	18	335	343	7	582	****
366	18	337	344	7	589	****
365	18	355	328	7	597	****
364	18	339	349	7	604	****
363	17	330	309	7	611	****
362	16	313	351	7	617	***
361	15	245	334	6	623	***
360	15	324	302	6	629	***
359	14	331	336	6	635	***
358	14	6	353	6	641	***
357	14	282	347	6	647	***
356	14	249	286	6	652	***
355	14	318	320	6	658	***
354	14	94	341	5	663	***
353	14	1	332	5	669	***
352	14	312	308	5	674	***
351	13	338	348	5	679	***
350	13	315	304	5	684	***
349	13	327	342	5	690	***
348	12	317	293	5	695	***
347	12	26	57	5	699	***
346	12	289	307	5	704	***
345	12	303	323	5	709	***
344	12	301	90	5	713	**
343	11	28	325	4	718	**
342	11	272	257	4	722	**
341	11	287	270	4	727	**
340	11	253	279	4	731	**
339	11	41	288	4	735	**

Troisième facteur :
Si l'on restreint l'ensemble des caractères morphologiques aux variables rendant compte de la taille du bâti : SURF, SMIN et SMAX, le troisième facteur apparait comme fonction parabolique du premier. On obtient sur le plan 1-3 (figures 2.1.4 4 et 5), le long de cette parabole, un classement des zones suivant la taille des éléments bâtis qu'elles contiennent. Ce facteur explique 8% de l'inertie totale du nuage et les trois variables citées y contribuent pour 52%.

Quatrième facteur :
Il rend compte principalement de deux caractéristiques du nuage:
1. L'écart important au centre de gravité des zones comportant de trés grands bâtiments. La modalité SUR6 représente 12,8% de l'inertie de l'axe et SMA5 9,2%.
2. L'opposition entre les zones où la valeur de l'environnement a une dominante foncée (VVD4: présence importante d'arbres) et une sous-dominante gris moyen (VVS3) d'une part, et celles ou la valeur dominante est plus claire (VVD2 ou 3) et la sousdominante foncée (VVS4) d'autre part. A ce deuxième type appartiennent en particulier les zones à vocation agricole situées à la périphérie de Marseille. Cette opposition, trés visible sur le plan factoriel 1-4, peut se schématiser comme suit:



Classification ascendante hiérarchique

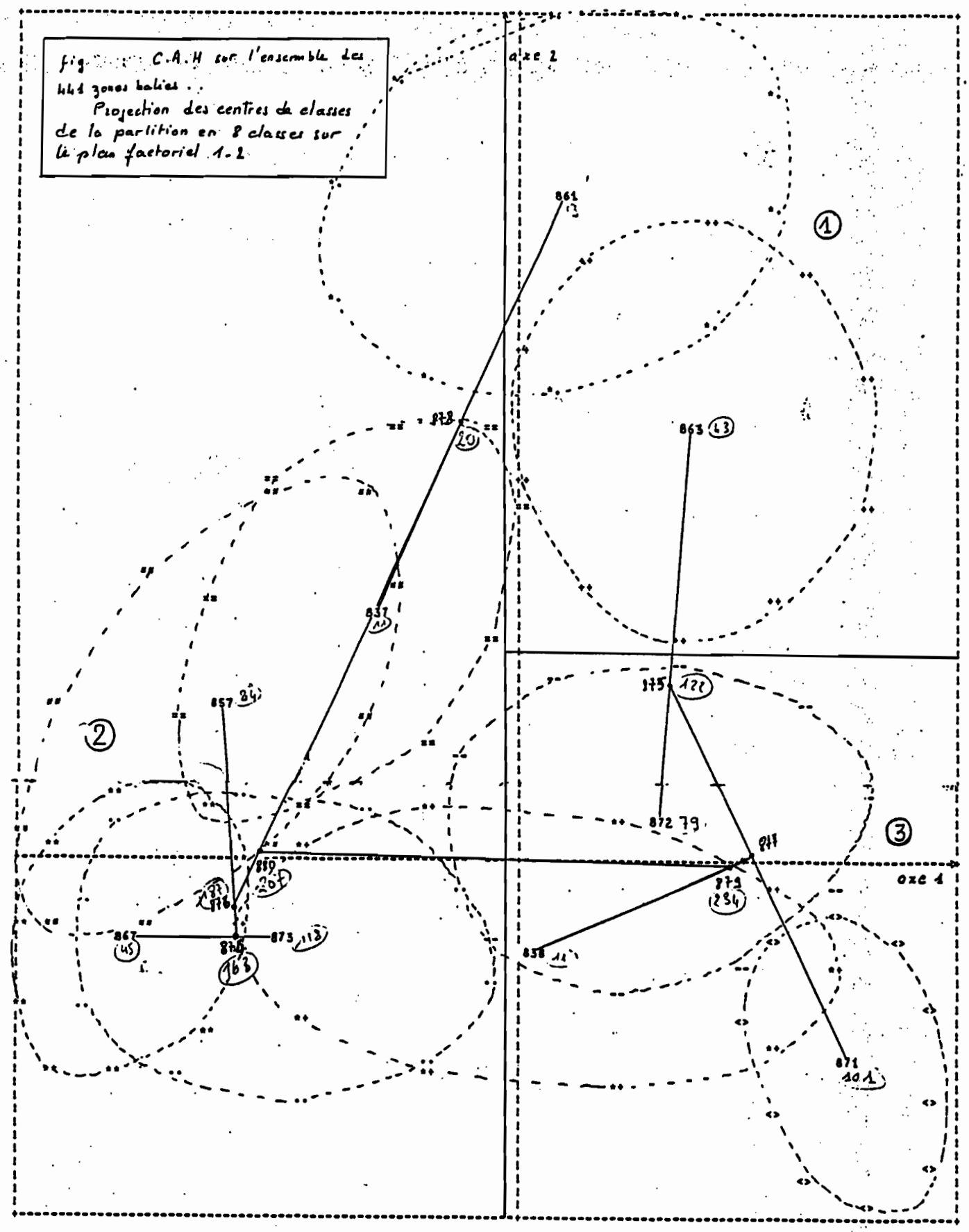
- Choix de la partition étudiée

Ce choix résulte d'un compromis effectué entre la logique de l'algorithme et deux contraintes imposées par les objectifs de la recherche :
1. L'histogramme des indices de niveaux de la hiérarchie présente des seuils significatifs aux passages de 14 à 13 classes, de 10 à 9, de 8 à 7 puis de 7 à 6, de 3 à 2 et enfin de 2 à 1 classes (voir figure 2.1.4 6). L'examen des partitions en deux et trois classes allant de soi, on a donc le choix ensuite de poursuivre l'interprètation jusqu'à 7, 8, 10 ou 14 classes.
2. L'objectif stratificatoire de l'analyse veut garantir une représentativité morphologique d'un échantillon de travail tiré dans la base de sondage des districts de recensement de l'INSEE. Ceux ci sont au nombre d'environ 4000 et le taux de sondage pressenti est de 2%. Pour avoir au minimum deux districts par strate il faut donc au moins cent districts dans toutes les strates, ce qui compte tenu de la taille moyenne des zones homogénes dèfinies lors de la photo-interprétation, nous contraint à n'étudier que les partitions où toute strate regroupe un minimum de 25 à 30 zones. La partition en 14 classes apparait d'ors et déjà trop fine au regard de cet impératif.

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE:DEMANDE : 1 - 9 CLASSES ET ELLIPSES D'INERTIE Fig. 2.1.4. 7

NOMBRE DE POINTS : 99

fig : C.A.H sur l'ensemble des 441 zones balisées.
Projection des centres de classes de la partition en 8 classes sur le plan factoriel 1-2



NOMBRE DE POINTS SUPERPOSES : 1

== (..)

3. La hiérarchie de partitions proposée par la C.A.H peut facilement être mise en relation avec celle qu'établit dans la série des descripteurs morphologiques l'analyse factorielle. C'est d'ailleurs par ce biais que l'on caractérisera en termes morphologiques chaque classe. Malheureusement, le pouvoir discriminant d'un descripteur morphologique donné, d'ailleurs en grande partie fonction des codages successifs de l'information, n'est pas toujours en rapport avec l'intérêt thématique qu'il présente du point de vue démographique. On mesurera ,dans la suite combien ce problème a pesé sur la détermination de la partition retenue. Disons tout de suite qu'aprés ce constat, nous nous sommes sentis autorisés à pratiquer certains regroupements entre classes qui contredisent la logique de l'agrégation hiérarchique. Prenant le risque de scandaliser certains taxinomistes, nous espérons en revanche que le résultat sera efficace pour le télédétecteur comme pour le démographe.

- Les partitions en sept et huit classes

Nous verrons par la suite qu'on ne peut pas retenir ces partitions comme fournissant une stratification acceptable de la base de sondage, par contre elles servent de base à la division de l'ensemble des zones en trois grands sous-ensembles sur lesquels, moyennant la pondération de certains descripteurs, on obtiendra par de nouvelles C.A.H. un résultat convenable. Dans ce qui suit on se contente de caractériser sommairement les classes, sans retourner au tableau de données, en examinant la position des centres de classes sur les plans factoriels principeaux : 1-2 et accessoirement 1-3 (figures 2.1.4 7 et 8).

Sur la figure 7, on caractérise facilement les deux classes de la première partition : d'une part le sous-ensemble des zones, dont les projections sont situées dans le demi-plan droit (premier facteur positif), caractérisé globalement par les modalités HAU1, SUR1, SUR2, SMI1, RAP1, RAP2 (classe 879), d'autre part le sous-ensemble situé à gauche du même plan (premier facteur négatif), groupé autour des modalités HAU2, VBD1, SUR3, SUR4, SMA4, SMA5, SMI4 et RAP4 (classe 880). Ce deuxième ensemble est ensuite scindé en deux classes : 878 et 876. La classe 876 conserve plus de 90% des zones qui composaient 880 et sa caractésation n'en diffère guère en moyenne. Néammoins elle est plus homogène du point de vue de la taille des éléments bâtis et de leur densité puisque la classe 878 regroupe les 20 zones ayant soit une trés grande taille de bâti (SUR6), soit la densité de bâti maximale (DEN5).
La partition en sept classes divise chacune des trois classes 879, 878 et 876 :

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

La classe 878 éclate en 861 et 837. On caractérise facilement ces deux classes sur les plans 1-2 et 1-3 : 837 regroupe les zones où les bâtiments ont la taille maximale (SUR6 et SMA5), tandis que 861

AXE HORIZONTAL(1)--AXE VERTICAL(3)--TITRE:DEMANDE : 2 - 9 CLASSES ET ELLIPSES D'INERTIE *Fig. 2.1.4. 8*

NOMBRE DE POINTS : 99

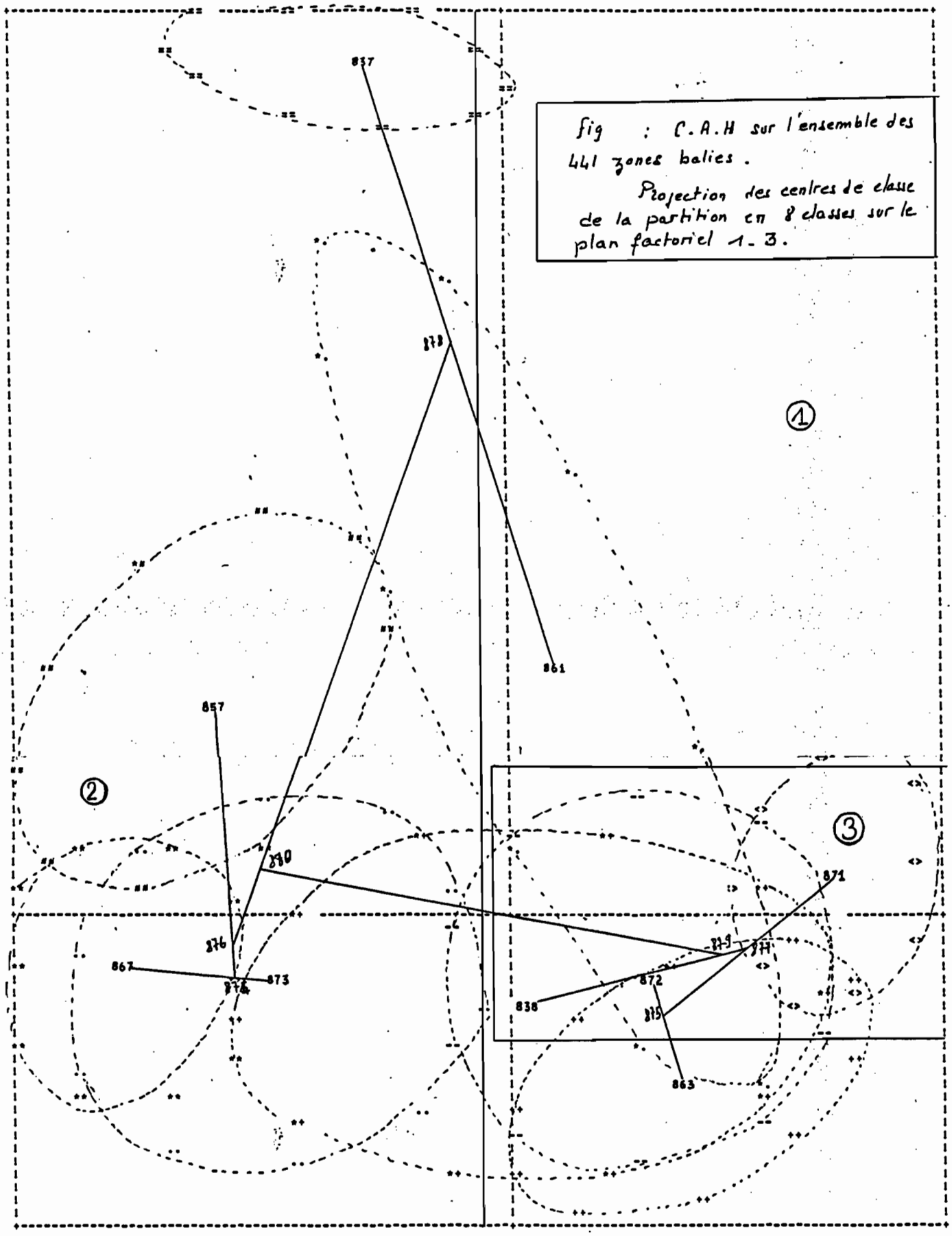


fig : C.A.H sur l'ensemble des 441 zones balies.
Projection des centres de classe de la partition en 8 classes sur le plan factoriel 1-3.

NOMBRE DE POINTS SUPERPOSES : 1

*, (<>)

contient celles où la densité du bâti est maximale. Soulignons au passage la variance trés élevée du troisième facteur dans cette classe : la taille des bâtiments y est donc trés variable.

La classe 876 fournit les classes 857 et 874. Sur les plans factoriels 1-2 et 1-3, 857 apparait comme groupant les zones ou le bâti est de grande taille : SUR4 et plus encore SUR5, SMI3 ou SMI4, RAP4, mais de densité variable (forte variance du second facteur dans la classe). 874 contient le restant des zones situées dans la moitié gauche du plan 1-2 (163 zones) caractérisées par SUR3 ou SUR4, RAP3 ou RAP4, SMI3 ou SMI4, HAU2 et VBD1.

Enfin la partition en huit classes présente l'intérêt de couper la classe 875, dont on soulignait plus haut l'hétérogénéité, en deux parties, les classes 863 et 872.
Dans 863 sont groupées 43 zones caracterisées par les modalités DEN4 et VBD3 (valeur sombre du bâti dominant),SUR2, RAP2 et SMI1. Les 79 zones restantes forment la classe 872 qui présente une élipse d'inertie légérement allongée le long de l'axe 1 : La surface des bâtiments y est variable de SUR1 à SUR3 et la densité également de DEN1 à DEN3.

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

Deux types de vérification de la typologie obtenue à l'issue de l'analyse du nuage des 441 zones ont été pratiqués :
En premier lieu, une mission à Marseille a permis de vérifier, sur un échantillon représentatif de chaque classe de la partition, la signification réelle de l'information receuillie sur les photographies aériennes. On disposait d'une centaine de zones test à partir desquelles on a pu étalonner plus précisement le codage de l'information visuelle. Ceci a naturellement conduit a une phase de corection du fichier de données et l'information morphologique relative à une quarantaine de zones s'est trouvée ainsi rectifiée. Bien sur les résultats présentés jusqu'à maintenant intégrent ces corrections.
La deuxiéme vérification consiste à cartographier le résultat de la classification pour qu'il puisse être superposé à la mosaïque photographique et soumis à critique visuelle, ce qui nous a conduit aux conclusions suivantes :

1. Les classes 861 et 863, bien qu'issues chacune d'une des deux classes de la première partition, peuvent être réunies sans inconvénients thématiques : elles représentent les quartiers les plus denses probablement trés spécifiques du point de vue démographique. Le seuil entre les modalités 4 et 5 de la variable "densité du bâti", qui crée la principale différence entre les profils des deux classes, n'est pas apparu suffisement robuste lors du codage pour justifier de séparer, dans cet ensemble, les zones où la densité serait maximale (DEN5), des zones ou elle est simplement trés forte (DEN4).
Par contre, la forte contribution au deuxième facteur des deux modalités a toujours permis de différentier clairement ce sous-ensemble dense du reste du nuage des zones : il s'agit bien en gros du centre ville plus ou moins ancien de Marseille.

2. La partition en huit classes propose une division de la classe 880, ensemble de zones caractérisées par les modalités de rangs élevés des variables décrivant les surfaces bâties, principalement fondée sur la

taille et la densité du bâti. Hormis la classe 861 évoquée ci dessus, apparaissent les classes 837, 857 et 874. La vérification visuelle permet d'affirmer que 837 regroupe les zones où dominent les plus grands bâtiments, qui sont toutes à vocation industrielle. Mais ces onze zones ne représentent qu' une minorité des quartiers industriels qui couvrent, au total une quarantaine de zones. Dans les classes 857 et 874, les trente zones industrielles restantes ne ssont pas distinguées des zones ou les immeubles d'habitation dominent : dans 857 il y 11 zones à dominante industrielle sur 24 et dans 874, 19 sur 163.

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

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

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

Deuxièmement : à l'intérieur des grands groupes définis ci dessus, la partition exploitée pêche par deux points aux yeux des thématiciens:
- Dans l'ensemble des zones majoritairement occupées par les grands bâtiments, elle ne distingue pas de façon sure celles à vocation industrielle.
- Le sous groupe des zones, généralement moins denses, occupées par les petits bâtiments, contient des zones ou l'habitat collectif domine sous forme de petits immeubles à étages et des zones à dominante pavillonnaire. D'autre part il éxiste également dans cet ensemble une différence importante dans la densité du bâti entre les zones que leur densité permet de considérer comme urbaine à part entière (DEN2 et DEN3) et des zones, généralement situées aux marges de l'agglomération, où l'espace bâti est marginal (DEN1).

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

pondération. Or l'élévation du bâti apparait bien déterminante pour séparer des zones ou les bâtiments ont une forme, une taille et une densité comparables, mais des vocations différentes : qu'il s'agisse de distinguer l'industriel du grand bâtiment d'habitation ou le petit immeuble du pavillon, la hauteur du bâtiment joue.
Pour valider cette hypothèse et améliorer la pertinance démographique de la typologie, on a donc pris les partis suivants :

Premièrement, conserver l'acquit de cette première analyse en séparant, d'aprés ses résultats, le fichier de départ en trois sous-ensembles :
- i. L'ensemble des zones où la densité du bâti est forte ou trés forte (DEN4 et DEN5), constitué des classes numéros 861 et 863 de la partition en huit classes. Cet ensemble contient 52 zones qui constituent dés maintenant la première strate.
- ii. L'ensemble des zones où dominent les grands bâtiments, du moins en surface occupée. Constitué des classes numéros 837, 857 et 874, cet ensemble compte 198 zones.
- iii. Les 191 zones où le bâti est majoritairement de petite taille contenues dans les classes numéros 871, 872 et 838.

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

-B : ANALYSE DES SOUS FICHIERS PONDERES ISSUS DE LA PREMIERE CLASSIFICATION

i : Le sous-fichier des zones à grands batiments.

Analyse factorielle

Premier facteur : En privilégiant, par multiplication par un facteur 3, la variable "élévation du bâti", on a évidement attiré artificiellement le premier facteur vers le dipôle HAUT1, HAUT2, qui explique 66% de l'inertie du facteur. Assez bien corrélées à la modalité HAUT1, les modalités SUR6 et SMA5, caractérisant les grands batiments, ont à elles deux, une contribution de 11% à l'inertie de l'axe. Ceci est du à la cooccurrence presque systématique dans le tableau de données, entre les plus grandes tailles de bâti et la faible élévation. L'axe peut donc être interprété comme le facteur qui permettra d'extraire les zones à dominante industrielle (pour le détail des contributions, voir le tableau de la figure 2.1.4. 9).

Deuxième facteur : Les cinq premières contributions à ce facteur sont le fait des modalité de rang 1 des variables suivantes: "surface de l'élément bâti majoritaire" (13,4% de l'inertie du facteur), "rapport

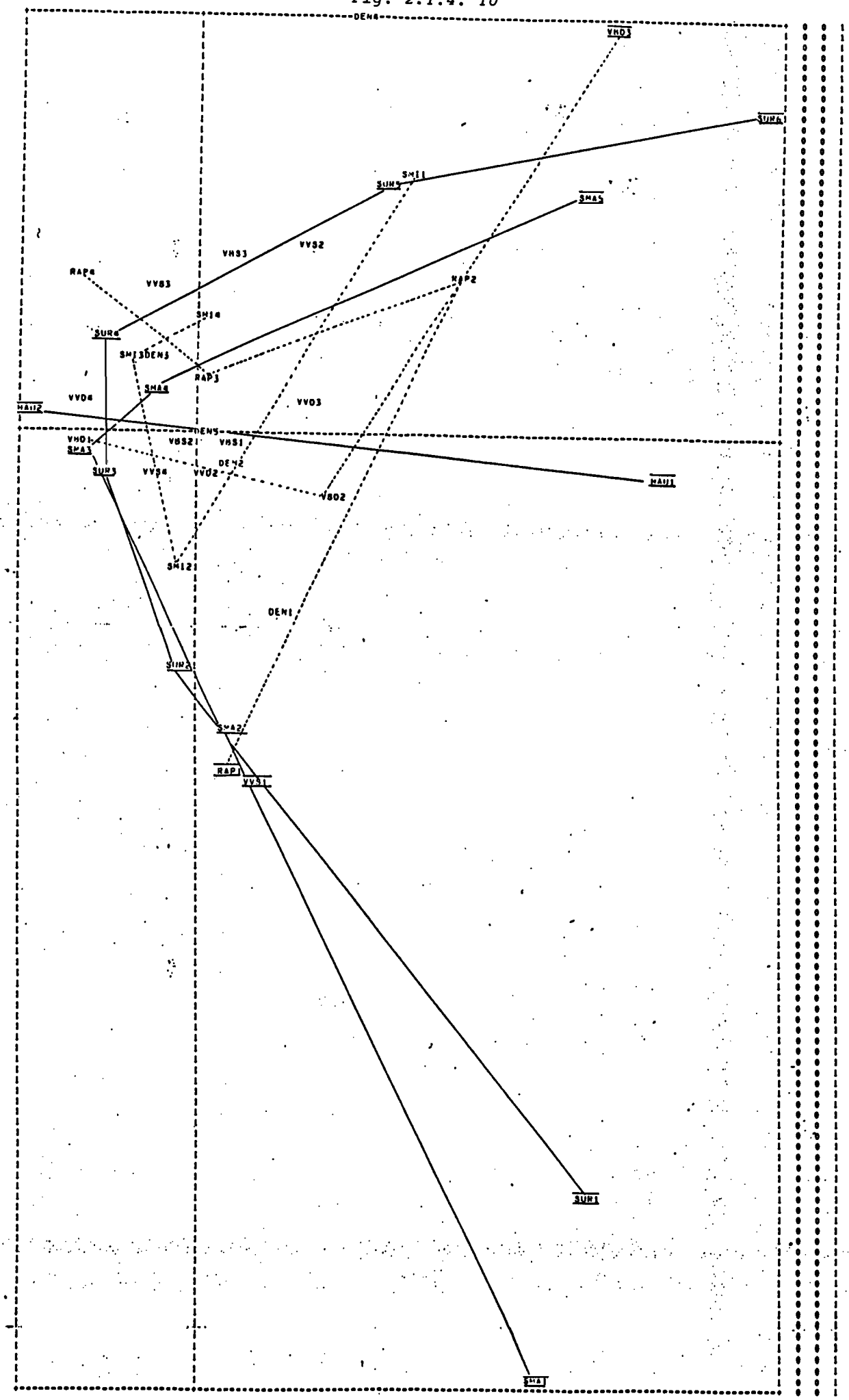
Fig. 2.1.4. 9

J1	QLT	POID	INR	1#F	COR	CTR	2#F	COR	CTR	3#F	COR	CTR	4#F	COR	CTR	5#F	COR	CTR	6#F	COR	CTR	7#F	COR	CTR
1 DEN1	657	14	28	221	10	2	-582	70	25	547	62	28	-385	31	15	254	13	7	-484	49	28	1426	422	252
2 DEN2	514	34	20	30	1	0	-118	9	2	-18	0	0	426	123	43	398	107	40	582	230	95	-252	43	18
3 DEN3	488	32	21	-181	20	3	239	36	10	-221	30	10	-140	12	4	-622	241	94	-386	93	40	-299	56	25
4 DEN4	213	3	32	478	10	2	1384	81	34	-50	0	0	-1297	71	40	856	31	19	-98	0	0	-704	21	14
5 DEN5	0	(0)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 SMI1	259	11	29	664	64	17	856	106	40	-244	9	4	-557	45	23	207	6	3	-449	29	18	-50	0	0
7 SMI2	390	41	17	-62	4	0	-452	196	44	-317	97	26	-71	5	1	98	9	3	99	9	3	271	70	26
8 SMI3	295	15	27	-261	15	3	218	11	4	343	26	11	684	104	50	-140	4	2	-621	86	48	-469	49	29
9 SMI4	248	17	27	-31	0	0	365	34	12	613	95	41	-96	2	1	-241	15	7	601	91	50	-204	11	6
10 SMA1	302	(2)	33	1024	22	5	-3061	193	83	-500	5	3	245	1	1	-644	9	5	-1522	48	32	-1089	24	17
11 SMA2	316	8	30	48	0	0	-971	94	37	1170	137	67	-838	70	38	180	3	2	-285	8	5	174	3	2
12 SMA3	368	29	22	-425	99	15	-78	3	1	-188	19	7	-475	124	47	159	14	6	375	77	34	-244	32	15
13 SMA4	488	31	21	-197	23	4	139	12	3	-193	22	7	682	277	102	152	14	5	-467	130	56	126	9	4
14 SMA5	611	13	28	1231	292	59	778	117	43	262	13	6	-96	2	1	-721	100	53	611	72	42	280	15	9
15 V8D1	453	47	14	-453	267	28	-47	3	1	135	24	6	249	81	21	206	55	15	99	13	4	92	11	3
16 V8D2	356	29	22	417	95	15	-226	28	8	-255	35	12	-178	17	7	-568	177	72	1	0	0	-81	4	2
17 V8D3	542	7	31	1343	158	35	1317	153	61	170	3	1	-967	82	45	1045	96	55	-695	43	27	-290	7	5
18 V8S1	279	35	19	53	2	0	-48	2	0	455	153	47	-98	7	2	-125	12	4	-53	2	1	-372	102	42
19 V8S2	321	41	17	-55	3	0	-55	3	1	-397	155	42	-110	12	4	234	54	17	-140	19	7	279	76	28
20 V8S3	326	7	31	57	0	0	588	30	12	45	0	0	1187	124	67	-775	53	31	1135	113	72	243	5	3
21 VVD1	0	(0)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22 VVD2	485	36	19	-16	0	0	-136	14	3	-204	32	10	-45	2	1	-680	355	127	-151	18	7	-289	64	26
23 VVD3	602	29	22	307	50	8	106	6	2	-553	163	57	22	0	0	643	221	91	348	65	29	423	96	45
24 VVD4	578	18	26	-461	59	11	101	3	1	1294	464	195	54	1	0	328	30	15	-256	18	10	-100	3	2
25 VVS1	501	13	28	163	5	1	-1136	249	91	506	49	22	-99	2	1	-492	47	25	-827	132	76	300	17	11
26 VVS2	511	20	25	277	24	4	624	121	40	-49	1	0	-714	159	71	768	184	88	-263	22	11	61	1	1
27 VVS3	534	18	26	-142	5	1	459	57	19	1041	292	123	737	146	68	-218	13	6	265	19	10	88	2	1
28 VVS4	516	32	20	-159	16	2	-159	16	4	-748	356	117	75	4	1	-146	14	5	360	82	35	-210	28	12
29 HAU1	932	76	70	1450	914	461	-155	10	10	-18	0	0	93	4	5	-20	0	0	-77	3	4	-49	1	2
30 HAU2	932	174	30	-630	914	200	67	10	4	8	0	0	-40	4	2	9	0	0	33	3	2	21	1	1
31 SUR1	538	4	32	1193	76	17	-2465	323	134	-742	29	15	1010	54	30	756	30	18	219	3	2	-651	23	15
32 SUR2	190	5	31	-80	0	0	-776	42	17	1124	89	44	60	0	0	-210	3	2	175	2	1	871	53	36
33 SUR3	509	37	19	-337	91	12	-162	21	5	98	8	2	-579	268	88	37	1	0	153	19	7	-357	102	41
34 SUR4	501	21	25	-361	44	8	305	31	10	-449	68	27	365	45	20	-344	40	19	-436	64	33	784	208	112
35 SUR5	490	11	29	586	50	10	834	101	38	-14	0	0	727	76	39	583	49	27	-572	47	29	-1075	167	105
36 SUR6	574	5	31	1844	219	50	1073	74	30	579	22	11	302	6	3	-452	13	8	1516	148	96	1191	92	62
37 RAP1	605	19	26	54	1	0	-1108	361	122	340	34	14	166	8	4	630	117	57	416	51	27	-335	33	18
38 RAP2	210	9	30	814	79	17	518	32	12	514	31	15	590	41	22	243	7	4	-400	19	12	-77	1	0
39 RAP3	467	34	20	19	0	0	154	16	4	-34	1	0	-598	248	86	-502	175	65	167	19	8	107	8	3
40 RAP4	425	21	25	-413	59	11	519	93	30	-458	73	29	561	109	48	141	7	3	-467	76	39	158	9	5
			10001			10001			10001			10001			10001			10001			10001			10001

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE: MARSEILLE : A . F . C . DU SOUS - TABLEAU GRANDS BATS . LE 13 / 02 / 86 3 PARA

NOMBRE DE POINTS : 80

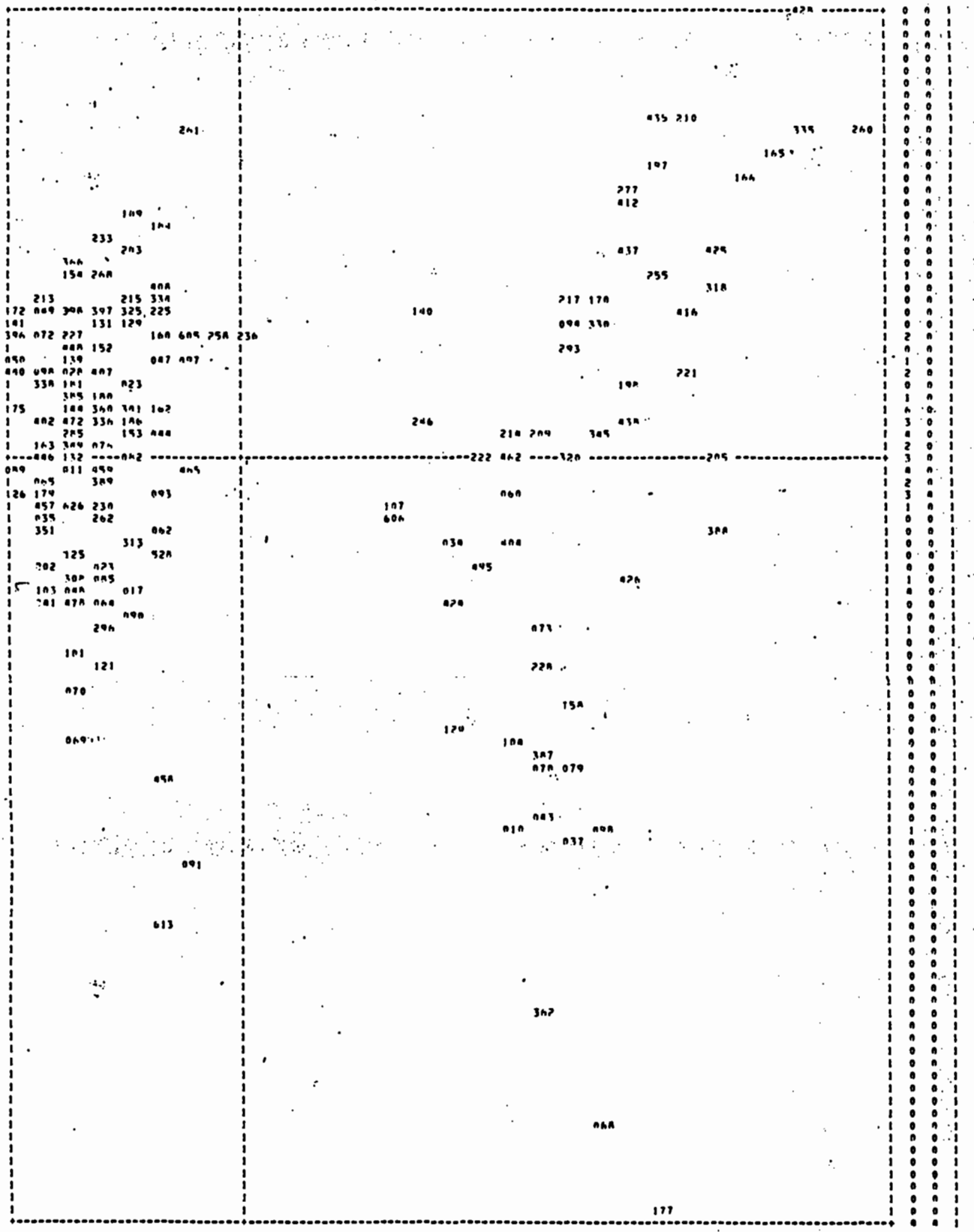
Fig. 2.1.4. 10



AXE HORIZONTALE(1)--AXE VERTICAL(2)--TITRE: MARSEILLE : A . F . C . DU SOUS - TABLEAU GRANDS HAUTS . LE 26 / 03 / 86 E GRAP

NOMBRE DE POINTS : 198

Fig. 2.1.4. 11



NOMBRE DE POINTS SUPERPOSES : 43

253 (184) 298 (194) 315 (072) 413 (168) 247 (139) 143 (028) 632 (094) 466 (385) 298 (162) 352 (301)
386 (348) 421 (175) 447 (179) 469 (184) 392 (336) 638 (072) 639 (334) 159 (153) 339 (153) 343 (285)
445 (209) 403 (076) 608 (076) 228 (132) 348 (132) 461 (328) 097 (011) 349 (011) 356 (011) 637 (465)
078 (065) 475 (349) 873 (068) 287 (179) 888 (179) 337 (238) 627 (085) 086 (088) 188 (088) 199 (183)
281 (088) 839 (296) 056 (010)

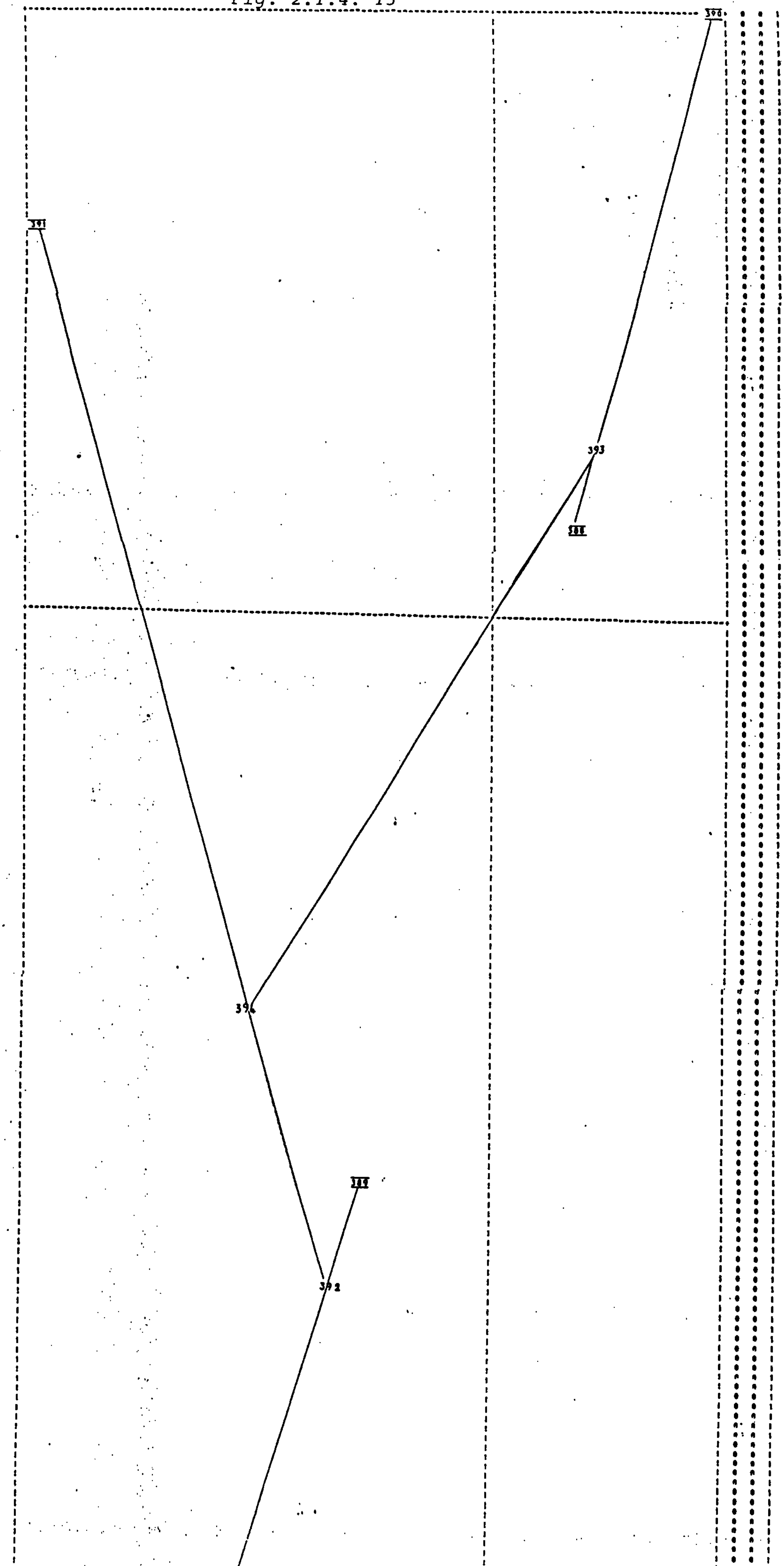
IMME DES INDICES DE NIVEAU 0.25000E+01 *Fig. 2.1.4. 12*

J	I(J)	A(J)	B(J)	T(J)	T(0)	HISTOGRAMME DES INDICES DE NIVEAU DE LA HIERARCHIE
3951	2361	3931	3941	941	941	
3941	1181	3911	3921	871	1421	
3931	1041	3881	3901	821	1831	
3921	901	3891	3771	361	2191	
3911	711	3611	3331	291	2481	
3901	681	3841	3731	271	2751	
3891	671	3801	3671	271	3021	
3881	661	3641	3061	261	3281	
3871	651	3581	3791	261	3541	
3861	511	3781	3851	201	3751	
3851	421	3721	3831	171	3911	
3841	401	3541	3621	161	4071	
3831	391	3811	3821	161	4231	
3821	371	3601	3761	151	4381	
3811	341	3521	3751	141	4511	
3801	341	3671	3741	141	4651	
3791	311	3561	3681	121	4771	
3781	281	3401	2921	111	4881	
3771	271	3571	3141	111	4991	
3761	261	3591	3701	101	5091	
3751	221	3501	3631	91	5181	
3741	211	3691	3711	91	5271	
3731	211	2901	3661	81	5351	
3721	211	2961	3651	81	5431	
3711	211	3291	2981	81	5521	
3701	201	3221	3461	81	5601	
3691	201	3451	2911	81	5671	
3681	191	3061	2991	71	5751	
3671	181	3351	3431	71	5821	
3661	181	3371	3441	71	5891	
3651	181	3551	3281	71	5971	
3641	181	3391	3491	71	6041	
3631	171	3301	3091	71	6111	
3621	161	3131	3511	71	6171	
3611	151	2951	3341	61	6231	
3601	151	3241	3021	61	6291	
3591	141	3311	3361	61	6351	
3581	141	61	3531	61	6411	
3571	141	2821	3471	61	6471	
3561	141	2491	2861	61	6521	
3551	141	3181	3201	61	6581	
3541	141	941	3411	51	6631	
3531	141	11	3321	51	6691	
3521	141	3121	3081	51	6741	
3511	131	3381	3481	51	6791	
3501	131	3151	3041	51	6841	
3491	131	3271	3421	51	6901	
3481	121	3171	2931	51	6951	
3471	121	261	571	51	6991	
3461	121	2891	3071	51	7041	
3451	121	3031	3231	51	7091	
3441	121	3011	901	51	7131	
3431	111	281	3251	41	7181	
3421	111	2721	2571	41	7221	
3411	111	2831	2701	41	7271	
3401	111	2531	2791	41	7311	
3391	111	411	2881	41	7351	
3381	111	2651	2561	41	7401	
3371	101	2771	2601	41	7441	
3361	101	3211	2751	41	7481	
3351	101	191	2691	41	7521	
3341	101	1311	1021	41	7561	
3331	101	2941	3161	41	7601	
3321	91	81	3261	41	7641	
3311	91	2761	2301	41	7671	
3301	91	3111	2711	41	7711	
3291	91	1271	3191	41	7751	
3281	91	2971	2641	41	7781	
3271	91	3101	711	41	7821	
3261	91	2781	2731	41	7851	
3251	91	2741	301	31	7891	
3241	81	2281	2621	31	7921	
3231	81	2251	2521	31	7951	
3221	81	2611	1491	31	7981	
3211	81	2291	3001	31	8021	
3201	81	2231	3051	31	8051	
3191	81	401	2811	31	8081	
3181	71	2361	2671	31	8111	
3171	71	1011	1261	31	8141	
3161	71	1211	2631	31	8161	
3151	71	2161	2951	31	8191	
3141	71	321	331	31	8221	
3131	71	2171	1731	31	8251	
3121	71	2321	2001	31	8281	
3111	71	2551	2371	31	8301	
3101	71	2511	1931	31	8331	
3091	71	2581	2481	31	8361	
3081	71	2401	2031	31	8381	
3071	71	441	2541	31	8411	
3061	71	21	2801	31	8441	
3051	71	651	2151	31	8461	
3041	71	1911	2341	31	8491	
3031	71	121	871	31	8521	
3021	61	691	2851	31	8541	
3011	61	1791	1191	31	8571	
3001	61	2101	2021	31	8591	
2991	61	1441	2661	31	8621	
2981	61	2351	1011	31	8641	
2971	61	211	2141	21	8671	
2961	61	141	2681	21	8691	
2951	61	2331	2841	21	8711	
2941	61	1821	1371	21	8741	
2931	61	1941	1971	21	8761	
2921	61	2591	2261	21	8781	
2911	61	2421	2431	21	8811	
2901	61	2501	1171	21	8831	
2891	51	2381	931	21	8851	
2881	51	431	1531	21	8871	
2871	51	831	861	21	8891	
2861	51	701	1471	21	8911	
2851	51	231	531	21	8941	
2841	51	2851	2011	21	8961	
2831	51	2091	491	21	8981	
2821	51	1511	2271	21	9001	
2811	51	451	1501	21	9021	
2801	51	51	71	21	9041	
2791	51	351	8471	21	9061	
2781	51	31	181	21	9081	
2771	51	1591	1251	21	9101	
2761	51	981	1121	21	9121	
2751	51	2411	1101	21	9141	
2741	51	161	201	21	9161	
2731	51	2811	111	21	9181	
2721	51	101	251	21	9201	
2711	51	2311	2041	21	9221	
2701	51	851	951	21	9241	
2691	51	241	271	21	9251	
2681	51	2181	291	21	9271	
2671	51	371	2441	21	9291	
2661	51	221	1541	21	9311	
2651	41	821	841	21	9331	
2641	41	2461	801	21	9341	
2631	41	2241	2071	21	9361	
2621	41	511	2391	21	9381	
2611	41	471	1341	21	9401	
2601	41	1201	1401	21	9411	
2591	41	811	911	21	9431	
2581	41	1831	2111	21	9441	
2571	41	1521	1641	21	9461	
2561	41	2121	1551	21	9471	
2551	41	1781	1411	21	9491	
2541	41	581	1181	21	9501	
2531	41	151	2281	21	9521	
2521	41	1001	1071	11	9531	
2511	41	661	721	11	9551	
2501	41	1041	1361	11	9561	
2491	41	1421	1631	11	9581	
2481	41	2221	1851	11	9591	
2471	31	1161	2041	11	9601	
2461	31	551	2131	11	9621	
2451	31	891	991	11	9631	
2441	31	1991	601	11	9641	
2431	31	681	751	11	9661	
2421	31	1041	1711	11	9671	
2411	31	1801	1721	11	9681	
2401	31	631	971	11	9691	
2391	31	1561	1651	11	9711	
2381	31	1761	1771	11	9721	
2371	31	1981	1571	11	9731	
2361	31	391	731	11	9741	
2351	31	311	341	11	9751	
2341	31	1201	1221	11	9771	
2331	31	881	2031	11	9781	
2321	31	421	591	11	9791	
2311	21	501	1351	11	9801	
2301	21	1601	1701	11	9811	
2291	21	1381	1621	11	9821	
2281	21	2191	1691	11	9831	
2271	21	1661	1751	11	9841	
2261	21	921	961	11	9851	
2251	21	791	1901	11	9861	
2241	21	1031	1131	11	9861	
2231	21	481	2061	11	9871	
2221	21	1141	1921	11	9881	
2211	21	41	91	11	9891	
2201	21	1431	1451	11	9901	
2191	21	361	461	11	9911	
2181	21	131	171	11	9911	
2171	21	1111	1961	11	9921	
2161	21	1291	1301	11	9931	
2151	21	671	781	11	9931	
2141	21	611	741	11	9941	
2131	21	561	641	11	9951	
2121	11	1151	1321	11	9951	
2111	11	1951	1871	11	9961	
2101	11	1671	1861	11	9961	
2091	11	381	541	11	9971	
2081	11	1231	1611	01	9971	
2071	11	1091	1891	01	9981	
2061	11	621	771	01	9981	
2051	11	1461	1741	01	9991	
2041	11	1241	1331	01	9991	
2031	11	1081	1881	0110001		
2021	11	1051	1061	0110001		
2011	01	1481	1681	0110001		
2001	01	1391	1581	0110001		
1991	01	521	761	0110001		

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE(DECANOR) = 5 CLASSES DE LA PARTITION

NOMBRE DE POINTS : 5

Fig. 2.1.4. 13



longueur/largeur de l'élément bâti majoritaire" (12,2%), "valeur sous dominante dans l'environnement du bâti" (9,1%), "surface maximum de l'élément bâti" (8,3). Enfin la modalité de rang 3 caractérisant la valeur dominante du bâti, a une contribution de 6,1% à l'inertie de l'axe. S'agissant des quatre premières, ce sont des modalités de poids relativement faibles, qui caractérisent une minorité de zones où dominent les petits bâtiments de forme carrée, à plusieurs étages. Outre qu'il met en évidence ce trait assez marginal du sous-ensemble, le second facteur est interessant lorsque on considere l'ensemble des modalités liées à la taille du bâti : pour les variables "surface de l'élément bâti majoritaire" et "surface maximum de l'élément bâti", le second facteur est une fonction monotone croissante du rang des modalités. C'est également vrai pour les modalités 2,3 et 4 de la surface minimum de l'élément bâti, mais sa modalité a une coordonnée nettement positive sur le deuxième axe, très proche de celle de SMA5. Ce phénomène n'apparaissait pas aussi clairement dans l'analyse du nuage complèt.

Plan factoriel 1-2 : (figures 2.1.4 10 et 11)
Deux faits marquants permettront, à partir de l'examen de la projection des centres de classes sur ce plan, de caractériser facilement les partitions obtenues par C.A.H. : (figure 2.1.4 10)
-1 L'axe 1 est quasiment aligné sur le dipôle HAU1, HAU2. La moitié droite du plan est le lieu exclusif des zones où le bâti est majoritairement bas. Soulignons également la proximité, dans le quart supérieur droit du plan, des modalités SUR5, SUR6, SMA5, VBD3 et SMI1 qui caractériserons les zones où dominent les bâtiments industriels.
-2 Les modalités des variables qui décrivent la taille moyenne et maximum des élément bâtis, forment sur le plan factoriel 1-2, une parabole axée sur le dipôle HAU1, HAU2. Cette situation est à rapprocher de celle décrite lors de l'examen du plan factoriel 1-3 dans l'analyse du nuage global. Elle lui est bien sur comparable mais en observant la projection du nuage des zone sur le plan factoriel 1-2 (figure 2.1.4 11), on constate que la pondération des modalités HAUT1 et HAUT2 éclate le nuage en deux parties chacune grossièrement parabolique et précisement localisée dans les deux moitiés du plan : bâtiments bas à droite, élevés à gauche. L'axe deux lui, apparaît comme une échelle de mesure de la taille du bâti, petits bâtiments du côté des coordonnés négatives, grands bâtiments du côté positif.

Classification hiérarchique

Au vu de l'histogramme des indices de niveaux de la hiérarchie de partitions proposée par l'algorithme (figure 2.1.4 12), les partitions en 2,3,4 et 5 classes représentent des étapes importantes (le seuil significatif suivant est celui de la partition en 10 classes, trop fine pour nos besoins). Nous allons nous intéresser maintenant à décrire la formation de la partition en cinq classes qui nous servira à établir la typologie définitive de ce sous-ensemble.
Comme on pouvait s'y attendre, la partition en deux classes rend compte de la dichotomie introduite entre les modalités HAU1 et HAU2 (figure 2.1.4 13). La classe 392 regroupe les 60 zones situées dans la partie droite du plan factoriel 1-2. Nous caractériserons précisément les classes de la typologie définitive au chapitre suivant. Disons simplement que cette classe contient l'essentiel des zones où dominent les bâtiments industriels qui constitueront la strate numéro 2. Plus de 90% des zones de cette classe possèdent la modalité HAU1 et il s'agit en majorité de zones contenant de très grands bâtiments (SUR5 ou SUR6).

Fig. 2.1.4. 14

	J1	OLT	POIO	IOR	1#F	COR	CTR	2#F	COR	CTR	3#F	COR	CTR	4#F	COR	CTR	5#F	COR	CTR	6#F	COR	CTR	7#F	COR	CTR
1	DEM1	922	59	45	74	3	1	-436	105	52	850	398	219	-240	32	20	310	53	41	-676	252	220	378	79	70
2	DEM2	356	20	30	170	8	2	-246	17	5	-601	100	37	81	2	1	-282	22	11	211	12	7	-838	195	114
3	DEM3	495	35	23	-442	121	24	164	17	4	-412	105	30	133	11	4	-151	14	6	595	220	100	112	8	4
4	DEM4	790	8	66	+2304	260	144	3225	491	359	-307	4	4	377	7	6	192	2	2	-742	26	34	102	0	1
5	DEM5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	SMI1	481	20	29	-301	26	6	191	10	3	-459	60	21	296	25	10	-786	176	89	330	31	17	734	153	88
7	SMI2	438	63	10	39	4	0	-91	21	2	-48	6	1	-192	93	14	234	138	25	-65	11	2	-256	166	34
8	SMI3	274	4	34	32	0	0	-381	8	3	1836	175	73	1257	82	39	-355	7	4	-246	3	2	-45	0	0
9	SMI4	367	1	34	2530	73	21	3664	170	64	2265	59	25	1526	27	13	1463	24	15	710	6	4	848	8	6
10	SMA1	220	30	24	-459	107	22	-35	1	0	238	29	9	-273	38	13	-31	1	0	-68	2	1	-288	42	20
11	SMA2	56	41	20	-32	1	0	-36	1	0	-55	3	1	78	5	1	189	31	11	46	2	1	-122	13	5
12	SMA3	352	9	33	-188	4	1	-32	0	0	-752	63	26	457	23	11	-217	5	3	545	35	23	1411	221	147
13	SMA4	442	6	33	+1843	365	100	337	12	4	-28	0	0	159	3	1	-680	50	29	-335	12	8	-16	0	0
14	SMA5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	VHD1	249	21	28	599	111	26	-628	122	37	-75	2	1	109	4	1	55	1	0	8	0	0	-169	66	15
16	VHD2	346	62	11	-237	129	12	242	135	16	21	1	0	-23	1	0	-75	13	3	11	0	0	980	66	45
17	VHD3	102	5	34	210	3	1	-323	7	3	-341	8	3	-58	0	0	462	15	9	217	3	2	-66	3	1
18	VHS1	195	38	21	-179	24	4	-126	12	3	-237	43	11	324	79	23	-108	17	6	-151	17	7	-194	30	12
19	VHS2	263	39	20	-172	24	4	-176	25	6	278	62	15	-310	77	22	85	2	1	237	45	18	-194	30	12
20	VHS3	491	11	32	+1158	190	60	1088	166	59	-369	19	8	28	0	0	278	11	6	-132	2	2	838	100	63
21	VVD1	667	4	34	1044	57	16	-1224	76	29	2292	275	114	979	50	24	-1062	112	66	756	30	20	-282	4	3
22	VVD2	482	20	28	268	20	5	459	61	19	14	0	0	-699	141	56	-175	9	4	-168	8	4	-914	242	136
23	VVD3	603	29	24	-221	28	5	118	7	2	334	55	16	-656	210	72	240	25	12	479	112	53	586	167	81
24	VVD4	804	35	22	-122	10	2	-204	27	7	-617	250	69	822	443	140	52	2	1	-330	71	31	39	1	0
25	VVS1	686	1	35	1529	39	11	-1090	20	8	3693	226	99	2300	88	44	-3866	250	154	1586	42	29	-1077	19	14
26	VVS2	221	14	31	-266	14	3	-428	35	12	176	6	2	370	26	11	-144	4	2	146	4	2	833	133	81
27	VVS3	730	32	23	-81	4	1	-315	56	15	-464	125	36	679	268	88	199	23	9	-591	203	92	-289	49	22
28	VVS4	803	40	20	81	5	1	447	166	37	129	14	3	-750	467	132	4	0	0	424	149	58	-48	2	1
29	HAU1	799	147	12	-310	475	49	203	204	28	89	39	6	111	61	11	48	11	2	-33	5	1	-23	3	1
30	HAU2	810	29	61	+1489	442	224	-1017	206	138	-598	71	53	-522	54	47	-301	18	19	307	19	22	42	0	0
31	SUR1	723	51	16	-503	341	45	232	72	13	-73	7	1	-195	51	11	-406	221	61	-149	30	9	-1	0	0
32	SUR2	726	26	26	331	46	10	-355	53	15	70	2	1	365	56	20	1022	436	197	509	126	63	-138	8	4
33	SUR3	363	10	32	+1272	218	58	-514	36	13	-432	25	10	-199	5	2	-464	29	16	-533	38	24	301	12	8
34	SUR4	613	1	34	3031	105	30	2967	100	38	4428	224	95	3245	120	58	-2053	48	29	1034	12	8	-583	4	3
35	SUR5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	SUR6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	RAP1	691	57	13	-397	286	31	68	8	1	-79	11	2	-179	58	11	-350	222	51	-221	89	23	94	16	4
38	RAP2	717	24	26	383	56	12	-37	1	0	200	15	5	672	174	65	804	248	115	717	198	102	-255	25	13
39	RAP3	391	7	33	+1960	288	79	-436	16	6	-390	12	5	-869	62	29	-95	1	0	-438	16	10	-26	0	0
40	RAP4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
				1000			1000			1000			1000			1000			1000			1000			1000

Le reste de l'ensemble (138 zones), dont 90% ont une majorité de bâtiments élevés, constitue la classe 393 qui éclate ensuite, donnant les classes 390 et 388. Cette séparation s'effectue principalement sur la taille des éléments bâtis comme le montre la projection des centres de classe sur le plan 1-2 : 390 regroupe 27 zones contenant des bâtiments élevés de petite tailles qui ont, pour la plupart, leur coordonnée sur le deuxième axe négative et constituent une partie de la strate numéro quatre, dite strate des petits immeubles d'habitation. Le gros du sous-ensemble étudié se trouve donc dans la classe 388, dont le centre est le plus proche du centre de gravité de l'ensemble. Ces 111 zones où dominent les grands bâtiments élevés du type grands ensembles d'habitat collectif, constituera la strate numéro 3 dite des grands ensembles d'habitation.
Dans la partition en quatre classes, la classe 394 a donné les classes 392 et 391. Les 11 zones de la classe 391, situées dans la partie inférieure droite du plan 1-2, sont des zones relativement atypiques dans le sous fichier puisqu'y dominent les petits bâtiments bas. Cette classe sera d'ailleur versée, après vérification, dans la strate numéro cinq, majoritairement issue du sous fichier iii des zones à petits bâtiments, dite strate du pavillonnaire.
Enfin, dans la partition en cinq classes, c'est à nouveau la classe 392 qui éclate en 389 et 377. La classe 377 est contituée de sept zones que singularisent leur forte densité de bâti (DEN4). Pour cinq d'entre elles, il s'agit d' immeubles élevés, les deux autres contenant des bâtiments bas. Ces deux zones seront maintenues dans la strate 2, tandis que les cinq autres forment pour l'instant une classe à part, dont le faible effectif pourait être compensé par un nombre de districts élevé dans chaque zone (à vérifier...).

ii : Le sous fichier des zones à petits bâtiments

Analyse factorielle

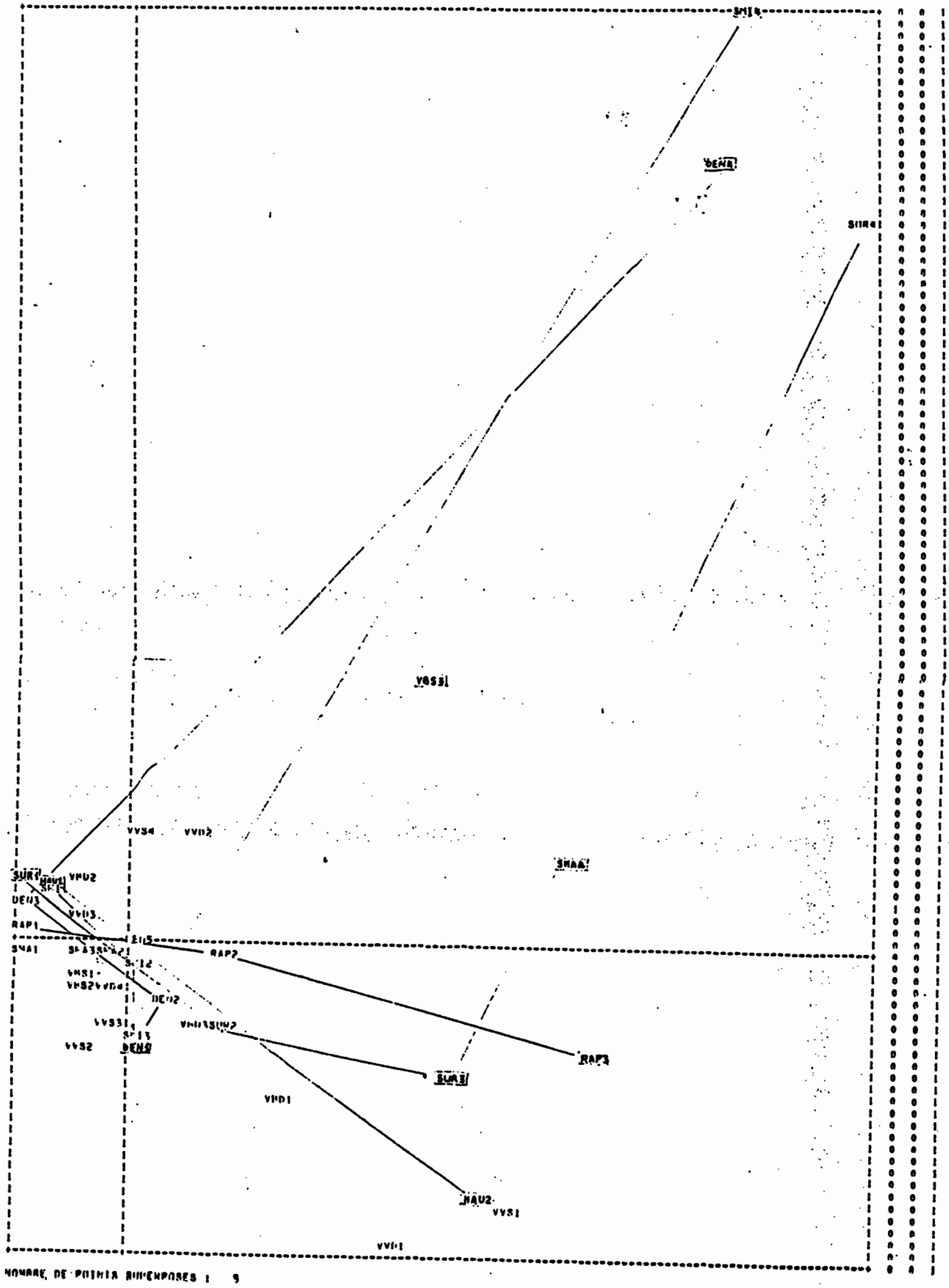
La pondération par un facteur 3 des modalités HAU1 et HAU2, DEN1 et DEN4, explique en grande partie le positionnement des trois premiers axes qui vont nous intéresser ici.

Premier facteur : La modalité HAU2, située du côté positif de l'axe et qui ne concerne que 29 des zones du sous-ensemble à l'étude, contribue pour 22,4% à l'inertie du premier facteur. Les autres modalités contributives sont également de poids faibles : on trouve, dans l'ordre des contributions décroissantes, et également du côté positif de l'axe, DEN4 (8 zones, CTR=14,4%), SMA4 (8zones, 10%), RAP3 (7 zones,7,9%), SUR3 (10 zones, 5,8%) et VBD3 (11 zones 5%). A l'opposé, le gros du nuage ne fournit qu'une modeste contribition au positionnement de l'axe, qui est le fait de modalités "lourdes" ayant des coordonnée+s légèrement négatives : HAU1 (147 zones, 4,9% de l'inertie de l'axe), SUR1 (51 zones, 4,5%) et RAP1 (57 zones, 3%).
En résumé, prés de 80% de l'inertie sur le premier facteur est due à l'opposition suivante : d'une part les modalités, minoritaires dans ce sous-ensemble, qui caractérisent les zones à bâtiments élevés (HAU2) ou de grande taille (SUR3, SMA4) et de formes allongées (RAP3)ou encore les zones à forte densité de bâti (DEN4), et d'autre part, le gros du nuage, regroupé autour de HAU1, SUR1 et RAP1; les sous groupes minoritaires expliquant environ 65% de l'inertie de l'axe.

AXE HORIZONTAL(1)--AXE VERTICAL(2)--TITRE: MARSEILLE : A . F . C . DU SUD - TABLEAU PETITS BAIS . PUNO - 2 - LE 21 / 02 /

NOMBRE DE POINTS : 40

Fig. 2.1.4. 15



NOMBRE DE POINTS SUPERPOSES : 9

NAU1(SHI1) SHA5(DEN5) SUR5(DEN5) SUR6(DEN5) RAP6(DEN5)

deuxième facteur :Prés de 50% de l'inertie de l'axe est due à l'opposition entre les huit zones prenant la modalité DEN4 (coordonnées positives, 36% de l'inertie de l'axe) et les 29 zones comprenant une majorité de bâtiments élevés (coordonnées négatives, 14% de l'inertie de l'axe). C'est d'ailleurs cette seule caractéristique qui servira à la description de la partition retenue.

troisième facteur : Les faits marquants et utiles par la suite, sont la forte contribution des zones trés peu denses (DEN1, 22% de l'inertie sur le facteur) et l'opposition entre les environnements à dominante claire (VVD1,coordonnée positive, 11,4%) et les environnements à dominante foncée (VVD4, coordonnée negative, 7%).
Pour plus de détail sur les contributions, consulter le tableau 2.1.4. 14.

Plan factoriel 1-2 : Sur la figure 2.1.4. 15, on distingue principalement trois régions sur le plan factoriel 1-2 :
-1 La partie supérieure droite où figurent 9 zones atypiques dans le sous fichier, caractérisées par des densités de bâti élevées (DEN4) ou des grandes tailles de bâtiments (SUR4, SMI4).
-2 La partie inférieure droite regroupe, autour de la modalité HAU2, les zones où les bâtiments sont majoritairement élevés, de taille moyenne et de formes plus ou moins allongées (SUR3, RAP3)
-3 Enfin, autour du centre de gravité, se trouve la majorité des zones à petits bâtiments bas.
Ce plan factoriel permet de repérer et de caractériser facilement les petit sous-ensembles originaux. On y projettera avec profit les centres des classes de la partition en trois classes.

Plan factoriel 1-3 (fig 2.1.4. 16) : Deux remarques permettront d'utiliser ce plan factoriel pour caractériser la partition en sept classes :
-1 A droite du plan, les modalités SUR4 et SMI4, de coordonnées positives sur le troisième facteur, s'oposent à la modalité DEN4, de coordonnée négative.
-2 Dans la partie centrale du nuage (faible valeur absolue du premier facteur), le troisième facteur opose les zones peu denses (DEN1) et de valeur claire (VVD1, VVS1) aux zones plus denses (DEN2 et 3) ou l'environnement du bâti est foncé (VVD4).

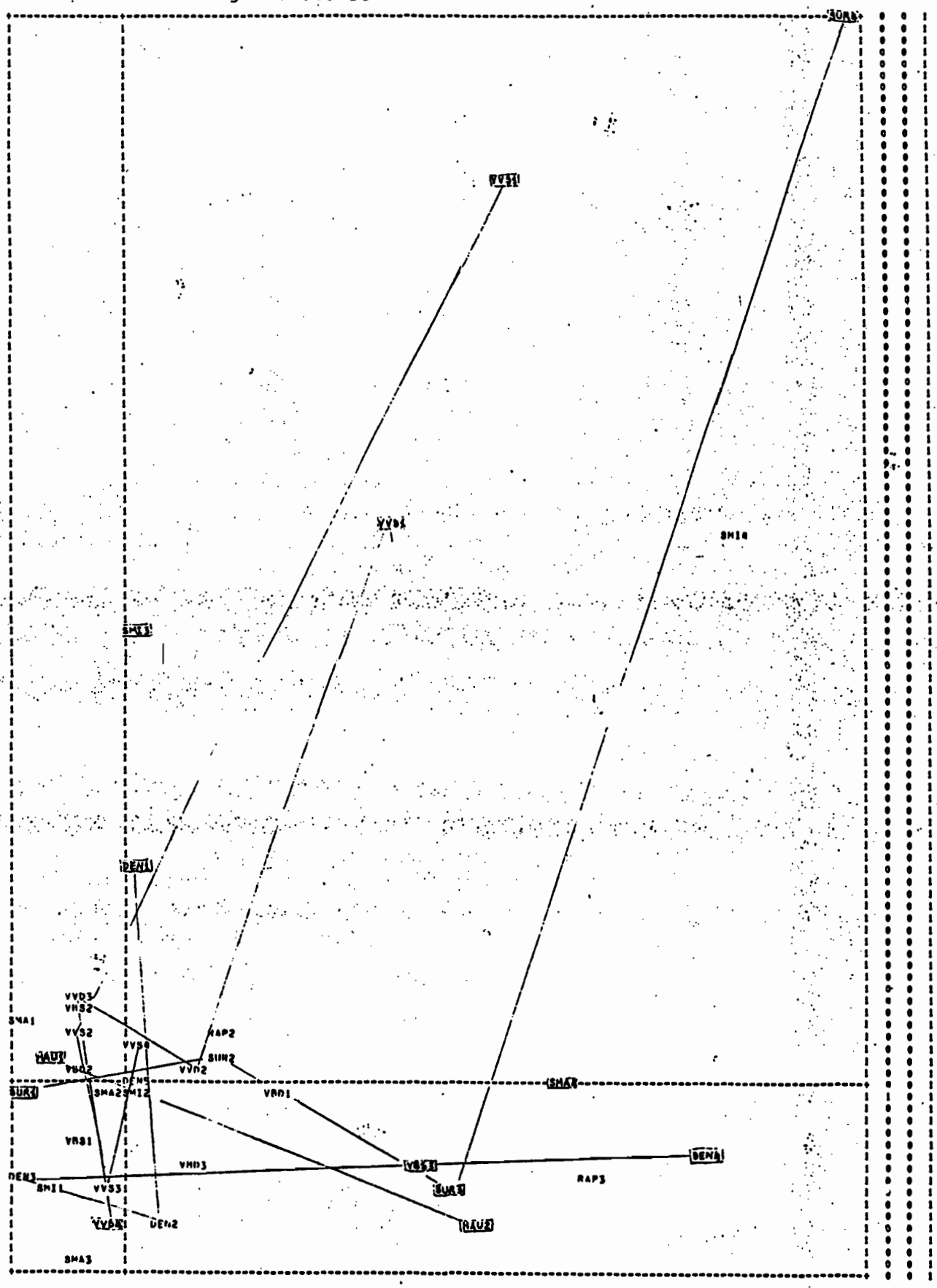
Classification hiérarchique

L'histogramme des indices de niveaux (fig 2.1.4 17) présente des seuils significatifs pour les partitions en 2, 3, 4 et 5 classes puis pour la partition en 7 classes. Plus loin, la partition en 11 classes présenterait trop de singletons. Les conclusions tirées de l'analyse factorielle permettent de caractériser clairement ces différentes partitions à partir de leur projections sur les plans factoriels 1-2 et 1-3.
Sur le plan 1-2 (fig 2.1.4 18), la position des centres des classes 372 et 369, issues de la premiére partition, est éloquente : les 9 zones de 369 apparaissent trés éloignées du centre de gravité , caractérisées par les modalités SMI4, DEN4 et SUR4. De même, les 25 zones de la classe 371, qui proviennent d'une nouvelle ponction dans 372, le gros du nuage étant maintenant regroupé dans 370, sont facilement caractérisables par les modalités HAU2, SUR3 et RAP3.

AXE HORIZONTAL(1)--AXE VERTICAL(3)--TITRE: MARSEILLE : A . F . C . DU SOUS - TABLEAU PETITS BATS POND = 2 - LE 21 / 02 /

NOMBRE DE POINTS : 48

Fig. 2.1.4. 16



NOMBRE DE POINTS SUPERPOSES : 5

SMA5(DENS) SUP5(DENS) SUR4(DENS) RAP4(DENS) PAP1(SUR1)

SOMME DES INDICES DE NIVEAU 0.24009E+01

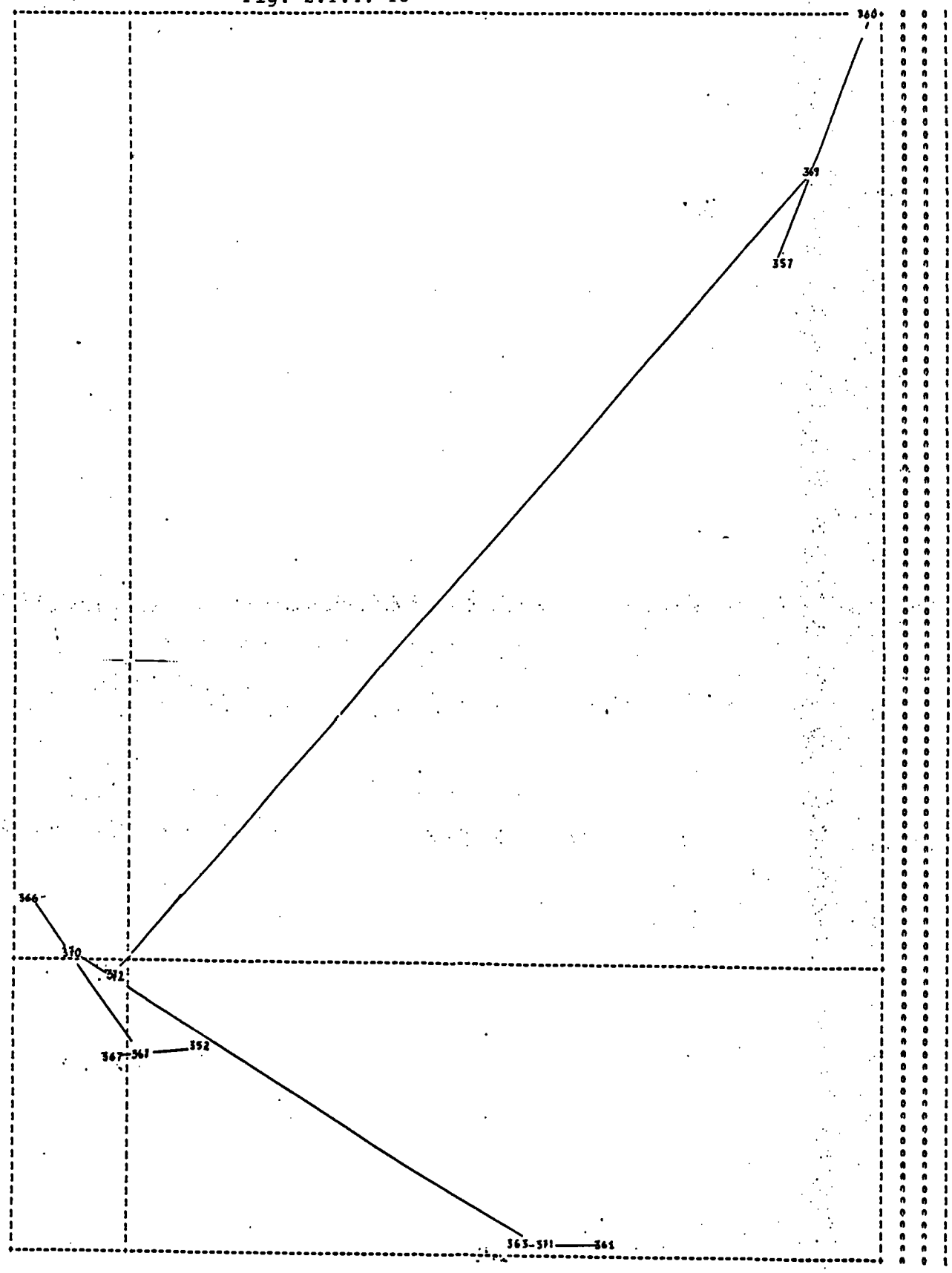
Fig. 2.1.4. 17

J	I(J)	A(J)	B(J)	T(J)	T(Q)	HISTOGRAMME DES INDICES DE NIVEAU DE LA HIERARCHIE
3731	2011	3721	3691	84	84	**********
3721	1531	3701	3711	64	147	**********
3711	1321	3611	3631	55	202	**********
3701	1131	3661	3681	47	249	**********
3691	961	3601	3571	40	289	**********
3681	921	3671	3521	38	327	**********
3671	711	3581	3641	30	357	**********
3661	681	3551	3651	28	385	**********
3651	641	3591	3621	27	412	**********
3641	641	3461	3401	27	439	**********
3631	491	3491	3531	21	459	**********
3621	461	3361	3441	19	479	**********
3611	431	3321	81	18	496	**********
3601	391	1371	1701	16	513	**********
3591	371	3261	3561	16	529	**********
3581	371	3481	3541	15	543	**********
3571	331	2731	3101	14	557	**********
3561	301	3341	3301	13	570	**********
3551	281	3501	3511	12	582	**********
3541	281	3421	2981	11	593	**********
3531	261	3451	2161	11	604	**********
3521	241	3231	3431	10	614	**********
3511	231	3281	3471	10	623	**********
3501	231	3041	3391	10	633	**********
3491	211	2941	2561	9	642	**********
3481	201	3191	3331	8	650	**********
3471	201	3071	3381	8	658	**********
3461	181	3121	3411	8	666	**********
3451	181	3271	3391	7	673	**********
3441	181	3151	2611	7	681	**********
3431	171	3101	3371	7	688	**********
3421	171	2531	3291	7	695	**********
3411	151	41	3311	6	701	**********
3401	141	3141	3241	6	707	**********
3391	141	3091	3251	6	712	**********
3381	141	3081	2711	6	718	**********
3371	141	3161	1361	6	724	**********
3361	131	2851	3221	6	730	**********
3351	131	3201	3051	5	735	**********
3341	131	2801	2841	5	740	**********
3331	121	2721	3131	5	746	**********
3321	121	51	91	5	751	**********
3311	121	2991	31	5	756	**********
3301	121	3021	2701	5	761	**********
3291	121	2901	3011	5	766	**********
3281	121	3111	2751	5	770	**********
3271	111	3171	2911	5	775	**********
3261	111	2931	2871	5	780	**********
3251	111	3211	351	5	785	**********
3241	111	3001	691	4	789	**********
3231	101	1581	2921	4	793	**********
3221	101	2821	3061	4	797	**********
3211	91	2791	2271	4	801	**********
3201	91	2571	1831	4	805	**********
3191	91	2041	2591	4	809	**********
3181	91	2771	3031	4	813	**********
3171	91	2671	2651	4	816	**********
3161	91	1451	2951	4	820	**********
3151	91	2581	2971	4	824	**********
3141	81	1611	2341	3	827	**********
3131	81	2481	2811	3	830	**********
3121	81	11	61	3	834	**********
3111	81	2401	2021	3	837	**********
3101	81	2461	2091	3	840	**********
3091	81	2501	2881	3	844	**********
3081	81	2261	2741	3	847	**********
3071	81	1111	2621	3	850	**********
3061	81	2491	2961	3	854	**********
3051	71	2211	2681	3	857	**********
3041	71	2541	2431	3	860	**********
3031	71	1271	2641	3	862	**********
3021	71	2661	2321	3	865	**********
3011	71	2551	2831	3	868	**********
3001	71	121	2781	3	871	**********
2991	71	21	71	3	874	**********
2981	61	1731	1101	3	876	**********
2971	61	2391	1261	3	879	**********
2961	61	2181	841	3	882	**********
2951	61	481	1301	3	885	**********
2941	61	2761	1331	3	887	**********
2931	61	2351	1291	2	889	**********
2921	61	1671	1651	2	892	**********
2911	61	1711	1341	2	894	**********
2901	61	2601	2071	2	896	**********
2891	61	1661	721	2	899	**********
2881	61	311	561	2	901	**********
2871	61	2521	2011	2	903	**********
2861	51	581	711	2	905	**********
2851	51	1301	2441	2	908	**********
2841	51	2451	2691	2	910	**********
2831	51	2301	2511	2	912	**********
2821	51	491	2401	2	914	**********
2811	51	1721	2331	2	916	**********
2801	51	2001	1971	2	918	**********
2791	41	101	761	2	919	**********
2781	41	241	1481	2	921	**********
2771	41	1461	1651	2	923	**********
2761	41	1781	1311	2	925	**********
2751	41	1391	1281	2	927	**********
2741	41	1121	1131	2	928	**********
2731	41	1441	1351	2	930	**********
2721	41	191	2471	2	932	**********
2711	41	1631	1231	2	933	**********
2701	41	2231	2631	2	935	**********
2691	41	2091	2021	2	937	**********
2681	41	1551	1411	2	938	**********
2671	41	1521	1211	2	940	**********
2661	41	2191	421	2	942	**********
2651	41	1741	1641	2	943	**********
2641	41	751	1011	2	945	**********
2631	41	2111	2201	2	946	**********
2621	41	601	1181	2	948	**********
2611	41	1941	2371	2	949	**********
2601	41	2241	961	2	951	**********
2591	41	2281	2291	2	952	**********
2581	41	2141	2221	1	954	**********
2571	31	101	2411	1	955	**********
2561	31	1421	1151	1	957	**********
2551	31	2051	1081	1	958	**********
2541	31	291	1071	1	959	**********
2531	31	571	971	1	961	**********
2521	31	2361	1771	1	962	**********
2511	31	1851	1621	1	963	**********
2501	31	221	231	1	965	**********
2491	31	1561	681	1	966	**********
2481	31	1901	2251	1	967	**********
2471	31	1931	1031	1	969	**********
2461	31	1861	1751	1	970	**********
2451	31	2311	871	1	971	**********
2441	31	1321	1691	1	972	**********
2431	31	1221	2381	1	973	**********
2421	31	1171	1191	1	974	**********
2411	31	381	1061	1	975	**********
2401	21	851	1241	1	976	**********
2391	21	2101	2121	1	977	**********
2381	21	1601	1511	1	978	**********
2371	21	2171	621	1	979	**********
2361	21	2131	1401	1	980	**********
2351	21	1871	1531	1	981	**********
2341	21	1161	1591	1	982	**********
2331	21	1201	1541	1	983	**********
2321	21	2151	1101	1	984	**********
2311	21	471	1981	1	985	**********
2301	21	1501	1571	1	985	**********
2291	21	941	2081	1	986	**********
2281	21	2061	931	1	987	**********
2271	21	701	731	1	988	**********
2261	21	1821	1841	1	988	**********
2251	21	1961	741	1	989	**********
2241	21	1921	1091	1	990	**********
2231	21	1951	791	1	991	**********
2221	21	2031	821	1	991	**********
2211	11	1251	1491	1	992	**********
2201	11	441	841	1	992	**********
2191	11	331	641	1	993	**********
2181	11	321	341	1	994	**********
2171	11	451	461	1	994	**********
2161	11	1431	1471	0	995	**********
2151	11	611	811	0	995	**********
2141	11	551	771	0	996	**********
2131	11	1811	1761	0	996	**********
2121	11	401	411	0	996	**********
2111	11	211	911	0	997	**********
2101	11	171	391	0	997	**********
2091	11	161	271	0	998	**********
2081	11	951	941	0	998	**********
2071	11	861	981	0	999	**********
2061	11	651	1081	0	999	**********
2051	11	501	1041	0	110	**********
2041	11	141	921	0	110	**********
2031	01	801	1911	0	110	**********
2021	01	901	1881	0	110	**********
2011	01	1791	1801	0	110	**********
2001	01	1891	1991	0	110	**********
1991	01	831	1021	0	110	**********
1981	01	631	1051	0	110	**********
1971	01	541	671	0	110	**********
1961	01	511	591	0	110	**********
1951	01	431	891	0	110	**********
1941	01	371	531	0	110	**********
1931	01	281	301	0	110	**********
1921	01	261	641	0	110	**********
1911	01	251	521	0	110	**********
1901	01	201	361	0	110	**********
1891	01	151	781	0	110	**********
1881	01	111	131	0	110	**********

AXE HORIZONTAL(1)--AXE VERTICAL(2)--IMPRIMANTE : 1 - 7 CLASSES DE LA PARTITION

NOMBRE DE POINTS : 7

Fig. 2.1.4. 18



Les séparations qui s'opèrent ensuite à l'intérieur des classes 371, 370 et 369, doivent être observées sur le plan 1-3 (fig 2.1.4 19). En premier lieu, la classes 371 est scindée quasi parallèlement au troisième axe, en 363 (coordonnée négative) et 361 (coordonnée fortement positive). La classe 363 regroupe 22 zones où dominent les bâtiments élevés de taille moyenne aux valeurs de toit parfois foncées. Cette classe ira rejoindre dans la strate numéro 4 les plus petits des immeubles élevés issus du sous-ensemble des zones à grands bâtiments. Située à l'oposé du troisième facteur par rapport à 363, la classe 361 contient 8 zones à bâtiments élevés, caractérisées par les modalités DEN1, VVD1 et VVS1. Il s'agit de zones trés peu denses à valeur très claire, où figurent quelques bâtiments isolés. Etant donnée la faible densité du bâti, elles iront grossir la strate 6 dont il sera question ci-dessous.
La partition en cinq classes permet ensuite l'éclatement de 370 en 366 et 368. Comme en témoigne la position des deux centres de classes le long de l'axe 3, le critère principal de cette séparation est la densité du bâti dans les zones : les 92 zones qui forment 366 sont regroupées autour des modalités SUR1, HAU1, VVD4, DEN2, DEN3, SMI1, SMA1, SMA2, etc... Il s'agit des zones, majoritaires dans le sous-ensemble à l'étude, occupées par de petits bâtiments bas, où la densité du bâti est moyenne et la valeur de l'environnement foncée. Elles constitueront la strate 5 qui rassemble ce qu'il est convenu d'appeler les zones pavillonaires. La classe 368, elle, est groupée autour de la modalité DEN1. La majorité des zones qui la constitue formeront le gros de la strate 6 : il s'agira des zones à faibles densités de bâti, située pour la plupart à la périphérie de l'aglomération; le plus souvent l'environnement des rares bâtiments qui s'y trouvent est à vocation agricole et de valeur assez claire.
Enfin, dans la partition en sept classes, ce sont les classes 369 et 368 qui éclatent : la classe 369, qui contenait des zones atypiques soit par la taille des bâtiments soit par leur densité, fournit les classes 360 et 357. 360 est constituée de deux zones à grands bâtiment élevés, qui seront reversées dans la strate 3 qui provient majoritairement du sous-ensemble des zones à grands bâtiments. Quant à 357, elle regroupe 7 zones à forte densité de bâti, mais dont les autres caractéristiques morphologiques sont variables. Elles seront versées dans la strate numero 7 qui contient un reliquat de zones difficiles a caractériser. Dans cette strate, on trouve également la classe 352 formée de 11 zones provenant de l'éclatement de 368. ces zones sont aussi assez atypiques mais le restant de la classe 368 (classe 367), caractérisé par la faible densité du bâti, constitue le gros de la strate 6 dite des zones péri-urbaines
peu denses.

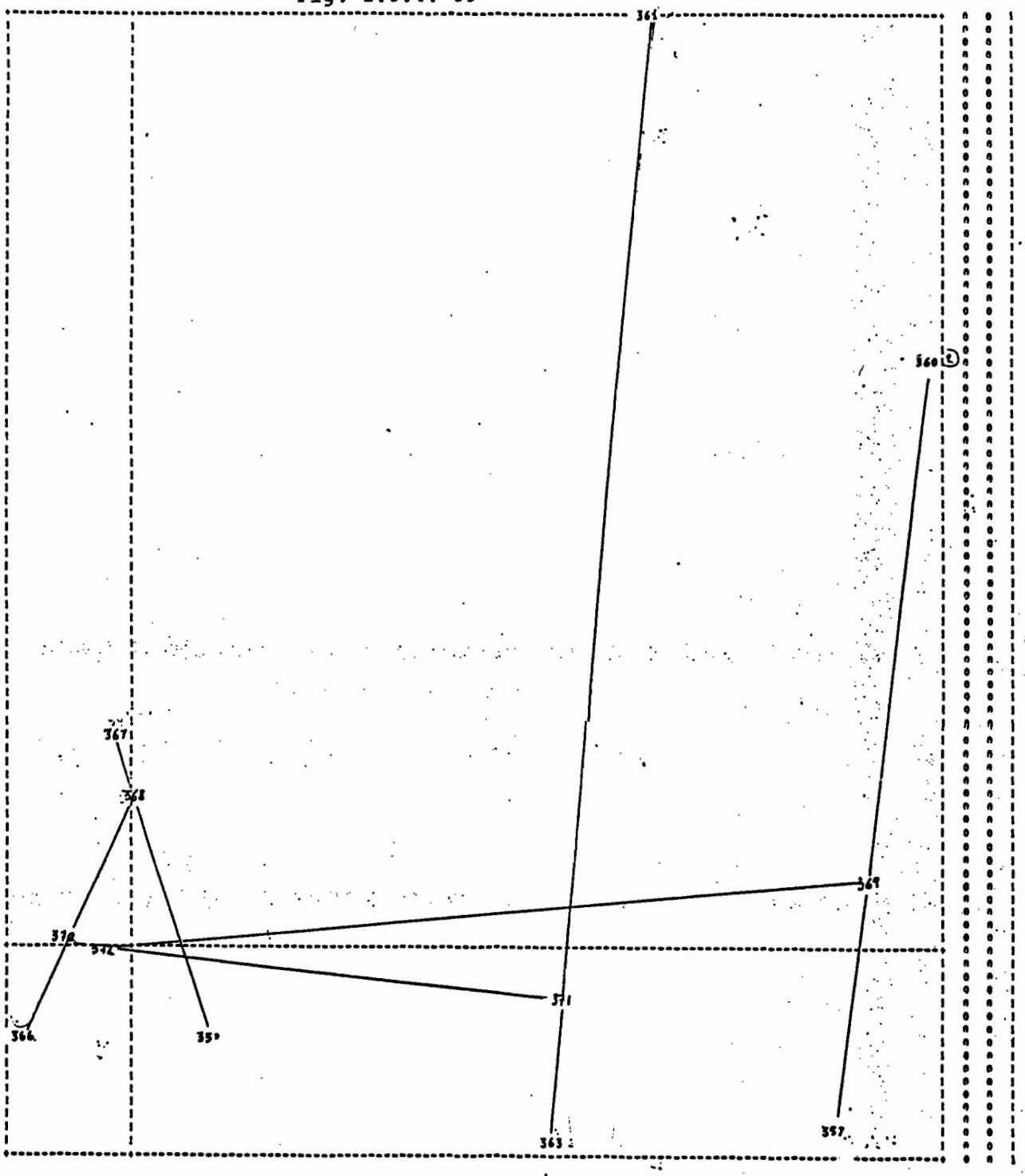
2..1.5. LA TYPOLOGIE EN HUIT CLASSES DE LA MORPHOLOGIE URBAINE

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

AXE HORIZONTAL(1)--AXE VERTICAL(3)--TITRE:DEMANDE 1 2 - 7 CLASSES DE LA PARTITION

NOMBRE DE POINTS : 7

Fig. 2.1.4. 19



1 . Chaque strate doit être composée d'un nombre suffisant de districts INSEE. On a vu que cet impératif correspond à un minimum de 30 zones par strate.
2 . L'application démographique du programme de recherche impose que la typologie rende compte des distinctions majeures des types d'habitat, par contre les subdivisions dans les zones non habitées importent beaucoup moins.
Dans ce qui suit on décrit successivement la typologie de deux points de vues : en caractérisant la morphologie de chaque strate à partir des tableaux donnant les fréquences des modalités descriptives, puis en examinant la répartition spatiale des types à partir de leur cartographie.

A : CARACTERISATION MORPHOLOGIQUE DES STRATES.

La synthèse des résultats de l'analyse est présentée sous forme de tableaux croisant les huit types morphologiques avec les modalités d'un descripteur donné.
On aura donc dix tableaux de ce type, dont la case courante (i, j) contient le nombre de zones de la strate j, ayant pris la modalité i du descripteur ainsi que les fréquences relatives en % par rapport au total général, au total de la modalité et au total de la strate (figures 1 à 5).

Strate 1 : Elle regroupe les 49 zones les plus denses du domaine d'étude. Ses principales caractéristiques morphologiques sont les suivantes :
- 85% des zones sont concentrées sur les modalités de rangs 4 ou 5 de la variable "densité du bâti" (DEN4, DEN5).
- Les éléments bâtis sont en majorité petits ou moyens et de formes peu allongées : SUR1, 2 ou 3 (98%), SMI1 ou 2 (100%) et la valeur du rapport longueur/largeur qui n'excède pas 5 est majoritairement comprise entre 1 et 2 (67%).
- La valeur du bâti est à dominante sombre : VBD3 (65%) et VBD2 (35%).
- L'environnement du bâti, ici très minoritaire en surface, a une valeur moyenne (31%) ou très foncée (63%).
Pour cette strate, on peut donc formuler une légende synthétique assez précise :

Strate 1("centre ville dense") : zones à densité de bâti forte ou très forte ou les petits et moyens bâtiments de formes peu allongées à toits sombres dominent.

Strate 2 : Il s'agit des 48 zones, ou dominent les bâtiments industriels, offrant la conjonction d'une grande taille des élément bâtis avec leur faible élévation. On y observe les fréquences suivantes :
- SUR4, SUR5 et SUR6 : 75% des zones. Au minimum, la taille des bâtiments correspond a la modalité SUR3 (25%).
- HAUT1 : 96% des zones.
- On trouve à peu prés toutes les formes de bâtis avec toutefois une majorité de formes allongées : RAP3 et RAP4 (63%).
- La densité du bâti y est moyenne : DEN2 et DEN3 (71%).
- La valeur de l'environnement est également moyenne : VVD2 et VVD3 (90%).
- toute les valeurs du bâti sont représentées : VBD1 (25%), VBD2 (44%), VBD3 (31%).

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

Strate 3 : 116 zones à grands bâtiments qu'on peut caractériser de manière assez précise :
- Grande taille du bâti : SUR3 (54%), SUR4 (35%), SUR5 (10%).
- HAU2 : 97%
- Deux modes de formes pour le bâti majoritaire : carrées (RAP1, 15%) ou allongées (RAP3 et RAP4, 78%).
- Valeurs claires ou moyennes du bâti : VBD1 (68%), VBD2 (31%).
- Valeur variable de l'environnement : VVD2 (44%), VVD3 (34%), VVD4 (21%).
- Densité moyenne du bâti : DEN2 ou DEN3 (85%).

Légende de la strate 3 ("grands ensembles d'habitation") : Zones à grands bâtiments hauts de formes carrées ou très allongées, la valeur du bâti est majoritairement claire.

Strate 4 : 49 zones où dominent les bâtiments de taille moyenne et majoritairement hauts :
- Taille moyenne du bâti : SUR1 (16%), SUR2 (41%), SUR3 (39%). La taille du bâti est assez homogène à l'intérieur des zones : SMI1 ou SMI2 (74%), SMA1 ou SMA2 (63%).
- Les constructions hautes dominent : HAU2 (86%), HAU1 (14%).
- Formes carrées (RAP1 : 49%) ou peu allongées (RAP2 ou RAP3 : 49%).
- Densité du bâti plutot faible : DEN1 (37%), DEN2 (39%) ou DEN3 (24%).
- Valeur claire du bâti : VBD1 (63%), VBD2 (37%).
- Valeur foncée de l'environnement : VVD3 et VVD4 (65%).

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

Strate 5 : 103 zones avec un bâti de petite taille et uniformément bas. Les modalités majoritaires sont les suivantes
- Petites tailles du bâti : SUR1 (69%), SUR2 (25%).
- HAU1 : 98%
- Formes majoritairement carrées : RAP1 (72%), RAP2 (24%).
- Densité plutot forte du bâti : DEN2 (35%), DEN3(61%).
- Environnement assez foncé : VVD3 (33%), VVD4 (42%).
- Valeurs gris moyens des toits : VBD2 (78%).

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

State 6 : 53 zones à faible densité de bâti généralement situées à la périphérie de l'aglomération.
- 90% des zones prennent la modalité DEN1.
- Petite taille du bâti : SUR1 et SUR2 (83%).
- Formes carrées ou peu allongées : RAP1 (70%), RAP2 et RAP3 (30%).
- Constructions basses : HAU1 (92%).
- Valeur gris moyen du bâti : VBD2 (77%).
- Toutes valeurs de l'environnement : VVD1 (17%), VVD2 (17%), VVD3 (38%), VVD4 (28%).

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

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

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

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

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

Légende de la strate 8 : 5 zones densément bâties de grands bâtiments élevés.
L'effectif, en districts INSEE, de la strate ne permettra probablement pas de la conserver comme telle malgrés sa bonne caracterisation morphologique (facile avec cinq zones !). Elle sera sans doute réagglomerée à la strate 3 qui lui ressemble le plus.

Fig. 2.1.5. 1

SAS 10:39 FRIDAY, MARCH 28, 1986 1

TABLE OF DENSR BY STRATE

DENSR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	0	9	16	18	4	48	4	0	99
	0.00	2.04	3.63	4.08	0.91	10.88	0.91	0.00	22.45
	0.00	9.09	16.16	18.18	4.04	48.48	4.04	0.00	
	0.00	18.75	13.79	36.73	3.88	90.57	22.22	0.00	
2	0	20	42	19	36	2	3	0	122
	0.00	4.54	9.52	4.31	8.16	0.45	0.68	0.00	27.66
	0.00	16.39	34.43	15.57	29.51	1.64	2.46	0.00	
	0.00	41.67	36.21	38.78	34.95	3.77	16.67	0.00	
3	8	14	55	12	63	3	4	0	159
	1.81	3.17	12.47	2.72	14.29	0.68	0.91	0.00	36.05
	5.03	8.81	34.59	7.55	39.62	1.89	2.52	0.00	
	16.33	29.17	47.41	24.49	61.17	5.66	22.22	0.00	
4	34	4	2	0	0	0	7	5	52
	7.71	0.91	0.45	0.00	0.00	0.00	1.59	1.13	11.79
	65.38	7.69	3.85	0.00	0.00	0.00	13.46	9.62	
	69.39	8.33	1.72	0.00	0.00	0.00	38.89	100.00	
5	7	1	1	0	0	0	0	0	9
	1.59	0.23	0.23	0.00	0.00	0.00	0.00	0.00	2.04
	77.78	11.11	11.11	0.00	0.00	0.00	0.00	0.00	
	14.29	2.08	0.86	0.00	0.00	0.00	0.00	0.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

SAS 10:39 FRIDAY, MARCH 28, 1986 8

TABLE OF HAUTR BY STRATE

HAUTR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	21	46	4	7	101	49	13	0	241
	4.76	10.43	0.91	1.59	22.90	11.11	2.95	0.00	54.65
	8.71	19.09	1.66	2.90	41.91	20.33	5.39	0.00	
	42.86	95.83	3.45	14.29	98.06	92.45	72.22	0.00	
2	28	2	112	42	2	4	5	5	200
	6.35	0.45	25.40	9.52	0.45	0.91	1.13	1.13	45.35
	14.00	1.00	56.00	21.00	1.00	2.00	2.50	2.50	
	57.14	4.17	96.55	85.71	1.94	7.55	27.78	100.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

Fig. 2.1.5. 2

SURFR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	18	0	0	8	71	30	9	0	136
	4.08	0.00	0.00	1.81	16.10	6.80	2.04	0.00	30.84
	13.24	0.00	0.00	5.88	52.21	22.06	6.62	0.00	
	36.73	0.00	0.00	16.33	68.93	56.60	50.00	0.00	
2	24	0	2	20	26	14	6	0	92
	5.44	0.00	0.45	4.54	5.90	3.17	1.36	0.00	20.86
	26.09	0.00	2.17	21.74	28.26	15.22	6.52	0.00	
	48.98	0.00	1.72	40.82	25.24	26.42	33.33	0.00	
3	6	12	60	19	6	8	3	3	117
	1.36	2.72	13.61	4.31	1.36	1.81	0.68	0.68	26.53
	5.13	10.26	51.28	16.24	5.13	6.84	2.56	2.56	
	12.24	25.00	51.72	38.78	5.83	15.09	16.67	60.00	
4	1	10	41	2	0	1	0	2	57
	0.23	2.27	9.30	0.45	0.00	0.23	0.00	0.45	12.93
	1.75	17.54	71.93	3.51	0.00	1.75	0.00	3.51	
	2.04	20.83	35.34	4.08	0.00	1.89	0.00	40.00	
5	0	13	12	0	0	0	0	0	25
	0.00	2.95	2.72	0.00	0.00	0.00	0.00	0.00	5.67
	0.00	52.00	48.00	0.00	0.00	0.00	0.00	0.00	
	0.00	27.08	10.34	0.00	0.00	0.00	0.00	0.00	
6	0	13	1	0	0	0	0	0	14
	0.00	2.95	0.23	0.00	0.00	0.00	0.00	0.00	3.17
	0.00	92.86	7.14	0.00	0.00	0.00	0.00	0.00	
	0.00	27.08	0.86	0.00	0.00	0.00	0.00	0.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

RAPTR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	10	7	17	24	74	37	8	0	177
	2.27	1.59	3.85	5.44	16.78	8.39	1.81	0.00	40.14
	5.65	3.95	9.60	13.56	41.81	20.90	4.52	0.00	
	20.41	14.58	14.66	48.98	71.84	69.81	44.44	0.00	
2	33	11	8	10	25	12	8	0	107
	7.48	2.49	1.81	2.27	5.67	2.72	1.81	0.00	24.26
	30.84	10.28	7.48	9.35	23.36	11.21	7.48	0.00	
	67.35	22.92	6.90	20.41	24.27	22.64	44.44	0.00	
3	6	21	48	14	4	4	2	4	103
	1.36	4.76	10.88	3.17	0.91	0.91	0.45	0.91	23.36
	5.83	20.39	46.60	13.59	3.88	3.88	1.94	3.88	
	12.24	43.75	41.38	28.57	3.88	7.55	11.11	80.00	
4	0	9	43	1	0	0	0	1	54
	0.00	2.04	9.75	0.23	0.00	0.00	0.00	0.23	12.24
	0.00	16.67	79.63	1.85	0.00	0.00	0.00	1.85	
	0.00	18.75	37.07	2.04	0.00	0.00	0.00	20.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

Fig 2.1.5. 3

SAS 10:39 FRIDAY, MARCH 28, 1986 3

TABLE OF SMAXR BY STRATE

MAXR / STRATE	1	2	3	4	5	6	7	8	TOTAL
1	8	0	0	5	35	24	3	0	75
	1.81	0.00	0.00	1.13	7.94	5.44	0.68	0.00	17.01
	10.67	0.00	0.00	6.67	46.67	32.00	4.00	0.00	
	16.33	0.00	0.00	10.20	33.98	45.28	16.67	0.00	
2	17	2	1	26	45	22	9	0	122
	3.85	0.45	0.23	5.90	10.20	4.99	2.04	0.00	27.66
	13.93	1.64	0.82	21.31	36.89	18.03	7.38	0.00	
	34.69	4.17	0.86	53.06	43.69	41.51	50.00	0.00	
3	13	8	52	7	17	2	1	5	105
	2.95	1.81	11.79	1.59	3.85	0.45	0.23	1.13	23.81
	12.38	7.62	49.52	6.67	16.19	1.90	0.95	4.76	
	26.53	16.67	44.83	14.29	16.50	3.77	5.56	100.00	
4	11	14	52	10	6	5	5	0	103
	2.49	3.17	11.79	2.27	1.36	1.13	1.13	0.00	23.36
	10.68	13.59	50.49	9.71	5.83	4.85	4.85	0.00	
	22.45	29.17	44.83	20.41	5.83	9.43	27.78	0.00	
5	0	24	11	1	0	0	0	0	36
	0.00	5.44	2.49	0.23	0.00	0.00	0.00	0.00	8.16
	0.00	66.67	30.56	2.78	0.00	0.00	0.00	0.00	
	0.00	50.00	9.48	2.04	0.00	0.00	0.00	0.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

SAS 10:39 FRIDAY, MARCH 28, 1986 4

TABLE OF SMINR BY STRATE

MINR / STRATE	1	2	3	4	5	6	7	8	TOTAL
1	27	11	13	6	27	5	7	1	97
	6.12	2.49	2.95	1.36	6.12	1.13	1.59	0.23	22.00
	27.84	11.34	13.40	6.19	27.84	5.15	7.22	1.03	
	55.10	22.92	11.21	12.24	26.21	9.43	38.89	20.00	
2	22	17	56	30	75	39	11	2	252
	4.99	3.85	12.70	6.80	17.01	8.84	2.49	0.45	57.14
	8.73	6.75	22.22	11.90	29.76	15.48	4.37	0.79	
	44.90	35.42	48.28	61.22	72.82	73.58	61.11	40.00	
3	0	10	23	5	1	9	0	0	48
	0.00	2.27	5.22	1.13	0.23	2.04	0.00	0.00	10.88
	0.00	20.83	47.92	10.42	2.08	18.75	0.00	0.00	
	0.00	20.83	19.83	10.20	0.97	16.98	0.00	0.00	
4	0	10	24	8	0	0	0	2	44
	0.00	2.27	5.44	1.81	0.00	0.00	0.00	0.45	9.98
	0.00	22.73	54.55	18.18	0.00	0.00	0.00	4.55	
	0.00	20.83	20.69	16.33	0.00	0.00	0.00	40.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

Fig. 2.1.5. 4

SAS 10:39 FRIDAY, MARCH 28, 1986 6

TABLE OF VVDOR BY STRATE

VDOR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	0	1	1	0	0	9	0	0	11
	0.00	0.23	0.23	0.00	0.00	2.04	0.00	0.00	2.49
	0.00	9.09	9.09	0.00	0.00	81.82	0.00	0.00	
	0.00	2.08	0.86	0.00	0.00	16.98	0.00	0.00	
2	15	18	51	17	26	9	5	3	144
	3.40	4.08	11.56	3.85	5.90	2.04	1.13	0.68	32.65
	10.42	12.50	35.42	11.81	18.06	6.25	3.47	2.08	
	30.61	37.50	43.97	34.69	25.24	16.98	27.78	60.00	
3	3	25	39	12	34	20	4	0	137
	0.68	5.67	8.84	2.72	7.71	4.54	0.91	0.00	31.07
	2.19	18.25	28.47	8.76	24.82	14.60	2.92	0.00	
	6.12	52.08	33.62	24.49	33.01	37.74	22.22	0.00	
4	31	4	25	20	43	15	9	2	149
	7.03	0.91	5.67	4.54	9.75	3.40	2.04	0.45	33.79
	20.81	2.68	16.78	13.42	28.86	10.07	6.04	1.34	
	63.27	8.33	21.55	40.82	41.75	28.30	50.00	40.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

SAS 10:39 FRIDAY, MARCH 28, 1986 7

TABLE OF VVSDR BY STRATE

VSDR STRATE

FREQUENCY / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	3	6	13	10	3	3	0	0	38
	0.68	1.36	2.95	2.27	0.68	0.68	0.00	0.00	8.62
	7.89	15.79	34.21	26.32	7.89	7.89	0.00	0.00	
	6.12	12.50	11.21	20.41	2.91	5.66	0.00	0.00	
2	26	19	26	3	14	12	2	2	104
	5.90	4.31	5.90	0.68	3.17	2.72	0.45	0.45	23.58
	25.00	18.27	25.00	2.88	13.46	11.54	1.92	1.92	
	53.06	39.58	22.41	6.12	13.59	22.64	11.11	40.00	
3	3	12	24	14	33	21	6	2	115
	0.68	2.72	5.44	3.17	7.48	4.76	1.36	0.45	26.08
	2.61	10.43	20.87	12.17	28.70	18.26	5.22	1.74	
	6.12	25.00	20.69	28.57	32.04	39.62	33.33	40.00	
4	17	11	53	22	53	17	10	1	184
	3.85	2.49	12.02	4.99	12.02	3.85	2.27	0.23	41.72
	9.24	5.98	28.80	11.96	28.80	9.24	5.43	0.54	
	34.69	22.92	45.69	44.90	51.46	32.08	55.56	20.00	
TOTAL	49	48	116	49	103	53	18	5	441
	11.11	10.88	26.30	11.11	23.36	12.02	4.08	1.13	100.00

Fig. 2.1.5. 5

SAS 10:39 FRIDAY, MARCH 28, 1986 4

TABLE OF VHDOR BY STRATE

VHDOR / STRATE FREQUENCY PERCENT ROW PCT COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	0 0.00 0.00 0.00	12 2.72 7.55 25.00	79 17.91 49.69 68.10	31 7.03 19.50 63.27	23 5.22 14.47 22.33	11 2.49 6.92 20.75	1 0.23 0.63 5.56	2 0.45 1.26 40.00	159 36.05
2	17 3.85 7.69 34.69	21 4.76 9.50 43.75	36 8.16 16.29 31.03	18 4.08 8.14 36.73	80 18.14 36.20 77.67	41 9.30 18.55 77.36	6 1.36 2.71 33.33	2 0.45 0.90 40.00	221 50.11
3	32 7.26 52.46 65.31	15 3.40 24.59 31.25	1 0.23 1.64 0.86	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.23 1.64 1.89	11 2.49 18.03 61.11	1 0.23 1.64 20.00	61 13.83
TOTAL	49 11.11	48 10.88	116 26.30	49 11.11	103 23.36	53 12.02	18 4.08	5 1.13	441 100.00

SAS 10:39 FRIDAY, MARCH 28, 1986

TABLE OF VHSDR BY STRATE

VHSDR / STRATE FREQUENCY PERCENT ROW PCT COL PCT	1	2	3	4	5	6	7	8	TOTAL
1	26 5.90 13.54 53.06	24 5.44 12.50 50.00	39 8.84 20.31 33.62	26 5.90 13.54 53.06	46 10.43 23.96 44.66	25 5.67 13.02 47.17	4 0.91 2.08 22.22	2 0.45 1.04 40.00	192 43.54
2	22 4.99 10.68 44.90	20 4.54 9.71 41.67	62 14.06 30.10 53.45	19 4.31 9.22 38.78	47 10.66 22.82 45.63	27 6.12 13.11 50.94	6 1.36 2.91 33.33	3 0.68 1.46 60.00	206 46.71
3	1 0.23 2.33 2.04	4 0.91 9.30 8.33	15 3.40 34.88 12.93	4 0.91 9.30 8.16	10 2.27 23.26 9.71	1 0.23 2.33 1.89	8 1.81 18.60 44.44	0 0.00 0.00 0.00	43 9.75
TOTAL	49 11.11	48 10.88	116 26.30	49 11.11	103 23.36	53 12.02	18 4.08	5 1.13	441 100.00

b) Répartition spatiale des strates

La diapositive n° 6 et le tableau 2.1.5.6 (système TIGER : requête sur la relation morphologie, attribut strate, pour l'ensemble de la zone d'étude) précisent la répartition spatiale de chacune des strates dans la zone d'étude. Trois phénomènes majeurs apparaissent :

- *les strates occupent des surfaces variables* : tandis que les strates "grands ensembles d'habitation" et "zones pavillonnaires" occupent des surfaces très importantes (environ 20 %), les strates 7 ("reliquat") et 8 ("zones densément bâties de grands bâtiments élévés") n'ont qu'une superficie très réduite (respectivement 3 % et 1 %), les strates "centre ville dense", "quartiers industriels" et "petits immeubles d'habitation"se trouvant en situation intermédiaire (6 à 10 %).

- *le degré de connexité des strates est, lui aussi, variable :* à la concentration de la strate "centre ville dense", et à un moindre degré de la strate "quartiers industriels", s'oppose la dispersion géographique des éléments composant la strate "petits immeubles d'habitation". On observe que certaines zones connexes distinguées visuellement sont regroupées dans une même strate.

- *la répartition géographique des strates rend compte des traits majeurs de l'organisation de l'espace marseillais* : opposition nord/sud, et disposition concentrique. La partie centrale de Marseille autour du vieux port correspond à la strate 1,"centre ville dense". Au nord-ouest de Marseille, la strate industrielle (2) voisine avec les grands ensembles de la strate 3 . Le croissant Sud et Est se caractérise par la juxtaposition des éléments appartenant aux strates 3 ("grands ensembles d'habitation"), 4 ("petits immeubles d'habitation") et 5 ("zones pavillonnaires". Les zones peu denses de la strate 6 sont toutes à la périphérie de l'agglomération.

c) Perspectives pour le tirage de l'échantillon de travail

Comme il l'a déjà été dit, la typologie en huit classes sert de stratification pour le tirage d'un échantillon de districts INSEE représentatif, en termes de surface, des types de morphologie urbaine présents dans la zone d'étude.
Afin de satisfaire cette condition de représentativité en surface, nous avons choisi la procédure suivante :

- *réaffecter la typlogie aux districts INSEE* (opération en cours de réalisation) : par le système TIGER, on obtient pour chaque district INSEE sa répartition (en % de surface) selon les types morphologiques. Si le district appartient intégralement à un seul type morphologique, il est automatiquement affecté à ce type morphologique; dans le cas où le district INSEE n'est pas morphologiquement homogène, l'affectation au type majoritaire dans le district n'est pas automatique, mais résulte d'une règle de décision basée sur l'observation de la répartition entre les différents types. Si une association de types morphologiques se répète pour un grand nombre de districts, nous en ferons un nouveau type morphologique, caractérisé par cette association de types.

Fig. 2.1.5.6

RECAPITULATIF DE LA REQUETE

Nom de la base : MARSEILLE
Point bas : Longitude 0 deg 10 min - Latitude 0 deg 2 min
Point haut : Longitude 0 deg 17 min - Latitude 0 deg 11 min

DEFINITION DE LA REQUETE

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

IMAGE PAR RELATION

MORPHOLOGIE Superficie totale : 185342

Code	Pixel	des carre	%	Strate
1	15482	2576.7180	0.08	1
2	19384	3226.1401	0.10	2
3	38848	6465.5938	0.21	3
4	10865	1808.2961	0.06	4
5	36014	5993.9219	0.19	5
6	14512	2415.2778	0.08	6
7	4985	829.6692	0.03	7
8	1388	231.0092	0.01	8
9	43864	7300.4219	0.24	non bati

- *calcul de la taille moyenne des districts INSEE par strate*, d'après la surface et le nombre de districts de chaque strate.

- *tirage de l'échantillon sur une grille de points superposée à la carte des districts INSEE mentionnant leur type morphologique*: pour chaque strate, tirage sans remise des districts INSEE auxquels appartiennent les points de la grille. Afin d'obtenir un taux de sondage constant en moyenne, la maille de la grille varie selon la strate, en fonction de la taille moyenne des districts INSEE appartenant à la strate.

Plusieurs essais seront faits, en faisant varier le calage de la grille sur la carte ; l'échantillon de travail retenu sera celui qui respecte le mieux la répartition en surface des types morphologiques. De plus, O. BARBARY procédera à d'autres tests sur les échantillons issus de la stratification morphologique, dans un but de recherche sur les sondages spatiaux en milieu urbain (voir chapitre 4).

2. 1. 6. ENSEIGNEMENTS METHODOLOGIQUES ET PERSPECTIVES POUR LA SUITE DE L'EXPERIMENTATION.

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

Dans un premier temps, il parait nécessaire de rappeler dans quelle cadre d'objectifs et de nature de l'information s'inscrit cette évaluation :

Les objectifs :La représentativité de l'échantillon de travail doit permettre aussi bien la mise au point de méthodes d'extraction de l'information morphologique à partir de l'image satellite, qu'une modélisation statistique des relations bâti/population. Pour ce second point il est indispensable de quantifier aussi précisément que possible les contributions des différents descripteurs à la différenciation des zones.

De l'information morphologique aux données traitées : Pour obtenir une information morphologique exhaustive et homogène sur l'ensemble de la zone d'étude le seul support duquel on sache pour l'instant l'extraire est la photographie aérienne. Ces documents correspondent déjà à une première transformation de l'information dépendante de différents paramètres : la visée verticale, l'émulsion utilisée, l'échelle de prise de vue et de restitution. La médiatisation se pousuit avec l'intervention du photo-analyste : privilégier des indicateurs visuels analytiques précis permet de mieux contrôler les effets de cette deuxième transformation. Mais quel que soit le mode de traitement de l'information extraite de photographies, ces deux transformations successives sont incontournables.

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

Elaboration du tableau de données

Nous avons exposé chapitre 2.1.2. les principes de la codification de l'information en un tableau descriptif de la morphologie des zones. La liste des descripteurs correspond à un choix subjectif mais qui respecte certaines règles : en résumé, les caractères morphologiques, définis rigoureusement, doivent permettre de discriminer des zones voisines différentes et de regrouper des zones semblables non adjacentes.

C'est à cette étape qu'est fixée la précision de l'information morphologique qui va être traitée : opter pour certain critères qualitatifs conduira à une dégradation des descripteurs métriques afin d'obtenir un tableau de données homogènes. Cette dégradation peut être de deux types : passage du métrique a une échelle de notes discrètes par seuillage des distributions ou passage au codage disjonctif où l'on ne rend plus compte du carractère ordonné des modalités.
Comme on l'a souligné lors de l'examen des distributions des caractères métriques, on ne perd pas grand chose à agglomérer les valeurs à la valeur centrale de chaque mode, du moins quand cette dernière est unique. Si l'on disposait par ailleurs d'échelles de notes permettant de coder les autres descripteurs jusqu'alors purement qualitatifs, on pourrait homogénéiser le tableau de description morphologique à un niveau d'information quasi-optimal compte tenu de la nature du phènomène étudié et de l'échelle d'observation.
Pour y parvenir, il faudra améliorer la méthode de codification adoptée pour la densité du bâti, sa hauteur et les valeurs de gris : en effectuant les mesures de chaque descripteur pour un certain nombre de zones étalons représentant toutes ses modalités, on pourrait associer à chaque modalité une valeur moyenne qui deviendrait la note attribuée à la zone.
Grâce à cette amélioration, on disposerait donc d'un tableau de notes qui, une fois dedoublé, se prête de façon classique à la même chaîne de traitement (AFC, CAH).

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

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

C'est un phénomène de même nature, mais incontournable étant donnée la nature actuelle des données, qui explique qu'on ait du multiplier par trois l'importance relative de la variable "hauteur du bâti" pour séparer les bâtiments industriels des grands ensembles.
A l'issue de l'analyse, on peut donc dire que le recodage disjonctif, suivant le nombre de modalités attribuées à tel ou tel descripteur, perturbe l'équilibre informationnel entre variables. Par la suite, le rétablissement de cet équilibre ou plus exactement l'établissement d'un équilibre qui satisfasse les objectifs thématiques fixés à la

typologie, a demandé un ajustement par pondération qui ne peut s'effectuer que par tatonnements successifs ce qui l'a rendu assez laborieux.
La proposition d'amélioration de la méthode faite ci-dessus, alliée a une meilleure précision de certains descripteurs, devrait permettre de réduire de beaucoup ces tatonnements : on partirait d'un tableau ou la qualité de l'information apportée serait moins variable d'un descripteur à l'autre et on suprimerait l'arbitraire lié au nombre de modalités disjonctives créées lors du recodage. L'établissement de la typologie devenant ainsi plus "automatique" la méthode serait plus facilement transposable à d'autre sites.

Conclusion

La démarche de classification adoptée pour parvenir aux résultats exposés peut paraitre très hétérodoxe au théoricien de l'analyse des données en ce qu'elle perturbe l'algorithme d'agrégation par divisions en sous ensembles du fichier d'origine et itération du processus sur des données pondérées. Elle peut également être taxée de coupage de cheveux en quatre bien inutile par le thématicien urbaniste ou géographe qui, visuellement, n'aurait eu aucune difficulté à operer beaucoup plus rapidement des regroupements semblables et serait même certainement parvenu à une description plus fine de l'espace urbain. Que dire, pour sa défense? A notre sens, trois arguments principaux peuvent être avancés :

1. On espère que la stratification résultant de cette démarche, pour grossière qu'elle soit, s'avèrera efficace comme stratification, tant du point de vue de la représentativité morphologique de l'échantillon des districts que pour l'estimation par sondage de la densité de population. Si c'est le cas, une partie du mérite en revient aux libertés prises par rapport à l'algorithme de classification. C'est en effet principalement pour augmenter cette efficacité que l'on n'a exploité que partiellement la classification globale et pondéré différentiellement sur chaque sous ensemble, les caractères morphologiques jugés thématiquement importants. Concédons toutefois que la méthode gagnerait peut être en précision et sûrement en rapidité, portabilité et élégance si l'on appliquait les améliorations proposées plus haut au codage de l'information morphologique.

2. Le cadre méthodologique de l'analyse typologique permet d'affirmer que la précision de la description (ou, si l'on préfère, le niveau de finesse de la partition retenue), reflète a peu prés exactement le niveau de précision de l'information morphologique qui a servi à sa mise au point. En clair, on ne risque pas d'avoir défini des regroupements visuellement justifiables mais que l'on ne pourrait pas caractériser à partir du tableau de description morphologique des zones. Or ceci est une condition indispensable à la poursuite du travail de modélisation mathématique des relations entre morphologie urbaine et démographie.

3. On évoque, dans le plan de travail pour la partie statistique de cette recherche, l'utilisation d'estimateur bayésien de l'effectif de population lors de la phase de production de données démographiques par sondages. L'utilisation de tels estimateurs repose sur des hypothèses dites bayèsiennes formulées dans des modèles qui expliquent la densité de population par diverses variables morphologiques. On parle ici de plusieurs modèles parce qu'on pressent que pour être efficaces, ces

hypothèses devront probablement prendre la forme de modèles spécifiques pour chaque strate. Dans ce cadre, on devra bien sûr étudier les valeurs centrales et les dispersions des caractères morphologiques aussi bien à l'intérieur de chaque strate, qu'entre les différentes strates.
La démarche adoptée a permis une quantification facile et assez précise de ces paramètres à l'intérieur de chaque strate, ou du moins pour sept d'entre elles exception faite de la strate numéro sept.

Ces premiers résultats devraient permettre, avant de mener l'inférence sur les paramètres des modèles spécifiques à chaque strate à partir des échantillons de districts INSEE, de proposer un modèle global explicatif de la densité de population moyenne des strates. Bien sûr il faut dire,(ce faisant on asséne au lecteur courageux parvenu jusqu'ici une énorme évidence), que ce modèle n'aura qu'une valeur symbolique du point de vue de l'estimation, même globale, de population. Néanmoins, il fournit un cadre général pour la suite et surtout une occasion de tester, sur des données réelles mais peu nombreuses, les méthodes de régression et d'estimation qu'on emploiera par la suite.

2. 2 . ANALYSE DE L' INFORMATION A INTEGRER DANS LA BASE

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.

2. 2. 1. 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

-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*

---

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

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

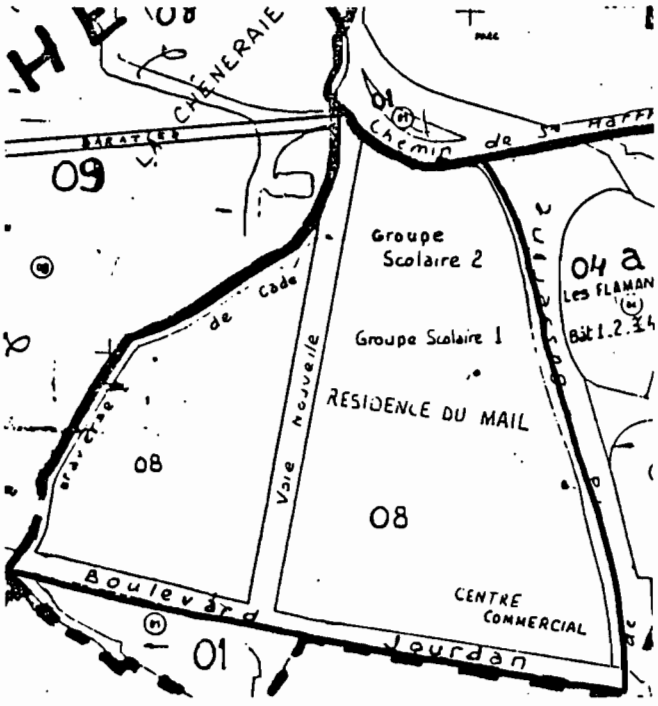
Sexe
Situation de famille
Date de naissance
Nationalité (francais, étrangers)
Département de résidence au 1.01.1975
Indicateur de scolarisation des enfants
Type d'activité (actif ayant un emploi, chomeur, élève ou étudiant,autre inactif)
Statut de la profession exercée (salarié, non salarié)
Caractère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

recensements de population . 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.

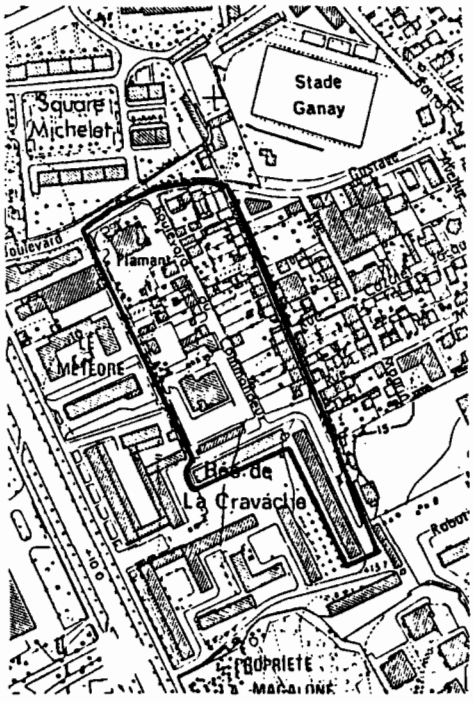
Une précision s'impose à propos de cette bande magnétique : demandée à l'Observatoire régional de l'INSEE à Marseille en Février 1985, elle ne nous a été livrée qu'en Octobre 1985. Jusqu'à cette date, nous ne disposions donc pour travailler que d'un listing des effectifs de population, que nous avons du faire saisir à Paris. Cette exemple est symptomatique du délai nécessaire pour obtenir des résultats exhaustifs du recensement à un niveau géographique fin : *3, 5 ans se sont écoulés entre le recensement et la production des résulats exhaustifs sur la ville de Marseille.*
Cette observation sur le délai que réclame l'exploitation exhaustive d'un recensement est en partie à l'origine du programme de recherche sur l'utilisation de la démographie pour l'observation des populations urbaines : ce délai est incompatible avec le rythme de croissance des villes des pays en développement. L'expérience que nous venons d'avoir sur Marseille confirme la nécessité de développer d'autres systèmes d'observation démographique pour les villes ; mais elle montre aussi la difficulté à mettre en oeuvre des recherches réclamant cette source de données pour mettre au point des nouvelles méthodes d'observation démographique.

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

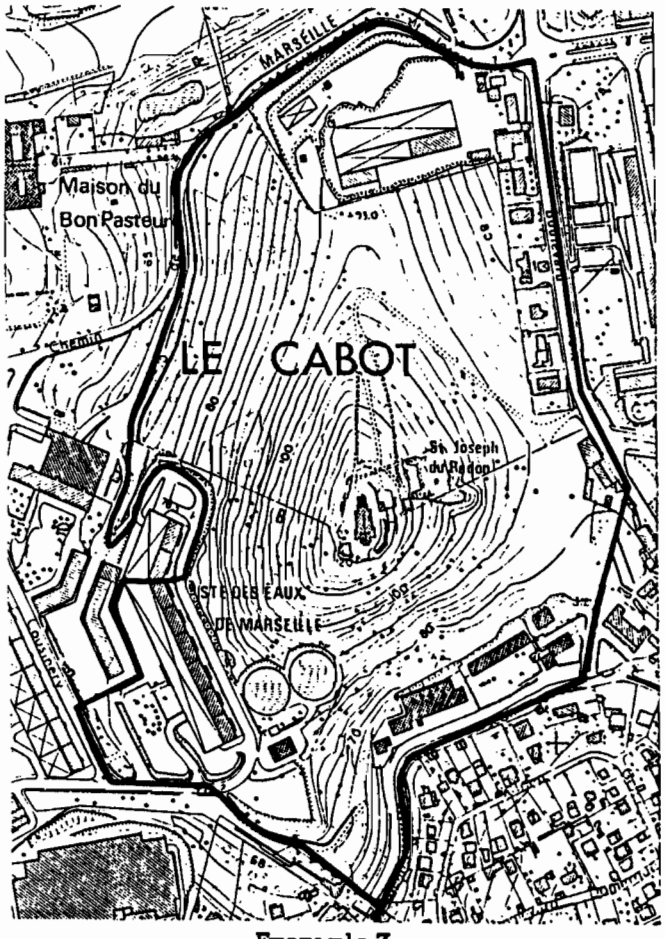
Figure 3 - Marseille . Découpage INSEE en districts de recensement



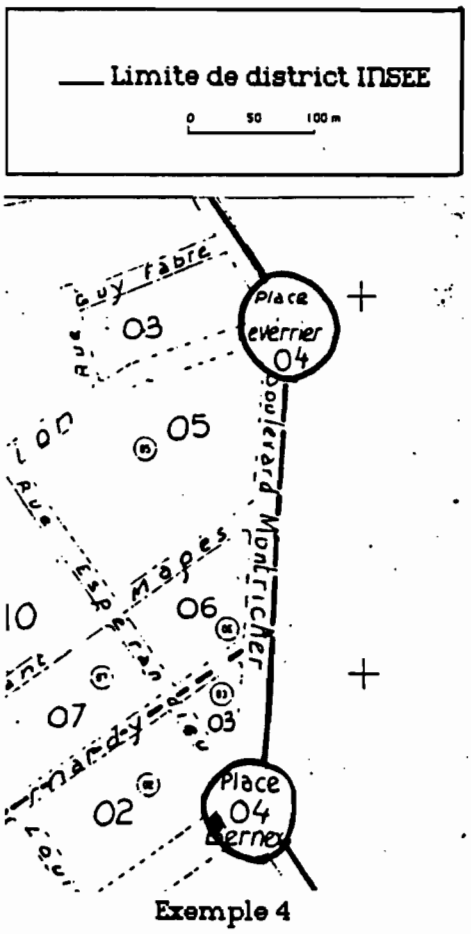
Exemple 1



Exemple 2



Exemple 3



Exemple 4

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 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'ilôt 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.

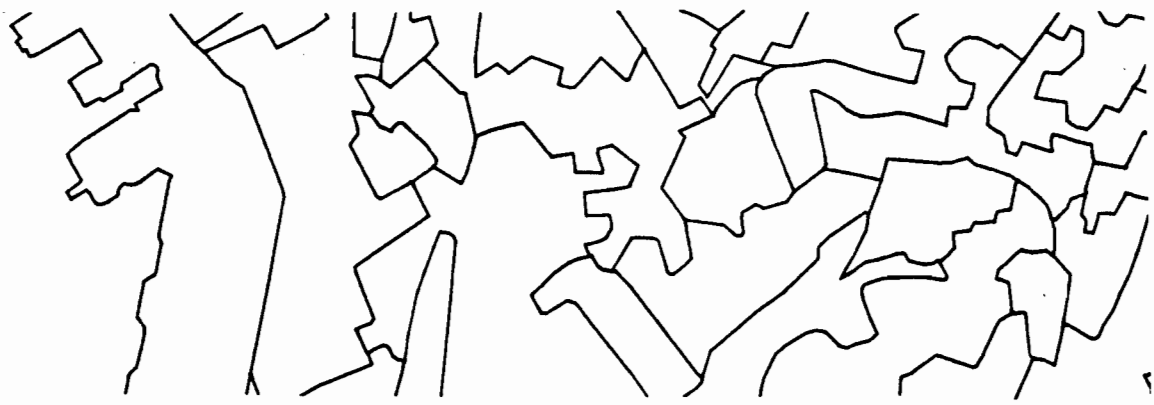
2.2.2. L'information sur la morphologie urbaine

La connaissance de la morphologie urbaine sur l'ensemble de la zone d'étude repose sur *l'analyse visuelle d'une mosaïque de photographies aériennes au 1/23000*(mission IGN d'Août 1982), avec contrôle terrain .
Comme nous l'avons vu au chapitre 2. 1 , *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 caractères morphologiques. 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*.

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

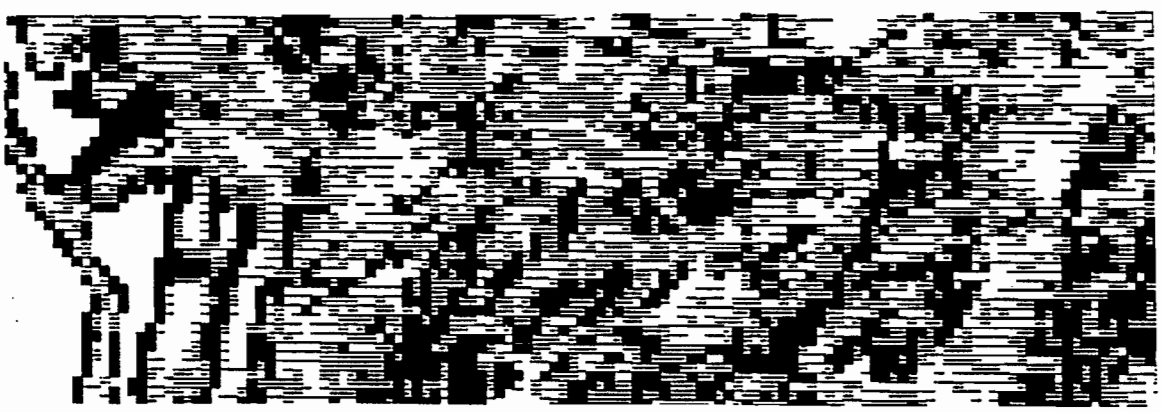


Image Thematic Mapper

0 250 500 m

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 aucunement compte du découpage établi par l'INSEE, ni de celui des ilôts.* Il est néanmoins certain que le tracé des rues introduit une discontinuité sur la photographie aérienne, et a donc tendance à influer sur le tracé dez zones homogènes.

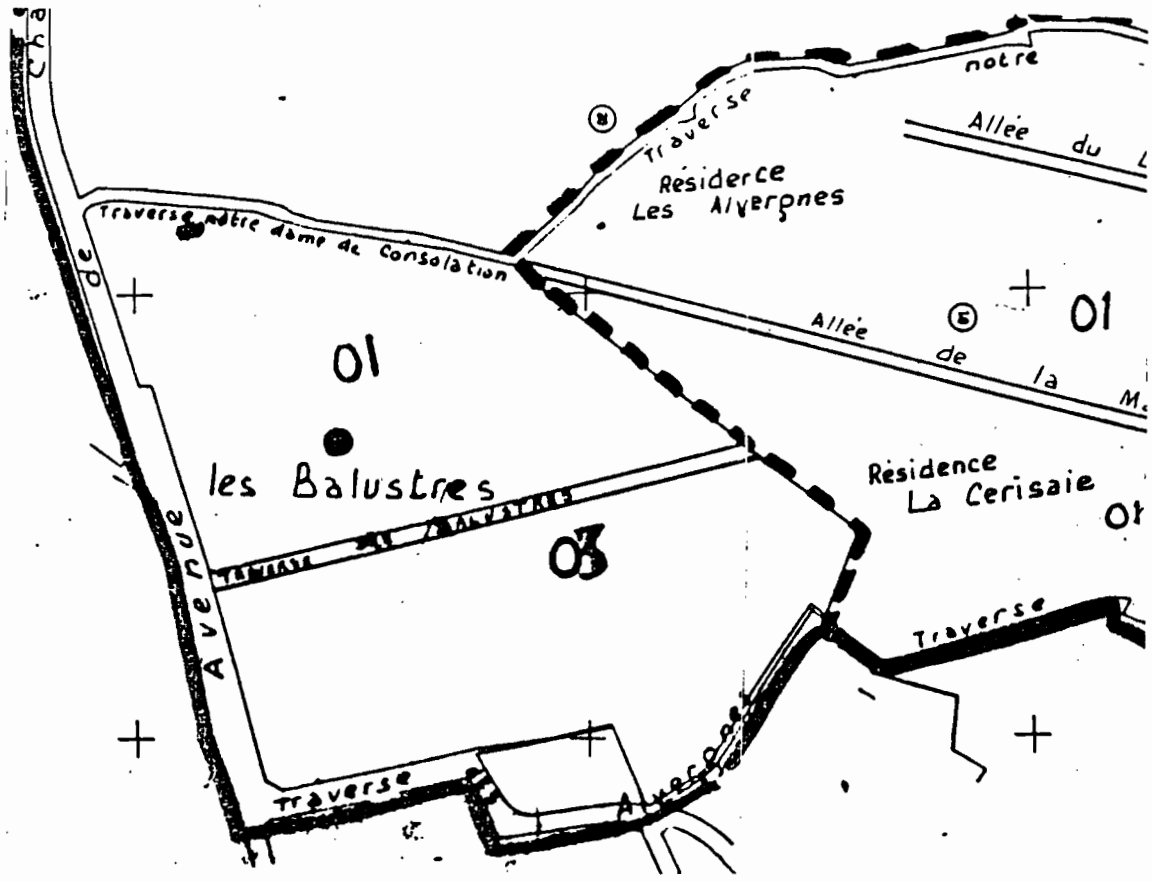
2. 2. 3. L'image Thematic Mapper

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

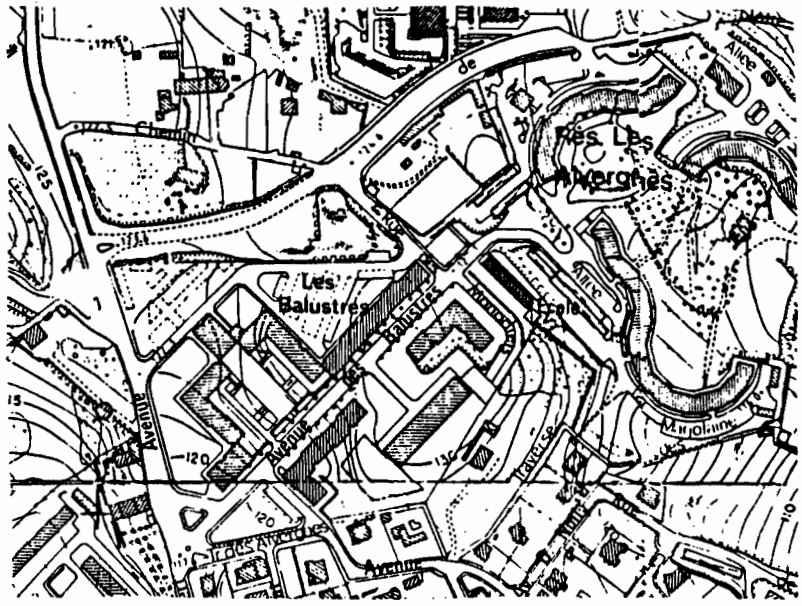
La principale caractéristique de l'information contenue dans les images satellite vient de leur mode d'enregistrement, qui se fait sur des unités élémentaires appelées "pixels", ayant une certaine surface (parallélogramme de 30 mètres de côté environ pour Thematic Mapper) : *les données satellitaires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

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 cartographiqe 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

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



INSEE



ICOREM

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 (au sens des bases de données relationnelles).

2 . 3 . INTEGRATION DES DONNEES DANS UNE BASE GEREE PAR LE SYSTEME TIGER

L'intégration des données dans la base comporte trois phases que nous aborderons successivement : la préparation des fonds cartographiques avant numérisation, la numérisation de ces fonds, et l'intégration de l'information thématique s'y rapportant.

2. 3. 1. Préparation des fonds cartographiques à numériser

L'intégration des fonds de cartes dans la base gérée par le système TIGER se fait par numérisation sur table à numériser BENSON. 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 signifiantes et objets d'étude : par exemple, nous utilisons le tracé de la voirie, issu de la carte des limites d'ilôts, 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 ; 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 (à même échelle) 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 70 : 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, au cours du premier semestre 1985 . Toute personne devant exploiter cartographiquement le recensement se trouve confrontée au même problème : on voit que celà 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 répétera 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 une 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 douzaine d'amers (points caractéristiques de la surface terrestre identifiés sur l'image satellite et localisés sur une carte topographique à grande échelle). L'opération de

(1) R. CLAIRIN,1981.Quelques réflexions sur la collecte démographique. STATECO n°25.Cité par G. THEODORE,1985.

recalage semble un préalable nécessaire avant toute intégration de données issues de la télédétection dans un système d'information géographique. C'est pourquoi il nous parait essentiel de tester sur Marseille la qualité de différentes procédures de recalage d'image satellite (voir premiers résultats au chapitre 5).

. 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*:

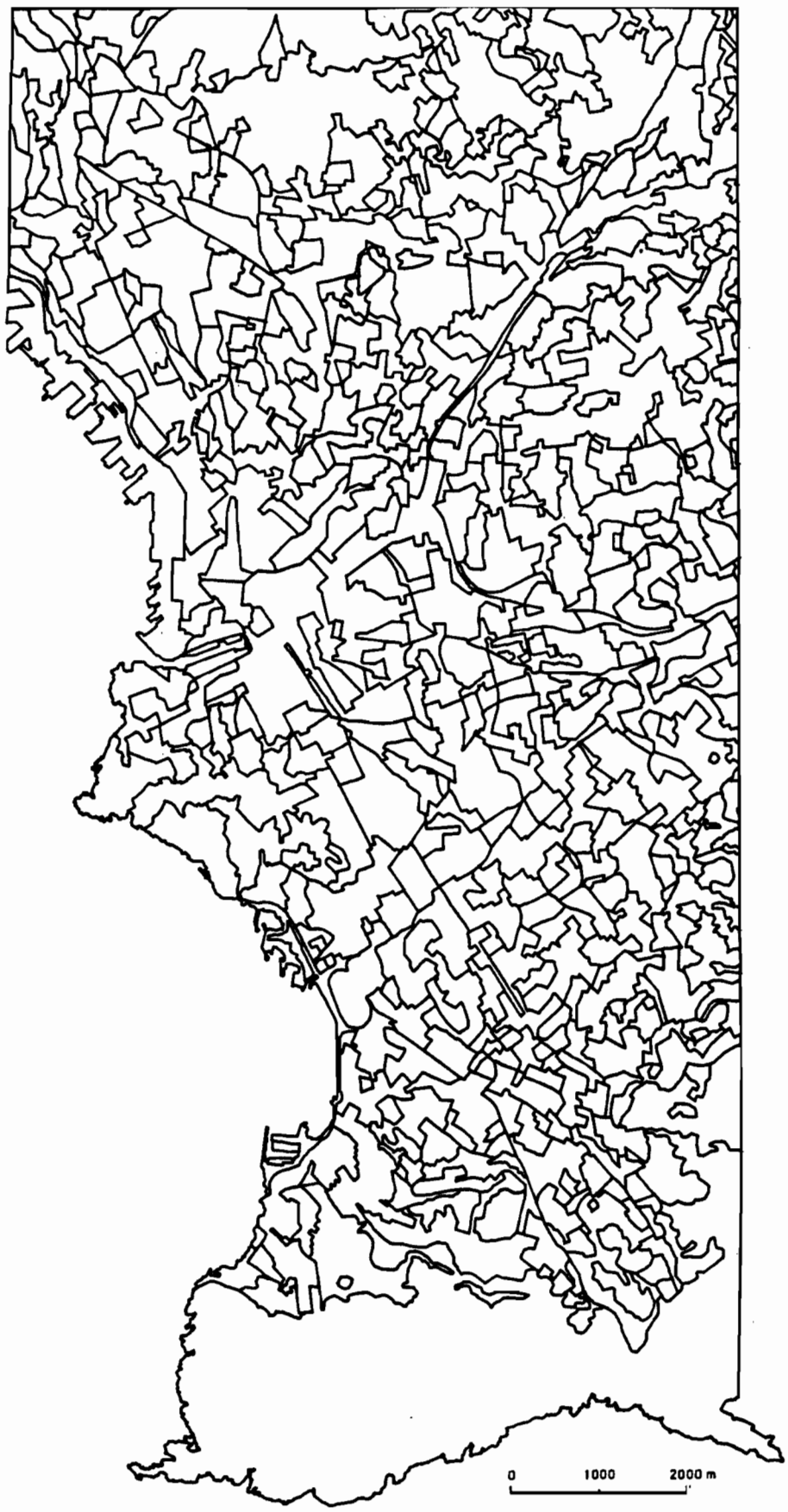
- *l'une est d'ordre purement cartographique* : le degré de précision d'un tracé cartographique est directement lié à son échelle. Plus on réduit l'échelle, plus le tracé est volontairement simplifié : les cartographes procédent ainsi, manuellement ou automatiquement maintenant, à une "généralisation" 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 pas les cartes ICOREM 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 (opération décrite au chapitre 3 de ce rapport) 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 : avoir réalisé le zonage morphologique sur une mosaïque photographique au 1/23000 induit directement un certain type de zonage

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



___ Limite de zone morphologiquement homogène

et un ensemble de descripteurs, différents de ceux que l'on aurait pu retenir en travaillant à une échelle de l'ordre du 1/5000 par exemple. *On ne peut donc pas exploiter les résultats relatifs à la morphologie à une échelle trop différente de celle à laquelle on a fait la photo-analyse*: ce serait une erreur de vouloir exploiter la typologie que nous avons établie pour travailler au 1/500, où, d'ailleurs, on perdrait la perception du tissu urbain pour ne plus observer que le bâtiment en tant qu'objet bien individualisé à cette très grande échelle.

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

Le système TIGER conserve l'information géographique en coordonnées géographiques : après la numérisation, les coordonnées table sont transformées en coordonnées géographiques à partir des coordonnés géographiques connues pour les trois ou quatre points de calage saisis en même temps que le fond cartographique. Cettre 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.

2 . 3. 2. Digitalisation des fonds de carte

La numérisation des cartes des districts de recensement et du découpage morphologique a été réalisée par l'Unité d'Infographie en une période de six mois environ (y compris contrôles et corrections). 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 à une échelle du 1/75000 : sur cet exemple, ne figurent que les arcs (contours de zones).

Les Figures 7 et 8 correspondent à des extraits du découpage INSEE (saisie réalisée au 1/2000 ou 1/5000 selon les quartiers de Marseille) :

Figure 7 - Découpage INSEE en districts de recensement . Restitution au 1/5000
(Unité d'Infographie. ORSTOM)

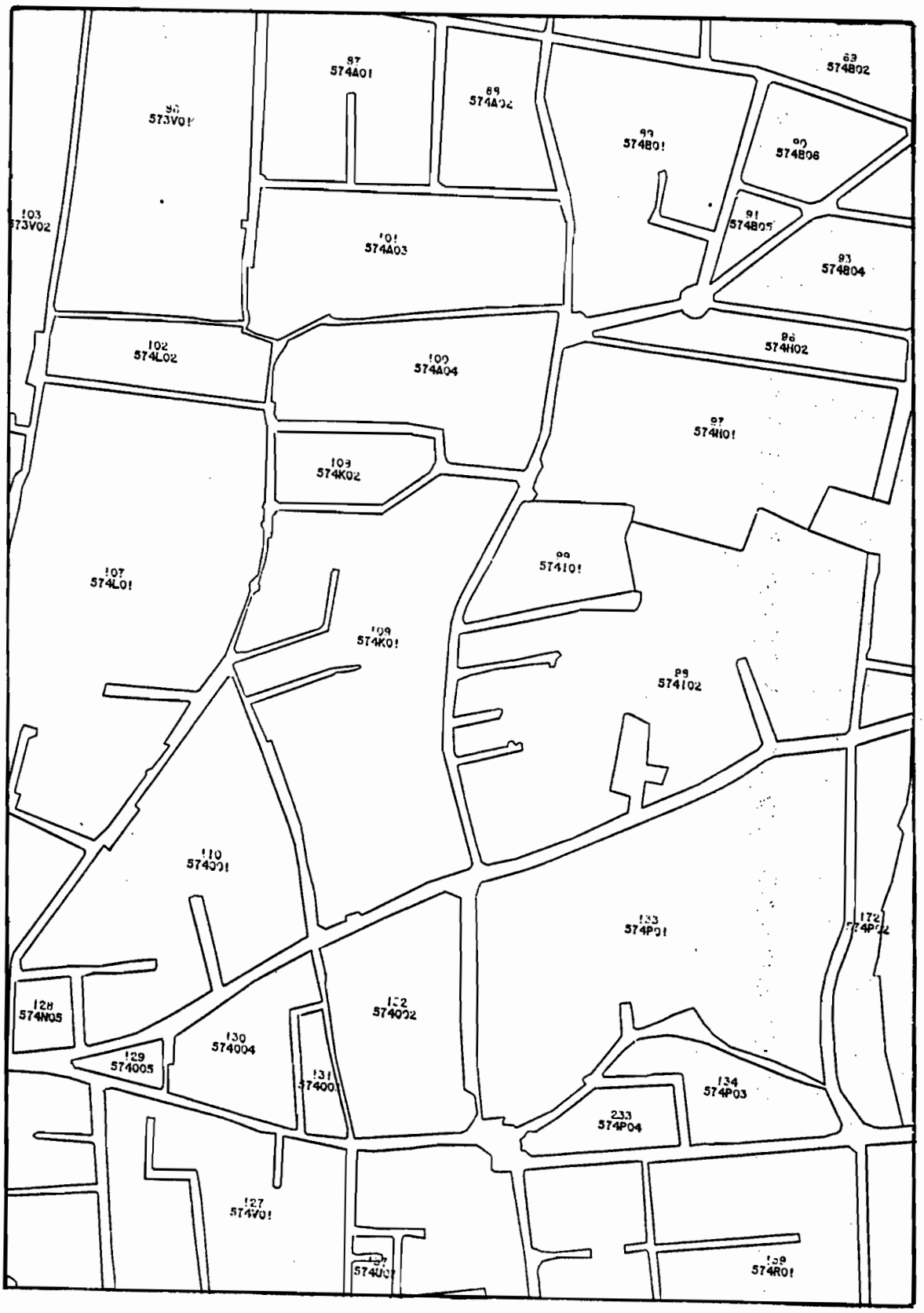


Figure 8 - Découpage INSEE en districts de recensement. Restitution au 1/15000
(Unité d'Infographie. ORSTOM)

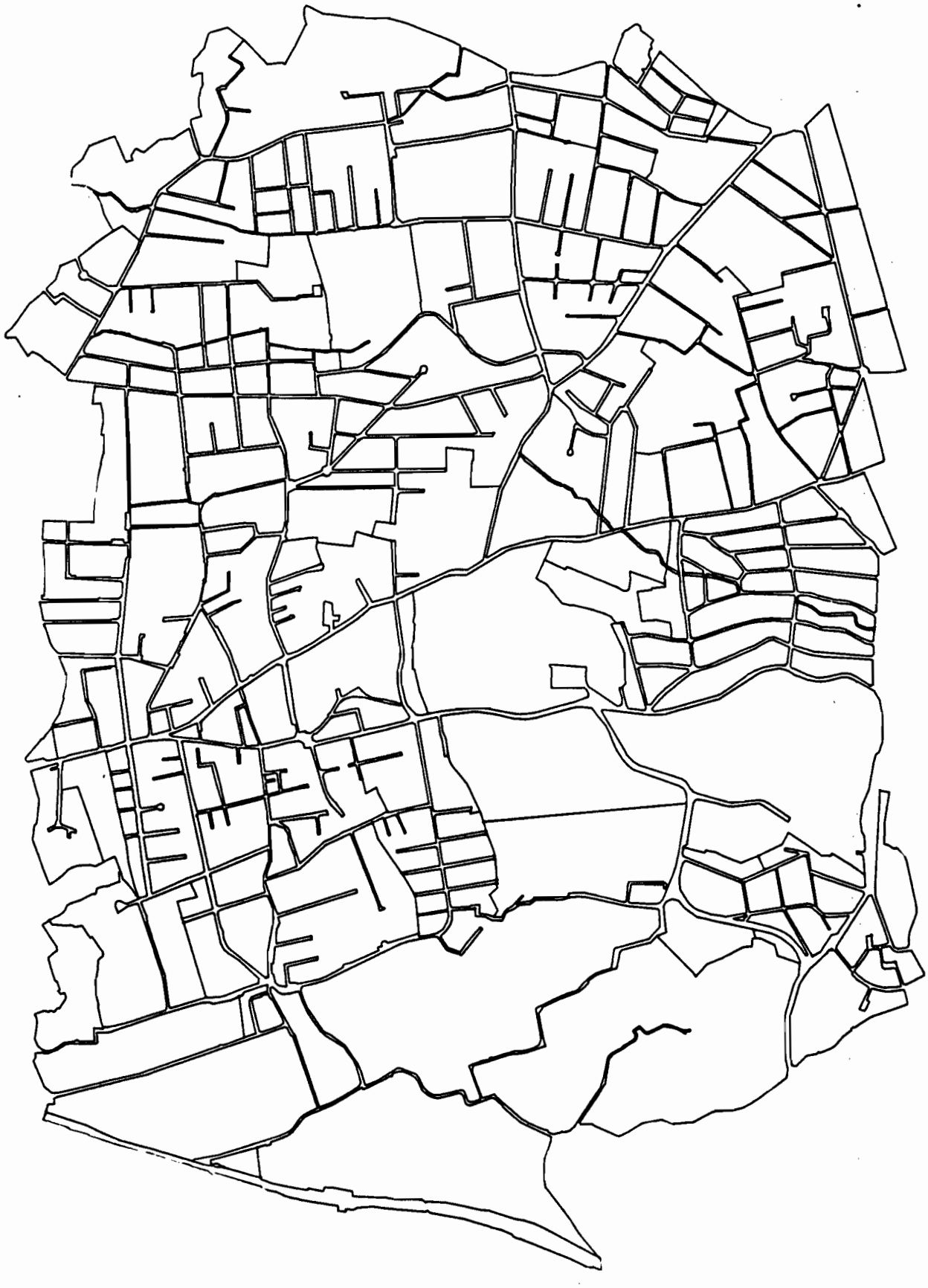


Tableau 1 - Relation morphologie. Liste des attributs

1 . Numéro de zone morphologique

2 . Densité

3 . Surface minimum de l'élément bâti

4 . Surface maximum de l'élément bâti

5 . Forme de l'élément bâti majoritaire

6 . Longueur de l'élément bâti majoritaire

7. Largeur de l'élément bâti majoritaire

8 . Valeur dominante du bâti

9 . Valeur sous-dominante du bâti

10 . Valeur dominante du non bâti non voirie

11 . Valeur sous-dominante du non bâti non voirie

12 . Hauteur du bâti

13 . Surface de l'élément bâti majoritaire

14 . Rapport longueur/largeur de l'élément bâti majoritaire

15 à 24 : Attributs 2 à 4, et 8 à 14 seuillées et recodés

25 . Type morphologique (8 types + non bâti)

Tableau 2 - Relation population. Liste des attributs

1. Population municipale

2 . Population sans double compte

3 . Nombre de ménages ordinaires

4 . Nombre de ménages collectifs

5 . Population comptée à part

6 . Nombre de résidences principales

7 . Nombre de résidences secondaires

8 . Nombre de logements vacants

9 . Nombre d'immeubles

- Figure 7 : restitution au 1/5000, avec le code de district INSEE et le numéro de zone attribué lors de la saisie
- Figure 8 : restitution au 1/15000, où ne figurent que les limites de zones.

2. 3. 3. Intégration de l'information thématique

L'information thématique relative aux fonds de carte numérisés constitue les "attributs", pour reprendre le vocabulaire des bases de données relationnelles : dans le cas de Marseille, on peut distinguer *deux relations (morphologie, et 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* (voir tableau 1) : vingt cinq variables décrivant analytiquement la morphologie, et un indicateur synthétique du type morphologique.
Ont été intégrées dans la base les variables brutes, telles qu'elles ont été produites par la photo-analyse, les variables dérivées (exemple : rapport longueur / largeur) ainsi que les variables seuillées et recodées. Nous avons donc une certaine redondance d'information, une même variable pouvant apparaitre trois fois dans la base, à des stades différents de recodage.

- *démographie* (voir tableau 2) : 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.

La base de données sur Marseille comporte donc maintenant les données morphologiques et démographiques nécessaires au programme de recherche ; l'intégration des données satellitaires (image TM recalée) vient compléter le dispositif. Cette intégration a nécessité des travaux de développement informatique (exposés au chapitre 3) sur les deux systèmes, : système TIGER de l'Unité d'Infographie, et Unité de traitement Numérique de l'Atelier de Télédétection .

3 . DONNEES DE TELEDETECTION ET SYSTEME D'INFORMATION GEOGRAPHIQUE

3.1 SYSTEMES D'INFORMATION GEOGRAPHIQUE ET BASES DE DONNEES.

Ce chapitre présente l'adaptation possible aux données localisées des systèmes de gestion de bases de données relationnelles, afin d'étendre à l'information géographique la puissance de ces systèmes. Pour cela, des opérations algébriques nouvelles seront introduites, ainsi que les méthodes effectives de réalisation informatique de ces opérations. L'élaboration du système d'information géographique TIGER sera alors présenté: l'image n'est que l'expression de l'attribut multidimensionnel de localisation et devient aussi bien le moyen que le résultat de l'interrogation de données géographiques.

3.1.1 Généralités sur les systèmes d'information géographique.

3.1.1.1 Terminologie.

Système d'information géographique (SIG) : ensemble de méthodes et de techniques permettant la manipulation et le traitement de l'information géographique.

Cartographie automatique : ensemble de techniques et d'outils permettant d'automatiser la conception graphique et la production cartographique.

Système de gestion de base de données (SGBD) : Ensemble de méthodes et de techniques informatiques qui permettent de gérer des données logiquement distinctes et de déterminer les accès aux données sans intervention manuelle.

Base de données (BD) : ensemble de données reliées et gérées par un SGBD.

Traitement d'image : ensemble de méthodes et techniques informatiques de manipulation d'images numériques (nombre fini de points, appelés pixels, à chaque pixel étant affecté une valeur numérique).

3.1.1.2 Concepts théoriques liés à l'informatisation des données localisées.

Données localisées: interprétation et espace de représentation.

Le terme entité utilisé dans la suite recouvre le concept d'ensemble d'objets d'un même type. Ces objets sont décrits, tels qu'ils existent dans le monde reel, par des variables élémentaires, appelées attributs de l'entité. Une entité géographique a la particularité d'associer une localisation graphique et temporelle à l'ensemble des attributs qui constituent la description non localisée de cette entité. La partie graphique est appelée **information graphique**, pour la distinguer de l'ensemble des autres attributs appelé **information descriptive**.

Il est important de noter qu'il n'y a pas de différence théorique entre information graphique et information descriptive, toutes deux constituent l'ensemble des attributs qui servent à décrire un objet de l'entité déterminée. Par contre, il existe toujours une dépendance fonctionnelle de l'une vers l'autre, qui motive cette distinction sémantique et qui se traduit dans les SIG par la séparation conceptuelle

du graphique et du descriptif. En effet, la localisation géographique et la donnée temporelle d'un élément de l'entité vont déterminer l'ensemble des valeurs des autres attributs de cet objet.

L'élément de base de l'espace réel est le point (au sens mathématique) de Rn*R (n=2 ou 3), et chaque objet d'une entité géographique devrait donc logiquement être un point de Rn*R. Toute entité géographique aurait alors un nombre infini d'éléments, impossibles à traiter matériellement et informatiquement. Il est donc nécessaire de représenter un nombre infini d'éléments par un nombre fini d'objets, de nature différente, c'est-à-dire de changer l'espace de représentation: ce passage est bien sûr à la base des problèmes de validité et des difficultés de traitement de l'information localisée. En fait, la description de l'entité, si ce n'est même sa définition, va être en partie déterminée par les méthodes de représentation utilisées, une même entité pouvant donner lieu à plusieurs descriptions en fonction du choix de l'espace de représentation et de l'opération de codage : c'est le cas d'une ville, qui peut être représentée par un **point** de l'espace à une certaine échelle et par une **aire** à une autre échelle, les attributs descriptifs étant alors également différents; nous aurons alors deux entités distinctes en fonction de l'échelle de description. Comme dans tout problème informatique, le choix de l'espace de représentation et le codage des données influence de manière déterminante la nature même de l'information représentée et donc l'utilisation qui en sera faite. La donnée n'existe pas, en fait, avant que soit défini quel sera l'espace de représentation utilisé pour la coder.

Tout le problème est donc de trouver un nouvel espace de représentation qui permette de réduire la description de l'espace à un nombre fini d'éléments : il faut passer des éléments de l'espace géographique (les points) à des sous-ensembles de cet espace et décrire la localisation de ces sous-ensembles avec un nombre fini de paramètres. Les entités auront alors comme objets non plus des points, mais des ensembles de points, et les valeurs des attributs descriptifs se rapporteront à ces ensembles. Dans ces conditions, il est évident que la méthode de définition de ces sous-ensembles va influencer, sinon déterminer, la définition des attributs descriptifs. Nous appelerons **entité dérivée** l'ensemble des objets décrits par ces attributs.

Deux méthodes différentes sont utilisées :

a) définition à priori des sous-ensembles (par exemple maillage de l'espace) :

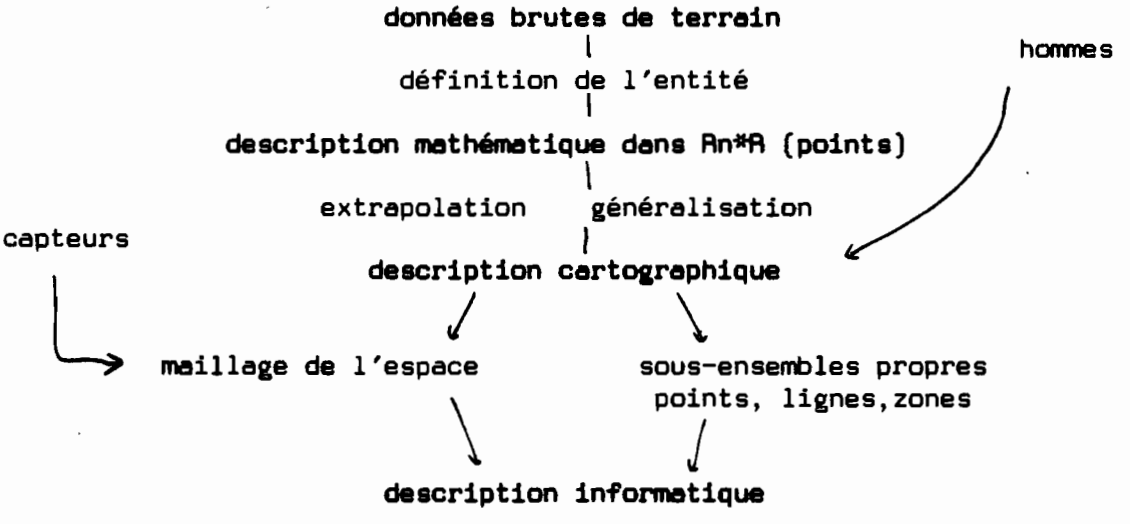
La description de la localisation des sous-ensembles est très simple, car ils sont justement choisis pour faciliter représentation et description avec le moins de paramètres possible. Par contre, on perd une partie importante de l'information descriptive, les objets de l'entité dérivée n'étant plus définis par les valeurs descriptives des points mais uniquement par les paramètres de leur localisation dans l'espace. Par exemple, si l'on prend comme définition des sous-ensembles les cellules d'un maillage, cela revient à dire que tous les points d'une même cellule auront les mêmes valeurs descriptives.
Cette méthode utilise principalement le découpage de l'espace en polygones réguliers : carrés, triangles, hexagones... Ces polygones sont en général repérés dans l'espace par un système de coordonnées propre à

leur définition (lignes, colonnes pour les carrés, etc).

b) définition des sous-ensembles en fonction des objets de l'entité :

La seconde méthode de description géographique utilisée est la définition des objets en fonction des valeurs descriptives de l'entité. On va ainsi créer des sous-ensembles que l'on classe habituellement en **zones, lignes, points**. Les sous-ensembles ainsi créés et qui ne sont pas des points peuvent être validés avec certitude (c'est le cas par exemple de zones administratives), ou être l'objet d'une certaine interprétation (zones définies à partir de relevés ponctuels,...). Cette interprétation est le plus souvent manuelle et représentée sur une carte avec plus ou moins de précision (interpolation).

En fait, le cheminement de l'information géographique est actuellement plus complexe que le passage direct point mathématique-description informatique. Aujourd'hui, l'information est rarement donnée directement en termes d'espace mathématique et de données brutes, mais après l'étape interprétation-extrapolation-cartographie: à partir de l'information brute de terrain, la donnée est interprétée et extrapolée manuellement puis représentée sur une carte qui, à son tour, est l'objet d'une représentation informatique. Ce qui signifie que l'on effectue manuellement une étape importante de la description géographique, avec tous les problèmes de validité sous-jacents. C'est seulement après cette étape qu'intervient la description informatique et le choix d'une méthode de représentation.



Si l'on veut néanmoins raisonner en terme de globalité géographique, il est nécessaire de revenir à un espace de référence qui ne peut être que l'espace mathématique. Autrement, on risque de comparer des objets géographiques qui n'ont pas la même échelle de validité. De plus, il est dangereux de vouloir raisonner exclusivement en termes de cartes : cela signifierait que l'on ne peut intervenir informatiquement qu'à partir de l'étape de mise en forme cartographique de l'information. Or c'est justement sur les traitements en amont qu'il faut actuellement faire un effort de réflexion. Par ailleurs, la tendance actuelle est d'intégrer le processus de recueil à celui de l'élaboration et du traitement de

l'information, ce qui est peu compatible avec une étape cartographique manuelle.

Reste qu'aujourd'hui, le plus souvent encore, les cartes et les tableaux de valeurs descriptives associés constituent la première étape de description et de représentation. C'est sur ces cartes que l'on trouve l'information nécessaire au codage et à la représentation informatique.

Avec la première méthode de représentation, l'attribution des valeurs descriptives au maillage prédéfini se fait en général par superposition d'une grille à une carte classique , ou par saisie automatique avec un pas déterminé grâce à un appareil à balayage (raster-scanner) (fig. 1).

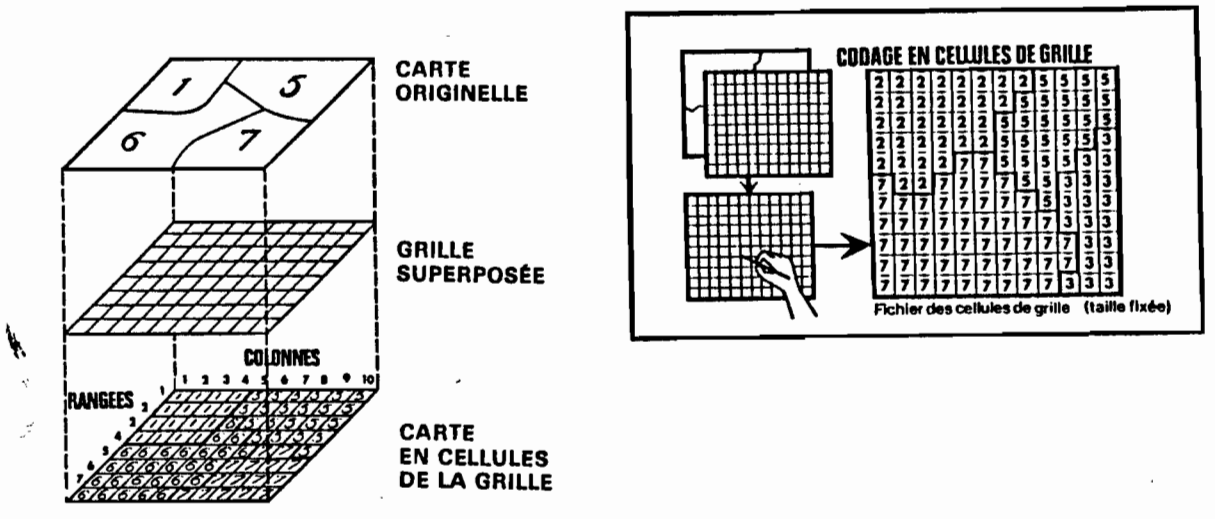


fig. 1 : Le codage des cellules d'une grille à partir d'une carte.
(d'après Dangermond 1983)

Outre la facilité de description de la localisation, les avantages de cette méthode sont dûs essentiellement à la facilité de traitement graphique: les opérations ensemblistes se font sur des cellules fixées. De nombreux systèmes s'inspirant des techniques de traitement d'images, ou provenant de la télédétection spatiale, utilisent cette représentation. Les inconvénients sont par contre importants: perte d'information descriptive, fixation à priori de la taille des objets et donc de l'échelle de validité et de traitement, important volume de stockage des données. Il existe cependant des méthodes permettant de réduire ce volume, par compactage ou structure arborescente.
Le principal inconvénient réside peut-être dans la difficulté d'évolution des systèmes qui utilisent ce mode de représentation: toutes

les données doivent être rapportées à priori à des cellules identiques pour être traitées. Une cellule peut recevoir alors un nombre important d'attributs descriptifs, au risque de perdre la notion d'échelle de validité, différente suivant les attributs en fonction du travail préliminaire d'interprétation des données brutes de terrain.

Avec la seconde méthode de représentation, les sous-ensembles définis en fonction des objets de l'entité sont habituellement classés en zones, lignes, points, suivant la dimension des sous-espaces qui les contiennent. Les points sont décrits tout simplement par leurs coordonnées (cartésiennes, sphériques, cylindriques...) dans un système de référence choisi. Les coordonnées sphériques correspondent à la longitude, colatitude, et sont donc pratiques et souvent utilisées. La précision informatique des coordonnées est un facteur important du codage, car il peut influencer la validité des données graphiques.

Les lignes sont décrites par un ensemble de points, appelés sommets : la ligne est alors représentée graphiquement par l'ensemble des segments définis par ces sommets. Le nombre de sommets en fonction du rayon de courbure détermine la précision du codage : elle dépend à la fois de la méthode de saisie, des matériels utilisés et de l'opérateur s'il existe.

Les zones sont repérées par leur contour, qui est considéré soit comme propre à chaque zone, soit comme un ensemble de frontières topologiques.

Dans le premier cas, la zone est décrite graphiquement par un ensemble de lignes fermées, appelées boucles. Si l'on conserve ces lignes telles quelles, il y a alors duplication d'une partie de l'information graphique, un contour étant en général commun à deux zones: globalement, ces lignes se retrouveront donc stockées deux fois.
Dans le second cas, les contours sont considérés comme un ensemble de frontières, chaque frontière étant décrite par une ligne. Les extrémités de chaque ligne sont alors appelées noeuds, les lignes elles-même étant nommées arcs. Chaque noeud appartient au moins à deux arcs, et l'ensemble des arcs d'une zone doit constituer une courbe fermée.

Des entités plus complexes peuvent être définies à partir d'entités dont les éléments sont des points, lignes ou zones. Par exemple, on peut définir des entités dont les objets sont des couples d'éléments d'autres entités. La représentation de ces objets est donc un peu plus complexe, car elle fait appel non seulement à la description graphique des éléments, mais également à la description des relations qui définissent les objets. L'exemple le plus courant est une relation dont les objets sont des couples ordonnés de points (description de flux, migrations, réseaux linéaires ...). La description de la relation peut s'apparenter à celle d'un graphe.

Les structures de représentation de l'information localisée.

Le choix de la structure de représentation se pose pour tout type de donnée: il doit être effectué en fonction des besoins ultérieurs et leur être adapté. Pour la partie graphique de l'information géographique, cette structure est surtout fonction de trois critères: facilité des manipulations métriques et topologiques, compacité, structuration.

Plusieurs niveaux de représentation peuvent ainsi être nécessaires suivant les traitements à effectuer. Des procédures d'identification et de reconnaissance permettent de passer d'un niveau de représentation à un autre, suivant une hiérarchie d'opérateurs de reconnaissance (de l'ensemble de points au contour, de la courbe à la forme caractéristique, etc.), qui permettent également de passer d'un ensemble de représentation infini à un ensemble fini.

Le premier niveau de représentation correspond aux coordonnées des points de l'espace décrivant les objets (par exemple pour une ligne, l'ensemble des points qui permettent l'approximation des éléments graphiques à une précision donnée). A un niveau supérieur, la représentation des éléments fait appel à des relations entre ces éléments, relations qui doivent, en principe, donner l'ensemble de l'information nécessaire aux problèmes à résoudre ultérieurement. On distingue ainsi:

* des relations unaires, ou formes primitives, qui donnent le type de l'élément et ses paramètres (trait et longueur ou orientation, par exemple);

* des relations binaires, numériques (basées sur des indices de ressemblance ou de dissemblance), ou topologiques (basées sur les notions de voisinage, de position, etc.). La représentation est souvent donnée sous forme de graphe dont les sommets sont les éléments, et les arcs entre sommets créés si la relation est vraie . Dans le cas numérique, les arètes sont valuées par les distances entre les sommets. Si le nombre de formes primitives est faible, le nombre de relations binaires ou n-aires possibles pour décrire les éléments graphiques est très grand : tout le problème consiste alors à choisir les bonnes relations permettant une description minimum, en fonction du but poursuivi (Simon 1984). La description d'éléments à caractère géographique utilise de préférence des relations invariantes par transformations isométriques.

Nous avons vu qu'un SIG peut contenir une grande quantité d'informations graphiques. Il est donc important que la structure de représentation choisie permette de compacter cette information pour pouvoir la stocker sur le moins de place possible, sans néanmoins que l'opération de transformation n'altère les performances du système. Les méthodes de compactage sont évidemment fonction des méthodes de représentation de l'information graphique (maille, vecteur), et doivent si possible permettre dans la forme compactée des manipulations graphiques ou des calculs métriques élémentaires.

D'autre part, l'interrogation des données d'un SIG consiste essentiellement en une opération de recherche et de sélection sur la localisation: sélection d'éléments dans une fenêtre, dans le voisinage d'un objet, etc. Il est donc nécessaire d'organiser et de structurer l'information par rapport à la localisation, de manière à ce que ces opérations soient exécutables en un temps acceptable par l'utilisateur.

Ce sont donc ces trois critères : facilité de manipulation graphique, compacité, structuration qui vont guider le choix des méthodes de représentation des éléments graphiques.

Dans la représentation des données géographiques en points, lignes, zones, nous avons vu que, si les points ne posent pas de problèmes particuliers (leurs coordonnées sont conservées directement), les éléments graphiques à prendre en compte sont des courbes habituellement représentées par des suites de points dont on connait les coordonnées. Les méthodes de compactage consistent à réduire l'espace mémoire nécessaire au stockage direct des coordonnées de l'ensemble des points.

Les méthodes de compactage de représentation de courbes sont pour la plupart basées sur un codage incrémental, appelé encore codage en chaine: les courbes sont approchées par un ensemble de points où chaque point est déduit du précédent par la donnée d'un paramètre (angle, distance, courbure,...).
Le principal inconvénient du codage en chaine provient de son caractère relatif: les coordonnées d'un point doivent être calculées à partir du début de la chaine. Ceci complique ou rend impossible à priori les manipulations graphiques sur les éléments codés (intersection, inclusion, calculs topologiques et métriques, ect.). Pour remédier à cette difficulté, on rajoute des informations, notamment sur les caractéristiques topologiques, le plus souvent sous forme de relations binaires entre primitives graphiques (la courbe est frontière de telles zones, deux zones sont connexes,etc.). Le système DIME est l'un des premiers à avoir incorporé des structures topologiques (Peucker 1975).

La représentation de l'espace géographique par un découpage arbitraire, utilisée comme nous l'avons vu dans de nombreux systèmes, produit un important volume d'information vu le nombre des cellules crées par cette méthode, dès que l'espace d'étude devient significatif. A moins de disposer de moyens de stockage importants et rapides (gros disques durs avec accès rapides), cette représentation ne peut être mémorisée telle quelle.
Le compactage consiste ici à essayer de regrouper les cellules de même valeurs : ces méthodes ne sont donc valables que si une valeur unique peut être attribuée à chaque cellule, c'est-à-dire que l'ensemble des valeurs des attributs de l'entité doit être fini.
La méthode élémentaire de compactage d'une grille de cellules est de balayer cette grille ligne par ligne, et pour chaque ligne de ne conserver que les séquences de cellules de même valeurs. La méthode peut être améliorée en testant plusieurs orientations de balayage, car les données peuvent être agencées suivant une direction privilégiée qui produira alors un compactage beaucoup plus performant (Amidon and Atkin 1971). La qualité du compactage dépend bien sûr de la complexité de l'information graphique, à savoir le nombre de composantes connexes. Ces méthodes sont faciles à mettre en oeuvre mais empêchent toute structuration des cellules entre elles et toutes opérations géométriques sous la forme compactée.
La seconde méthode de compactage consiste à changer de représentation en conservant les zones homogènes décrites par la suite des mailles formant le contour : les méthodes de représentation de contours sont alors les même que celles vues pour les courbes quelconques (codage en chaine, codage topologique). Cette représentation, très compacte par rapport au volume de données initiales, peut être obtenue directement à partir d'une numérisation des contours (manuelle ou automatique), ou par extraction des contours par des opérateurs de reconnaissances, notamment si l'image est fournie en pixels par un procédé automatique (caméra de télévision, raster-scanner, capteur à balayage). Le compactage est très

efficace par rapport au volume initial des données, et la représentation s'apparente aux méthodes vues précédemment pour les données vectorielles. Malheureusement, on perd ainsi le principal intérêt de la représentation en cellules, à savoir la facilité de manipulation graphique (comparaison de cellules, union, intersection...).

Pour conserver l'intérêt de ce type de représentation (la structuration en cellules identiques) et réduire le volume nécessaire au stockage, des méthodes de structuration ensembliste ont été développées, de manière à ne pas perdre les relations de voisinage des cellules entre elles.

Une autre approche plus récente est basée sur des représentations arborescentes par subdivision successive de l'espace, supposé être ici un ensemble de 2n sur 2n cellules (Rosenfeld and Samet 1980). Si la région homogène à représenter ne couvre pas l'ensemble de l'espace, on subdivise cet espace en quatre carrés de 2n-1*2n-1 cellules, en recommençant ainsi pour chaque carré aussi souvent que nécessaire jusqu'à ce que l'on obtienne des blocs homogènes (qui peuvent être de simples cellules) entièrement inclus dans la région ou disjoints d'elle. Cette décomposition peut être représentée par un arbre 4-aire (**quadtree**) : la racine représente l'espace tout entier, les quatre fils d'un noeud représentent les quatre carrés issus de la division, et les feuilles les blocs pour lesquels aucune division ultérieure n'est nécessaire. Il existe de nombreux algorithmes pour la construction et la manipulation des quadtrees (Samet 1984).

Cette représentation est assez compacte et bien adaptée aux opérations graphiques et aux calculs des propriétés géométriques. Son principal intérêt est d'être de résolution variable, car un détail est représenté par une précision suffisante sans que les grandes régions homogènes ne nécessitent beaucoup de mémoire, tout en conservant une structure entre régions qui facilite les calculs graphiques.

<u>La structuration et l'indexation de données multidimensionnelles.</u>

La structuration des données géographiques en vue d'améliorer les opérations de recherche en mémoire de stockage est un problème plus général qui concerne l'ensemble des données graphiques de la base, quelsque soient les modes de représentation choisis. Il s'agit ici des méthodes de partage ou d'indexation des objets, méthodes basées sur leur localisation absolue au niveau de l'espace initial de représentation (R2 ou R3). La nécessité d'une telle structuration est évidente car l'interrogation d'une base de données localisées va faire intervenir en priorité la localisation; une interrogation classique, dans un SIG, sera en effet du type: retrouver dans telle zone tous les objets de tel type qui répondent à telles conditions. Cette interrogation se décompose d'une manière optimale en:

-retrouver tous les objets dans une fenêtre contenant la zone d'étude.
-sélectionner les objets qui répondent aux critères de sélection.

La première opération, la sélection spatiale, est de loin le critère le plus contraignant, car il repose sur un attribut multidimensionnel (n=2 ou 3), face aux opérations de sélection sur les attributs monodimensionnels. Devant s'effectuer en priorité pour optimiser ces

temps de recherche, elle nécessite l'accès possible à l'ensemble des données de la base sur les mémoires de stockage et représente l'opération la plus coûteuse en temps d'exécution. L'organisation des données sur les disques doit donc être conçue par rapport à l'attribut de localisation.

Les méthodes classiques de gestion de données ont pour objet de réduire le nombre des opérations d'accès aux mémoires de stockage. Elles sont basées sur des techniques d'indexation d'attributs monodimensionnels ordonnés (accès aléatoire, séquentiel indéxé) et ne sont pas applicables directement à l'attribut de localisation, qui est par nature multidimensionnel sans relation d'ordre compatible avec la structure euclidienne (Knuth 1973, Lee 1980) : on peut certes organiser l'indéxation d'un ensemble d'objets par leurs noms, grâce à l'ordre lexicographique par exemple, les regrouper en paquets grâce à la relation d'ordre, et stocker les objets de chaque paquet sur un secteur contigu des disques magnétiques, de manière à pouvoir accéder au paquet en une opération d'accès mémoire; en revanche, il n'est pas possible, pour le cas le plus simple de données ponctuelles, de structurer ces points de la même manière, avec un ordre sur les abscisses ou sur les ordonnées, indépendamment l'un de l'autre, sans perdre toute relation de voisinage entre deux points (il n'existe pas de relation d'ordre total dans les espaces euclidiens de dimension supérieure à 1). Une telle structuration est donc très inéfficace dans le cas qui nous intéresse, à savoir retrouver rapidement tous les objets d'une même fenêtre ou tous les objets proches d'un objet donné : ceci impose donc un rapprochement logique (ou mieux physique sur les mémoires de stockage) des objets basé sur leur localisation. Il est nécessaire de développer pour cela des méthodes de gestion adaptées aux données localisées, ou plus généralement multidimensionnelles, méthodes bâties sur une structuration basée sur la notion de voisinage et de sous-ensemble borné pour la métrique choisie dans l'espace de représentation.

L'idée la plus simple, calquée sur les techniques du séquentiel indéxé (Gardarin 1983), est de décomposer l'espace en sous-ensembles bornés, appelés feuilles, d'indexer les objets par rapport à ces feuilles et d'effectuer les recherches séquentiellement dans les feuilles sélectionnées pour la fenêtre géographique d'étude.
Pour optimiser le temps d'accès à la mémoire de stockage, la taille d'une feuille sera choisie de manière à être lue en une seule opération d'entrée sortie, ce qui impose un stockage de tous les objets de la feuille sur un secteur contigu de la mémoire de masse. Le nombre des objets pouvant être très variable d'un endroit à l'autre de l'espace géographique, cette taille ne peut être fixée à priori sans perdre une place importante de la mémoire (les feuilles se trouvant dans des zones de faibles densité d'objets seraient alors presque vides). La taille d'une feuille doit donc être variable en fonction de la densité d'objet: si le nombre des objets d'une feuille dépasse le maximum, fonction de la taille de la page mémoire, cette feuille sera divisée en sous-feuilles, et chaque objet réaffecté à l'une de ces sous-feuilles. Cette division peut se faire de manière à faciliter le choix de la sous-feuille, pour une localisation donnée, et donc naturellement par dichotomie: la feuille est alors divisée en 4^m (pour n=2) ou 8^m (pour n=3) sous-feuille, où m n'est pas trop élevé pour ne pas créer un grand nombre de feuilles presques vides, et donc une importante place perdue sur les disques (le cas fréquent est m=1).

L'éclatement d'une feuille en sous-feuilles est une opération complexe, qui impose également la réorganisation des index basés sur les numéros de feuilles.
L'organisation des feuilles et les différents niveaux d'éclatement peuvent être représentés par un **m-quadtree** (**m-octree** pour n=3). Chaque noeud non terminal contient la description de la division et pointe vers ses fils, et chaque noeud terminal contient l'adresse de la feuille dans la mémoire de stockage. Un tel quadtree est appelé **"région quadtree"**, par opposition aux quadtrees vus plus haut, servant à représenter directement un espace découpé en cellules.

La recherche sur la localisation (sélection géographique) se fait donc par descente dans l'arbre: à chaque niveau sera choisi le fils qui contient la feuille cherchée, jusqu'au noeud terminal qui donne l'adresse de la feuille. Si m est choisi de manière à ce que la lecture d'un noeud nécessite qu'une entrée-sortie (la recherche du fils se faisant en mémoire centrale), l'adresse de la feuille est déterminée en quelques entrées-sorties (E/S). Par exemple, pour m=3, l'adresse d'une feuille parmi 16.106 se fera en 4 E/S, soit environ 0.2 secondes: sur un espace géographique de la taille du Brésil (625 degrés carrés), la recherche d'une feuille de 30 mètres de coté nécessiterait moins d'une demi-seconde. Pour m=1, ce temps est au maximum d'une seconde.
La recherche des feuilles voisines fait appel à des cheminements dans le quadtree, et ne demande donc elle aussi que quelques opérations d'E/S (Samet 1984).

L'inconvénient majeur de la partition en "region quadtree" est le coté arbitraire du découpage en quatre régions égales, ce qui peut provoquer un important deséquilibre de l'arbre si les données ne sont pas réparties de manière relativement homogène dans l'espace, et donc une baisse notable des performances du système. Pour remédier à ce problème, on peut utiliser un découpage récursif qui tient compte de la position des objets de manière à optimiser le nombre des objets présents dans chaque partie du découpage, à l'image des **k-d tree** (Matsuyama 1984). Bien sur, le découpage doit alors être mémorisé car il n'est plus implicite comme dans les "region quadtree"; la recherche des feuilles du k-d tree contenues dans la fenêtre d'étude se fait également par descente dans un arbre binaire, mais en consultant pour chaque noeud la manière dont l'espace a été découpé.

La recherche d'un objet doit pouvoir se faire également sur la valeur de l'une de ses clés, sans connaître la localisation : par exemple, un nom de ville doit permettre de retrouver rapidement l'adresse de la feuille qui la contient. Ce type de recherche nécessite la mémorisation pour chaque objet de l'adresse de la feuille auquel il appartient: ceci revient à créer pour la clé choisie un index dense, et d'organiser cet index de manière classique en arbre B+ ou en paquets: quelques opérations d'E/S suffiront à retrouver l'adresse de la feuille qui contient l'objet. Les objets voisins se trouveront alors dans la feuille sélectionnée ou dans les feuilles adjacentes, faciles à déterminer grâce à l'indexation principale sur la localisation en cheminant dans les arbres associés. Ces index, nécessairement denses, peuvent occuper une place importante.

Comme nous l'avons vu plus haut, l'éclatement et l'agrégation de feuilles imposent la réaffectation des objets dans les feuilles et donc

la réorganisation des index. Toutes ces opérations sont relativement complexes à mettre en oeuvre.

L'affectation d'un objet dans une feuille est simple pour les données ponctuelles, car il n'existe jamais qu'une seule feuille qui contienne le point. Tout se complique lorsque ces objets sont des lignes ou des zones, car ils peuvent alors ne pas être contenus strictement (au sens de l'inclusion) dans la feuille et l'unicité du choix d'affectation n'est plus assurée. Diverses solutions peuvent être apportées à ce problème :
- affecter un centroïde à chaque objet et à stocker l'objet dans la feuille qui contient le centroïde. Une telle solution a l'inconvénient majeur d'éliminer d'une feuille tous les objets qui n'y ont pas leur centroïde, même s'ils ont une partie importante incluse dans la feuille.
- diviser les objets en sous-objets en introduisant le bord de la feuille, mais cette méthode complique beaucoup les manipulations nécessaires au stockage et à l'extraction des données.
- conserver avec chaque feuille le numéro des enregistrements de tous les objets totalement ou partiellement contenus dans la feuille: la sélection d'une feuille par descente dans le k-d tree ou le "region quadtree" permet de determiner l'ensemble des feuilles à consulter ainsi que le numéro des objets dans chaque feuille (Matsuyama 1984).
- conserver une feuille à chaque niveau de la structure arborescente, où ne seront stockés que les objets qui ne peuvent être contenus entièrement dans aucunes des feuilles des niveaux suivants. Si l'on arrive à résoudre le problème de l'affectation des petits objets mal placés par rapport à la division de l'espace, cette structure multi-niveau introduit une hiérarchie dans l'encombrement des objets qui peut également constituer un critère de sélection intéressant (Frank 1981).

Notons pour conclure ce chapitre que très peu de systèmes mettent en oeuvre ces techniques de structuration, qui relèvent encore pour la plupart du domaine de la recherche et de l'expérimentation. De nombreux systèmes ne sont pas aptes à fournir des performances valables lorsque la taille de la base de données géographiques est importante, et n'offrent pas alors les possibilités d'interactivité qui semblent néanmoins nécessaires à un système performant.

3.1.2 **Bases de données et information géographique : l'extension du modèle relationnel.**

La conception relationnelle des données permet de définir des méthodes et des techniques de gestion et de manipulation très puissantes. Etendre cette conception aux données localisées représente une avancée importante quant aux possibilités offertes par la manipulation de ces données pour la gestion et la compréhension de l'espace. Les données localisées, ce sont avant tout des cartes et des légendes associées : c'est du moins ainsi qu'elles sont perçues par la plupart. Pour raisonner en terme de gestion de données, il va être nécessaire d'inverser cette approche, et de considérer les données géographiques localisées comme des collections d'éléments décrits par un certain nombre de paramètres dont la localisation est un entre tous. C'est, en quelque sorte, remettre la carte à sa place comme l'expression d'un paramètre, comme la visualisation d'un attribut au même titre qu'une liste de valeurs numériques sur une imprimante. Nous allons donc, dans un premier temps, rappeler des notions fondamentales sur les systèmes de gestion des bases de données relationnelles.

3.1.2.1 Notions classiques sur les SGBD.

Un SGBD, c'est avant tout un interface entre l'usager et les mémoires de stockage lui donnant l'illusion d'avoir des données stockées et assemblées comme il le souhaite et d'être le seul à les utiliser. C'est également et surtout un outil de gestion permettant de rechercher, modifier, insérer efficacement des données dans une grande masse d'informations, partagée par tous les usagers suivant leurs droits d'accès, chacun travaillant sur sa vision et sa propre structuration de l'information.

Ainsi, à l'inverse des systèmes de gestion de fichiers (SGF), les SGBD permettent la description des données (définition des noms, formats, caractéristiques) de manière indépendante de leur utilisation (recherche, mise à jour). Ils se composent grossièrement de trois couches successives de fonctions, depuis les mémoires de stockage jusqu'à l'usager :

- le premier niveau (SGF) assure la gestion des fichiers et relie organisation physique (sur les disques) et organisation logique (fichiers). Ce niveau est inaccessible à l'usager ;

- le second niveau assure la gestion des données stockées, le placement, l'assemblage, les liens entre les données (pointeurs), et les structures permettant de les retrouver rapidement (schémas) : c'est le système d'accès aux données (SGBD interne) ;

- le troisième niveau est chargé de la présentation des données aux applications et aux usagers, de l'analyse, de l'interprétation et de l'optimisation des requêtes, et de la mise en forme des données à échanger avec le monde extérieur (SGBD externe).

Chaque niveau utilise des langages adaptés : langage de description de données (LDD), langage d'interrogation (LID), langage de manipulation (LMD).

Comme nous l'avons dit, l'objet principal des SGBD est l'utilisation partagée d'un même ensemble de données par plusieurs utilisateurs n'ayant ni la même vision, ni les mêmes applications. Un SGBD doit donc assurer la non redondance, la sécurité, la cohérence des données stockées. Ces objectifs impliquent :

- l'indépendance physique entre structure de donnée et structure de stockage.
- l'indépendance logique entre structure de donnée modélisant le monde réel et structure de donnée d'une application.

Il est nécessaire de mettre en place plusieurs niveaux de description pour les données :

Le niveau conceptuel correspond à la structure canonique des données telles qu'elles apparaissent dans le monde réel. Plusieurs modèles peuvent être utilisés pour définir le schéma conceptuel. On utilise en général le modèle entité association qui permet de définir :

- les types de données élémentaires définissant les attributs des objets : nom de département, valeur de pH, type de sol, nombre de lits ;
- les types de données composées correspondant aux objets ou entité du monde réel : département, horizon, pédologie, hôpital ;
- les types de données composées permettant de décrire les objets correspondant à des associations du monde réel. Ces types contiennent des attributs provenant d'autres entités : malade, décrit par des attributs nom de personne, nom d'hopital, date d'entrée, etc.

Le niveau conceptuel est complété de règles de cohérence entre les attributs d'une entité ou d'une association.

Le second niveau de description correspond aux données telles qu'elles sont percues par l'application ou l'usager : c'est le niveau externe. A ce niveau peuvent bien sûr coexister plusieurs schémas externes, chacun se rapportant à une application.
Le troisième niveau correspond à la structure de stockage (fichiers, chemins d'accès aux données).

C'est le SGBD qui va faire le lien entre les différents niveaux de description : il a besoin pour cela de toutes les descriptions de données et des règles de correspondance entre ces descriptions, qui sont conservées dans ce que l'on appèlle le **dictionnaire des données**.

Prenons un exemple.
Considérons la gestion de données cadastrales. Nous avons un certain nombre d'entités : la parcelle, le propriétaire, la construction, le plan d'occupation des sols, le lotissement, les réseaux aériens et souterrains, le quartier ..., et des associations : le permis de construire, le certificat de conformité...
Chaque entité ou association possède des attributs : pour la parcelle, nous pourrions avoir le numéro de parcelle, le numero de section, le nom du propriétaire, le classement au P.O.S., le coefficient d'occupation, le prix au mètre carré; pour la construction, le numéro de la parcelle, la section, la surface habitable, le nombre d'étages, la fonction, le type, pour le permis, le numero de permis, le nom du propriétaire, le numéro de la parcelle, la section, la date de dépôt, etc.

Une gestion des parcelles n'aura pas besoin de l'ensemble de ces attributs. Le schéma externe présenté par le SGBD à l'application fera apparaître, par exemple, une structure contenant :
parcelle, avec numéro de parcelle, section, nom du propriétaire, permis demandés, permis obtenus, constructions sur la parcelle, fonction de la construction, type.

La description des données va indiquer le type (qualitatif, quantitatif, nominal, ordinal, numérique), les formats, ainsi que des contraintes de validité et d'intégrité.

3.1.2.2 Le modèle relationnel.

Le modèle relationnel consiste à représenter les entités et associations du monde réel par des relations, ensembles de valeurs d'attributs qui décrivent les objets des entités ou associations. Chaque relation est caractérisée par son schéma (ensemble d'attributs), porte un nom, et les éléments de la relation sont appelés des tuples.

Une mauvaise perception du réel et une mauvaise conception des entités et associations engendrent des problèmes dans les relations qui leur correspondent. Par exemple, si l'on définit une relation canton avec les attributs nom de canton, nombre d'habitants, nom de département, taux de croissance du département, des informations vont être redondantes car les données sur le département seront répétées pour tous les cantons d'un même département.

Il est nécessaire pour bien définir les relations d'étudier les propriétés sémantiques des données et de définir les dépendances entre attributs qui en résultent. Ces dépendances se classent en trois types :

a) Les dépendances fonctionnelles : un attribut B dépend fonctionnellement d'un attribut A si à toute valeur de A correspond une valeur unique de B, soit
A --> B <=> ((xy R et xy' R) => y=y')
On peut introduire ici la notion fondamentale de clé d'une relation : c'est un ensemble minimal d'attributs qui détermine de manière unique tous les autres (tous les attributs sont en dépendance fonctionnelle de la clé) .

b) Les dépendances multivaluées qui caractérisent l'indépendance de deux ensembles d'attributs corrélés à un même troisième : B dépend de A, si toute valeur de A détermine un ensemble de valeurs de B indépendamment des autres attributs de la relation, soit
A ->-> B <=> ((xyz R et xy'z' R) => (xy'z R et xyz' R))

c) les dépendances de jointures qui permettent de décrire les relations entre sous-ensembles d'attributs d'une relation.

A partir de l'ensemble des attributs et de leurs dépendances, des algorithmes pourront déterminer les entités et associations canoniques du réel et fournir des relations qui ne souffrent pas d'anomalies : cette opération se fera par décomposition successive jusqu'à l'obtention de relations normalisées (Gardarin 1983). Ces relations seront alors facilement manipulables grâce aux opérations de l'algèbre relationnelle,

qui sont à la base des langages d'interrogation et de manipulation des SGBD relationnels :

- l'union de deux relations A et B de même schéma : la relation résultante est de même schéma que A et B et a pour tuples l'union des tuples de A et B ;

- le différence de deux relations A et B de même schéma : la relation résultante est de même schéma que A et B et a pour tuples la différence ensembliste des tuples de A et B ;

- le produit cartésien de deux relations de schémas quelconques : la relation résultante a pour schéma la concaténation des schémas de A et B et pour tuples le produit cartésien des deux ensembles de tuples de A et B ; cette opération, comme on peut s'en douter, est des plus couteuses ;

- la projection d'une relation sur un certain nombre de ces attributs : la relation résultante a pour schéma les attributs sur lesquels la projection est faite, les tuples étant obtenus par élimination des valeurs des attributs n'appartenant pas au schéma résultant, ainsi que par élimination des tuples en double.

- la restriction d'une relation A par une qualification Q : la relation résultante est de même schéma et a pour tuples les tuples de A qui satisfont la qualification Q.

- la jointure de deux relations A et B selon une qualification Q : la relation résultante est la restriction du produit cartesien de A et B par la qualification Q. Cette opération est fondamentale, car elle permet de relier deux relations sur un ou plusieurs de leurs attributs et de créer une troisième relation résultant du croisement. Par exemple, la relation de schéma numéro de parcelle, nom de section, nom de propriétaire, type de construction, classement au POS résulte de la projection d'une jointure entre la relation parcelle, de schéma numéro de parcelle, nom de section, nom de propriétaire, classement au POS..., et la relation construction, de schéma nombre d'étages, type, numero de parcelle, nom de section, ..., jointure selon la qualification nom de section = nom de section et numéro de parcelle = numéro de parcelle : si la relation parcelle contient les tuples

parcelle	numero de parcelle	nom de section	nom de propriétaire	P.O.S
	10	H	Durand	UC
	25	G	Leroy	Na
	110	F	Martin	Nb

et la relation construction les tuples

construction	nombre d'étages	type	numéro de parcelle	nom de section
	2	immeuble	12	H

1	pavillon	10	H
2	immeuble	10	H
3	immeuble	110	P
2	pavillon	25	G

la jointure indiquée aura pour tuples

numéro de parcelle	nom de section	nom de propriétaire	type de construction	classement au POS
10	H	Durand	pavillon	UC
10	H	Durand	immeuble	UC
25	G	Leroy	pavillon	Na

Il est important de noter que la jointure ne conserve pas la notion de clé : si A1 est clé de la relation R1 et A2 clé de la relation R2, A1 ne sera vraisemblablement plus clé de la relation R3 résultat d'une jointure entre R1 et R2. L'ensemble (A1,A2) sera une clé de la relation R3 s'il fait parti de cette nouvelle relation.

Les interrogations de la base de données peuvent être exprimées directement en terme d'opérations relationnelles (ex. SQL), ou grâce à des langages d'interrogation basés sur la vérification de formules dont les variables sont soit des tuples (ex. INGRES), soit des valeurs d'attributs (ex. QBE).

Ainsi, la question "imprimer les noms des propriétaires de parcelle contenant une construction de deux étages ou plus" sera décomposée séquentiellement en :

- une restriction de la relation construction sur la qualification (nombre d'étages >= 2),
- une projection de la relation construction sur les attributs numéro de parcelle et nom de section,
- une jointure entre les relations construction et parcelle comme dans l'exemple précédent,
- une projection de la relation résultante sur l'attribut nom de propriétaire.

3.1.2.3 L'intégration des données localisées.

Objectifs.

Intégrer la variable de localisation dans le schéma relationnel, c'est bénéficier sur cet attribut de toute la puissance du modèle, c'est simplifier les opérations de manipulation et les interrogations utilisant la localisation dans l'espace en mettant ces opérations au même niveau que les opérations sur les relations non localisées.

Nous allons donc compléter l'algèbre relationnelle par de nouvelles opérations algébriques liées au caractère multidimentionnel de

l'attribut de localisation :

- la restriction spatiale, qui correspond à la sélection d'objets par rapport à la localisation,
- la jointure spatiale, qui correspond à la mise en relation de deux objets par rapport à leurs localisations respectives,
- la projection spatiale, qui correspond à l'opération d'impression de l'attribut de localisation et qui est une opération de cartographie.

Ces nouvelles opérations, et surtout les deux premières, vont permettre de manipuler la localisation grâce à des opérations algébriques et logiques : c'est dans ce sens que l'image, considérée habituellement comme un résultat final, va devenir également le résultat intermédiaire d'une séquence d'opérations de gestion et de manipulation de données qui peut très bien ne pas avoir de résultats finals graphiques.

D'autre part, et c'est fondamental, la localisation redevient un attribut logique, c'est-à-dire non lié à telle ou telle entité : toute relation localisée pourra ainsi être mise en relation avec toute autre relation localisée par une opération de jointure spatiale. Nous verrons comment la représentation de la localisation influence et limite cette opération.

L'intégration de la localisation dans le schéma relationnel est donc prometteur, aussi bien par la puissance de manipulation qu'elle apporte que par les contraintes et les problèmes de représentation et de validité de l'information localisée qu'elle met en évidence. Dans ce cadre, elle va permettre de réfléchir à de nouvelles approches de la représentation géographique.

Les opérations algébriques nouvelles.

Pourquoi de nouvelles opérations ?

Supposons que l'on ait par exemple des relevés pédologiques, des données de couverture végétale définies sur photographies aériennes, et un découpage administratif en départements. Questions : sélectionner les profils se trouvant dans tel ou tel département, établir une liste des départements avec pour chacun d'eux le pourcentage de la surface occupée par la forêt, etc.
Si chaque profil ne contient pas le nom du département auquel il se rapporte, les opérations classiques des SGBD relationnels ne pourront répondre à la première question. Si l'on veut de plus poser la même question par rapport aux types de couverture végétale, il faudrait indiquer pour chaque profil le type de couverture végétale, etc. La réponse à la seconde question sera impossible à réaliser, à moins que les zones de forêts ne constituent un sous-découpage des départements et ne contiennent chacunes le nom du département auquel elles se rattachent : ce découpage en zones de végétation serait quelque peu arbitraire et ne pourrait être conçu que par rapport au découpage administratif, ce qui n'est pas souvent le cas. En règle générale, toute liaison avec une information localisée non prévue au départ sera impossible.

Toutes les opérations que nous avons envisagées portent pourtant sur l'expression de la localisation dans l'espace, et donc intrinsèquement sur le même attribut. Si l'on dispose d'une référence unique pour

l'attribut de localisation, et pour chaque relation de la description de la localisation des objets par rapport à cette référence, tout objet pourra être mis en relation avec n'importe quel autre objet ainsi localisé, indépendamment d'un système de référence attaché à telle ou telle entité.

Le système de référence unique est bien sûr l'espace mathématique euclidien : l'attribut de localisation est alors de dimension deux ou trois, sans relation d'ordre canonique (liée à la métrique définissant la structure de l'espace).
La qualification des opérations de restriction ou de jointure porte sur un attribut de dimension deux ou trois. Elle n'est plus liée à une structure d'ordre mais à la structure métrique de l'espace euclidien.

Quelque soit le type de représentation utilisée pour la localisation (zones, lignes, points, mailles,...), il va falloir formellement passer de nouveau par l'espace euclidien pour mettre les objets en relation sur leur localisation : il s'agit maintenant de repasser de l'espace de représentation à l'espace euclidien, de l'entité dérivée à l'entité initiale, à l'inverse de la démarche exposée en 3.1.1.2.

L'attribut formel de localisation sera noté dans la suite loc, et sa valeur représente l'ensemble des points de l'espace euclidien attaché à un objet. Il est, de par son caractère formel, indépendant du type de représentation utilisée pour la localisation des objets dans la base (zones, lignes, points, cellules,...) et des méthodes effèctives de réalisation des opérations algébriques. Comme nous l'avions déjà remarqué en 3.1.1.2, cet attribut formel, associé s'il y a lieu à un attribut temporel, détermine l'ensemble des attributs descriptifs de la relation : la localisation spatio-temporelle constitue alors une clé d'une relation localisée. Nous l'appelerons clé graphique.

La restriction spatiale.

C'est l'opération correspondant à la sélection des objets par rapport à un domaine défini de l'espace, noté D. Ce domaine peut être défini soit directement (fenêtre d'étude), soit par rapport aux objets d'une autre relation (1) : zone autour d'un point, le long d'une ligne,... Ainsi, la sélection des profils de sols dans tel département correspond à la restriction de la relation profils sur l'attribut loc dans le domaine défini par le département choisi.

Plusieurs types d'opérations de restriction spatiale peuvent être envisagées quand les objets sont des ensembles de points, en prenant une qualification ensembliste plutôt que métrique : si A désigne un objet et D le domaine de restriction :

$A \subset D$	l'objet A est inclus dans le domaine D
$A \cap D \neq \emptyset$	l'objet A intersecte le domaine D
$A \cap D = \emptyset$	restriction par exclusion, A et D sont disjoints
$a \in D$	le centroïde de A appartient à D.

---

(1) comme nous le verrons plus tard , cette opération s'apparente en fait à une semi-jointure, puisque l'ensemble D des valeurs de l'attribut de qualification est connu.

Notons que la restriction spatiale, tout comme la restriction classique, ne modifie pas les objets et n'en crée pas de nouveaux. Il n'y a donc pas, pour cette opération, de problèmes liés au type de représentation utilisé. Seule la réalisation de l'opération de restriction elle-même sera fonction du type de représentation.

La jointure spatiale.

De toutes les opérations algébriques que nous avons vues en 3.1.2.1, la jointure est certes la plus importante en pratique, car, comme nous l'avons souligné, c'est elle qui permet la mise en relation de deux tuples par le biais d'un attribut commun, en créant ainsi un nouvel objet ayant les caractéristiques des deux objets répondant à la qualification de jointure, et correspondant dans la nouvelle relation à un tuple formé des valeurs d'attributs des deux tuples mis en relation.

La jointure spatiale est donc l'opération de jointure portant sur l'attribut de localisation loc. La qualification de jointure s'exprime par rapport à la distance :

a1 <= d(loc1,loc2) <= a2

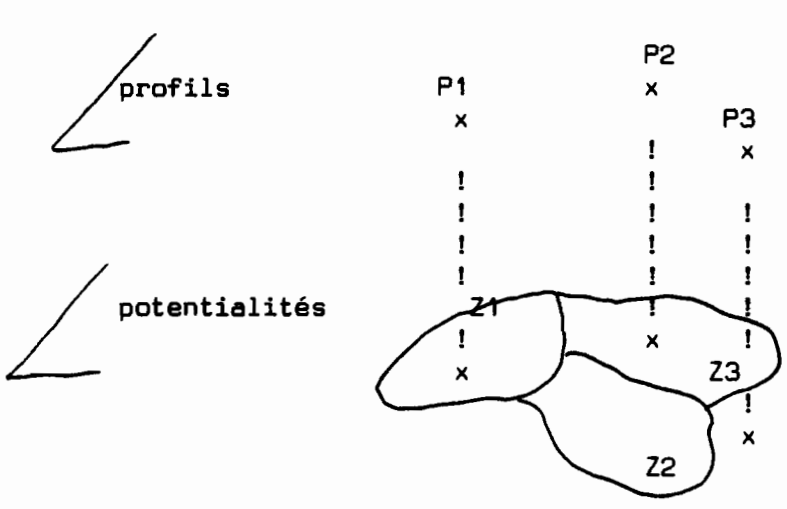
par exemple, a2 = 0 signifie que deux points de l'espace sont mis en relation s'ils ont la même localisation.

Pour bien comprendre cette opération, il est nécessaire de revenir à l'espace de référence et de repasser du type de représentation des objets en zones, lignes, points aux objets élémentaires que sont les points de l'espace euclidien avec leur localisation intrinsèque. L'opération de jointure spatiale s'effectue, en théorie, en considérant que les seuls objets logiques sont les points de l'espace.

Le résultat de l'opération de jointure pourra à nouveau faire l'objet d'un changement de représentation du type passage entité principale entité dérivée. Le cheminement théorique d'une jointure spatiale est donc :

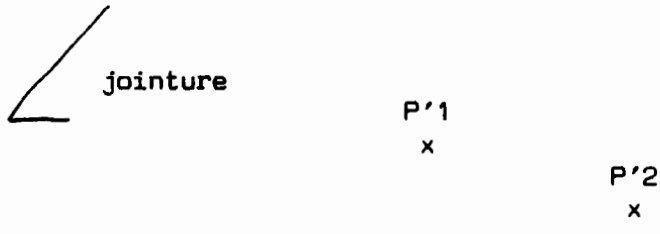
espace de représentation	-->	espace euclidien	-->	jointure sur les points de l'espace euclidien	-->	espace de représentation (si nécessaire)

Ainsi, la jointure spatiale de la relation zonale potentialités (pente, pierrosité, salinité) et de la relation ponctuelle profils (couleur,pH) sur la qualification d(loc1,loc2)=0 aura pour résultat une relation nouvelle ayant pour attributs (pente, pierrosité, salinité, couleur, pH) et dont les objets sont l'ensemble des points appartenant à la fois à la relation potentialités et à la relation profils (fig. 2).
Si la qualification de jointure est d(loc1,loc2) < a, le résultat est plus complexe puisqu'il correspond à un ensemble d'objets pour lesquels la localisation ne constitue plus une clé : un point peut apparaître plusieurs fois avec des descripteurs différents s'il répond à la qualification de jointure pour plusieurs objets de chaque relation. La notion de clé graphique n'est pas conservée par la jointure spatiale, comme c'est d'ailleurs le cas pour la jointure classique.



	couleur	pH
P1	rouge	7.5
P2	jaune	5.8
P3	ocre	6.5

	pente	piérrosité	salinité
Z1	10-20	faible	faible
Z2	20-30	moyenne	trè faible
Z3	10-20	très faible	moyenne



Le résultat de la jointure est composé des points P'1 et P'2 :

	pente	pierrosité	salinité	couleur	pH
P'1	10-20	faible	faible	rouge	7.5
P'2	10-20	très faible	moyenne	jaune	5.8

fig. 2 : jointure entre zones et points sur la qualification d(loc1,loc2)=0

Des exemples de jointures entre relations de type zone sont donnés par les figures 3 et 4. La qualification d(loc1,loc2)=0 correspond à l'intersection des zones et conserve la localisation comme clé de la relation résultante. La qualification a < d(loc1,loc2) < b, avec a non nul, donne un résultat plus complexe et difficile à représenter graphiquement dans son intégralité, puisque la localisation n'est plus clé de la relation résultante.

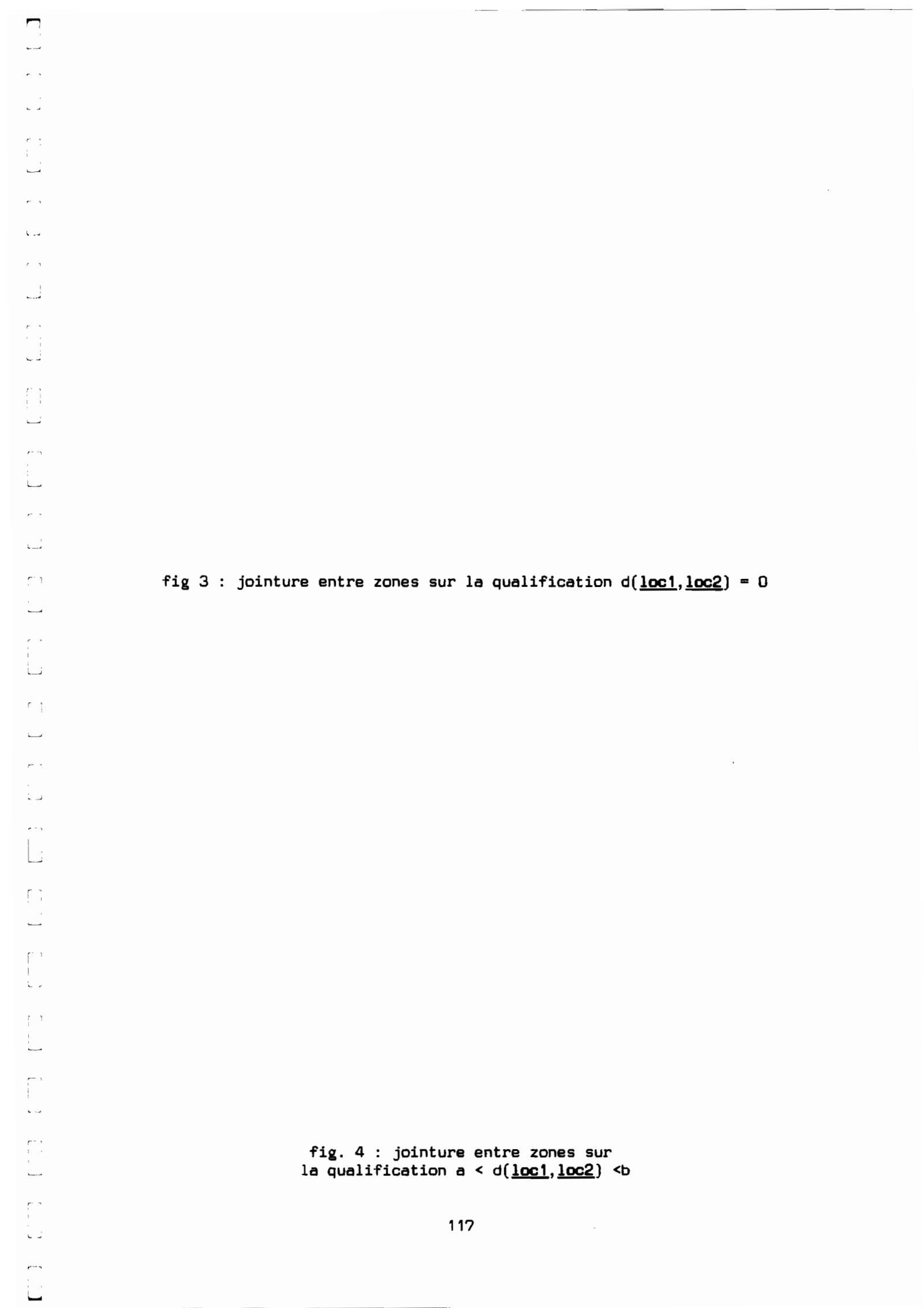


fig 3 : jointure entre zones sur la qualification d(**loc1**,**loc2**) = 0

fig. 4 : jointure entre zones sur
la qualification a < d(**loc1**,**loc2**) <b

La jointure spatiale est une opération intéressante dans la mesure où elle permet de comparer des objets sur leur localisation, sans avoir à spécifier et décrire cette localisation. Le résultat doit d'avantage être considéré au niveau de cette mise en relation qu'au niveau du résultat graphique formé par les tuples de la nouvelle relation : ainsi, on met en relation parcelles (numéro, nom de propriétaire, nombre de constructions) et bornes d'incendie (numéro, débit) sur la qualification d(loc1,loc2) < 100 mètres surtout pour avoir la liste des numéros de parcelles et le nombre de contructions associés à une borne, plutôt qu'une carte des parcelles ou portions de parcelles et bornes résultant de la jointure. Comme nous l'avons déjà souligné, la représentation graphique d'un résultat de jointure spatiale peut être difficile car nous n'avons plus dépendance fonctionnelle des attributs descriptifs par rapport à la localisation : on peut avoir des tuples différents formés du même ensemble de points mais ayant des valeurs descriptives différentes.

A la différence de l'opération de restriction, la jointure spatiale crée de nouveaux objets; la description de la localisation de ces objets pose les mêmes problèmes que la description de la localisation des objets initiaux : les nouveaux objets sont des points ou des ensembles de points de l'espace euclidien, qu'il est nécessaire de regrouper et de décrire d'une manière ensembliste avec les choix possibles que nous avons vu au paragraphe 3.1.1.1. Bien sur, les méthodes de réalisation de la jointure et de description de la localisation des résultats seront fonction du type de description des objets initiaux.

Certaines opérations que nous avons définies comme des opérations de restriction sont en fait des jointures : la restriction d'une relation sur la localisation d'une autre relation est en fait une semi-jointure avec qualification d(loc1,loc2)=0 : on connait à priori l'ensemble des valeurs de l'attribut de qualification, représenté par le domaine D. L'interêt de cette opération réside néanmoins plus souvent dans le résultat de la restriction apportée à l'une des relations que dans la mise en relation effective des attributs des deux relations. C'est pour cette raison que nous l'avons classée comme restriction spatiale :

sélection --> définition d'un --> semi-jointure --> projection de
dans R1 domaine D de R1 par rapport à D la jointure sur
les attributs
de R2

La jointure spatiale est une opération dangereuse si les objets mis en relation n'ont pas la même echelle de validité, ou si la validité de la localisation par rapport aux données descriptives est trop faible pour permettre le passage de l'espace de représentation à l'espace euclidien (pour des données à caractère statistique par exemple, ou n'ayant qu'un intérêt de représentation graphique).

3.1.3 La mise en oeuvre des concepts théoriques.

Le système TIGER, développé par l'unité d'infographie de l'Orstom, est un système d'information géographique organisé suivant le schéma des systèmes de gestion de bases de données relationnelles, et enrichi pour cela d'opérations algébriques nouvelles pour la manipulation relationnelle de la localisation. Il réunit ainsi les possibilités des systèmes de gestion de base de données et celles des systèmes d'information géographique, en mettant en oeuvre les concepts théoriques que nous avons exposés précédemment.

3.1.3.1 Architecture et structures de données.

L'architecture du système TIGER est celle d'un calculateur base de données simplifié : il possède un dictionnaire des données, indiquant les relations et attributs ainsi que leurs caractéristiques (types, définition du schéma interne associé), un dictionnaire des accès permettant de gérer les niveaux externes (par l'allocation de droits et la définition des accès aux données), un langage de commande permettant d'interroger et de manipuler les données. Chaque opération utilise ces trois structures pour accéder au niveau interne et au système de gestion de fichier (fig.5).

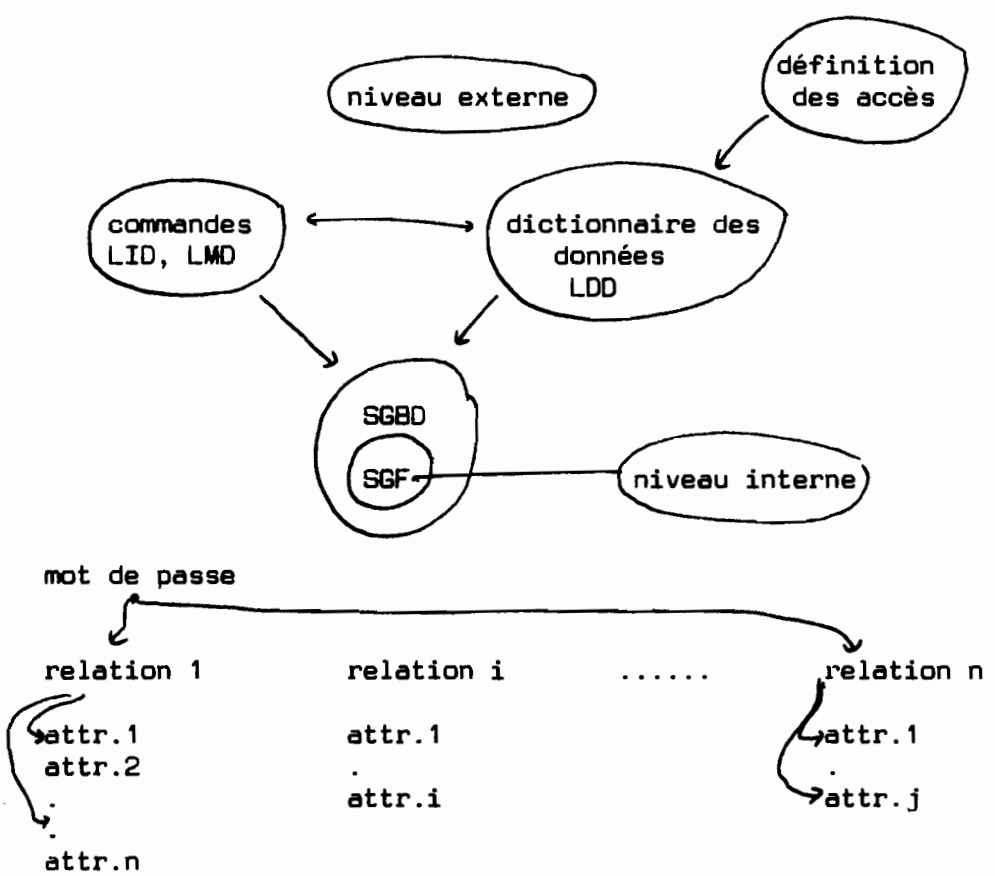


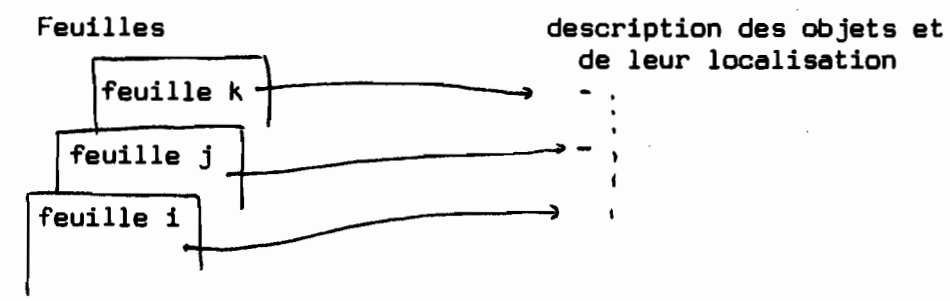
fig. 5 : architecture du système.

Les données sont structurées en relations correspondant aux entités du monde réel. Ces relations sont de plusieurs types : relations non localisées, relations localisées (ponctuelles, réseaux, zonales).

L'information graphique associée aux relations localisées est conservée sous forme ponctuelle ou vectorielle : coordonnées géographiques pour les points, ensemble d'arcs de points pour les réseaux et pour les zones. Tous les points sont décrits par leurs coordonnées géographiques (longitude, colatitude par rapport à un point de référence propre à la base), et donc toutes les localisations sont compatibles pour l'ensemble des objets des relations d'une même base.

Indexation des données dans le système :

Les relations localisées sont indexées sur la localisation par la notion naturelle de feuille, chaque feuille correspondant à une coupure et comprenant au plus 1500 objets. Toute recherche dans une relation localisée passe par recherche des feuilles concernées par le territoire d'étude, chaque feuille contenant la description de sa localisation et donnant l'adresse des objets qu'elle contient, permettant ainsi d'y accéder rapidement. Chaque relation localisée possède son propre ensemble de feuille, puisque cette indexation dépend essentiellement de la densité des objets propre à une relation donnée.



La structure interne des relations zonales :

L'information graphique associée aux relations zonales est constituée de l'ensemble des arcs décrivant les zones. Chaque arc est décrit par les numéros des zones dont il est frontière et par la suite des points qui le constitue.

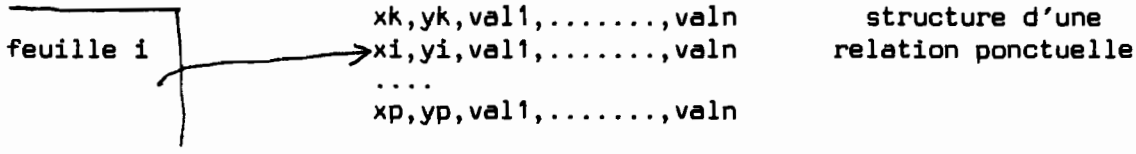
```
nz1    nz2   x1,y1,......., xn,yn
....
nzi    nzj   x1,y1,.....,xp,yp
....
```

L'information descriptive est conservée dans un fichier séparé, contenant pour chaque zone l'ensemble des descripteurs suivant le schéma donné dans le dictionnaire des données. Chaque feuille indique son point d'entrée dans ce fichier descriptif, ainsi que dans le fichier des arcs correspondant.

Structure interne des relations ponctuelles et réseaux :

Cette structure est plus simple que pour les relations zonales car les

fichiers descriptifs et graphiques n'ont pas besoin d'être séparés :



3.1.3.2 La réalisation des opérations spatiales.

La transformation vectoriel-matriciel :

Si les données graphiques sont structurées et stockées dans la base sous forme vectorielle, le système va utiliser pour la réalisation de certaines opérations une structure matricielle sous forme d'une image numérique de pixels associée à un fichier descriptif. Chaque pixel est relié au fichier descriptif par sa valeur numérique qui permet de retrouver le tuple descriptif qui lui correspond.
La taille de la matrice de pixels est paramétrable jusqu'à 1024*1024; la résolution du pixel dépend alors de la fenêtre géographique utilisée. Elle est en général plus faible que l'échelle de validité de la description de la localisation : le pixel peut être alors considéré comme simulant le point de l'espace euclidien. Le passage en matriciel permet donc de mettre en oeuvre concrètement les opérations spatiales, en apportant la facilité de manipulation propre à cette description tout en conservant la structure relationnelle des données. Les éléments vectoriels étant conservées en coordonnées géographiques, le passage en matriciel dans le plan se fait suivant une projection géographique donnée. Cette projection peut être choisie par l'utilisateur.

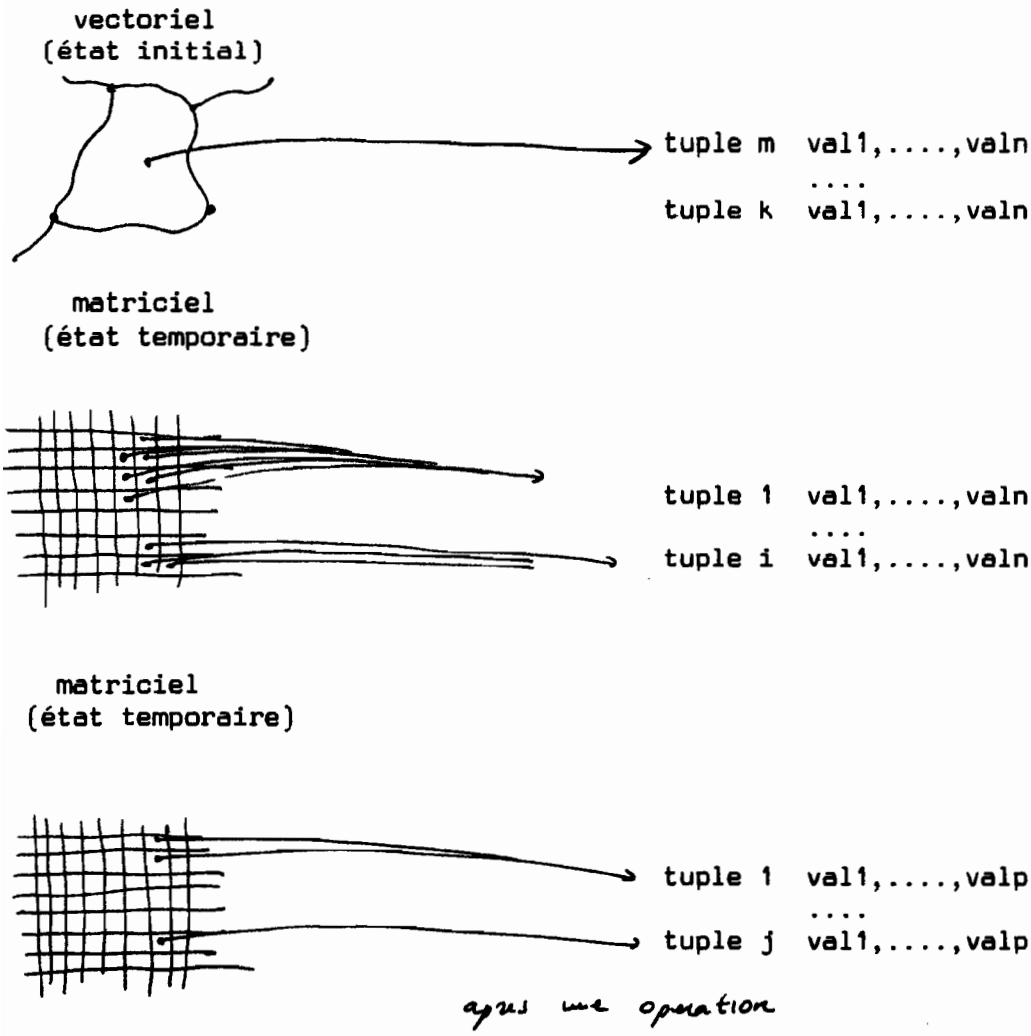
Le passage de vectoriel à matriciel est un des éléments importants du système; néanmoins, il reste au niveau interne et s'effectue d'une manière interactive lors de l'opération qui le requiert. Cette double structure (stockage vectoriel et manipulation matricielle), associée au paramétrage de la résolution matricielle, confère au système une souplesse importante dans la réalisation des différentes opérations de gestion et de manipulation de la localisation.

Ainsi, la description de la localisation va être transformée lors d'une session d'interrogation des données :

- état initial vectoriel ou ponctuel. Cet état n'est pas modifié par les opérations de manipulation des attributs descriptifs (indépendants du mode de représentation de la localisation), ainsi que par les opérations de restrictions spatiales ne s'apparentant pas à une semi-jointure : seuls les fichiers descriptifs sont alors modifiés.

- état temporaire matriciel. Cet état est automatique pour les opérations de projection et de jointure géométrique. Dans cet état, chaque pixel de la matrice est relié au(x) tuple(s) des valeurs d'attributs descriptifs qui lui correspond(ent) : la séparation du graphique et du descriptif est maintenue. Bien sur, on ne considère qu'un tuple de valeurs descriptives pour l'ensemble des pixels qui ont les mêmes valeurs descriptives. L'état matriciel de la localisation d'une relation sera conservé jusqu'à la fin de la session

d'interrogation et l'impression des résultats. Il n'y a pas de retransformation matriciel-vectoriel après chaque phase de l'interrogation.
Une nouvelle interrogation, ou tout changement de la fenêtre d'étude élimine tous les états temporaires matriciels, qui ne sont conservés par l'utilisateur qu'à des fins de stockage ou d'archivage d'un résultat hors du système de gestion.



La réalisation des opérations spatiales :la restriction

La restriction d'une relation par rapport à une fenêtre d'étude se fait simplement en sélectionnant les feuilles intersectant la fenêtre, et en sélectionnant dans chaque feuille les objets se trouvant dans la fenêtre : chaque ligne ou chaque zone possède dans le fichier de ses descriptifs la fenêtre qu'elle occupe, qui permet ainsi une sélection rapide sans avoir à consulter les arcs eux-mêmes.

La fenêtre d'étude peut être définie de plusieurs manières :

- directement en coordonnées géographiques, en indiquant son point bas gauche (point de colatitude maximum et de longitude minimum) et son

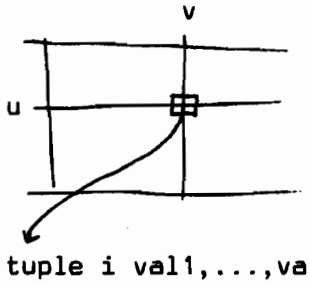
point haut droit (point de colatitude minimum et de longitude maximum).
- en indiquant une feuille donnée d'une relation. La fenêtre correspondra alors au domaine géographique de la feuille.
- en indiquant le point bas gauche et la résolution du pixel en métres dans les images matricielles générées lors du passage vectoriel matriciel.
- en indiquant des valeurs d'attributs d'une relation. La fenêtre sera alors le domaine maximum comprenant l'ensemble des objets ayant ces valeurs.

Notons que le passage en matriciel se fait, quelque soit taille de la fenêtre, à la résolution donnée: ainsi, réduire la fenêtre d'étude augmente la précision des objets dans la matrice, puisque la taille du pixel diminue (le fenestrage ne provoque alors pas seulement un effet de zoom comme c'est le cas pour les systèmes stockant leurs données graphiques en mailles).

La réalisation de la jointure spatiale :

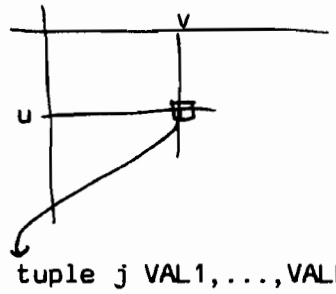
La réalisation de la jointure spatiale consiste à comparer les images matricielles créées à partir de l'information graphique vectorielle et à créer les tuples descriptifs et l'image codée correspondant aux tuples résultant de la jointure. La jointure sur une qualification d(loc1,loc2)=0 est d'une réalisation simple car elle conserve la clé graphique de la relation : la relation image-tuple reste fonctionel; la réalisation de la jointure se déroule alors en plusieurs phases :
1) création d'une nouvelle image par calcul pour chaque pixel d'un code résultant des valeurs du même pixel dans les images des deux relations,
2) création des tuples descriptifs correspondant aux combinaisons rencontrées dans la première phase (ce qui revient à éliminer les tuples en double),
3) établissement de la relation image-tuples.

relation 1



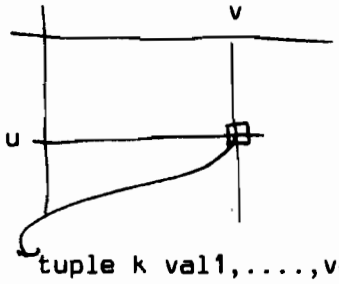
tuple i val1,...,valn

relation 2



tuple j VAL1,...,VALk

relation 3 : jointure



tuple k val1,....,valn,VAL1,....,VALk

3.1.3.3 L'intégration d'images numériques.

Les données localisées extérieures au système et codées sous forme d'images numériques (du type des images satellitaires), que nous appelerons **données images**, s'y intègrent facilement grâce à la phase matricielle de description de la localisation des données de la base. Le seul problème technique pour pouvoir mettre en relation ces images et l'information gérée par le système **TIGER** consiste à faire coincider les paramètres des deux matrices de pixels : projection géographique, résolution du pixel, localisation dans le plan de projection. C'est bien sur au système de gestion de s'adapter aux différentes données images qui lui sont intégrées. Cette adaptation se fait naturellement par les choix de la projection géographique, de la résolution du pixel et de la fenêtre d'étude, choix rendus possibles par le passage temporaire de la structure vectorielle à la structure matricielle.

La mise en relation de données images et de données existantes structurées suivant un schéma conceptuel précis est bien sur la première phase d'un processus visant à extraire, à partir de ces données images, l'information nécessaire à la définition ou la caractérisation d' objets nouveaux associés à une entité et un schéma conceptuel donné. Cette phase ultérieure de création ou de modification de relations est l'objet même de l'intégration de données images dans le système d'information géographique, qui ne doit pas avoir pour objet la gestion d'un ensemble d'images.

Un de nos objectifs est par contre de formaliser la phase d'extraction d'information et d'intégration par rapport à un schéma conceptuel connu : la démarche expérimentale d'extraction d'information à partir de données images pourrait donner lieu à la détermination de règles de décision et permettre l'utilisation de techniques d'intelligence artificielle dans ce dommaine. Nous pourrions alors envisager la construction d'un système expert basé sur de telles règles et possédant comme base de fait aussi bien le résultat d'opérateurs de reconnaissance de formes sur l'image que le résultat de l'interrogation des données existantes dans la base.

3.1.3.4 Description fonctionnelle du système **TIGER**.

Les modules du système **TIGER** seront à l'issue de son développement :

a) un module de saisie graphique par digitaliseur :

La saisie de ces arcs est contrôlée de manière interactive pour éviter diverses erreurs :
-erreurs de manipulations, oublis.
-non fermeture des zones, problèmes topologiques et graphiques (connexité, simplicité, intersections d'arcs).
La saisie se fait sur une fenêtre déterminée par l'opérateur, la juxtaposition des fenêtres et la reprise en cours de manipulation étant gérée par le système.

b) un module de description et de saisie de l'information descriptive;

c) un module d'intégration des données dans la base :
recalage des données localisées en coordonnées géographiques, en

fonction de la projection du document digitalisé, intégration des données descriptives;

d) un module d'interrogation de la base de données :

- choix de la projection géographique de travail et de restitution (MAP) : cylindrique, azymutale, conique.

- restrictions géographiques par fenestrage (WIND) : sélection d'une coupure, sélection directe en coordonnées géographiques, sélection par la taille du pixel, etc.

- recherches thématiques (SELEC) : sélections sur les données descriptives.

- jointures classiques (JOINT) : croisements de données sur des critères algébriques.

- jointures géométriques (GJOINT) : croisements de données sur des critères spatiaux.

- projection et impression de résultats (PROJEC et PRINT).

Les résultats sont donnés sous forme de liste ou d'image, suivant les questions posées et le choix de l'utilisateur. Le dialogue avec le système se fait au moyen d'un langage de manipulation spécifique et de menus associés.

Des modules synthétiques regroupent un certain nombre d'opérations algébriques de base, en facilitant l'interrogation et en optimisant les traitements : THEM (classification, sélection, projection, création de l'image et impression des résultats, croisement de deux images thématiques, reclassification et agrégation de classes).

d) un module de cartographie automatique (CART) :
sélection de contours, de symboles, adjonctions de textes, etc. Ce module crée des segments graphiques indépendants des organes de sortie (écran ou traceur) et qui seront interprétés par les modules de visualisation propres à ces organes.

e) des modules de visualisation (GIXI, BENSON) :
La visualisation des images se fait de manière interactive sur écran graphique ou sur traceur. On peut ainsi déterminer l'échelle de tracé (pour les traceurs), de nouvelles agrégations (redéfinition des classes thématiques), choisir les couleurs, les symboles, les caractères ...

f) un module de calculs statistiques (STAT) :
statistiques univariées (moments, histogrammes), ou bivariées (corélations, régressions). Ces calculs peuvent se faire à tous moments de l'interrogation et permettent ainsi de modifier si besoin est le cours de l'interrogation (pour le choix des seuils de classifications per exemple).

g) un module de création d'attributs :
- par classification (CLAS) : passage de numérique à nominal, ou de nominal à nominal par regroupement de valeurs.

- par calcul numérique (sur les attributs numériques) ou par calcul logique (sur les attributs nominaux) (CRIS).
- par jointure et statistique (GSTAT) : de nouveaux attributs sont créés par calcul statistique sur le résultat d'une jointure classique ou géométrique : il est ainsi possible d'affecter à une zone la moyenne d'un attribut de points se trouvant à une distance donnée de la zone, ou le pourcentage de sa surface occupée par la valeur d'un attribut propre à un autre découpage, etc.

h) un module d'intégration de données satellitaires :
Ce module assure la mise en relation d'une image satellite recalée et de l'information de la base correspondante, ainsi que des comparaisons et des calculs graphiques entre ces deux informations (calculs statistiques sur les valeurs radiométriques par rapport aux classes thématiques, etc.).

3.1.3.3 Description matérielle.

L'ordinateur utilisé actuellement est un MINI 6-Bull; le système utilise trois terminaux, un disque dur de 80 Mo, une imprimante, un lecteur de bande magnétique. Le matériel graphique est composé de :

- une table à digitaliser BENSON 6301;
- un écran graphique couleur GIXI (Radiance 320): 256 couleurs affichables sur une palette de seize millions;
- une table traçante BENSON 1625, 8 plumes.

Bibliographie.

Abel D.J., 1984 — A B+-tree structure for Quadtrees, Computer Vision, Graphics, and Image Processing, vol. 27, no 1,p. 19-31

Amidon E.L. and Atkin G.S., 1971 — Algorithmic selection of the best method for compressing map data string : Communications, ACM, vol. 14, no 12, p. 769-774.

Capson D.W., 1984 — An improved algorithm for the sequential extraction of boundaries from a raster scan, Computer Vision, Graphics, and Image Processing, vol. 28, no 1, p.109-125

Dangermond J., 1981 — Some triends in the evolution of GIS technologie, Marble Ed., Kensington Workshop, p.25-57

Dangermond J., 1983 — Classification des éléments des logiciels utilisés habituellement dans les systèmes d'information géographique, fascicule no 96 du Comité Français de Cartographie, p. 7-20

DIME, 1970 — Technical description of the DIME System, U.S. Bureau of Census : Census use study, the DIME Geocoding System, Report no 4, Washington D.C., p. 25-30

Frank A., 1981 — Application of DBMS to Land information systems, CH 1701-2, IEEE 81, p.448-453

Gardarin G., 1983 — Bases de données : les systèmes et leurs langages, Ed. Eyrolles, Paris.

Lee D.T. and Wong S.K., 1980 — Quintary trees : a file structure for multi-dimensional Database systems, ACM Transactions on DBMS, vol. 5, no 3, p. 339

Matsuyama T., Le Viet H., and Nagao M., 1984 — A file organisation for Géographic Information Systems based on Spatial Proximity, Computer Vision, Graphics, and Image Processing, vol. 26, no 3, p. 303-318

Rosenfeld A. and Samet H., 1980 — Tree structure for region representation, Aangeenbrug, Autocarto 4, vol. 1, p.108-118

Samet H., 1980 — Region representation : quadtrees from boundary codes, Communication ACM, vol. 23, no 3, p. 163-170

Samet H, 1984 — The quadtree and related hierarchical data structures, ACM Computing surveys, 16 (2), juin 1984

Samet H. and Rosenfeld A., 1980 — Quadtree representation of binary images, 5th. International Conf. on Pattern Recognition, Miami, p. 815-818

Simon J.C., 1984 — La reconnaissance des formes par algorithmes, Masson ed., Paris.

3.2. INTEGRATION DONNEES DE TELEDETECTION/DONNEES EXTERIEURES

3.2.1. LES DONNEES DE DEPART.

Avant de voir comment s'effectue l'intégration Données de télédétection/ Base de données localisées, nous avons à regarder de plus près quels sont les types de données que nous devons relier.

La totalité des informations concernant le zonage morphologique sur la mosaïque photo aérienne (limite de zone et les 10 attributs) demande 350 000 octets. L'image de la même zone de Marseille en données Thematic Mapper nécessite 1 500 000 octets. La même surface couverte par SPOT demandera 2 500 000 octets. Cette différence provient de deux conceptions différentes de la spatialisation.

Les données des bases que gère le système TIGER repose sur une saisie et un stockage vectoriel, c'est-à-dire que l'on ne considère que le périmètre d'une zone. A cette zone sont attribués un numéro et les différentes variables associées. La donnée est considérée sous la forme spatiale sous laquelle elle est recueillie. On connaît à priori l'homogénéité de la répartition spatiale de la donnée (rappelons que dans le cas du zonage morphologique la zone est définie par son homogénéité vis à vis des différents descripteurs).

Les données de télédétection sont, par nature, matricielles. La surface du sol est découpée par un carroyage régulier, la taille de chaque carré (pixel) est une caractéristique du capteur (30 m. de côté pour Thematic Mapper). La localisation et le périmètre de la zone sont implicites. Certes il est toujours possible de diviser un carré en deux, quatre, six, ...etc. Mais cette opération n'est parfaitement légitime que si la répartition de la radiométrie est homogène sur toute la surface. En urbain, on est à peu près certain que cette condition n'est jamais remplie.

3.2.2. LES ATELIERS ET LES SYSTEMES.

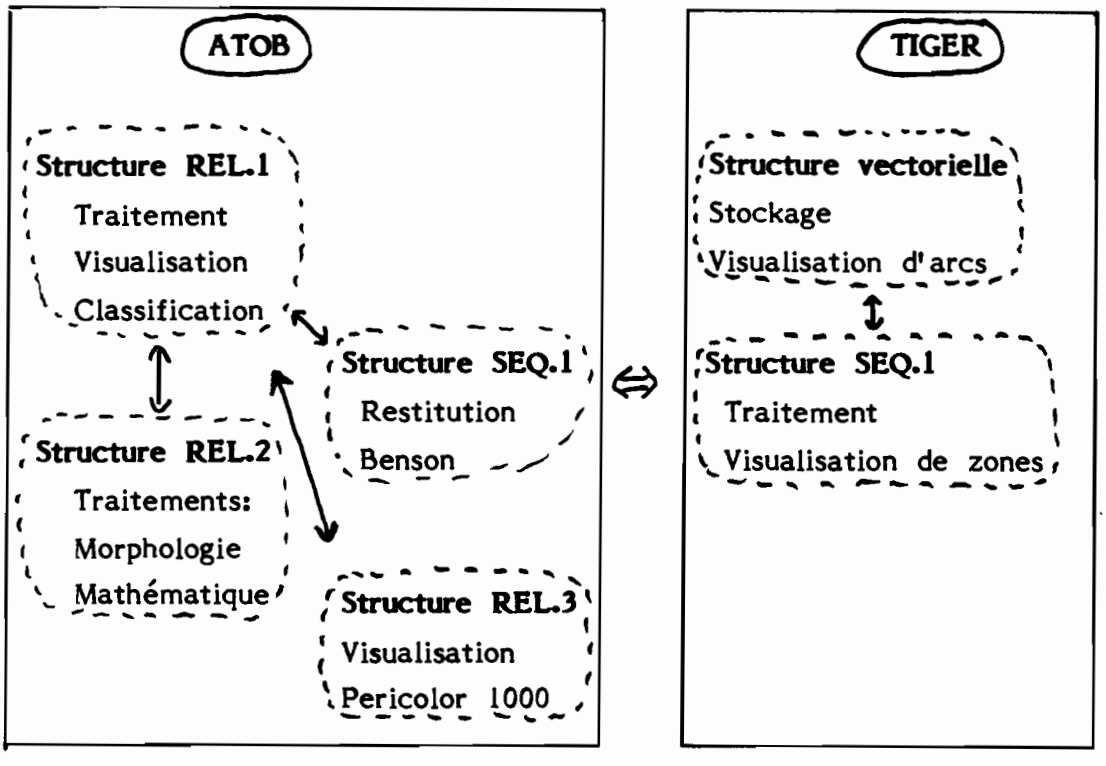
Pour gérer ces deux types de données, deux ateliers différents ont été installés à l'Orstom pour développer deux systèmes.

L'Atelier de Télédétection de Bondy a commencé l'utilisation des données satellite en 1977. Un ensemble de systèmes de visualisation, classification, restitution s'est développé. Utilisant au départ des classifications supervisées interactives, il a intégré ensuite des algorithmes d'analyses ACP et des classifications par discrimination bayesienne, puis des développements ont été effectués en direction de la morphologie mathématique. La vocation de

l'Atelier à traiter des données de satellites a conduit naturellement à conserver la structure matricielle pour le stockage et le traitement.
Le système TIGER s'est développé à l'atelier d'infographie à partir de 1982. Conçu dès le départ pour gérer des données de provenance très diverses il adopte d'emblée un système et une structure moins rigides.

Des structures de fichiers différentes :

Si le système TIGER se présente comme un logiciel intégré, l'ATOB a plutôt développé un nombre de programmes, maintenant important (une cinquantaine), s'articulant autour d'une structure matricielle de fichiers définis dès le départ.



Fichier REL.1 : les données sont groupées par ligne de pixels correspondant à une ligne de balayage du scanneur du satellite. Les canaux, au nombre de six, sont entrelacés (une ligne canal 1, une ligne canal 2 etc..).
Fichier REL.2 : les données sont groupées par ligne de pixels. Un fichier comprend 12 sous-fichiers contenant chacun un canal.
Fichier REL.3 : Matrice de 256x256. Chaque point est adressé séparément.
Fichier SEQ.1 : la matrice est décrite linéairement par trois variables (un code, l'adresse de début, l'adresse de fin du code).

Le schéma montre que le passage des données de télédétection dans le système TIGER est réalisé, pour ce qui est de la structure des données, de façon naturelle. En effet chaque système comporte évidemment des passages entre les différents types de fichiers internes au système (REL.1, REL.2, REL.3, SEQ.1 pour l'ATOB; SEQ.1, Vectorielle pour TIGER). L'adaptation a consisté simplement à rendre identiques les fichiers SEQ.1.

Cette identité des fichiers, obtenue de façon concertée depuis 1983, permet d'obtenir ainsi deux systèmes communicants sans programme véritablement spécifique à ce passage.

Cette communication permet d'intégrer dans les bases gérée par le système TIGER des données satellite, au départ matricielles jusque sous une forme vectorielle si on le désire (de préférence alors sous une forme simplifiée). Elle permet aussi d'amener dans les fichiers gérés par le système ATOB sous forme de pseudo-canaux, toute donnée en provenance du système TIGER (après transformation sous une forme matricielle).

Cette procédure va nous permettre d'utiliser l'un ou l'autre des systèmes selon les disponibilités des machines ou la spécificité des traitements à réaliser.

3.2.3. LA CONCORDANCE GEOMETRIQUE DES DEUX TYPES DE DONNEES.

La "jonction" des données s'effectue dans une structure matricielle par les système TIGER. Les données, stockées selon leurs coordonnées géographiques sont projetées sur une matrice de taille définie, à une échelle définie, selon une projection "géographique" définie, d'après les coordonnées, dans la projection utilisée, du point bas/gauche de la fenêtre que l'on veut extraire. C'est ce système de gestion qui va permettre la mise en relation géographique avec une matrice de données de télédétection dont on connait les paramètres spatiaux : taille du pixel, projection utilisée etc...

Les données de Thematic Mapper, corrigés géographiquement par l'IGN selon un modèle linéaire avec une dizaine d'amers correspondent à une cartographie Lambert III. Le fichier résultant se présente selon une matrice de 1000x1000. Cela constitue un fichier trop important pour les moyens actuels de l'ATOB et nous avons extrait un fichier de 359x591. Le traitement efectué par l'IGN nous donne les coordonnées Lambert III d'un point de l'image desquelles on déduit facilement celles du point bas gauche de l'image extraite.

Par exemple, pour les traitements réalisés jusqu'à maintenant : taille de matrice 740x740, échelle définie par un pixel de 15 m., projection Lambert III, coordonnée X : 842750, coordonnée Y : 105050. Nous avons quadruplé le pixel Thematic Mapper pour minimiser la taille du tracé des contours. En effet si les limites des contours n'ont pas d'épaisseur dans la base de donnée, leur représentation sur écran Radiance ou Péricolor nécessite un pixel au minimum. La taille du pixel de travail ou de visualisation est choisi en fonction du rapport : nombre de pixels compris dans la zone/ nombre de pixels a cheval sur la limite. Lorsque la projection utilisée est identique il suffit d'un point commun aux deux fichiers pour que ceux-ci soient en concordance. Comme en témoigne la diapositive A, la jonction réalisée de cette façon, présente un résultat satisfaisant. La précision du repérage dans toute la partie sud de la fenêtre semble inférieure au quart de pixel. Par contre la partie nord semble présenter un décalage d'un ou deux pixels. Cependant le manque de points de repère indubitables ne nous permet pas d'apporter une réponse définitive sur la seule observation visuelle. Le "mauvais" résultat peut s'expliquer de deux façons différentes :

-nous avons rapporté les coordonnées des pixels à leur coin nord-ouest. Si, lors du recalage géographique de l'IGN les coordonnées d'un pixel concernent le centre de ce pixel l'ensemble est translaté d'un demi pixel et "l'erreur" minimum de calage est diminuée.

-l'image Thematic Mapper semble présenter une rotation d'un ou deux milliradians. Ceci est perceptible le long de la côte où les amers sont fiables. Les diapositives 3 à 8 montre que sur tous les canaux le contraste eau/minéral est suffisant. Plus vers l'est, vers l'intérieur de la ville, les points de repère ne sont plus aussi certains et nous ne pouvons affirmer qu'il y a rotation et non pas un simple décalage local.

Le recalage de l'IGN consiste à opérer une rotation (et une translation) avec une méthode du plus proche voisin. Le pixel de la nouvelle matrice n'est pas une moyenne des pixels homologues de l'ancienne matrice mais prend la valeur du pixel le plus proche. La nouvelle matrice est orientée nord-sud. L'erreur résiduelle est théoriquement d'un pixel au maximum.

Une premiere rectification a été faite à l'ATOB. La diapositive n° 35 est une composition colorée des canaux 1,4,7 d'une partie du port de Marseille elle montre clairement le décalage existant entre les canaux 1 et 4, en bleuté et le canal 7 en rouge. Cette translation mise en évidence sur cette portion d'image est générale et apparemment constante. Nous avons donc simplement décalé la matrice d'une colonne pour les canaux 7 et 5.

Par la suite, deux types d'opérations seront faites:

-rotation de l'image de quelque milliradians par le programme Atob de RECTIFication.

-modification des coordonnées du point bas/gauche lors de la jointure dans le système TIGER.

Après chaque opération nous vérifierons la rectitude géométrique par simple visualisation, mais aussi par mesure de la variance dans quelques zones test. L'amélioration du calage amenant une diminution de la variance de chaque zone.

3.2.4. UTILISATION DES DONNEES DE LA BASE LORS DE L'ANALYSE DES DONNEES DE TELEDETECTION.

Nous disposons donc d'un système qui permet l'intégration des données de télédétection dans le système de gestion d'information géographique. Il est clair cependant que les données de télédétection sont difficilement utilisable en leur état initial. Pratiquement cela ne présente que peu d'intérêt d'encombrer la base de données que l'on ne sait pas exploiter. Ne seront introduit que certains résultats de traitement qui pourront présenter un intérêt informatif. Pour rechercher quels sont ces traitements l'apport de données, extérieures à la télédétection, nous parait indispensable et c'est grâce à l'apport de données localisées que pouront se faire les investigations mais aussi les validations des différent traitements. Dans toute cette phase du travail, le courant d'information viendra du système TIGER vers le système ATOB plutôt que des données télédétection vers la base de données localisées.

Ces données exogènes bien localisées vont permettre de développer de nouvelles méthodes d'analyse. Classiquement l'analyse des données de télédérection s'effectue soit sur toute l'image, soit sur un échantillon de l'image par tirage systèmatique, soit sur des portions d'image constituant des zones d'entrainement dont les limites sont sans rapport avec le contenu de la zone elle-même (cf. Chap. 5). Le fait de caractériser un pixel par son appartenance à une zone géographique nous ouvre une nouvelle voie d'investigation.

3.2.4.1. l'analyse par variable ou attribut :

TIGER permet de raccorder les six canaux Thematic Mapper et les différentes variables provenant de l'analyse visuelle des photos aériennes. La diapositive n° 14 montre les six modalités de l'attribut "densité du bâti". Ce fichier est obtenu par une "requête" dans le système TIGER et simplement considéré dans le système ATOB comme un "canal" supplémentaire.

RECAPITULATIF DE LA REQUETE

Nom de la base : MARSEILLE
Point bas : Longitude 0 deg 9 min - Latitude 0 deg 2 min
Point haut : Longitude 0 deg 18 min - Latitude 0 deg 8 mi

DEFINITION DE LA REQUETE

Relation	Attribut	Classe	Valeurs
MORPHOLOGIE	Densite	non bati	
		tres faible	
		faible	
		moyenne	
		forte	
		tres forte	

IMAGE PAR RELATION

MORPHOLOGIE Superficie totale : 283475

Code	Pixel	des carre	%	Densite
1	95810	5988.1250	0.34	non bati
2	32027	2001.6882	0.11	tres faible
3	50110	3131.8765	0.18	faible
4	74456	4653.5000	0.26	moyenne
5	21423	1338.9380	0.08	forte
6	9649	603.0627	0.03	tres forte

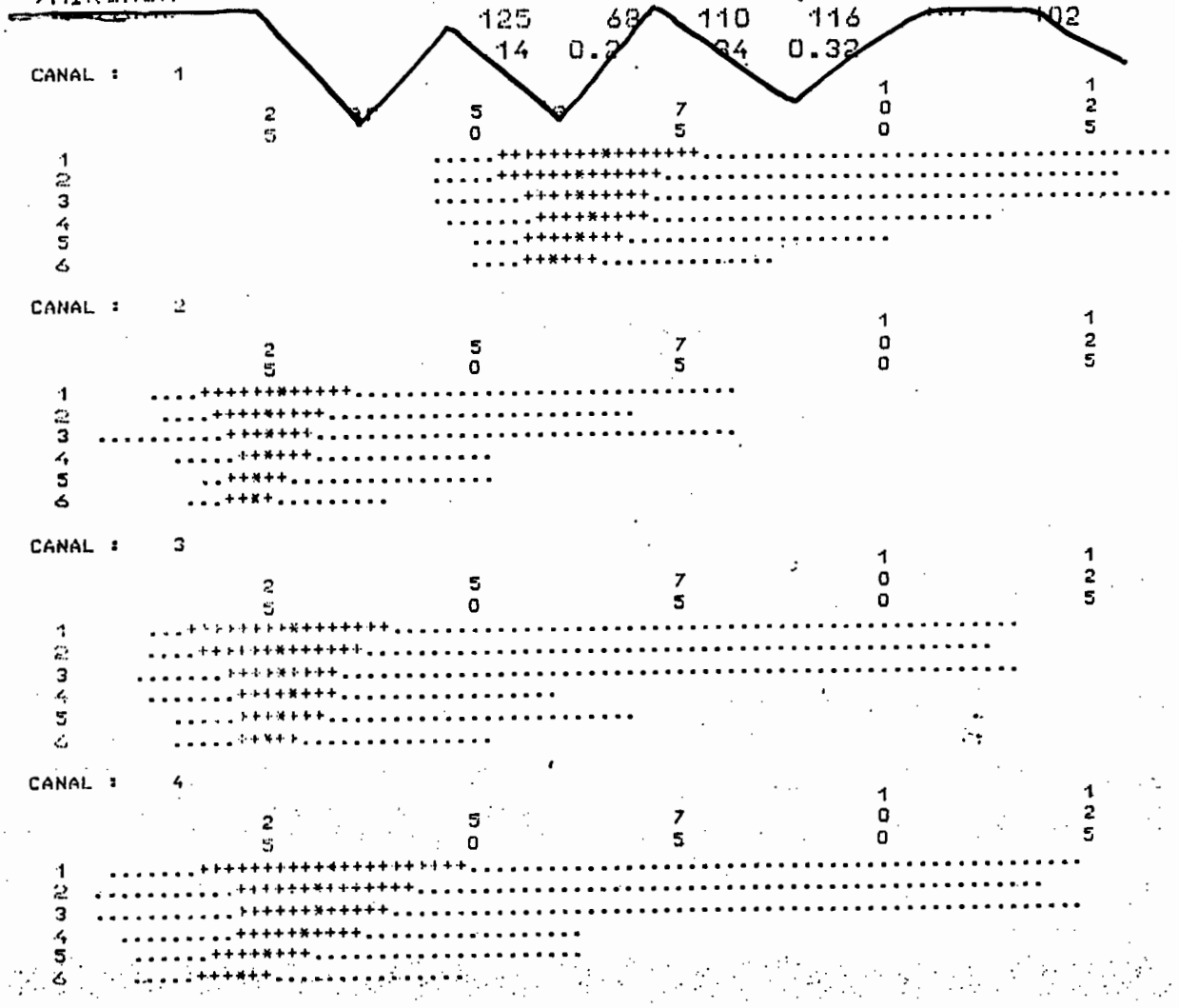
Le programme STADES (dans le système ATOB) nous donne ci-dessous des renseignements statistiques pour chaque modalité : nombre de pixels, histogramme, minimum, maximun, moyenne, écart-type. La zone est ici prise dans le sens "ensemble des pixels appartenant à la modalité n de l'attribut Densité".

,Canal dans image de reference 3
Zone numero : 1

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	95810	95810	95810	95810	95810	95810
,MOYENNE	62.86	25.08	26.89	31.59	42.80	22.17
,ECART TYPE	11.89	7.73	11.71	15.02	28.04	15.00
,MINIMUM	44	12	11	7	4	0
,MAXIMUM	143	79	113	120	210	106
,CDV	0.19	0.31	0.44	0.48	0.66	0.68

,Canal dans image de reference 3
Zone numero : 2

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	32027	32027	32027	32027	32027	32027
,MOYENNE	60.51	24.05	25.66	30.63	37.11	20.53
,ECART TYPE	8.57	5.61	8.63	9.85	15.97	9.98
,MINIMUM	44	13	11	5	4	0
	125	68	110	116	[illegible]	102
	14	0.2	84	0.32		



En croisant les six canaux Thematic mapper et l'attribut : "numéro de zones", le même programme nous donne des renseignements sur chacunes des 500 zones issues, par une requète du système TIGER, de l'analyse visuelle des photos aériennes.

,Canal dans image de reference 4
Zone numero : 151

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	841	841	841	841	841	841
,MOYENNE	59.23	23.56	25.14	33.36	37.66	20.13
,ECART TYPE	4.47	2.89	4.33	5.71	8.06	5.14
,MINIMUM	49	17	15	18	14	5
,MAXIMUM	77	36	42	51	62	36
,CDV	0.08	0.12	0.17	0.17	0.21	0.26

,Canal dans image de reference 4
Zone numero : 152

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	347	347	347	347	347	347
,MOYENNE	63.55	25.50	27.25	34.20	40.86	21.63
,ECART TYPE	7.83	4.34	5.82	8.40	10.78	6.41
,MINIMUM	52	18	15	13	17	9
,MAXIMUM	110	49	54	62	84	48
,CDV	0.12	0.17	0.21	0.25	0.26	0.30

,Canal dans image de reference 4
Zone numero : 153

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	3019	3019	3019	3019	3019	3019
,MOYENNE	56.92	21.78	22.41	30.30	33.70	17.29
,ECART TYPE	6.35	4.19	6.34	8.74	13.29	7.45
,MINIMUM	45	16	13	11	9	4
,MAXIMUM	89	44	56	63	105	55
,CDV	0.11	0.19	0.28	0.29	0.39	0.43

,Canal dans image de reference 4
Zone numero : 154

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	204	204	204	204	204	204
,MOYENNE	61.46	25.81	28.96	34.51	43.67	24.58
,ECART TYPE	6.76	6.19	10.17	10.03	17.25	12.18
,MINIMUM	50	18	16	16	13	8
,MAXIMUM	77	41	59	58	88	59
,CDV	0.11	0.24	0.35	0.29	0.40	0.50

,Canal dans image de reference 4
Zone numero : 155

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	2142	2142	2142	2142	2142	2142
,MOYENNE	58.35	22.67	24.27	27.96	33.94	19.50
,ECART TYPE	4.11	2.62	3.95	5.54	7.37	4.72
,MINIMUM	49	17	14	14	11	8
,MAXIMUM	79	37	40	46	61	39
,CDV	0.07	0.12	0.16	0.20	0.22	0.24

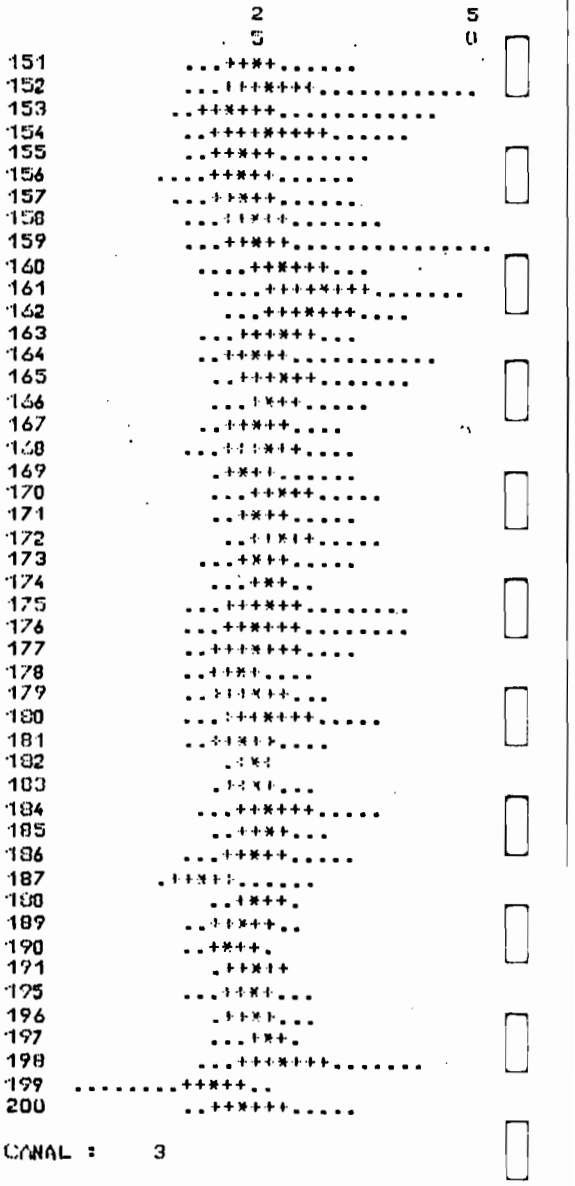
,Canal dans image de reference 4
Zone numero : 156

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
,NOMBRE DE PIXELS	3792	3792	3792	3792	3792	3792
,MOYENNE	50.43	22.95	24.99	27.24	34.98	20.67
,ECART TYPE	4.41	3.08	4.88	7.19	10.53	6.74
,MINIMUM	48	15	13	9	7	3
,MAXIMUM	76	35	41	50	77	42
,CDV	0.08	0.13	0.20	0.26	0.30	0.33

,Canal dans image de reference 4
Zone numero : 157

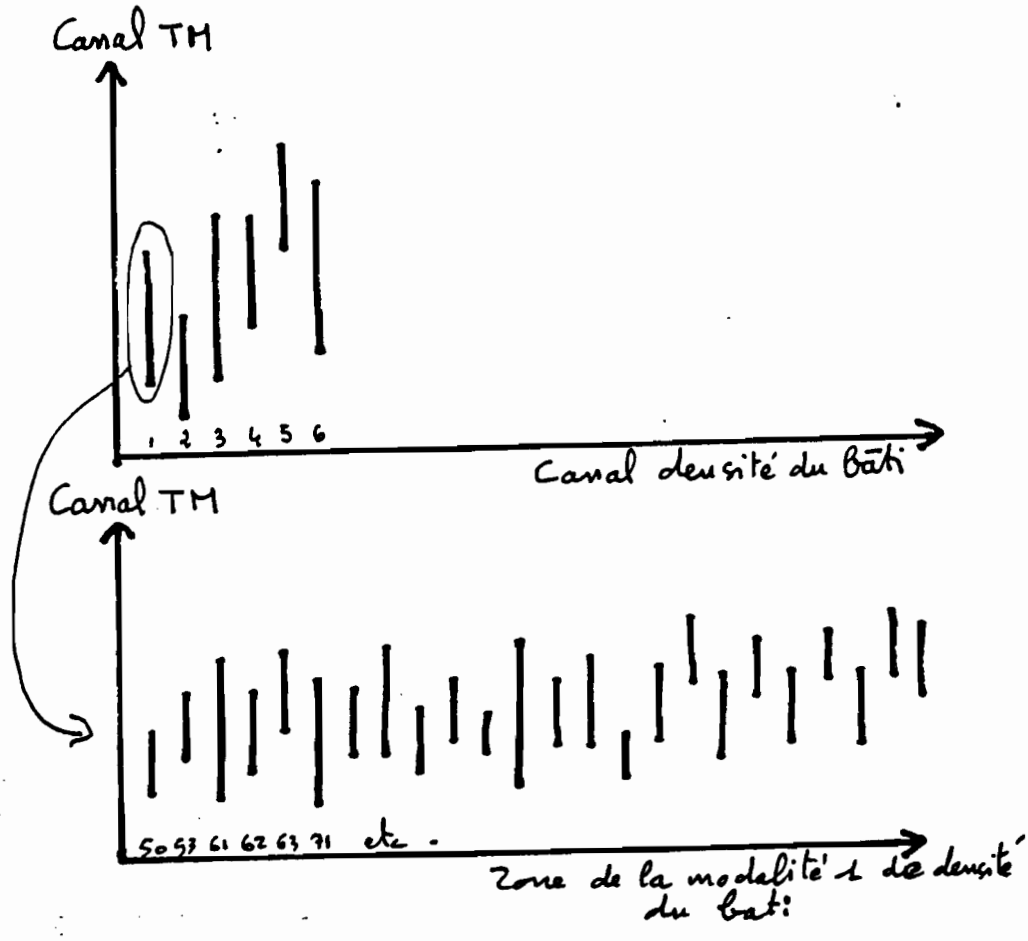
	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7

CANAL : 2



CANAL : 3

En croisant l'attribut "densité du bâti", l'attribut "numéro de zone" et les six canaux du Thematic mapper nous avons une description statistique détaillée des différentes zones composant la modalité n de l'attribut "densité du bâti".



Nous avons ainsi un outils d'analyse précis qui nous permettra, car cette étape n'a commencé que très récemment, de faire un lien précis entre la radiométrie et l'état du terrain. Cette précision permettra d'expliquer pourquoi certaines zones font croître la variance d'un canal radiométrique et, par là, de dégager des corrélations relativement indépendantes de l'image de Marseille. Cette analyse va nous permettre de rechercher les diverses combinaisons linéaires de canaux TM qui offriront des taux de corrélation important avec différents attributs de la base qui décrivent bien la morphologie urbaine.

3.2.4.2. L'analyse des textures.

On sait (cf chap. 5) que la voie radiométrique ne sera pas suffisante pour

obtenir une classification en types d'occupation du sol et en types de bâti à partir de l'imagerie spatiale. Nous avons à développer "l'analyse de structure" des pixels par les voies des filtres ou de la morphologie mathématique. L'apport de la base de données exogènes, par exemple les zones morphologiquement homogènes sur photo aérienne, les différents ilôts, nous permettra de suivre une voie zonale. Le pixel n'est plus seulement considéré par rapport à ses 8 ou 24 voisins mais comme appartenant à une zone définie. La notion d'agencement peut se poser en terme mieux définis et mesurables.

3.2.4.3. Les méthodes d'analyses et de vérifications sur échantillons.

L'analyse par zones de toute l'image se fera, pour des raisons simplement matérielle dans le système ATOB. La coexistance des deux systèmes, TIGER et ATOB, dans un même atelier dépasse pour le moment les capacités de chacun des deux ateliers. La méthode d'échantillonnage d'ilôts décrite en 2.1.5. et chapitre 4 réduit considérablement le nombre de données à traiter. Lorsque nous tirerons 80 ilôts sur les 3500 qui composent le secteur étudié, le nombre de pixels de l'image TM sera ramené à 4000. La plupart des traitements d'analyse pourra se faire directement dans le système TIGER.

Ainsi, si le système de gestion de base de données TIGER nous apparait comme pouvant, à terme, réaliser l'ensemble des projets d'utilisation de la télédétection en démographie urbaine : l'analyse radiométrique et structurale, les différentes classifications, l'utilisation des informations issues de la télédétection. L'ensemble des deux systèmes garde suffisament de souplesse pour surmonter les difficultés inérantes aux systèmes informatiques en développement : le manque de place mémoire, la relative lenteur des traitements qui limite le nombre d'utilisateurs, le manque de moyens de programmation. D'ores et déja les outils indispensables à l'avancement du projet "utilisation de la télédétection pour le suivi de la morphologie et la démographie" sont en place, la mise en place définitive des différents modules de traitement et de capacités de traitement supplémentaires se fera progressivement.

4 . RELATIONS BATI/POPULATION ET SONDAGES SPATIAUX EN MILIEU URBAIN

Ce chapitre reprend l'essentiel du projet de thèse d'O. BARBARY, allocataire MRT en accueil à l'ORSTOM, sur le sujet de thèse suivant :

Elaboration de plans de sondage pour l'estimation de données spatio-temporelles, en prenant en compte les informations fournies par la télédétection . Application aux données démographiques en milieu urbain des pays en développement.

4. 1 . INVENTAIRE EXPLORATOIRE DES METHODES MATHEMATIQUES ET STATISTIQUES APPLICABLES

Le champ des méthodes mathématiques potentiellement applicables à la problématique définie ci-dessus apparait très vaste. Elles relèveront en tous cas de trois domaines :
1) L'analyse des données, comprise ici dans l'acception la plus large du terme, non restreinte à l'une ou l'autre des écoles.
2) Les techniques de régression, ajustement et modélisation mathématique.
3) La théorie des sondages et plus spécifiquement des sondages aréolaires.
Un premier inventaire non exhaustif permet de citer les principales méthodes dans chacun des domaines.

L'analyse des données

_ L'analyse exploratoire des grands tableaux de données au sens défini par J.P. BENZECRI : analyse factorielle des correspondances appliquée aux tableaux de contingence ou aux tableaux mis sous forme disjonctive complète (correspondance multiple), classification ascendante hiérarchique et l'ensemble des méthodes d'aide à l'interprétation dont les programmes sont regroupés par l'ADDAD.

_ Les méthodes de classification non hiérarchique comme les nuées dynamiques (E. DIDAY) ou les méthodes anglo-saxonnes de cluster analysis.

_ L'analyse de similitude (C. FLAMANT) et plus généralement les méthodes combinatoires d'analyse de données : classification, analyse booléenne, méthodes de sériation.

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

Les techniques de régression et de modélisation mathématique.

_ Méthodes de régressions linéaires ou polynomiales.
_Analyse de la variance.
_Discrimination logique.
_Analyse discriminante.

La théorie des sondages
- Sondages aréolaires stratifiés ou en grappes.
- Estimation bayésienne.
- Méthode d'estimation des variables régionalisées et krigeage.

Un des objectifs du travail sera donc de sélectionner à chaque étape les méthodes susceptibles d'être appliquées compte tenu du volume et de la nature des données, et de comparer entre eux les résultats qu'elles fournissent.

4 . 2 . OBJECTIFS GENERAUX . MARSEILLE. JUIN 1985-JUILLET 1986 : PHASE DE MISE AU POINT DE LA METHODE

Pour le projet de recherche Télédétection et démographie, l'ORSTOM a choisi deux sites de travail qui correspondent en fait à deux phases du projet : la phase de mise au point de la méthode portera sur Marseille, la phase de première application sur une ville de pays en développement, Quito (Equateur)(voir deuxième partie du rapport).

Pour mettre au point et valider une nouvelle méthode de collecte de données démographiques fondée sur l'estimation par sondage, on se propose d'utiliser l'information sur la morphologie urbaine fournie par les images satellite. Dans la phase de mise au point sur Marseille, l'objectif principal sera d'estimer les effectifs de population.
Nous recherchons dans cette expérience la réponse à deux questions:

- 1 : Quelle est l'information morphologique reliée à la population et comment l'intégrer dans la définition du plan de sondage pour optimiser sa précision ?

-2 : Cette information morphologique est elle contenue dans l'image satellite ? si c'est le cas, quelle chaine de traitement permet de la mettre en évidence?

Seule la première question fait l'objet du travail dans le cadre de ma thèse, la deuxième ayant trait à l' interprétation de l'image satellite. On peut fixer quatre objectifs principaux à ce travail d'étude des relations statistiques entre morphologie urbaine et démographie.
Dans ce qui suit on tente, pour chacun des quatre objectifs, de faire l'inventaire des problèmes qu'ils soulèvent pour pouvoir proposer des méthodes mathématiques susceptibles de les résoudre.

- Premier objectif : *Définition d'une typologie et d'une stratification de l' espace urbain sur des critères morphologiques*

1 : Recenser les caractères morphologiques observables soit sur les photographies

aériennes soit au sol et proposer pour chacun une mesure ou un codage numérique.

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

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

- Deuxième objectif : *Mesurer le gain de précision qu'apporte une stratification morphologique dans l'estimation par sondage de données démographiques.*

En testant un certain nombre de méthodes classiques de sondages aréolaires, sur listes ou spatiaux, stratifiés ou non, il faudra comparer les variances d'estimation et déterminer quelle est l'expression numérique de la répartition de la population, densités de population ou effectifs, dont la précision d'estimation bénéficie le plus d'une stratification morphologique.

- Troisième objectif : *Proposer un modèle mathématique des relations entre morphologie et population en milieu urbain.*

Ici les principaux problèmes seront d'évaluer si l'importance des perturbations du modèle légitime ou non le raisonnement inférentiel, et de confronter les variables démographiques métriques aux variables morphologiques principalement qualitatives.

- Quatrième objectif : *Tester les possibilités d'estimation bayésienne des données démographiques.*

La question est de savoir si le modèle proposé dans la phase précédente peut fournir la base d'hypothèses bayésiennes expliquant les variables démographiques à l'aide des variables morphologiques. Si la réponse est positive, quel gain de présision permettent les estimateurs bayésiens par rapport aux estimateurs sans biais classiques.

- Remarque : Dans l'expérience à Marseille, les tests des différents plans de sondage n'impliquent pas la réalisation d'enquêtes démographiques. La base de sondage sera toujours l'ensemble des ilôts INSEE et les résultats sont obtenus en appliquant les estimateurs aux données du recensement sur les ilôts de l' échantillon. Cela permet de mesurer réellement la précision des différents plans mais, en contrepartie, on n'a pas la liberté du choix de l'unité spatiale à enquêter: on est tenu au découpage en districts de recensement fait par l'INSEE.

4. 3. ETAPES ET METHODES DU TRAVAIL STATISTIQUE A MARSEILLE

Ce qui suit ne constitue pas le plan de la thèse : il s'agit de faire des propositions méthodologiques pour atteindre les principaux objectifs de la mise au point d'une méthode à Marseille. La chronologie de ce plan de travail statistique est imposée dans ses grandes lignes par les étapes du développement du projet de recherche ORSTOM : il faut qu'il fournisse au fur et à mesure des demandes , des réponses aux questions qui conditionnent l' avance d'un projet auquel participent plusieurs chercheurs. Le plan de thèse lui-même sera issu de ce plan de travail après réorganisation des chapitres en fonction des domaines mathématiques et statistiques abordés. Certains points pourront être développés dans la thèse au delà des strictes besoins du projet et faire l'objet de conclusions plus générales.
Etant donné le stade d'avancement des travaux, certaines des propositions méthodologiques ci après apparaissent comme des esquisses exploratoires donnant les grandes lignes du travail à effectuer.

4 . 3 . 1 . Analyse des données pour la modélisation des relations entre variables morphologiques et démographiques en milieu urbain

Cette phase du travail vise à atteindre les trois premiers objectifs fixés à l'expérience de Marseille :
- 1 : Définition d'une typologie morphologique de l'espace urbain.
- 2 : Mesure du gain de précision dû à cette typologie.
- 3 : Modélisation mathématique des relations entre morphologie et population.

. Analyse typologique et classification automatique pour la stratification de l'espace urbain d'après des critères morphologiques. Juin 85 - Avril 86.

Pour modéliser les relations morphologie/population aussi bien que pour identifier l' information morphologique sur l'image satellite, on a besoin d'un échantillon représentatif d'ilôts INSEE sur lequel confronter l' information spectrale, une information morphologique fine issue des photographies à grande échelle et d'enquête de terrain et l' information démographique. La représentativité de l'échantillon doit être comprise ici relativement à une typologie des ilôts suivant leur caractéristiques morphologiques. L'établissement d'une telle typologie présente donc un double intérêt :
- Premièrement, permettre de définir une statification de la base de sondage qui garantisse une représentativité morphologique de l'échantillon de travail.
- Deuxièmement, permettre le test de l'efficacité de cette stratification dans l'estimation de population.
La démarche suivante doit aboutir à la typologie et au tirage de l'échantillon de travail. On comprendra que l'on accorde à l'exposé de cette phase une importance particulière (les résultats des étapes 1 à 4 de la démarche sont exposés au chapitre 2.1. de ce rapport).

1 : F. DUREAU a effectué un *zonage de la mosaïque des photographies aériennes* au 1/23 000 en 542 unités visuellement homogènes du point de vue morphologique, les contours de ces unités n'ayant a priori pas de rapport avec ceux des ilôts INSEE.

2 *: On a quantifié et codé les caractéristiques morphologiques de chaque zone* dans un tableau croisant les 542 zones avec 10 variables descriptives (voir Tableau 1).

3 : *Analyse statistique monodimentionnelle de chaque descripteur*: On cherche à rendre homogènes les caractéristiques statistiques des variables par transformation des variables métriques en variables qualitatives et recodage disjonctif complet du tableau.

4 : *Analyse typologique du tableau recodé* : étant donné la taille et les caractéristiques du tableau de données et le caractère exploratoire de cette première analyse, on applique dans cette phase les méthodes de l'analyse des donnée développées par BENZECRI et ses collaborateurs et des techniques de classifications hiérarchiques et non hiérarchiques : analyse des correspondances et classification ascendante hierarchique, nuées dynamiques ou cluster analysis pour des classifications supervisées.
On obtient une stratification de l'espace urbain en ensembles de zones morphologiquement semblables. L'analyse permet également d'identifier les indicateurs morphologiques les plus discriminants ainsi que la méthode de classification la mieux adaptée à ce type de données et aux objectifs poursuivis.

5 : *Réaffectation de la typologie obtenue aux ilôts INSEE*: On obtient ainsi, grâce au système TIGER, une base de sondage composée d'ilôts INSEE, stratifiée selon le type morphologique.

6 : *On peut envisager deux types de représentativité de l'échantillon :*
- 1 : *L'allocation dans chaque strate est proportionnelle au nombre d'ilôts de la strate*: On obtiendra ce résultat grâce à un tirage systématique sur liste dans chaque strate avec le même taux dans toutes les strates, ou avec un tirage des ilôts grâce à une grille de points superposée à la carte. La maille de la grille varie alors dans chaque strate en fonction de la taille moyenne de l'ilôt dans la strate pour obtenir un taux constant en moyenne.
- 2 : *L'échantillon est représentatif des surfaces couvertes par chaque strate*. Là aussi les deux types de procédure sont possibles.

7 : *Comparaison de plusieurs échantillons de travail :*
Au stade de l'échantillon de travail, une méthode de tirage ne pourra être déclarée meilleure qu'une autre que si elle respecte mieux les caractéristiques d'une ou plusieurs strates. Pour pouvoir comparer entre elles plusieurs techniques, il faut donc réaffecter aux ilôts l'ensemble des caractéristiques morphologiques des zones auxquelles ils apartiennent ou tout au moins des variables qu'on a jugées déterminantes dans l'analyse. C'est en comparant pour ces variables et pour chaque strate, les moyennes obtenues pour différents échantillons aux moyennes calculées sur l'ensemble des ilôts, que l'on choisira un des échantillons.

Enfin, s'il s'avère qu'aucun échantillon n'est meilleur que tous les autres sur toutes les strates, on choisira celui qui permet la meilleure représentativité des strates les plus denses.

. Test d'efficacité d'une stratification morphologique dans l'estimation de population par sondage. Mars 86 - Mai86.

La typologie des quartiers de Marseille réaffectée aux ilôts constitue une première information à croiser avec la population. On se propose donc de profiter de la nécessité du tirage de l' échantillon de travail pour mesurer, sur divers plans de sondages classiques , sondages sur liste ou tirages de point dans une grille, le gain de précision apporté à l' estimation par une stratification de la base de sondage d'après des critères morphologiques.
Le principe de comparaison de divers plans de sondage dans l'expérience à Marseille sera toujours le suivant : on effectuera à partir des données de population connues sur l'ensemble de la base de sondage, le calcul et la comparaison des variances des estimateurs ascociés à chaque plan.
Ainsi pour apprécier le gain de précision lié à la stratification morphologique, on comparera les variances associées à des plans stratifiés et les variances des estimateurs correspondants dans des plans non stratifiés. Enfin, au cours de cette phase, on s'intéressera aussi à voir quelles techniques de tirage bénéficient le plus de la stratification.

. Méthodes statistiques pour la mise au point de modèles explicatifs de la densité de population. Avril 86 - Juillet 86.

Il s'agit là de formuler et de valider des hypothèses de dépendance généralement linéaire ou affine, de la densité de population par rapport à un certain nombre de variables morphologiques.
La densité de population définie sur la surface entière de l'ilôt peut être vue comme le produit de la "densité de bâti" (rapport surface bâtie/ surface totale) , par la "densité d'occupation du bati" (rapport population/ surface batie).
On pense généralement dans les milieux de télédétection pouvoir estimer assez précisément le premier terme de ce produit (densité du bâti) à partir de l' image satellite. On s'attachera donc dans un premier temps à mesurer la corrélation existante entre densité de population et densité de bâti, puis on essaiera de proposer un modèle explicatif du second terme en l'approximant à l'aide d'une fonction de variables morphologiques. La recherche d'un tel modèle comprend deux étapes : la formulation des hypothèses du modèle, puis leur validation.

a- Formulation des hypothèses

Dans cette phase, il faudra disposer d'informations morphologiques plus précises

que celles issues de la photo-interprétation de la mosaïque au 1/23 000. On travaillera donc sur l'échantillon d'ilôts défini lors de la phase précédente, sur lequel les photographies au 1/8 000 et les vérifications sur le terrain permettent de décrire précisement la morphologie de chaque ilôt. Compte tenu du taux envisagé pour cet échantillon, environ 2 %, le tableau de données résultant aura environ 80 lignes mais le nombre des descripteurs des ilôts pourra être plus elevé que lors de l'analyse typologique. L'éventail des méthodes d'analyse de données et d'ajustement applicable à un tel tableau est plus vaste que dans la phase précédente :

1: *Analyse des données pour la sélection des variables morphologiques pertinentes.*
A nouveau il s'agit de tester plusieurs méthodes adaptées aux diverses formes que peut prendre le codage de l'information morphologique : analyse de similitude et analyse booléenne, analyses factorielles, analyse de la variance et/ou analyse bayésienne (méthodes inférentielles d'analyse de données). Sous réserve que cela s'avère efficace, on pourra également utiliser les variogrammes et covariogrammes de la théorie des variables regionalisées de G.MATHERON, les variables démographiques et morphologiques étant vues comme des régionalisations dont on doit comparer les variations spatiales.

2: *Ajustement linéaire entre densité de population et variables morphologiques.*
Là encore le caractère métrique ou qualitatif des variables explicatives peut conduire à privilégier une méthode parmi les suivantes : régression multiple, analyse discriminante, discrimination logique.

b-Validation des hypothèses

Des outils de statistique paramétrique et issus de la théorie des tests d'hypothèses peuvent être appliqués aux paramètres des modèles d'ajustement définis dans l'étape précédente. Les tests se feront sur l'échantillon quand on voudra conserver la précision maximum de mesure des caractères morphologiques. Sur des modèles simplifiés, intégrant une information morphologique plus grossière, on pourra se satisfaire de la réaffectation aux ilôts INSEE de l'information morphologique issue de la mosaïque des photographies aériennes au 1/23 000. Il sera alors possible de tester des hypothèses sur l'ensemble des ilôts.

4 . 3. 2 . Plans de sondages aréolaires stratifiés et estimation bayésienne de la population

L'aboutissement logique du schéma d'expérience sur Marseille est l'intégration de l'information morphologique dans la stratégie de sondage. Selon la théorie classique des sondages (voir, par exemple, J. DESABIE ou CH. GOURIEROUX) , cette intégration peut se faire principalement à deux niveaux :
-1) Dans la définition de la stratification et la stratégie d'allocation aux strates.

-2) Dans le choix d'estimateurs bayésiens.

. Définition des strates et allocation

L'expression classique de la variance d'un estimateur sans biais (dit d'HORVITZ-THOMPSON) du total dans un sondage stratifié est :

$$V(\hat{T}_{h.t.}) = \sum_{h=1}^{k} Nh(Nh-nh).\, Sh^2 \,/\, nh \;\ldots$$

où k est le nombre de strates, Nh le cardinal de la strate h, nh l'effectif de l'échantillon contenu dans la strate h, (nh / Nh est donc le taux de tirage dans la strate h), enfin $Sh^2 = (N / N-1).\, \sigma^2$, où N est le cardinal de la population et σ^2 la variance du caractère à estimer.

De cette formule, on dégage les règles suivantes à appliquer à la définition des strates et aux taux de tirage dans les strates :
- 1 Les variances dans chaque strate, du caractère observé (population totale ou densité de population), doivent être minimisées. On retiendra donc la stratification construite sur des critères morphologiques qui minimise la somme des variances de la population à l'intérieur des strates.
- 2 L'allocation à une strate : nh doit être proportionnelle à la variance σ_h^2 du caractère dans la strate. On choisira de sonder chaque strate avec un taux proportionnel à cette variance, surtout si les strates hétérogènes sont densément peuplées puisque leur contribution à la population totale est importante.

. Choix d'estimateurs bayésiens

En supposant que la phase de mise au point d'un modèle explicatif de la densité de population ait abouti de manière concluante, on disposera d'une information permettant de traiter la valeur prise par une variable démographique donnée, comme la réalisation d'un vecteur aléatoire Y (qui possède autant de coordonnées qu'il y a d'ilôts INSEE à Marseille), expliqué par le modèle linéaire :

$$\underset{(N,1)}{Y} = \underset{(N,p)}{X} \cdot \underset{(p,1)}{\beta} + \underset{(N,1)}{u} \qquad \text{avec } E(u)=0 \text{ et } V(u)=\sigma^2\Omega,$$

où X est la matrice correspondant à la donnée, pour chacun des N ilôts INSEE de p caractères morphologiques, le vecteur des p paramètres du modèle (coefficients des $X_1 \ldots X_p$ dans l'équation $Y = X_1\beta_1 + X_2\beta_2 \ldots + X_p\beta_p + u$) et u, le résidu ou perturbation du modèle dont on connait grâce à la phase précédente de l'expérience, la matrice de variance - covariance .
L'estimateur bayésien correspondant à cette situation est :

$T = t(Y_s) + {}^t1_{\bar{s}} \cdot X_{\bar{s}} \cdot \beta$, où $t(Y_s)$ désigne le total observé sur l'échantillon s. ${}^t1_{\bar{s}}$ le transposé du vecteur unité sur le complémentaire de

l'échantillon ($N-n$ coordonnées toutes égales à 1) et $X_{\bar{s}}$ la matrice X restreinte au complémentaire de l' échantillon, qui représente donc la connaissance de la morphologie des $N-n$ ilôts n'appartenant pas à l'échantillon.
Il s'agira dans la phase de test de ces estimateurs de comparer les risques bayésiens associés à divers plans de sondage, les risques étant exactement calculables grâce aux données du recensement.

5. LA TELEDETECTION EN MILIEU URBAIN

5 . 1 . LE POINT SUR LA TELEDETECTION URBAINE A TRAVERS LA BIBLIOGRAPHIE FRANCAISE ET ANGLO-SAXONNE SUR LE SUJET

5. 1. 1. L'état des connaissances et le bilan des méthodes

Une fiche de dépouillement standart a été utilisée pour analyser de façon homogène tous les articles (voir page suivante).

. Liste des 35 articles analysés

ANUTA P.E., BARTOLUCCI M., DEAN M.E., LOZANO D.F., MALARET E., Mc GLLEN C.D., VALDES J.A., VALENZUELLA C.R., 1984 - Landsat 4 MSS and Thematic Mapper data quality and information content, IEEE Transaction on Geoscience and Remote Sensing, Vol 22, N°3.

BALLUT, A., LECUP J., LENCO M. - Traitement en classification supervisée de l'occupation de l'espace de l'image Ile de France du 23/3/1973 sur station précablée active. Utilisation d'image satellite pour décrire l'état et le suivi de l'occupation du sol et de la végétation en Ile de France, IAURIF, pp 22-52.

BALLUT A., 1979- Les limites de la télédétection en milieu urbain, thèse de 3ième cycle, EHESS, Paris.

BARDINET C., 1982- Morphologie urbaine et télédétection : les cas d'Annaba (Algérie) et de N'Djaména (Tchad) par Landsat, Journées de télédétection en milieu urbain, 13 Mai 1982, CNRS IGN, Paris, pp100-120.

BOASSON E., HOWARTH P.J., 1983- Landsat digital enhancement for change detection in urban environment, Remote sensing of environment, N°13, pp 149-160.

BOUILLOT J., CAMPAGNE P., FLOURIOT J., LAPORTE J.M., SOUDOPLATOFF S., 1984 - Recherche méthodologique sur l'utilisation de la télédétection en aménagement urbain, Simulation SPOT Niamey.

BOZET M., DOTU M., INSTALLE M., WILMET J., 1980 - Filtrage spatial des données Landsat en vue de la cartographie urbaine, Annales des Mines, Avril, Mai 1980.

BOZET M., DOTU M., INSTALLE M., WILMET J., 1978 - Méthode de classification et d'interprétation des données du satellite Landsat pour l'étude des structures urbaines en Europe Occidentale, Bulletin trimestriel de la Société Belge de Photogrammétrie et de télédétection, N°131-132, pp 3-16.

CARTER P., GARDNER W.G., JACKSON M.J., SMITH T.H., 1980 - Urban land mapping from remotely sensed data. Photogrammetric Engineering and Remote Sensing, Vol 46, N°8, pp1041-1050.

CUSHNIE J.L., 1984- Improving the accuracy of computer classification of Thematic Mapper data, Satellite Remote Sensing, Proceeding of the 10 th anniversary, International conference held in Readin, UK, pp329-339.

COLWELL R.N., POULTON C.E., 1985- SPOT simulation for urban monitoring : a comparison with Landsat 4 TM and MSS imagery and with high altitude color

Fiche de dépouillement bibliographique

TOLL D.L. 1985 Landsat 4 Thematic Mapper scene characteristics of suburban and rural area.
Photogrammetric Engineering and Remote Sensing. Vol 51, N°9, pp 1471-1482

Objectifs

Evaluation des propriétés spécifiques de discrimination spectrale de TM
Estimer des méthodes sélectives pour améliorer les performances cartographiques (modes d'occupation du sol)

Documents de Référence

Photo aérienne infra rouge couleur du 13 juillet 1982 Echelle de restitution 1/40000
Carte de l'occupation du sol déduite par photo interprétation de la photo aérienne IRC

Description de la zone étudiée
Site : Beltsville - Laurel (Nord Est de Washington DC)

Type d'images

TM 1, 2, 3, 4 du 29 juillet 1982 corrigées radiométriquement et géométriquement
TM 1,2,3,4,5 et 7 du 2 Novembre 1982 sans nuages
MSS 1,2,3 et 4 du 2 Novembre 1982 " "
MSS 4,5,6 et 7 du 11 juillet 1981 (Landsat 2)
} 500 x 500 pixels 28,5 m² résolution au sol pour TM

Typologie de la classification

classification en 7 postes : eau, forêts, terres agricoles, carrières, routes principales, commerces et industries, habitats résidentiels
la zone Beltsville et Laurel située au Nord de Washington DC comprend : { des habitats résidentiels multi et uni familiaux / des centres commerciaux / des complexes industriels (pas d'industrie lourde)
Sondage avec stratification : 75 pixels/classe soit 525 pixels échantillon
(non explicité)

Prétraitements

- Recherche des meilleures combinaisons de canaux (utilise classification selon le max de vraisemblance) ordre décroissant :
1 canal : 5/6/3/2/4/1/7 ; 2 canaux 2,3/3,4/1,4/1,5/2,5/3,5/5,6/ ; 3 canaux 2,4,6/1,4,5/2,4,5/3,4,6/1,4,6/1,3,4/1,2,4/2,3,4/3,5,6/
- Analyse canonique (60 groupes spectraux) : Automne TM 1er Axe : 2ème Axe : Opposition visible/PIR, 3ème Axe : Opposition MIR/visible ; 4ème Axe discrimine [illegible]
Eté 4 bandes TM 1er Axe : visible ; 2ème Axe PIR ; 4ème Axe absorption chlorophyllienne.
- Filtrages : 1 filtre médian et 1 filtre de lissage ont été effectués pour réduire la variance intra classe (Matrice 5 x 5)

Méthodologie de la Classification

Classification selon le maximum de vraisemblance (Gaussien). Supervisée
9 classifications ont été testées sur l'échantillon (Intervalle de confiance 95 %)

Résultats

Classif		Précision
1	(sans transformation)	79,2 %
2	filtre médian	89,1 %
3	filtre lissage	90,5 %
4	Transformation canonique	83,4 %
5	"supervisée"	83,2 %
6	Multidate 29/07 et 2/11/82	87,2 %
7	29/07 (4 bandes)	77,5 %
8	2/11 (4 bandes)	72,9 %
9	2/11 (7 bandes)	79,2 %

Conclusions

Les bandes TM suffisent pour discriminer le mode d'occupation du sol
il existe une redondance à l'intérieur des zones spectrales, on peut donc supprimer sans perte significative de précision les canaux TM2 et TM6
- TM permet d'identifier des masses d'eau trois fois plus petite que ne le permettait MSS

Matrice de corrélation TM (2/11/82)

	1	2	3	4	5	6	7
1	1						
2	.96	1					
3	.93	.89	1				
4	.57	.62	.65	1			
5	.76	.81	.79	.70	1		
6	.84	.87	.84	.60	.95	1	
7	-.05	-.11	-.16	.05	-.09	-.20	1

Remarques

Remarque : contradiction apparente entre le fait de retirer TM2 et TM6 qui entre dans les

infra red photography, Photogrammetric Engineering and Remote Sensing, Vol 51, N° 8, pp 1093-1101.

DEANGELIS R.N., MILAZZO V.A., 1984- Application of simulated SPOT data to mapping land cover patterns and changes in urban fringe environments, SPOT Simulation Application handbook, Proceedings of the 1984 SPOT Symposium, May 20-23, Scottsdale, Arizona, pp 177-186.

DINSTEIN I.H., HARALICK R.M., SHANMUGAM K., 1973- Textural features for image classification, IEEE Transaction on Systems, Man and cybernetic, Vol 3, N°6, pp 610-621.

DOLAN G.M., MARTIN S.R., WARNICK L.J., 1984 - Comparative evaluation of simulated SPOT, Landsat TM and NHAP CIR data for urban land cover and impervious surface identification, SPOT Simulation Application handbook, Proceedings of the 1984 SPOT Symposium, May 20-23, Scottsdale, Arizona, pp 148-156.

FERIER D., LOODTS J., - Analyse multispectrale de Bruxelles, Bulletin trimestriel de la Société Belge de Photogrammétrie et de Télédétection, N°133-134, pp 7-41.

FORSTER B.C., 1985- Principle and rotated component analysis of urban surface reflectance, Photogrammetric Engineering and Remote Sensing, Vol 51, N°4, pp 475-477.

KENNARD R.E., TOLL D.L., 1984 - Investigation of SPOT spectral and spatial characteristics for discrimination land use and land cover in Prince Georges'County, Maryland, SPOT Simulation Applicatopn handbook, Proceedings of the 1984 SPOT Symposium, May 20-23, Scottsdale, Arizona, pp 165-170.

FUEKI T., HONG J. K., SHIBANO K., YANG L.S., 1985 - Analysis of urban spatial structure of Marseille by Landsat Thematic Mapper.

HAACK B.N., 1984 - Multisensor data analysis of urban environment, Photogrammetric Engineering and Remote Sensing, Vol 50, N° 10, pp 1471-1477.

HUSSON A., OLIVA P., 1982 - Classification supervisée des ensembles urbains et péri urbains de la région d'Aix en Provence, Journées de télédétection en milieu urbain, 13 Mai 1982, CNRS IGN, Paris, pp 28-42.

JENSEN J.R., 1978 - Spectral and textural features to classify elusive land cover at urban fringe, The Professionnal Geographer, Vol 31, N04, pp 401-409.

LAPORTE,J.M., 1983 - Etude de la texture sur des simulations d'image du satellite de télédétection SPOT, Thèse de 3ème cycle, Paris VII.

LIMA ROSARIO DA TRIDALE F., 1985 - Etude par télédétection de l'organisation de l'espace dans la commune de Carbonne, DESS, Paris VI, 37P.

LO C.P., - Some problems of computer assisted mapping of land use from Landsat data :the Hong Kong case.

MADEC V., PEBAYLE J., 1984 - Intérêt du futur satellite SPOT pour l'étude du milieu urbain, IAURIF, Rapport de stage.

NGUYEN P.T., 1984 - La simulation SPOT pour l'étude du milieu urbain,

PEBAYLE J., 1986 - L'intérêt en géographie des données du satellite SPOT, Thèse de doctorat 3ème cycle, Université de Paris I, 232p.

QUATTROCHI D.A., 1985 - Analysis of Landsat 4 Thematic Mapper data for classification of the Mobile, Alabama metropolitan aera, Seventeenth International Synposium on Remote Sensing of Environment, Ann Harbor, Michigan, May 9-13, pp1393-1402.

RICHARD J.A., 1984 - Thematic Mapping from multitemporal image data using the principal component transformation, Remote sensing of environment, N° 16, pp 35-46.

RIMBERT S., 1982 - Expérience CNES 81/ 0208, Rapport APT CNRS N°1080, LTC Strasbourg 30 P + figures.

SOUDOPLATOFF S.,1984 - Apport de la simulation SPOT aux études urbaines, Rapport IGN.

SOYER J., WILMET J., 1982 - Lubumbashi et le Sud Est du Haut Shaba: interprétation de données Landsat, Bulletin de la Société Belge d'Etudes Géographiques, N°1, pp87 - 100.

TOLL D.L., 1985 - Landsat 4 Thematic Mapper scene characteristics of suburbain and rural aera, Photogrammetric Engineering and Remote Sensing, Vol 51, N°9, pp 1471-1482.

TOLL D.L., 1985 - Effect of Landsat Thematic Mapper sensor parameters on land cover classification, Remote Sensing of Environment, Vol 17, pp129-140.

TOLL D.L., 1984 - An evaluation of simulated Thematic Mapper data and Landsat MSS data for discrimination suburban and regional land use.Photogrammetric Engineering and Remote Sensing, Vol 50, N°12, pp 1713-1724.

. Synthèse des articles

a) LES DOCUMENTS DE REFERENCE

Cinq grands types de documents servent couramment de référence pour choisir les sites d'entraînement dans l'optique d'une classification supervisée ou de l'identification de signatures spectrales, ainsi que pour vérifier les résultats dans les zones test aprés classification. Il s'agit :

- *D'enquêtes sur le terrain des zones échantillon.* Dans cinq articles, les auteurs ont défini leur typologie ou ont procédé à des vérification par ce seul moyen.

- *De photographies aériennes.*
  ◇ Panchromatiques : 23 auteurs. L'échelle de restitution varie de 1/100 000 à 1/10 000 . la moyenne se situe aux alentours de 1/24 000. Deux auteurs seulement ont profité des possibilités qu'offre la stéréoscopie pour l'interprétation du milieu urbain.
  ◇ Infra rouge couleur : 8 auteurs. L'échelle de restitution des photographies varie entre 1/20 000 et 1/40 000. Un seul auteur a tiré profit de la lecture stéréoscopique.

-*De photographies terrestres.* Un auteur a utilisé des phorographies couleur au sol.

- *De cartes topographiques ou thématiques déduites de l'interprétation de photographies aériennes* : 30 auteurs. Leur échelle varie entre 1/5000 (carte de l'occupation du sol issue du MOS) et 1/500 000 (carte géologique). Des cartes provenant d'organismes aussi divers que l'USGS, l'IGN, Michelin, et le Ministère des Travaux Publics ont servi de base à ces travaux.

- *D'études statistiques*. Deux auteurs ont puisé des informations susceptibles de les aider à mettre au point leur nomenclature dans des recueils statistiques divers.

- *D'autres données*. Quatre auteurs se sont servi du MOS, un autre a travaillé à partir d'une base de données géographiques (CUSHNIE J.L., 1984).

Commentaires

En ce qui concerne la procédure par enquête, on regrettera le manque de détails caractérisant la méthode de sondage suivie qui ne figure que dans peu d'articles dépouillés, ainsi que le petit nombre d'informations concernant le nombre de zones ou de pixels composant l'échantillon (vérification ou entrainement)
Les indications relatives aux cartes (échelle, date, organisme) sont souvent incomplètes. Par contre, la date de prise des photographies aériennes (IRC ou Panchromatique) est généralement très proche de la date d'acquisition des images satellite.

b) IMAGE MSS

Objectifs

Cinq grands objectifs sont au centre des préoccupations des auteurs:

- Tester les potentialités du satellite pour une classification des différents modes d'occupation du sol.
- Evaluer les capacités de MSS à détecter les changements d'affectation pour éventuellement actualiser une banque de données géographiques.
- Mettre au point une typologie, de préférence basée sur des critères radiométriques, qui réponde à la fois aux besoins des aménageurs et des télédétecteurs.
- Déterminer l'influence de traitements (filtres, ACP, indices, etc...) ou l'apport de paramètres texturaux sur l' amélioration des classifications, et comparer la fiabilité des différents types de classifications.
- Réaliser de façon automatique une carte thématique de l'occupation du sol.

Type d'image

Ce sont toutes des images enregistrées par le satellite Landsat.
Peu d'auteurs précisent la qualité des images (BARDINET C., 1982), (BALLUT et Al, 1982). BOZET (et Al, 1978) attirent l'attention du lecteur sur l'importance des phénomènes de pollution. Dans (FERIER et Al , 1979) est expliquée en détail la méthode de correction radiométrique choisie. CARTER (et Al, 1980) indiquent que compte tenu des perturations climatiques, une couverture LANDSAT MSS complète sur leur zone de travail (l'Angleterre) n'est possible qu'à une ou deux occasions dans l'année.

De nombreux auteurs font part de leur difficultés pour superposer une carte thématique sur une image MSS (BALLUT A., 1979) . Aprés correction géométrique, CARTER (et Al, 1980) constatent une erreur standard de 1 pixel. Pour superposer deux images MSS, RICHARD J.A. (1984) fonde sa méthode sur le calcul de l'interpolation cubique (30 amers ont été retenus) et obtient moins d'un quart de pixel de différence (moins de 15 m).

. Typologies de la classification

La typologie de l'occupation du sol choisie par les différents auteurs varie suivant le document qui est à l'origine de leur travail:

- Typologie fonctionnelle en 19 postes du Mode d'Occupation du Sol réalisée par l'IAURIF.
- Typologie issue de l'interprétation de photographies aériennes : 26, 8, 12, 5, 7,18, 6, 8, 6 classes.
- Typologies issues des classifications non assistées : 21, 15, 6, 16 classes (regroupées ultérieurement en 6 classes synoptiques) , 49 taxons (regroupées en 5 classes).

Dans son article, FORSTER B.C. (1985) décrit la contenance moyenne d'un pixel urbain dans sa zone de travail : herbe (36 %), toiture (23 %), arbre (16 %), bitume (14 %), béton (5 %), autre (6 %).

Prétraitements

Le plus simple de ces prétraitements, et qui donne de bons résultats est la *détermination d'indices*:

- MSS4-MSS6, (MSS4+MSS5+MSS6)/MSS4 (BARDINET C., 1982);
- MSS7 / MSS5, MSS5 / MSS4 (LO C.P.) ;
- L' indice de végétation qui met en évidence la limite urbaine ainsi que les axes routiers principaux (BOASSON et Al, 1983).

ou *la création de néo canaux*:

- Transformation de KARHUNEN-LOEWE (on retient les 2 premières composantes principales) (BOZET et Al, 1978,1980), (FERIER et Al, 1979), ou les 4 premières dans le cas d'une analyse diachronique (RICHARD J.A., 1984).

Les auteurs s'intéressent aussi à la répartition spatiale des pixels. Deux grands axes sont explorés :

- Affectation du pixel à une cellule par *filtrage spatial* (BOZET et Al, 1980) ce qui a pour effet d'améliorer le résultat des classifications non supervisées (moins de pixels isolés , plus de classes) et d'accroître la netteté des contours.

- Calcul de *paramètres de texture*: matrice de cooccurence des niveaux de gris (14 paramètres) (DINSTEIN et Al, 1973) , (JENSEN J.R., 1978)

- *Transformée de FOURRIER* (BARDINET C., 1982)

Des approches originales sont tentées :

- Calcul de la diminution du désordre dans l'image, que provoque la correction radiométrique, à l'aide de la théorie de l'information ou application du modèle numérique des catastrophes, qui s' est soldé par un

échec compte tenu de la trop grande fluctuation des radiances entre pixels voisins, (FERIER et Al, 1979).

- "Etude des formes contenantes" et en particulier extraire des réseaux à l'aide de filtres (variance mobile et Laplacien), séparer une texture plus grossière d'une texture plus fine, et rechercher des formes squelette en contractant l'information spatiale (RIMBRT S., 1981).

Dans (FORSTER B.C., 1985), *l'Analyse en Composantes Principales* (ACP) exprime les contributions de chaque composante pour chacune des classes thématiques et l'on peut visualiser les positions des différentes classes sur le premier plan factoriel . Ce type d'analyse fait quelquefois apparaitre des noyaux (regroupementsde pixels proches de part leurs caractéristiques spectrales).
De plus, on peut créer des néo-canaux par calcul de paramètres de texture, et réaliser une ACP sur l'ensemble des 7 canaux TM originaux et des néo-canaux de texture : les premières composantes principales issues de l'ACPcorrespondent à un "compactage" de l'information intégrant les données spectrales brutes et les paramètres texturaux. (WILMET J., 1982).

Méthodologie des classifications et résultats

BALLUT A. (1979) emploie une classification supervisée de type géométrique, et il fait varier la caractéristique de calcul des distances, ainsi que le mode de choix des pixels.

CARTER (et Al, 1980) fait appel à une classification supervisée selon le maximum de vraisemblance. La cause principale des mauvais classements provient selon lui d'une rectification géométrique insufisante, d'une résolution spatiale et spectrale inadaptée à la problématique, et du manque de sophistication du logiciel de classification.

DINSTEIN (et Al, 1973) calcule les matrices de cooccurence des niveaux de gris puis procède au classement selon une classification supervisée, avec hypothèse statistique pour la photographie aérienne digitalisée, et selon les hyper-plans discriminants pour l'image satellite.

JENSEN J.R. (1978) crée un quatrième canal composé respectivement de l'Ecart Type Local (ETL), du Contraste (C), du Filtrage Haute Fréquence (FHF) et du Second Moment Angulaire (SMA). Une moyenne prenant en compte les 4 directions principales est ensuite effectuée.

LO C.P. teste différentes classifications.
Dans les trois premières, il fait appel à la distance euclidienne comme mesure Les traitements effectués sont :
- Transformation par une Analyse Canonique.
- Classification selon deux Néo Canaux : MSS5/MSS4, MSS7/MSS5.
Dans la classification suivante, il pose une hypothèse statistique (Maximum de Vraisemblance).

Les autres auteurs ne quantifient pas les résultats de leur classification, la plupart d'entre eux comparent visuellement les classes obtenues avec une carte thématique ou une photographie aérienne de leur zone d'étude. BARDINET C. (1982) propose une méthode de critique de la classification fondée sur l'analyse discriminante.

Seul l'article (BOZET et Al) combine deux classifications. La première, non assistée, est suivie d'un seconde, assistée; cependant l'identification de l'industrie lourde

demeure impossible. Dans leur article, HUSSON et Al (1982) indiquent les limites d'une classification, même supervisée.

AUTEUR	CANAUX	TRAITEMENT	POSTES	FIABILITE	ECHANTILLON
BALLUT85					
BALLUT79	4 MSS	Sebesteyn	19	22,4 %	zone de 621 pix
	4MSS	Euclidienne	19	24,0 %	zone de 621 pix
	4MSS	Xi et ecart	19	26,9 %	zone de 621 pix
	4MSS	sans	19	24,6 %	zone entière(1)
	4MSS	(2)	19	41,7 %	zone entière(1)
	4MSS	(2) (resp 2)	19	51,7 %	zone entière(1)
	4MSS	(2)(resp 3)	19	53,8 %	zone entière(1)
	4MSS	sans	8	60,0 %	zone entière(1)
CARTER	4MSS	sans	5	80,0 %(3)	6 zones test
DINSTEIN	4MSS	sans	8	77,0 %	624(4)
	4MSS	texture	8	83,0 %	624(4)
	photo	sans	7	83,0 %	624(4)
JENSEN	MSS4,5,6	sans	6	référence	
	MSS4,5,6	ETL	6	+37,0 %	(5)
	MSS4,5,6	C	6	+70,0 %	(5)
	MSS4,5,6	FHF	6	+80,0 %	(5)
	MSS4,5,6	SMA	6	+55,0 %	(5)
LO	4MSS	DE	9	73,4 %	(6)
	4MSS	DE	8	77,0 %	(7)
	4MSS	AC	8	81,0 %	(7)
	4MSS	NC	8	83,8 %	(7)
	4MSS	MV	8	72,4 %	(7)

BOASSON et Al détectent les changements intervenus entre deux dates sans mettre en oeuvre le mécanisme un peu lourd d'une classification. Une composition colorée de l'indice de végétation à ces deux dates fait apparaitre les zones nouvellement construites ainsi que la délimitation urbain /rural.

Conclusion

1 Image de 256 036 pixels

2 On ne prend en compte que les pixels ayant 1, (res 2), (res 3) couronne(s) de pixels de même MOS.

3 Ce chiffre ne concerne que les zones urbaines. 85 à 95 % des zones rurales ont été bien classées.

4 La moitié pour l'entrainement et l'autre moitié pour les tests.

5 Les pixels qui ont changé d'affectation sont vérifiés par enquête sur le terrain.

6 La classification obtenue est comparée avec une photographie aérienne échelle de restitution 1/ 100 000.Le nombre de pixels test n'est pas précisé.

7 La classification est comparée avec une photographie aérienne ; échelle de restitution 1/25 000. Le nombre de pixels test n'est pas précisé.

- Les informations recueillies à l'aide de la télédétection sont de nature biophysique et celà pose des *problèmes de correspondance avec les typologies* qui sont de nature essentiellement fonctionnelle (BALLUT et Al, 1980) ; il est donc urgent de définir une nomenclature adaptée.
- *La définition spatiale de MSS est inadaptée à l'étude des villes.* Compte tenu de la taille moyenne des éléments urbains , se pose le problème d'interprétation de la valeur du pixel : est ce la valeur générale du pixel ou bien la moyenne de la reflectance de deux objets très contrastés ? (BALLUT A., 1979).
- La *pollution urbaine* introduit un facteur perturbant la réflectance (BOZET et Al, 1978), mais ce type de contradiction urbaine est analysable par télédétection au même titre que les micro climats ou l'état de la végétation (BARDINET C., 1982).
- Le *choix des sites d'entrainement* est plus important que le choix de la classification (LO C.P.); cependant la classification selon la *distance euclidienne* semble la plus fiable, et c'est de plus la moins coûteuse. L'*analyse diachronique* de données multispectrales est à conseiller car elle permet de lever certaines ambiguïtés dans l'interprétation (SOYER et Al, 1982).

Perspectives

- *Tenir compte de la texture* dans le processus de classification, à l'aide de paramètres (BALLUT A, 1979), (CARTER et Al, 1980) ou de filtres spatiaux (BOZET et AL, 1980). DINSTEIN (et Al, 1973) dressait déja en 1973 un bilan positif de l'apport des paramétres texturaux pour les classification, confirmé par (JENSEN J.R., 1979) , mais constatait aussi qu'un travail d'amélioration restait à fournir en particulier pour déterminer la taille idéale des fenêtres.
- *Utiliser le canal infra rouge thermique de TM* (BALLUT A., 1979) ou les bandes radar (FERIER et Al, 1979).
- Mettre au point et calculer un *indice de minéralisation* (BALLUT A., 1979)

c) IMAGE TM

. Objectifs

Quatre objectifs principaux se dégagent de la littérature :

- Tester l'influence du filtrage / lissage sur la précision de la classification.
- Etudier les caractéristiques des 7 canaux de Thematic Mapper.
- Rechercher les combinaisons optimales pour classer ou cartographier l'espace urbain.
- Estimer l'utilisation, l'apport des données multispectrales pour la description du milieu urbain.

. Type d'image

Il s'agit le plus souvent d'images enregistrées par le satellite. Seul HAACK B.N. (1984) a mêlé des images TM simulées avec des images radar simulées à l'aide d'un capteur embarqué sur avion. La plupart des auteurs n'ont pas travaillé sur la bande 6 (infra rouge thermique de résolution 120 métres).
Dans (CUSHNIE J.L., 1984) ,il est rappelé l'influence de l'heure d'enregistrement de l'image et plus particulièrement de la position du soleil sur les phénomènes de zones d'ombre potentiellement nombreuses dans les villes.
Il est rarement précisé si ces images ont fait l'objet d'un prétraitement visant à les corriger géométriquement et radiométriquement.
Seul TOLL D.L. (1985) précise brièvement la qualité de l'image (présence ou non de nuages, brume, etc).

Un tiers des articles dépouillés ne présente pas de description, même sommaire du site étudié.

Typologie des classifications

Il est majoritairement fait appel à une optique fonctionnelle pour définir les thématiques de classification des différentes utilisations / couvertures du sol.
Le nombre de classes varient de 5 à 23. Dans certaines typologies, le bâti n'est caractérisé que par 2 classes en fonction de sa densité (CUSHNIE J.L., 1984) ou du type d'activité (commerce et industrie / habitat) (TOLL D.L., 1985). FUEKI (et Al., 1985) propose parallélement à une typologie fonctionnelle en 23 postes, une typologie des matériaux de construction comportant 10 classes, ainsi qu'une correspondance entre elles deux.
Ces typologies sont déduites de l'interprétation de photographies aériennes, préalablement à la classification, mais le type de classification choisie rend quelquefois indispensable l'adoption d'une nouvelle légende.

Prétraitement ou post-traitement

Trois grands types de traitements ont été effectués dans le but d'améliorer les résultats des classifications :

-Les *filtrages spatiaux*. Il s'agit de dégrader l'image (atténuation du contraste à l'aide d'un lissage) qui entrainera une réduction de la variance intra classe (TOLL D.L., 1985), (CUSHNIE J.L., 1984).

- Le *calcul d'indices*. LIMA ROSARIO DA TRIDALE F. (1985) fait appel à deux indices , l'indice de végétation et l'indice de brillance; le résultat de ces deux classifications est uniquement visuel.

- Les analyses :

* *Analyse multicanal*. TOLL D.I. (1985) et QUATTROCHI D.A. (1983) en extraient les corrélations entre les différents canaux.[6]
* *Analyse discriminante*. Elle permet de calculer le pouvoir, le degré de séparabilité de deux classes. On peut en déduire les combinaisons de canaux optimales : TM 4 / 4,5 / 3,4,5 / 2,3,4,5 / [CUSHNIE J.L., 1984], (HAACK B.B., 1984) : TM4/4,5/2,1,6/1,2,3,6, ou [TOLL D.L., 1985] : TM 5 / 2,3 / 2,4,6 / 3,4,5,7 .
* *Analyse canonique*. FUEKI et Al 1985 la développent pour rechercher la combinaison de canaux optimum face à un thème déterminé, cartographie de l' utilisation du sol : TM 2,3,4, et cartographie des matériaux de construction : TM 4,5,7 . Dans (TOLL D.L., 1985), la transformation canonique, qui est une transformation linéaire est mise à contribution pour réduire la variance intra classe tout en perdant le minimun de variance inter classes.

---

6 Matrice de corrélation des canaux TM.

	1	2	3	4	5	6	7
1	1,0						
2	0,96	1,0					
3	0,83	0,89	1,0				
4	0,57	0,62	0,65	1,0			
5	0,76	0,81	0,79	0,70	1,0		
6	0,84	0,87	0,84	0,60	0,95	1,0	
7	-0,05	-0,10	-0,16	0,05	-0,09	-0,10	1,0

* *Analyse statistique descriptive* : tableau présentant la moyenne et l'écart type pour chaque zone dans chaque canal (HAACK B.N., 1984).

Méthodologie de la classification et résultats

Nous ne parlerons ici que des classifications ayant donné lieu à une évaluation chiffrée des résultats. Les autres (méthodes de l'histogramme bidimentionnel et analyse discriminante) ont été uniquement évaluées par les auteurs en fonction de critères visuels : comparaison avec une carte thématique ou une photographie aérienne.
Différents types de classifications supervisées multicanal sont habituellement employées, mais la plus usitée est celle du maximum de vraisemblance (CUSHNIE J.L., 1984), (TOLL D.L., 1985). Une méthode non supervisée a été testée par QUATTROCHI D.A. (1983).

AUTEUR	CANAUX	TRAITEMENT	POSTES	FIABILITE	ECHANTILLON
CUSHNIE	TM3,4,5	Filtre moyen	5	93,3%	(7)
	TM1,3,4,5	Filtre médian	5	92,0%	(7)
	7 TM	sans	5	89,0%	(7)
QUATTROCHI	7TM	sans	11	89,18%	(8)
TOLL	7TM	Lissage	7	90,5%	75pix/classe
	7TM	Filtre médian	7	89,1%	75pix/classe
	11TM	Multidate	7	87,2%	75pix/classe
	7TM	Trans Canon	7	83,4%	75pix/classe
	7TM	sans	7	79,2%	75pix/classe

Conclusions

QUATTROCHI (1983) obtient, après classification, des résultats qui lui permettent d'affirmer qu'il est possible de discriminer les habitats pavillonnaires des autres types d'habitat. On retiendra surtout l'*amélioration de la classification due à l'utlisation de filtres spatiaux* (CUSHNIE J.L., 1984), en particulier du filtre médian qui réduit la proportion de pixels non classés. L'originalité de sa méthode tient au fait qu'il intégre lors de la classification, un canal filtré avec des canaux TM intacts, ce qui combine les avantages d'un filtrage avec une perte moindre de résolution. TOLL (1985) constate qu'il existe une redondance à l'intérieur des zones spectrales, et que l'*on peut sans perte significative de précision supprimer les canaux TM2 et TM7.*

7 Les pixels d'entrainement sont choisis au hasard par fenêtres de (5x5), ceux de test par fenêtres de (3x3). Suburbain (125 pixels d'entrainement , 63 pixels de test), urbain dense (50,27), eau (125,63), forêt (125,63), Agriculture (250,126).

8 138 polygones échantillon ont été vérifiés sur le terrain. La moitié pour identifier les classes, l'autre moitié pour évaluer la fiabilité de la classification.

d) COMPARAISON TM/MSS/SPOT/PHOTOGRAPHIES AERIENNES

Typologie de la classification

Toutes les thématiques à l'exception de (ANUTA et Al 1984) sont d'inspiration fonctionnelle.
Ces typologies issues de l'interprétation de photographies aériennes comportent entre 7 et 12 postes, mise à part celle de (DOLAN et Al,1984) dont le nombre de classes n'est pas précisément spécifié et qui s'inspire à la fois de la typologie de l'USGS Land use and Land Cover Classification et de la Howard County Classification Summary. Ces deux typologies très semblables possèdent 4 niveaux de précision. Aprés une classification de type non supervisée, les auteurs totalisent 51 classes pour l'image SPOT et 46 classes pour l'image TM ; de même, ANUTA (et Al, 1984) obtiennent 42 classes pour l'image TM et 21 classes pour l'image MSS à la suite d'une classification de type nuées dynamiques (classification non supervisée).
L'article de COLWEL R.N. et POULTON C.E. (1984) ne se base pas sur une réelle typologie puisque l'objectif de cette étude consiste en une détection d'objets ,urbains et ruraux, bien précis (maisons, églises, golfs chênes, etc...)

Traitements

ANUTA et Al. (1984) procèdent à l' Analyse en Composantes Principales d'une image TM et d'une image MSS représentant la même zone (Chicago) ce qui leur permet d'affirmer que, sur ce site, on peut se contenter des 4 premières composantes pour TM et des 2 premières pour MSS sans perte significative d'information (variance expliquée respectivement 98,50% et 96,78%).
Les mêmes auteurs déterminent, à l'aide de l'*Analyse Discriminante* qui mesure le pouvoir de séparabilité des canaux par le calcul de la distance statistique entre chaque classe, les *combinaisons de canaux* les plus discriminantes qui sont : TM 5 / 4,5 / 3,4,5 / 3,4,5,6. On peut constater, en comparant cette combinaison avec celles présentées dans le paragraphe précédent que les combinaisons "optimales" diffèrent, bien que les canaux 4 et 5 prédominent dans tou les cas.
Cette différence peut s'expliquer : d'une part ces auteurs ne poursuivent pas des objectifs strictement identiques, d'autre part la nature des régions observées (proportion plus ou moins grande d'espaces ruraux) favorise l'émergence de tel ou tel canal dans l'analyse discriminante.

Dans ses deux articles, TOLL dégrade la résolution spatiale (convolution par filtre) ainsi que le niveau de quantification (division par 4 de la valeur de chaque pixel) de TM afin de pouvoir évaluer, sur une même base, le gain ou la perte de discrimination imputable à TM.

Méthodologie de la classification et résultats

En ce qui concerne les classifications supervisées, qui sont les seules à avoir été suivies de mesures de fiabilité, la méthode dominante demeure la classification selon le maximum de vraisemblance.

AUTEUR	CANAUX	TRAITEMENT	POSTES	FIABILITE	ECHANTILLON
ANUTA	7 TM	sans	7	95,7 %	X
	TM 3,4,5,6	sans	7	92,6 %	X
	MSS4,5,6,7	sans	7	67,4 %	X
	7 TM	SECHO[9]	7	97,9 %	X
KENNARD	SPOT	sans	8	45,4 %	10
	TM 2,3,4	sans	8	49,2 %	1000[10]
	7 TM	sans	8	54,5 %	1000[12]
TOLL 1985	4 MSS	sans	7	74,8 %	75pi/postes
	7 TM	sans	7	83,2 %	75pi/postes
TOLL 1984	7 TM	simul[11]	12	70,4 %	100[12]
	4 MSS	sans	12	64,6 %	100

Conclusions

Tout d'abord, en ce qui concerne la *comparaison entre TM et MSS.* Selon (ANUTA et Al., 1984), *la classification des données TM est meilleure quelque soit la classe choisie.* Des classes inexistantes ou inséparables sur une image MSS peuvent être prises en compte avec TM (plages, routes, etc...). Cette affirmation est quelque peut tempérée par TOLL D.L. (1985) qui, s'il s'accorde à constater un gain de 8,9% du à l'addition des canaux TM1,5 et 7, ainsi qu'un gain de 5,2% du à la quantification sur 8 bits de TM, constate aussi une perte de précision due à l'augmentation de la résolution spatiale qui augmente la variance intra classe lors du processus de classification. Il note de plus que les canaux TM1,2,3 sont suffisants pour discriminer l'habitat urbain mais rappel l'intérêt du canal TM6 (infra rouge thermique) pour différencier les commerces des industries.

Les conclusions de l'article de KENNARD (et Al, 1984) sont très semblables en ce qui concerne *SPOT* . En effet si la meilleure résolution spatiale de SPOT réduit les problèmes de pixels frontières ou mixtes, elle augmente sensiblement la variance intra classe et de ce fait accroit à la fois le nombre de pixels mal classés et le nombre de pixels non classés. Par contre si les logiciels de classification tiennent compte de la répartition spatiale des pixels (ou de la texture) cet inconvénient devient un avantage.
Se fondant sur des critères uniquement visuels pour retrouver sur leurs photographies (SPOT, TM, MSS, et IRC échelle 1/24 000) un certain nombre d'éléments, COLWEL R.N. (et Al 1985) remarquent les meilleures possibilités d'interprétation des images SPOT (comparables en précision aux photographies aériennes) par rapport aux images TM ou MSS pour les grandes échelles. Les auteurs concluent leur article en rappelant la nécessité de mettre au point une typologie hiérarchisée (6 niveaux) commune à tous les télédétecteurs. C'est aussi

9 Supervised Extraction and Classification of Homogeneous Objects (algorithme prenant en compte la texture).

10 1000 pixels purs par classes pour l'entrainement et pour le test. Les pixels purs sont ceux dont le voisinage est constitué de pixels appartenant à la même classe thématique.

11 Simulation TM ; l'image a été dégradée pour obtenir une résolution spatiale de 30m.

12 100 pixels non contigues par classe.

l'avis de DOLAN G.M. (et Al 1984) qui concèdent que les données SPOT ou TM peuvent offrir une alternative aux photographies aériennes pour l'actualisation des cartes au 1/24 000. et que les images SPOT sont comparables aux photographies aériennes IRC pour l'interprétation du niveau 3 de l'USGS Land Use and Land Cover Classification System for Use with Remote Sensor Data.

e) SPOT

. Objectifs

Trois grands objectifs ressortent de la littérature consacrée au milieu urbain:

- Evaluer les potentialités de SPOT ; quel type d'information est il possible d'extraire des données SPOT, en particulier pour cartographier l'utilisation/couverture du sol (BOUILLOT et Al ,1984), (NGUYEN P.T., 1984) ou pour dégager une typologie de l'espace urbain (MADEC et Al, 1984), (SOUDOPLATOFF S., 1984).
- Evaluer les capacités de ce satellite pour détecter les changements d'occupation du sol (DEANGELIS R.M., 1984).
- Etudier précisément les signatures spectrales et leurs éventuelles correspondances fonctionnelles (PEBAYLE. J., 1986).

. Type d'image

Seules des simulations réalisées avec un scanner DAEDALUS embarqué sur avion ont été utilisées.

. Typologie de la classification

BOUILLOT J. (et Al, 1984) ainsi que SOUDOPLATOFF S. (1984) ont défini, à l'aide de photographies aériennes, sur le site de Niamey une typologie en 19 postes (dont 7 spécifiquement urbains) pour le mode multispectral et une seconde en 15 postes uniquement urbains pour le mode panchromatique. Ces typologies tiennent compte à la fois du type de matériau et de l'âge de la construction. La typologie proposée par l'USGS a été utilisée par DEANGELIS (et Al, 1984) qui a structuré sa photo interprétation en fonction du niveau III. MADEC (et Al, 1984) de même que NGUYEN P.T. (1984) ont opté pour une légende comportant respectivement 3 et 7 postes.

. Traitements

Les néo-canaux décorrélés résultant de l 'Analyse en Composantes Principales,ont fait apparaitre des réseaux de voirie initialement inexistants (SOUDOPLATOFF S., 1984) et ont facilité l'identification des différentes classes et de leurs signatures spectrales (NGUYEN P.T., 1984) . Ce dernier auteur a opéré un lissage par la médiane des canaux S1, S2, S3, suivi d' une segmentation de l'image (canal panchromatique). Dans un second temps, il a affecté à chacun des pixels la valeur moyenne des points du segment auquel il appartient.

L'indice de végétation (S3-S2)/(S3+S2) a été calculé avec profit par MADEC (et Al, 1984) et PEBAYLE J. (1986) ; c'est en effet un estimateur intéressant de l'urbanisation qui permet en outre de faire apparaitre clairement l' habitat pavillonnaire .

. Méthodologie de la classification et résultats

Deux sortes de classifications ont été utilisées:

-Une *classification non supervisée, type nuées dynamiques* [BOUILLOT J. et Al, 1984], qui a partagé l'image de Niamey en 50 classes suggérant 4 grands thèmes : urbain, végétation sols nus et eau. Les auteurs indiquent q'une classification supervisée a été testée , et qu'elle a donné sensiblement les mêmes résultats.

- Des *classification supervisées*:
  * de type *parallélépipédique modifié* (DEANGELIS R.H. et Al, 1984) . Un tel traitement a détecté, en nombre, 83 % des changements d'affectation et en a correctement reclassé 74 % en fonction de la typologie de l'USGS niveau III.
  * de type *agrégation autour d'un centre mobile* (MADEC J. et Al, 1984), (PEBAYLE J., 1986).Sur la même région , l'Ouest de Paris, les premiers partitionnent l'image en typologie de 2 à 5 postes selon la proportion plus ou moins grande de végétation, alors que le second auteur définit 17 classes. Cette classification donne aux dires des auteurs de bons résultats.
  * de type *maximum de vraisemblance* sur une image lissée et segmentée (NGUYEN P.T., 1984). Cette classification , dont on ne connait pas la légende thématique, discrimine l'habitat individuel, mais ne différencie pas les zones d'habitat collectif.
  * SOUDOPLATOFF S. (1984) obtient une légende en 5 postes (plus ou moins densément bâti) à la suite d'une classification d'un type non précisé.

On notera qu'aucune de toutes ces classifications supervisées n'a donné lieu par la suite à une estimation chiffrée de sa fiabilité. Pour la plupart, les auteurs se contentent d'une estimation visuelle des résultats de leur traitement.

. Conclusion

La plupart des auteurs s'accordent à affirmer que les données SPOT sont utilisables, visuellement, au même titre que des photographies aériennes à condition toutefois de ne pas descendre à l'échelle du bâti ; aussi, dans un premier temps *l'interprétation visuelle des images*, méthode bien connue et fiable , peut apporter plus d'information que ne le ferait un traitement sophistiqué (SOUDOPLATOFF S., 1984).
Il est donc raisonnable de dresser une typologie de l'espace urbain à l'itérieur de la commune (MADEC et Al, 1984) ou de constituer une classification fiable des unités d'une taille supérieure à 1,25 Ha pour le niveau II de l'USGS voire même ponctuellement pour les niveaux III et IV (DEANGELIS et Al, 1984).
RIMBERT S. (1982) introduit la notion de "*mixels*", pixels mixtes dont les signatures sont identifiables, et qui résulte pour le type pavillonnaire par exemple de l'association tuile + jardin + rue + lotissement. Elle signale que ces signatures ne bénéficient pas de la stabilité, dans le temps, que l'on pourrait espérer (variation diachronique), ce qui lui permet d'aborder "l'étude des cycles et des processus" en fonction des variations saisonnières ou horaires.
Pour NGYUEN P.T. (1984), SPOT semble cependant pouvoir constituer un puissant outil de cartographie automatique .
Si SPOT et TM sont des outils complémentaires (PEBAYLE J., 1986), il n'en reste pas moins que des paramètres de *texture* doivent être pris en compte dans les classifications afin de profiter pleinement du principal avantage de SPOT qui est sa bonne résolution spatiale et non pas d'en subir les inconvénients (DEANGELIS R.M., 1984).

5 . 1 . 2. Les axes de recherche à développer

L'étude de la littérature nous a montré l'intérêt, mais aussi les limites dans certains cas, d'un certain nombre d'opérations préliminaires à la classification proprement dite :
- détermination des combinaisons de canaux pertinentes
- calcul d'indices (végétation et minéralisation, notamment)
- utilisation de filtres classiques ou morphologiques
- calcul de paramètres de texture
- exploitation, conjointe de deux images enregistrées à des dates différentes

Pour l'exploitation de l'image TM de Marseille dans le cadre de ce programme de recherche, *nous utilisons ces méthodes dans un environnement particulier, qui est celui de la base de données gérée par le système TIGER*.
Nous avons vu au chapitre 1. 2 pourquoi il est nécessaire de mettre en relation les données spectrales de l'image TM et l'information sur la morphologie urbaine, que ce soit pour le tirage d'un échantillon représentatif des types morphologiques, pour la mise au point de méthodes d'interprétation de l'image TM, ou pour l'évaluation des résultats des classifications.
Du fait de cette mise en relation, les traitements classiques (analyse statistique descriptive, calcul d'indices ...etc) exposés dans la littérature sont faits ici dans un cadre nouveau, avec une démarche zonale, qui a été développé au chapitre 3.2.

Nous nous appuyons sur trois types de documents pour déterminer la *combinaison de canaux* qui, pour chacun des descripteurs morphologiques, nous parait la plus pertinente :

* Les histogrammes bidimensionnels.
* L'analyse en composantes principales.
* *Le tableau des signatures spectrales*, issu du croisement de l'image TM (6 canaux) et des images produites par le système TIGER représentant les différentes zones morhologiquement homogènes (une image par descripteur morphologique).

Après cette première analyse statistique, nous travaillerons sur échantillon, en développant les axes de recherche suivants :

- *Calcul d'indices*, de végétation ou de minéralisation, qui comme nous l'avons vu précédemment apportent un élément de réponse en ce qui concerne la densité du bâti.

- *Evaluation de l'apport du canal TM 6* (infra rouge thermique) pour discriminer les activités industrielles.

- *Test de filtres*, morphologiques (ouvertures fermetures) ou classiques pour dégrader l'image (réduction de la variance intra classe).

- *Calcul de paramètres de texture*. Dans un premier temps, nous calculerons les paramètres classiques (matrices de coocurrence des niveaux de gris) puis nous testerons les méthodes développées par SERRA, basées sur la morphologie mathématique.

- *Exploitation de l'image TM de Marseille enregistée à une autre date* pour procéder à une analyse diachronique.et lever ainsi un certain nombre d'ambiguïtés quant à l'affectation des pixels.

5 . 2 . DEVELOPPEMENTS INFORMATIQUES

Un certain nombre de programmes déjà existants à l'unité de traitement informatique de l'ATOB ont été utilisés pour traiter l'image TM de Marseille.
Mais, d'une part l'exploitation d'une image TM (6 canaux à résolution 30 m), d'autre part la mise en relation avec d'autres données localisées, gérées ou produites par le système TIGER posent des problèmes nouveaux. En effet, nombre de programmes existants ne sont pas adaptés à un nombre aussi élevé de canaux que celui de Thematic Mapper, et correspondent à une analyse uniquement statistique des valeurs radiométriques.
C'est pourquoi plusieurs programmes ont déjà été écrits par B. LORTIC et A. MICHEL : la liste de ces programmes est donnée page 166.
Le fait de caractériser un pixel par son appartenance à une zone géographique dont on connait les principaux paramètres de la morphologie urbaine ouvre une nouvelle voie d'investigation (voir chapitre 3.2) : dans les mois qui viennent, le principal axe de développement informatique sera de prendre en compte cette dimension.

LISTE DES PROGRAMMES DEVELOPPES DANS LE SYSTEME ATOB LORS DE LA PREMIERE PHASE

ECHANT : Réalise le tirage d'un échantillon de pixels par tirage systématique d'un pixel sur n.
LANSEQ : Formate un échantillon pour une exploitation .CIRCE.

Ces deux programmes permet de travailler sur des fichiers résidents au CIRCE. Nous avons donc réaliser une partie des traitements avec les logiciels ADDAD implantés au CIRCE.

TIGLAN : assure le passage de fichiers Séquentiels SEQ.1 à des fichiers Relatif LAN.
LANTIG : assure le passage fichiers Relatifs LAN à des fichiers séquentiels SEQ.1.

ZORRO : permet la translation d'un canal d'un fichier LAN par rapport aux autre canaux.
STADES : réalise la description statistique des différent canaux ou néo-canaux (indices) de chaque zone.

5 . 3. ANALYSE STATISTIQUE DESCRIPTIVE DES DONNEES SPECTRALES DE L'IMAGE T.M.

L'analyse statistique porte sur 6 canaux de l'image T.M. de 1983 recalée par l'IGN, le canal 6 (infra-rouge thermique) étant exclu, pour le moment, de nos travaux : en effet, nous ne pouvons pas établir de références de terrain relatives à cette bande spectrale.
La fenêtre d'étude correspond à la fenêtre minimale encadrant la zone d'étude définie par la figure 1 du chapitre 2.1 : lignes 275 à 865, et colonnes 325 à 683, soit 212169 pixels.

D'autre part, il faut préciser que nous travaillons sur une image masquée : afin d'éliminer la mer de nos traitements, ont été mis à 0 dans tous les canaux , tous les pixels ayant une valeur comprise entre 0 et 9 dans le canal 4. Ce seuil radiométrique de 10 dans le canal 4 a été établi au vu des histogrammes, et du test de différents niveaux de coupure par visualisation du canal 4 sur PERICOLOR.
Toutes les conclusions présentées dans ce chapitre 5.5. portent donc sur une image masquée, ne comprenant pas le canal 6.

5 . 3 . 1. Statistique univariée

La figure 1 montre les histogrammes unidimensionnels de chacun des six canaux. Trois faits importants apparaissent à la lecture de ces histogrammes :

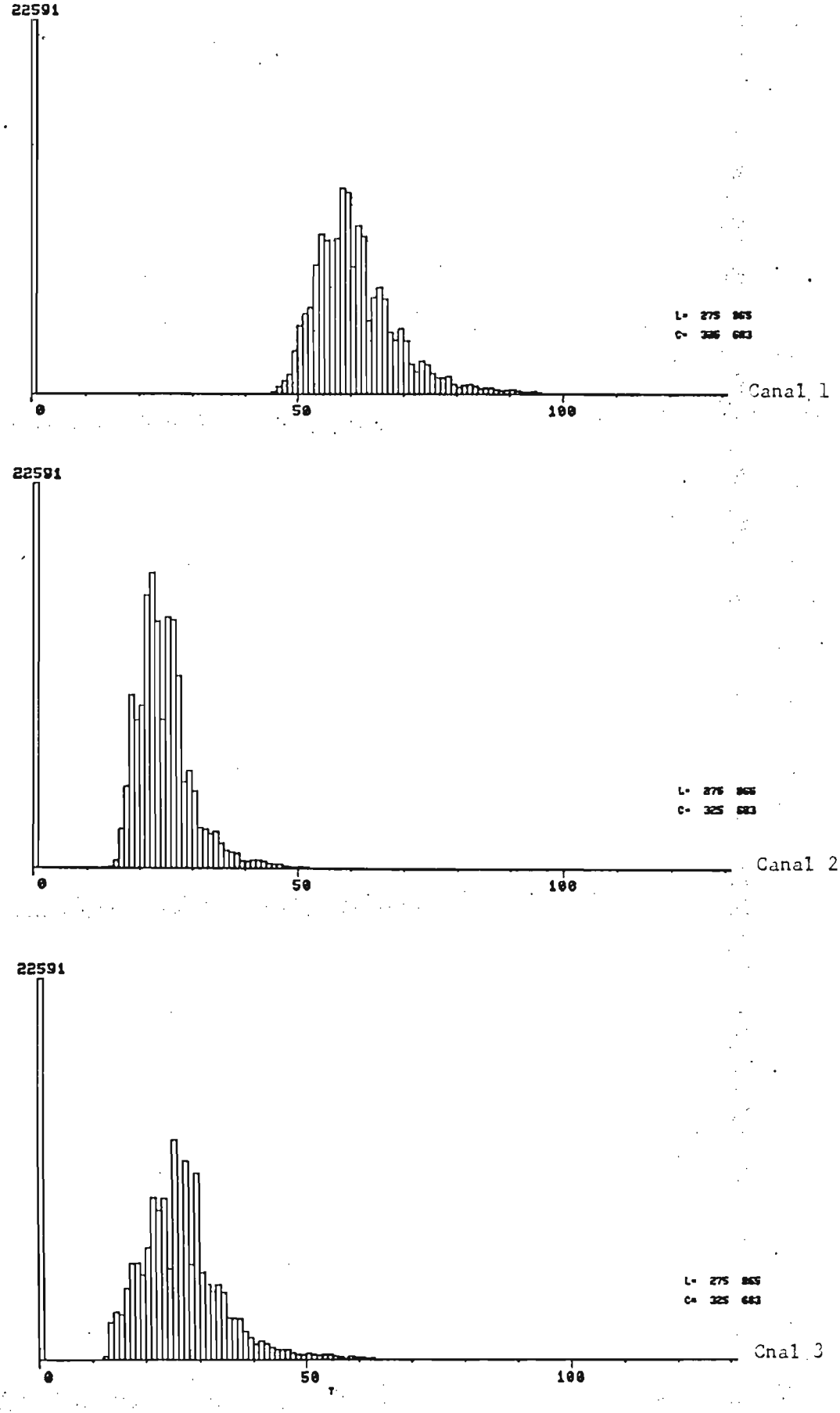
- *Les valeurs radiométriques dans chacun des canaux sont faiblement étalées* : l'étendue statistique (maximum - minimum) est de l'ordre de 50 pour tous les canaux , excepté le canal 5 dont l'étendue atteint 100.

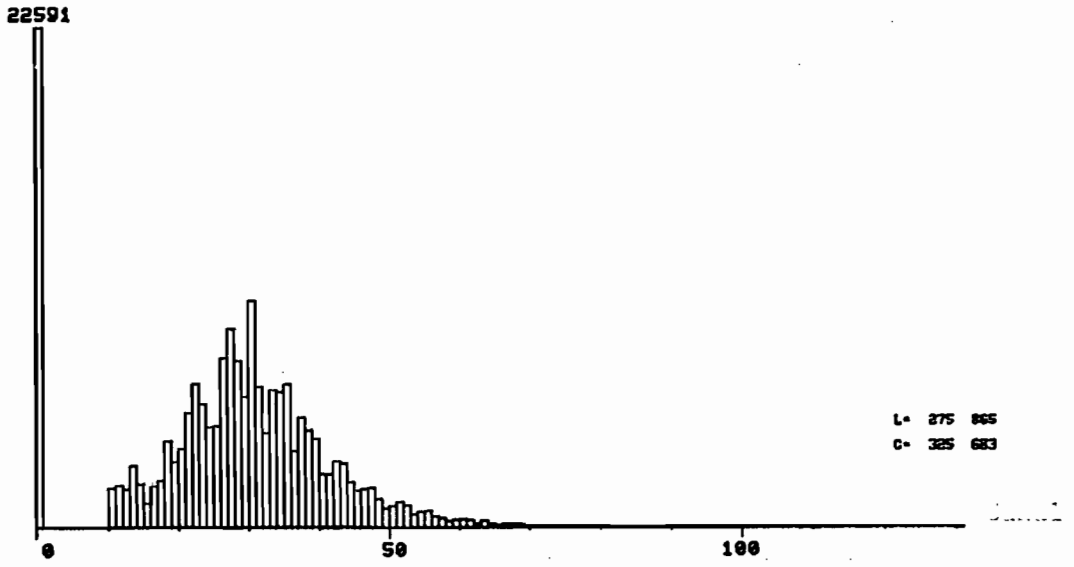
- A l'intérieur de ces étendues réduites des six séries statistiques, *le degré de concentration des valeurs est élevé.* On observe toutefois des différences sensibles selon les canaux : dans le canal 4, 70 % des pixels sont concentrés sur 10 valeurs radiométriques autour du mode (valeurs 20 à 30), alors que 49 % des pixels sont compris dans un intervalle équivalent (30 à 40) dans le canal 5.

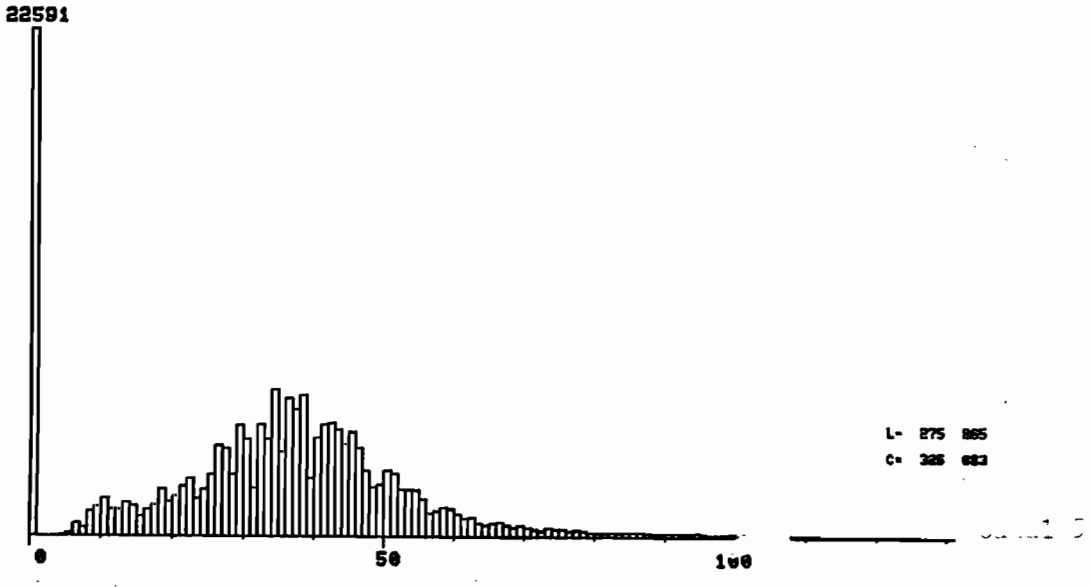
- *Globalement, les distributions sont unimodales et relativement symétriques, à l'exception du canal 5.* Il n'apparait pas de coupures nettes dans les histogrammes, mais seulement des minima locaux plus ou moins marqués selon les canaux.

Ces trois remarques sont à mettre en relation avec la date d'enregistrement de l'image, et l'origine de cette image. Il faut également souligner que le fait d'avoir masqué l'image fait disparaitre la partie gauche des histogrammes avec son mode : c'est le masque qui crée le caractère unimodal des distributions.
L'image a été enregistrée en *Février* : l'activité chlorophyllienne est peu importante à cette date de l'année, ce qui contribue à la faible dynamique de l'image que l'on vient de souligner.
D'autre part, il est fort probable que la station de traitement des images joue un rôle important dans la forme des séries statistiques des valeurs radiométriques : les valeurs radiométriques enregistrées par le satellite sont soumises à des

Figure 1 - Histogrammes unifimensionnels. Image TM1 Février 1983. Marseille.




22591
L= 275 865
C= 325 683
0
50
100


22591
L= 275 865
C= 325 683
0
50
100

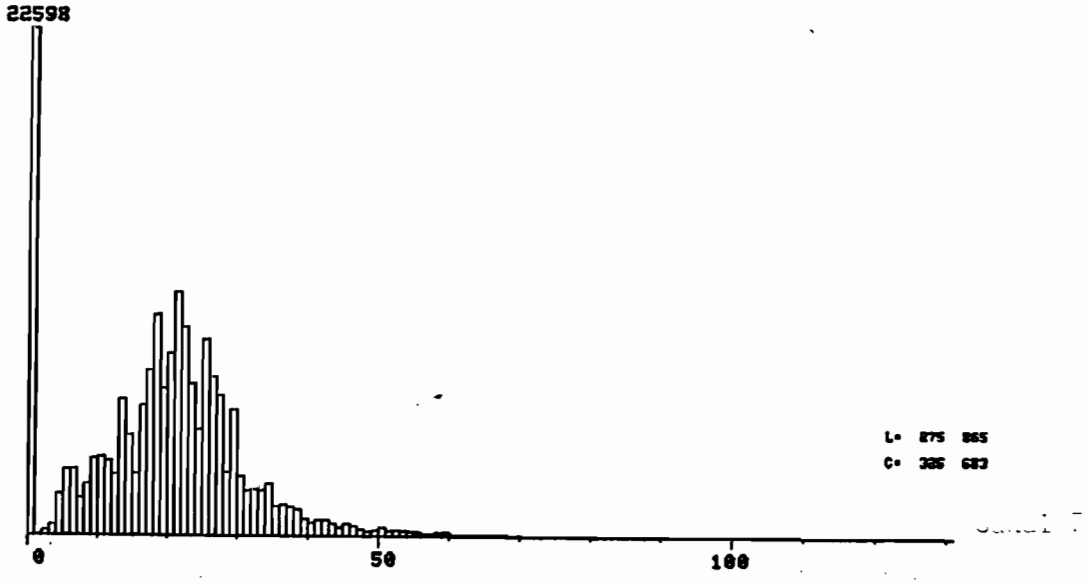

22598
L= 275 865
C= 325 683
0
50
100

Tableau 1- Analyse en composantes principales
Image TM Février 1983 Marseille.

MATRICE DES CORRELATIONS

(TOUS LES COEFFICIENTS SONT MULTIPLIES PAR 1000)

	CAN1	CAN2	CAN3	CAN4	CAN5	CAN7
CAN1	1000					
CAN2	950	1000				
CAN3	947	970	1000			
CAN4	646	740	754	1000		
CAN5	770	814	837	851	1000	
CAN7	816	847	876	761	945	1000

LES VALEURS PROPRES VAL(1)= 5.18262

NUM	ITER	VAL PROPRE	POURCENT	CUMUL	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
1	0	5.18262	86.377	86.377	**********
2	1	0.48216	8.036	94.412	******
3	1	0.22479	3.747	98.159	***
4	1	0.05092	0.849	99.008	*
5	1	0.03306	0.551	99.559	*
6	4	0.02647	0.441	100.000	*

	J1	GLT	POID	INR	1#F	COR	CTR	2#F	COR	CTR	3#F	COR	CTR	4#F	COR	CTR	5#F	COR	CTR	6#F	COR	CTR
1	CAN1	1000	0	167	923	852	164	346	120	249	-40	2	7	147	21	422	-65	4	129	-27	1	28
2	CAN2	1000	0	167	957	915	177	226	51	106	-107	11	51	-66	4	85	117	14	414	-67	4	169
3	CAN3	1000	0	167	968	937	181	190	36	75	-62	4	17	-91	8	162	-32	1	32	119	14	530
4	CAN4	1000	0	167	848	719	139	-438	192	398	-296	87	388	1	0	0	-42	2	54	-24	1	21
5	CAN5	1000	0	167	935	874	169	-269	72	150	163	34	149	99	10	194	86	7	226	54	3	112
6	CAN7	1000	0	167	942	887	171	-103	11	22	295	87	388	-84	7	137	-69	5	145	-60	4	137
				1000			1000			1000			1000			1000			1000			1000

traitements visant à corriger les biais d'enregistrement, mais aussi à des *traitements visant à améliorer la dynamique de l'image*. Selon la nature et l'ampleur des étalements effectués, variables selon les pratiques des différentes stations, les caractéristiques statistiques des valeurs radiométriques des images livrées au public varient. *Le degré de dispersion des valeurs radiométriques ne devrait donc pas être interprété de façon systématique en termes de qualité de l'image*: des étalements faibles comme ceux que nous observons ici ne signifient pas forcément que cette image de Marseille n'est pas bonne, mais plutôt qu'elle n'a pas sans doute pas fait l'objet d'amélioration de la dynamique.
Ce second point nous pousse à une grande prudence dans les études multi-temporelles, et dans l'utilisation des méthodes d'interprétation (méthodes de classification notamment) reposant, comme c'est quasi-systématiquement le cas, sur des calculs de variance: l'expérience de la typologie de la morphologie urbaine (voir chapitre 2. 2.) à partir des descripteurs observés sur photographies aériennes nous a montré l'influence très importante du nombre de modalités de chacune des variables sur les résultats de la classification.

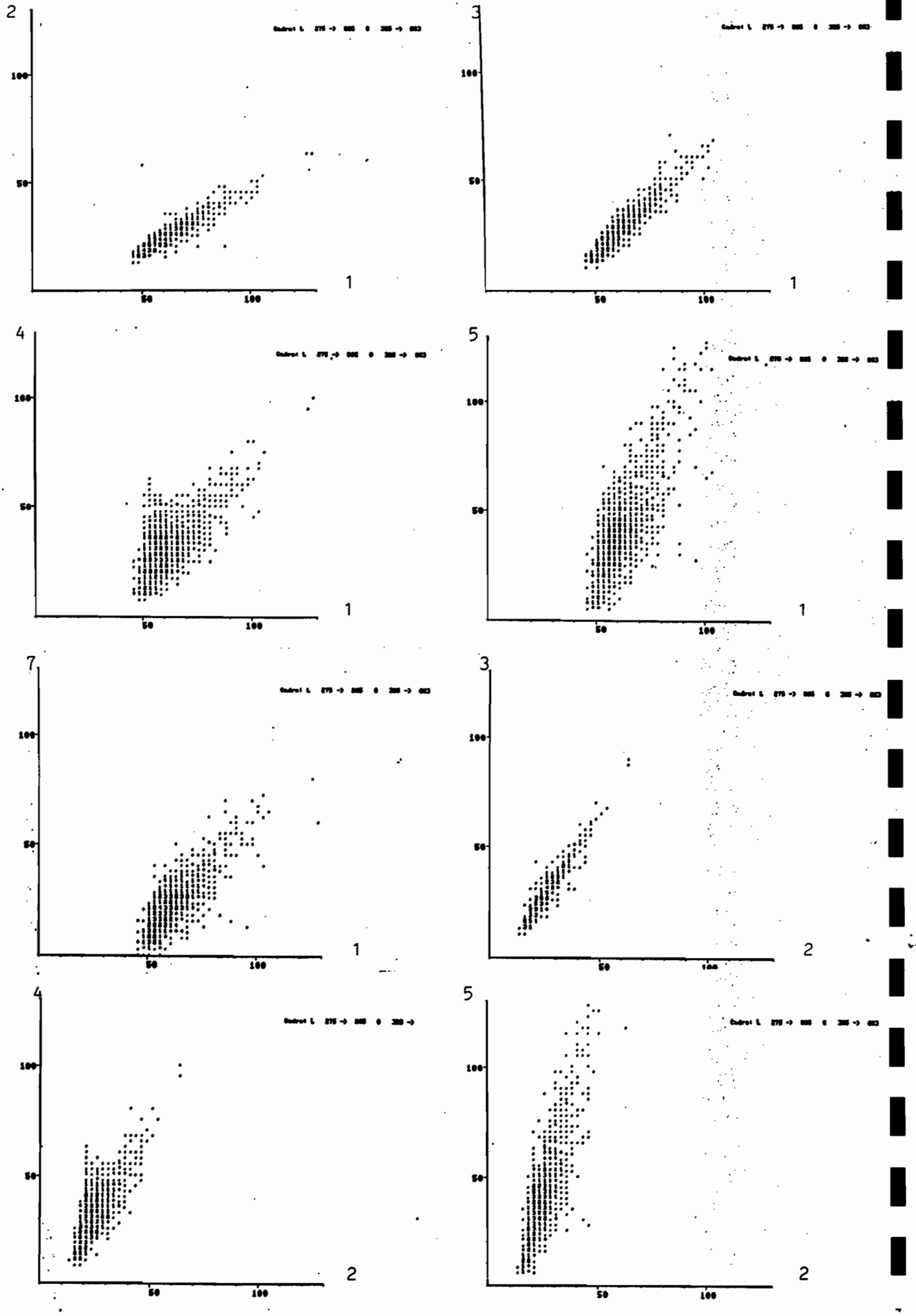
5. 3. 2. Corrélations entre canaux

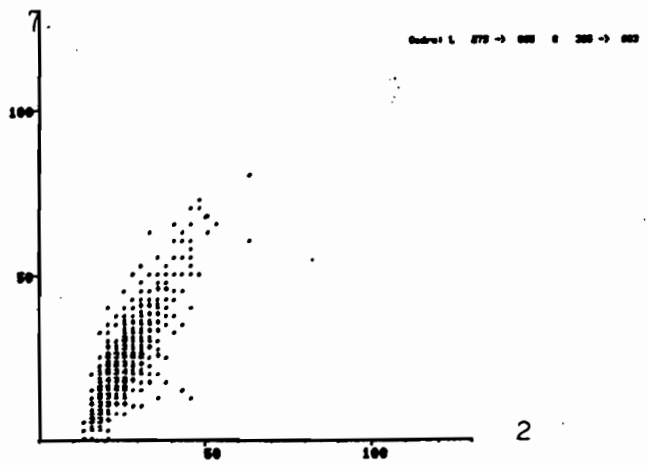
La figure 2 présente les histogrammes bidimensionnels entre les 6 canaux de l'image TM 1983 : le nombre de points est représenté par un signe conventionnel. Au vu de ces graphiques et de la matrice de corrélation (tableau 1), plusieurs remarques peuvent être faites :

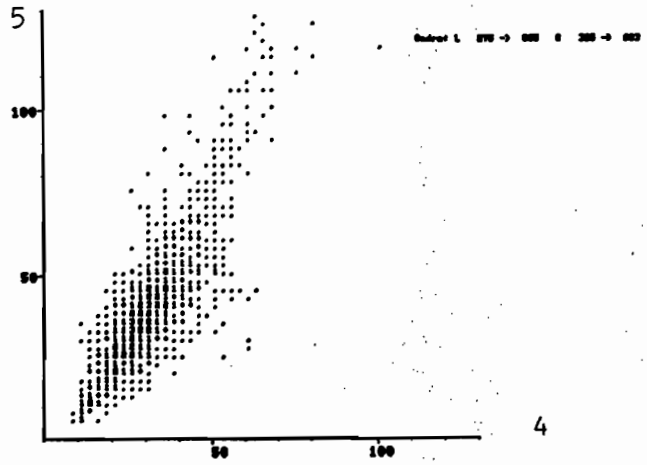
- Il existe un *fort niveau de corrélation entre les différents canaux de l'image TM* . aucun canal n'apparait indépendant d'un autre, les coefficients de corrélation sont toujours supérieurs à 600.
- *Les trois canaux du visible (1, 2, 3) sont très fortement corrélés*: leurs coefficients de corrélation varient entre 947 et 950.
- *Les canaux du visible (1, 2, 3) sont moins corrélés avec le canal 4* pourtant le plus proche spectralement, (coefficients varaiant entre 646 et 754), *qu'avec le canal 5*(770 à 837) *et surtout le canal 7*(816 à 876).
- *Les canaux 4 et 7 sont relativement peu corrélés*; leur coefficient étant de 761.

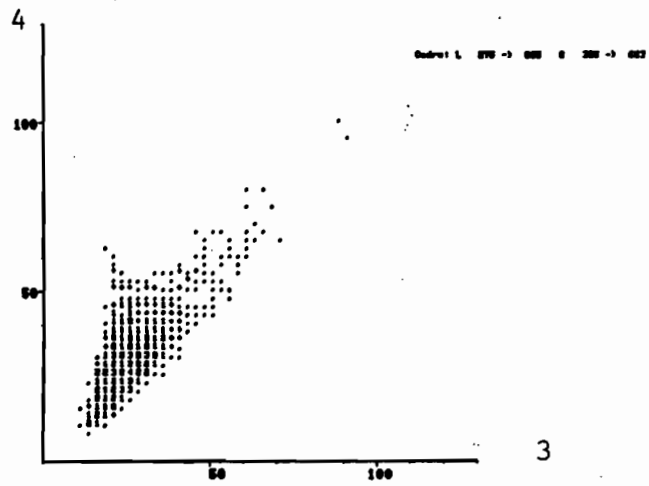
Il est intéressant de comparer ces résultats à ceux présentés dans trois articles où les auteurs ont utilisé des images TM en milieu urbain ou suburbain : les résultats de cette comparaison sont consignés dans le tableau 2.
Ce tableau met en évidence une forte cohérence des quatre matrices de corrélation si l'on raisonne sur le niveau relatif des coefficients de corrélation entre couples de canaux : *quelque soit le site étudié, les canaux du visibles sont les plus corrélés entre eux, le canal 4 est le moins corrélé à tous les autres canaux, et les canaux 4 et 7 montrent une très faible corrélation* (sauf dans l'article de QUATTROCHI).
Si l'ordre des niveaux de corrélation entre canaux paraît stable, *les coefficients de corrélation eux-mêmes présentent des valeurs très différentes* : TOLL (1984) et QUATTROCHI (1984) observent des coefficients de corrélation qui descendent jusqu'à, respectivement, 110 et 230 , tandis que le coefficient minimum observé par TOLL (1985) est du même ordre que celui de Marseille, 570 (646 à Marseille).

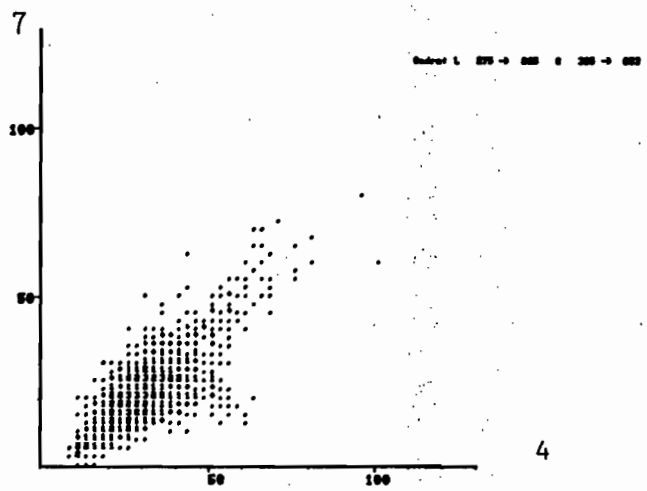
Figure 2 - Histogrammes bidimensionnels . Image TM Février 1983. Marseille.

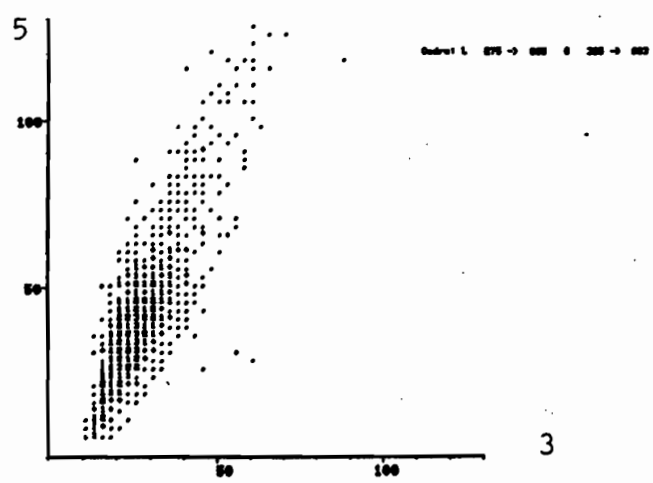


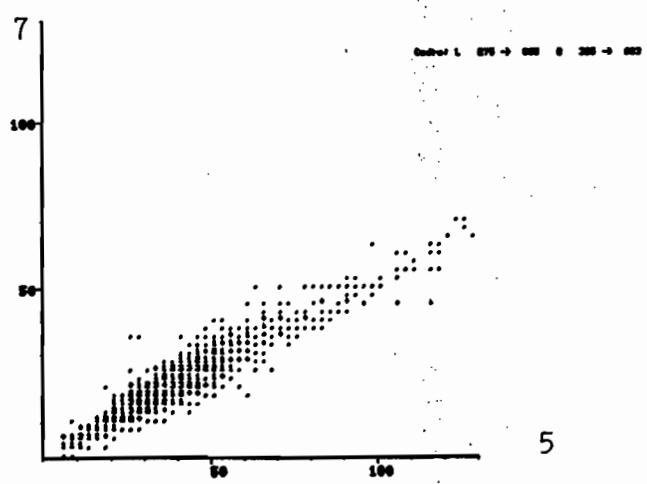

7
100
50
50
100
2


5
100
50
50
100
4


4
100
50
50
100
3


7
100
50
50
100
4


5
100
50
50
100
3


7
100
50
50
100
5

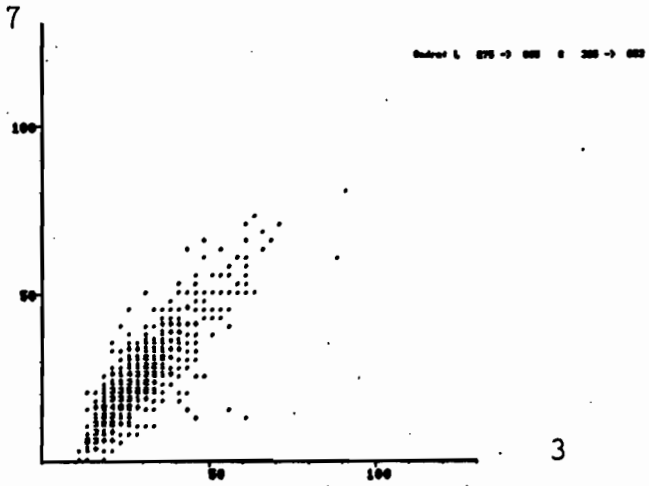

7
100
50
50
100
3

Tableau 2 - Coefficients de corrélation des canaux TM

CANAUX	Rang du coef. Ma.	T85	T84	QU	MARSEILLE	TOLL 85	TOLL 84	QUATTROCHI
2 - 3	1	3	1	1	970	890	970	950
1 - 2	2	1	3	3	950	960	950	940
1 - 3	3	7	1	4	947	830	970	890
5 -7	4	2	4	1	945	950	930	950
3 - 7	5	5	5	5	876	840	900	880
4 -5	6	11	11	6	851	700	300	810
2 - 7	7	4	7	7	847	870	840	780
3 - 5	8	9	8	8	837	790	820	770
1 - 7	9	5	6	9	816	840	870	680
2 - 5	10	8	10	10	814	810	730	650
1 - 5	11	10	9	12	770	760	780	530
4 - 7	12	14	15	10	761	600	110	650
3 - 4	13	12	13	13	754	650	150	470
2 - 4	14	13	12	14	740	620	170	370
1 - 4	15	15	14	15	646	570	120	230

Comment interpréter cette double observation, constance de l'ordre des coefficients de corrélation, et variation importante des valeurs prises par les coefficients de corrélation?
Trois hypothèses peuvent être avancées. D'une part, les images analysées par TOLL (1984) et QUATTROCHI (1984) ont été enregistrées respectivement en Juin et Octobre, à des dates où la végétation est plus importante que dans le cas de Marseille (Février) ou de TOLL (1985) (Novembre). D'autre part, les sites comportent une proportion plus ou moins importante de zone non urbanisée : d'où une variabilité des distributions statistiques des valeurs radiométriques dans chacun des canaux , qui peut conduire directement à une augmentation de la variance dans chaque canal, et à une diminution des corrélations entre canaux. Enfin, il est probable que toutes les images n'ont pas fait l'objet des mêmes traitements initiaux sur la radiométrie.

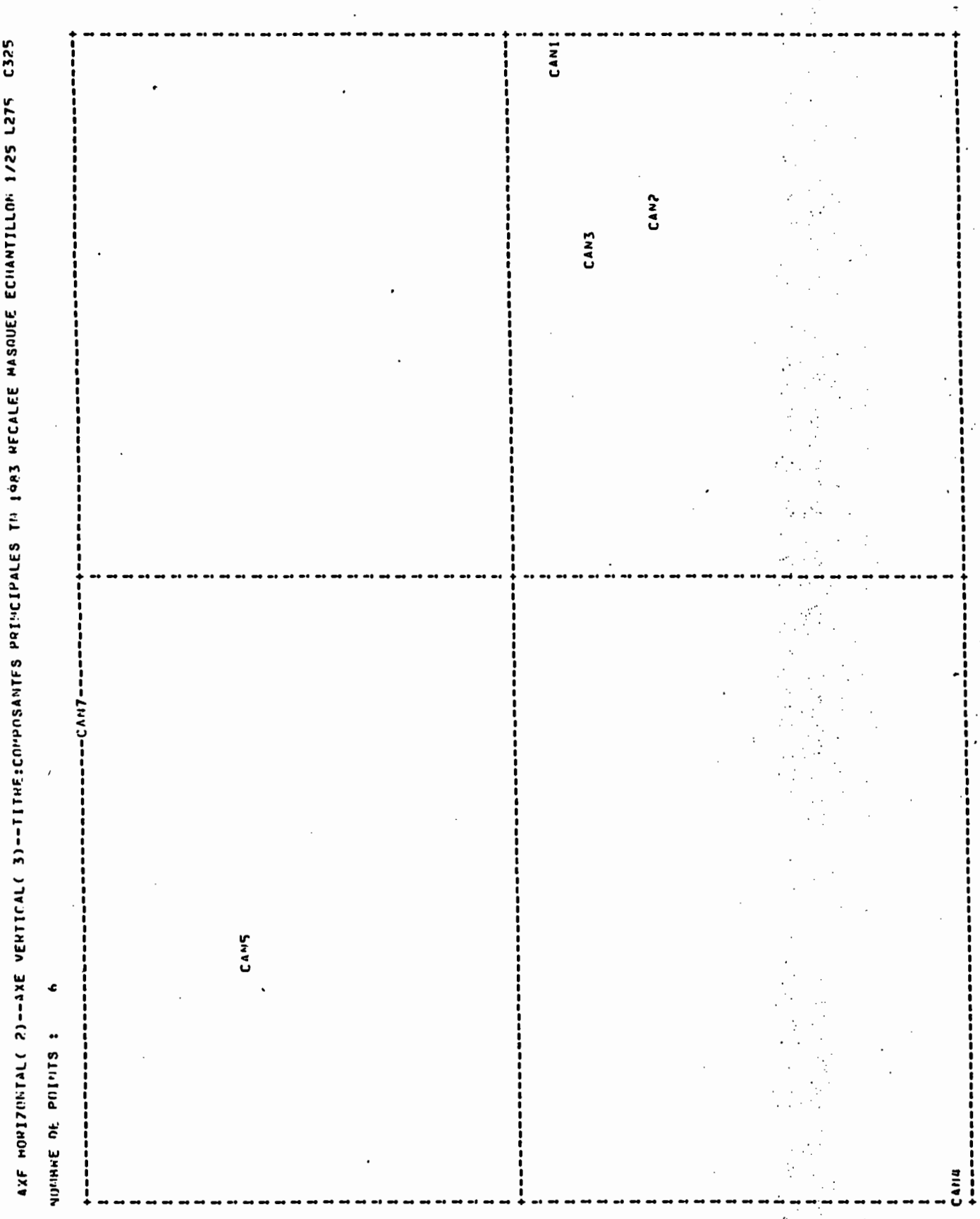
Cette comparaison des résultats sur Marseille avec ceux exposés dans la littérature scientifique renforce la conclusion du 5.3.2. : *il ne faut pas considérer le niveau absolu des variances intra canal, ou des corrélations entre canaux, mais leur valeur relative, c'est-à-dire l'ordre des niveaux de variance ou de corrélation des canaux.*

5 . 3 . 3. Analyse en composantes principales

L'objectif de cette Analyse en Composantes Principales (ACP) est d'avoir des informations sur la structure des données, grâce à la matrice de corrélation, que nous venons d'analyser, et à la détermination des axes principaux.
Le fichier soumis à l'analyse correspond à la fenêtre d'étude indiquée au début de ce chapitre. Après avoir transformé le fichier image masqué en fichier séquentiel, nous avons extrait 1 pixel sur 25 par tirage systématique (avec un pas constant) sur le fichier séquentiel ; du fichier séquentiel masqué et échantillonné, ont ensuite été éliminés tous les pixels ayant été masqués à 0. A l'issue de ces transformations de fichiers, nous sommes donc en possession d'un fichier séquentiel échantillon comprenant 6688 pixels répartis sur toute la fenêtre (L 275-865, C 325-683), à l'exception de la mer et des quelques autres zones masquées (zones d'ombre au sud de l'image).
Le logiciel utilisé pour l'ACP est ANCOMP de l'ADDAD (logiciel utilisé au CIRCE): avant l'algorithme d'ACP proprement dit, ce programme centre et réduit les séries statistiques soumises à l'analyse.

Le tableau 1 présente les principaux résultats de l'ACP : matrice des corrélations, tableau des valeurs propres et tableau des coordonnées et contributions des canaux aux composantes principales.
Première observation : trois axes seulement suffisent à expliquer 98,159 % de la variance totale, le premier axe expliquant à lui seul 86,377 % de l'inertie du nuage, le second 8,036 % et le troisième 3,747 %. Ce résultat est à rapprocher de celui obtenu par ANUTA (1984) :
- Axe 1 = 71,84 %, Axe 2 = 18,24 %, Axe 3 = 7,38 %, Axe 4 = 1,05 %.
- % variance expliquée par axe 1,2 et 3 = 97,45 %.

Figure 3 - Analyse en composantes principales. Représentation des canaux sur le plan 2 x 3. Image TM Février 1983. Marseille.



Dans les deux cas par conséquent, *l'axe 1 apparait absolument prépondérant et trois axes suffisent à expliquer environ 98 % de la variance* ; le % de variance expliquée par le premier axe à Marseille est toutefois plus élevé.

Si l'on s'intéresse maintenant à la signification des composantes principales (tableau 1 et figure 3), on observe les faits suivants :

- *tous les canaux contribuent fortement à la formation de la première composante principale* : seul le canal 4 se distingue légèrement, avec des coordonnées plus faibles (848, alors que les autres canaux ont des coordonnées >920) et une contribution moins élevée (139) que les autres canaux (>164).
- *La deuxième composante principale oppose les canaux du visible* (coordonnées positives) aux autres canaux (coordonnées négatives), *et plus particulièrement le canal 1* (coordonnée 346 et contribution 249) *au canal 4* (coordonnée -438 et contribution 398).
- Les canaux du visible contribuent peu à la formation de la troisième composante principale (contributions <51) : cette composante rend compte d'une opposition entre le canal 4 (coordonné -296, contribution 388) et le canal 7 (coordonnée 295, contribution 388), et, à un moindre dégré le canal (coordonnée 183, contribution 149). *Il s'agit donc, pour cette troisième composante principale, d'une opposition de l'infra-rouge proche (4) aux deux canaux de l'infra-rouge moyen (5 et 7).*

Si l'on désire créer un néo-canal conservant l'essentiel de la variance des 6 canaux TM, la première composante principale peut donc être utilisée ; la combinaison 1-4-7, ou 1-5-7, qui en termes de varaince est la plus discriminante, devra être testée sur le plan de sa signification thématique.

2 . LES TRAVAUX PREPARATOIRES SUR QUITO

Outre les travaux réalisés sur Marseille, l'équipe a préparé la deuxième phase du programme de recherche , qui doit porter sur la ville de Quito (EQUATEUR) :

1) En mai 1985, M. SOURIS et F. DUREAU ont effectué une mission à Quito afin de :
- Evaluer les possibilités d'équipement informatique des organismes équatoriens participant au projet d'Atlas de Quito, et proposer des configurations matérielles compatibles avec les besoins exprimés
- Evaluer les collaborations équatoriennes en télédétection.

2) En décembre 1985, M. SOURIS s'est rendu en Equateur à la demande des autorités équatoriennes, afin de présenter ses propositions d'équipemement informatique pour l'Institut Géographique Militaire.

Dans cette deuxième partie, nous ferons un bilan rapide de ces travaux préparatoires, qui permettent de préciser les conditions de travail à Quito et les objectifs de recherche qui pourront être développés.

Il faut souligner qu'une analyse précise des données du recensement de 1982 a été réalisée par D. COURET, allocataire MRT affectée à Quito par le Département D de l'ORSTOM : "Recensement de la population et logements"(Quito, septembre 1985).

1 . ANALYSE DE L'INFORMATION EXISTANTE

1.1. Image satellite

En réponse à notre demande de Février 1986 auprès d'EOSAT (USA), nous avons appris qu'il n'existe toujours pas au catalogue d'image satellite *Thematic Mapper* sur la ville même de Quito.
Nous allons donc commander l'enregistrement d'une image Thematic Mapper sur la ville, en passant par le GDTA.

Nous avons soumis en Juin 1984 un projet PEPS sur le thème "L'utilisation de la télédétection pour l'observation démographique en milieu urbain". Ce projet, non retenu par Spot Image dans le cadre du PEPS, a été sélectionné par le CNES, qui finance donc quatre scènes *SPOT* sur la ville de Quito, soit les deux scènes nécessaires pour couvrir l'ensemble de la ville, en panchromatique et multispectral, visée verticale. L'enregistrement a été demandé pour la période août-novembre 1986.
L'acquisition d'images en visée oblique, initialement demandée dans le PEPS, repose sur un financement ORSTOM, du fait que deux scènes sont nécessaires pour couvrir toute la ville de Quito.

Sauf impossibilité technique d'enregistrement, nous devrions donc disposer, au deuxième semestre 1986, d'images SPOT et T.M. sur la ville de Quito.

1.2. Photographies aériennes

Cinq couvertures aériennes sont disponibles sur l'ensemble de la ville de Quito :

- 1956, 1963, 1976 : couverture à moyenne échelle (1/60.000)
- 1981 : couverture à grande échelle (1/20.000)
- 1984 : couverture à très grande échelle (1/5.000 agrandie au 1/1.000)

La ville de Quito est donc bien couverte, exhaustivement, en photographies aériennes à une échelle adéquate pour analyser précisément la morphologie urbaine.
L'accès à ces photographies aériennes ne devrait pas poser de problèmes, puisqu'elles sont propriété de l'IGM, institution membre de l'APIQ. Mais, vu la charge de travail du laboratoire photographique de l'IGM, il faudra faire réaliser les traitements particuliers (agrandissements, mosaïques...etc) à l'extérieur de l'IGM.

1.3. Données démographiques

Les données démographiques exhaustives les plus récentes sont issues du recensement de la population, réalisé en novembre 1982. Auparavant, des

recensements avaient déjà été réalisés en 1963 et en 1974. C'est l'INEC, institut équatorien de la statistique qui est responsable de tous ces recensements. Seul le fichier informatisé du recensement de 1982 est disponible : les recensements antérieurs à 1974 n'ont pas été exploités informatiquement; et le fichier du recensement de 1974 n'est plus disponible à l'heure actuelle, pour des raisons de maintenance.

Notre mission nous a permis d'éclaircir plusieurs points importants à propos du recensement de 1982 :

-Cartographie des aires de recensement : les plans de Quito ont été dressés à partir des cartes topographiques de l'IGM au 1/10.000, réactualisées par relevés terrain. La voirie et les différents découpages géographiques afférant au recensement figurent sur ces plans. Le zonage le plus fin correspond à la "manzana", c'est à dire le pâté de maisons.
Les cartes disponibles en mai 1985 correspondent à la cartographie élaborée avant le recensement : des corrections doivent y être portées par l'INEC pour tenir compte des quelques modifications apparues au moment même de la collecte de données démographiques sur le terrain.
La qualité de ces plans est tout à fait bonne : contrairement aux plans fournis par l'INSEE, ils ne nécessitent a priori pas de nouveau tracé avant digitalisation.

- Fichier des résultats : Le recensement de 1982 a été exploité informatiquement par l'INEC, qui pourra nous fournir une copie de la bande magnétique des résultats par manzana et/ou une copie de la bande magnétique du fichier individuel, qui correspond à la saisie directe des questionnaires (l'accord relatif à l'obtention de cette bande ayant été signé par l'INEC et l'IGM en février 1986, nous devrions avoir une copie à Paris dans des délais assez brefs). Sur ce second fichier, le numéro de manzana a été saisi, et permet donc une exploitation finement localisée de l'information, tous les croisements entre variables étant rendus possibles par l'accès à l'information brute relative aux individus.

Outre l'intérêt du questionnaire (voir pages suivantes), qui comprend de *nombreuses variables sur le logement et les caractéristiques démographiques et socio-économiques des individus,* l'information du recensement de 1982 présente l'avantage de pouvoir être localisée avec une assez grande précision, grâce aux identifiants d'ilôts. Le recensement présente à ce niveau toutes les qualités nécessaires à son exploitation dans le cadre de notre programme de recherche.

Il existe néanmoins *deux limites à son utilisation* :
-*sa date de réalisation* : trois ans et demi se seront écoulés entre le recensement et l'enregistrement satellitaire. De ce fait, les estimations de population issues de l'enquête par sondage réalisée par tirage d'un échantillon spatial sur image satellite ne pourront pas être comparées directement aux résultats du recensement ; leur pertinence ne pourra donc être évaluée que par comparaison avec des projections de population (sous forme de "fourchettes") basées sur les chiffres du recensement de 1982. Le recensement de Quito ne sera donc pas utilisé de la même façon que celui de Marseille, pour lequel on a concordance de date avec l'image satellite.

- *sa fiabilité :* des évaluations réalisées par D. DELAUNAY (économiste ORSTOM, Quito) et D. COURET, il ressort que le recensement aurait sous-estimé la population de fait de la ville de Quito : date du recensement et comportement de la population quiténienne auraient contribué à sous-estimer la population de fait à Quito. Pour les responsables de l'INEC, il s'agirait plutôt d'une surestimation de la population de droit, les enquêteurs ayant eu tendance à indiquer systématiquement le chef-lieu comme lieu de résidence habituelle (il s'agit d'un biais classique des recensements démographiques). C'est pourquoi l'INEC a choisi d'exploiter le recensement sur la population de fait, et non la population de droit comme on le fait généralement.
Les chiffres du recensement comportent donc très certainement un biais et fournissent une estimation *minimale* de la population quiténienne.

Questionnaire du recensement de la population - Equateur - 1982

VICEPRESIDENCIA DE LA REPUBLICA

"LA INFORMACION SOLICITADA ES ESTRICTAMENTE CONFIDENCIAL"

INSTITUTO NACIONAL DE ESTADISTICA Y CENSOS

REPUBLICA DEL ECUADOR

IV CENSO DE POBLACION Y III DE VIVIENDA

14 de Noviembre de 1982

Nº DE VIVIENDA (Orden de visita) .. - DE -

I.- UBICACION GEOGRAFICA

1.- PROVINCIA ____

2.- CANTON ____

3.- PARROQUIA ____

4.- NOMBRE DE LA CIUDAD O CABECERA PARROQUIAL ____

5.- NOMBRE DE LA LOCALIDAD ____

6.- ZONA Nº ____ 7.- SECTOR Nº ____ 8.- MANZANA Nº ____

9.- AREA DE EMPADRONAMIENTO Nº ____

10.- DIRECCION DOMICILIARIA:

CALLE Y Nº ____

LOTE Nº ____

PISO Nº ____ DEPARTAMENTO Nº ____

CAMINO O CARRETERA ____

II.- VIVIENDA

1.- TIPO DE VIVIENDA

PARTICULAR		COLECTIVA	
Casa o Villa	☐ 01	Hotel, Pensión o Residencial	☐ 11
Departamento	☐ 02	Cuartel Militar o de Policía	☐ 12
Cuarto en casa de Inquilinato	☐ 03	Hospital, Clínica, etc.	☐ 13
Mediagua	☐ 04	Convento o Institución Religiosa	☐ 14
Rancho o Covacha	☐ 05	Otros	☐ 15 16
Choza	☐ 06	SI LA VIVIENDA ES COLECTIVA	
Otros	☐ 07	PASE A PREGUNTAS	
Locales no destinados para Vivienda	☐ 08	DE POBLACION	

2.- CONDICION DE OCUPACION DE LA VIVIENDA

- Ocupada con personas presentes ☐ 1
- Ocupada con personas ausentes ☐ 2
- Desocupada ☐ 3
- En construcción ☐ 4

3.- MATERIALES PREDOMINANTES EN LA VIVIENDA

A.- TECHO O CUBIERTA

- Losa de hormigón ☐ 1
- Eternit, Ardex o similares ☐ 2
- Zinc o similares ☐ 3
- Teja ☐ 4
- Paja o similares ☐ 5
- Otros materiales ☐ 6

B.- PAREDES EXTERIORES

- Hormigón, ladrillo o bloque ☐ 1
- Adobe o tapia ☐ 2
- Madera ☐ 3
- Caña revestida ☐ 4
- Caña no revestida ☐ 5
- Otros materiales ☐ 6

C.- PISO

- Entablado ☐ 1
- Parquet, Baldosa o Vinil ☐ 2
- Ladrillo o cemento ☐ 3
- Caña ☐ 4
- Tierra ☐ 5
- Otros materiales ☐ 6

4.- TENENCIA DE LA VIVIENDA

¿ LA VIVIENDA ES ? :

- Propia ☐ 1
- Arrendada ☐ 2
- Gratuita ☐ 3
- Por servicios ☐ 4
- Otros (Ej. Anticresis) ☐ 5

5.- EL ABASTECIMIENTO DE AGUA EN LA VIVIENDA ES :

¿ DE RED PUBLICA ?		¿ DE OTRAS FUENTES ?	
Dentro de la vivienda	☐ 1	Pozo o vertiente	☐ 4
Fuera de la vivienda pero en el edificio	☐ 2	Río o acequia	☐ 5
Fuera del Edificio	☐ 3	Carro repartidor	☐ 6
		Otro	☐ 7

6.- ¿ QUE DISPONIBILIDAD DE SERVICIO HIGIENICO (W. C.) TIENE LA VIVIENDA ? :

- Excusado de uso exclusivo de la vivienda ☐ 1
- Excusado de uso común ☐ 2
- Letrina ☐ 3
- Ninguno ☐ 4

7.- ¿ EL SISTEMA DE ELIMINACION DE AGUAS SERVIDAS DE LA VIVIENDA ES ? :

- Por red pública de alcantarillado ☐ 1
- Pozo ciego ☐ 2
- Ninguno ☐ 3

8.- ¿EL SERVICIO ELECTRICO QUE DISPONE LA VIVIENDA ES ? :

- De red pública ☐ 1
- De planta privada ☐ 2
- Ninguno ☐ 3

9.- ¿CUANTOS CUARTOS O PIEZAS EN TOTAL TIENE LA VIVIENDA ?

(No incluya cocina ni baño) ____

10.- ¿DE ESTOS, CUANTOS SE UTILIZAN SOLO PARA DORMIR ? ____

11.- ¿TIENE LA VIVIENDA CUARTO DE COCINA INDEPENDIENTE ?

SI ☐ 1

NO ☐ 2

12.- ¿CUAL ES EL PRINCIPAL COMBUSTIBLE QUE SE UTILIZA PARA COCINAR ? :

- Kerex o Gasolina ☐ 1
- Leña o Carbón ☐ 2
- Gas ☐ 3
- Electricidad ☐ 4
- Otro ☐ 5
- Ninguno (No cocina) ☐ 6

13.- ¿SE UTILIZA PARTE DE LA VIVIENDA PARA ARTESANIA, PEQUEÑA INDUSTRIA, COMERCIO U OTRAS ACTIVIDADES ?

- Sólo Artesanía ☐ 1
- Sólo Pequeña Industria ☐ 2
- Sólo Comercio ☐ 3
- Comercio y Artesanía ☐ 4
- Comercio y Pequeña Industria ☐ 5
- Otras actividades ☐ 6
- Sólo vivienda ☐ 7

RESUMEN DE POBLACION

Nº DE HOMBRES ____

Nº DE MUJERES ____

TOTAL ____

____ NOMBRE DEL EMPADRONADOR

____ FECHA

____ NOMBRE DEL JEFE DE SECTOR

____ FECHA

OBSERVACIONES: ____

III. POBLACION

PREGUNTAS	Nº	PRIMERA PERSONA	SEGUNDA PERSONA	TERCERA PERSONA

A. Características generales. PARA TODAS LAS PERSONAS

¿CUAL ES EL NOMBRE Y APELLIDO DE CADA UNA DE LAS PERSONAS QUE PASARON LA NOCHE DEL ... DE NOVIEMBRE EN ESTE LUGAR? No olvide anotar a los recién nacidos y a los ancianos.	1			
¿QUE PARENTESCO O RELACION TIENE CON EL JEFE DEL HOGAR? Marque la casilla correspondiente	2	JEFE ☐ 0	Cónyuge ☐ 1; Padres o Suegros ☐ 5; Hijo/a ☐ 2; Otros parientes ☐ 6; Yerno o Nuera ☐ 3; Otros no parientes ☐ 7; Nieto/a ☐ 4; Empleado/a doméstico/a ☐ 8	Cónyuge ☐ 1; Padres o Suegros ☐ 5; Hijo/a ☐ 2; Otros parientes ☐ 6; Yerno o Nuera ☐ 3; Otros no parientes ☐ 7; Nieto/a ☐ 4; Empleado/a doméstico/a ☐ 8
¿ES HOMBRE O MUJER? (Marque la casilla correspondiente)	3	HOMBRE ☐ 1 MUJER ☐ 2	HOMBRE ☐ 1 MUJER ☐ 2	HOMBRE ☐ 1 MUJER ☐ 2
¿CUANTOS AÑOS CUMPLIDOS TIENE? Anote la edad de su último cumpleaños. Para menores de un año anote "00".	4	Años cumplidos	Años cumplidos	Años cumplidos
¿DONDE NACIO? Si nació en la parroquia rural o ciudad donde es censado marque la casilla "AQUI", en caso contrario, anote el nombre de la parroquia rural o ciudad y provincia. Si nació en el extranjero, anote sólo el país.	5	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99
¿EN QUE PARROQUIA RURAL O CIUDAD VIVE HABITUALMENTE? Si el lugar donde vive habitualmente es la parroquia rural o ciudad donde es censado marque la casilla "AQUI" en caso contrario anote el nombre de la parroquia rural o ciudad y provincia. Si vive en el extranjero, anote sólo el país.	6	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Aquí ☐ 00; Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99
¿CUANTO TIEMPO VIVE EN EL LUGAR INDICADO? Anote el tiempo que vive en el lugar indicado en la pregunta 6. Si vive menos de un año anote "00"	7	Siempre ☐ 98 → Pase a pregunta 9; Nº de años; Se ignora ☐ 99	Siempre ☐ 98 → Pase a pregunta 9; Nº de años; Se ignora ☐ 99	Siempre ☐ 98 → Pase a pregunta 9; Nº de años; Se ignora ☐ 99
¿EN QUE PARROQUIA RURAL O CIUDAD VIVIO ANTES? Anote el nombre de la parroquia rural o ciudad y provincia. Si vivió en el extranjero, anote solo el país.	8	Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99	Parroquia rural o ciudad; Provincia; País extranjero; Se ignora ☐ 99

B. Características educacionales. PARA TODAS LAS PERSONAS DE 6 AÑOS Y MAS

¿SABE LEER Y ESCRIBIR? (Si sólo lee o solo escribe marque la casilla "NO")	9	SI ☐ 1 NO ☐ 2	SI ☐ 1 NO ☐ 2	SI ☐ 1 NO ☐ 2
¿ASISTE A ALGUN ESTABLECIMIENTO DE ENSEÑANZA REGULAR (CENTRO DE ALFABETIZACION, PRIMARIA, SECUNDARIA O SUPERIOR)?	10	Si asiste ☐ 1; No asiste ☐ 2; Se ignora ☐ 9	Si asiste ☐ 1; No asiste ☐ 2; Se ignora ☐ 9	Si asiste ☐ 1; No asiste ☐ 2; Se ignora ☐ 9
¿CUAL ES EL ULTIMO GRADO, CURSO O AÑO QUE APROBO EN EL CENTRO DE ALFABETIZACION, PRIMARIA, SECUNDARIA O SUPERIOR? Anote en la línea correspondiente el número del último grado, curso o año aprobado.	11	Ninguno ☐ 00; Centro de Alfabetización 1; Primaria 2; Secundaria 3; Superior 4; Se ignora ☐ 99	Ninguno ☐ 00; Centro de Alfabetización 1; Primaria 2; Secundaria 3; Superior 4; Se ignora ☐ 99	Ninguno ☐ 00; Centro de Alfabetización 1; Primaria 2; Secundaria 3; Superior 4; Se ignora ☐ 99

C. Características económicas. PARA TODAS LAS PERSONAS DE 12 AÑOS Y MAS

¿ES O FUE AFILIADO AL SEGURO SOCIAL?	12	SI ☐ 1 → Aporta actualmente (SI ☐ 3, NO ☐ 4); NO ☐ 2; Se ignora ☐ 9	SI ☐ 1 → Aporta actualmente (SI ☐ 3, NO ☐ 4); NO ☐ 2; Se ignora ☐ 9	SI ☐ 1 → Aporta actualmente (SI ☐ 3, NO ☐ 4); NO ☐ 2; Se ignora ☐ 9
¿QUE HIZO LA MAYOR PARTE DE LOS DIAS COMPRENDIDOS EN TRE EL 7 AL 13 DE NOVIEMBRE? Haga las preguntas en el orden indicado: ¿TRABAJO?, ¿NO TRABAJO PERO TENIA TRABAJO?, ETC.; al recibir una respuesta, marque la casilla respectiva y pase a la pregunta que corresponda.	13	¿Trabajo? ☐ 0; ¿No trabajo pero tenía trabajo? ☐ 1; ¿Busco trabajo habiendo trabajado antes? ☐ 2 → PASE A 14; ¿Busco trabajo por primera vez? ☐ 3; ¿Solo estudiante? ☐ 4; ¿Solo quehaceres domésticos? ☐ 5; ¿Solo jubilado? ☐ 6; ¿Solo pensionista? ☐ 7; Otros ☐ 8; Se ignora ☐ 9 → MUJERES PASE A 17, HOMBRES PASE A 21	¿Trabajo? ☐ 0; ¿No trabajo pero tenía trabajo? ☐ 1; ¿Busco trabajo habiendo trabajado antes? ☐ 2 → PASE A 14; ¿Busco trabajo por primera vez? ☐ 3; ¿Solo estudiante? ☐ 4; ¿Solo quehaceres domésticos? ☐ 5; ¿Solo jubilado? ☐ 6; ¿Solo pensionista? ☐ 7; Otros ☐ 8; Se ignora ☐ 9 → MUJERES PASE A 17, HOMBRES PASE A 21	¿Trabajo? ☐ 0; ¿No trabajo pero tenía trabajo? ☐ 1; ¿Busco trabajo habiendo trabajado antes? ☐ 2 → PASE A 14; ¿Busco trabajo por primera vez? ☐ 3; ¿Solo estudiante? ☐ 4; ¿Solo quehaceres domésticos? ☐ 5; ¿Solo jubilado? ☐ 6; ¿Solo pensionista? ☐ 7; Otros ☐ 8; Se ignora ☐ 9 → MUJERES PASE A 17, HOMBRES PASE A 21
SOLO PARA LOS QUE CONTESTARON 0, 1 o 2 EN PREGUNTA 13 — ¿CUAL ES LA OCUPACION QUE LE PROPORCIONO LOS MAYORES INGRESOS EN LOS DIAS COMPRENDIDOS ENTRE EL 7 AL 13 DE NOVIEMBRE O CUAL FUE SU ULTIMA OCUPACION SI ESTUVO CESANTE? Anote con algún detalle el trabajo que realiza (E): Profesor de Primaria	14			
¿A QUE SE DEDICA PRINCIPALMENTE EL ESTABLECIMIENTO DONDE DESEMPEÑO LA OCUPACION INDICADA? Anote por ejemplo: Educación primaria, consultorio médico, fábrica textil, restaurantes, hacienda ganadera, etc.	15			
¿CUAL FUE LA CATEGORIA O POSICION EN LA OCUPACION QUE INDICA? Haga las preguntas en el orden indicado y al recibir una respuesta, marque la casilla correspondiente.	16	Patrono o Socio Activo ☐ 1; Cuenta Propia ☐ 2; Empleado o Asalariado: Del Estado ☐ 3, Del Sector privado ☐ 4; Trabajador familiar sin remuneración ☐ 5; Otros ☐ 6; Se ignora ☐ 9	Patrono o Socio Activo ☐ 1; Cuenta Propia ☐ 2; Empleado o Asalariado: Del Estado ☐ 3, Del Sector privado ☐ 4; Trabajador familiar sin remuneración ☐ 5; Otros ☐ 6; Se ignora ☐ 9	Patrono o Socio Activo ☐ 1; Cuenta Propia ☐ 2; Empleado o Asalariado: Del Estado ☐ 3, Del Sector privado ☐ 4; Trabajador familiar sin remuneración ☐ 5; Otros ☐ 6; Se ignora ☐ 9

D. Características de fecundidad y mortalidad. SOLO PARA MUJERES DE 15 AÑOS Y MAS

¿CUANTOS HIJOS NACIDOS VIVOS HA TENIDO EN TOTAL?	17	Ninguno ☐ 00 → Pase a pregunta 21; Número; Se ignora ☐ 99	Ninguno ☐ 00 → Pase a pregunta 21; Número; Se ignora ☐ 99	Ninguno ☐ 00 → Pase a pregunta 21; Número; Se ignora ☐ 99
¿CUANTOS DE ESOS HIJOS ESTAN VIVOS ACTUALMENTE?	18	Ninguno ☐ 00; Número; Se ignora ☐ 99	Ninguno ☐ 00; Número; Se ignora ☐ 99	Ninguno ☐ 00; Número; Se ignora ☐ 99
¿EN QUE FECHA TUVO SU ULTIMO HIJO NACIDO VIVO?	19	Fecha: Mes, 19 Año	Fecha: Mes, 19 Año	Fecha: Mes, 19 Año
¿ESTA VIVO SU ULTIMO HIJO?	20	SI ☐ 1; NO ☐ 2; Se ignora ☐ 9	SI ☐ 1; NO ☐ 2; Se ignora ☐ 9	SI ☐ 1; NO ☐ 2; Se ignora ☐ 9

E. Estado civil o conyugal. PARA TODAS LAS PERSONAS DE 12 AÑOS Y MAS

¿ESTA USTED SOLTERO, CASADO, DIVORCIADO, VIUDO O SEPARADO?	21	Soltero ☐ 1; Casado ☐ 3; Divorciado ☐ 4; Viudo ☐ 5; Separado ☐ 6; Se ignora ☐ 9	Soltero ☐ 1; Casado ☐ 3; Divorciado ☐ 4; Viudo ☐ 5; Separado ☐ 6; Se ignora ☐ 9	Soltero ☐ 1; Casado ☐ 3; Divorciado ☐ 4; Viudo ☐ 5; Separado ☐ 6; Se ignora ☐ 9

2. EQUIPEMENT INFORMATIQUE

La mission de Mai 1985 a montré que l'IGM avait des projets de deux ordres : automatisation de la photogrammétrie avec stockage sur support informatique des fonds de carte saisis, et mise en place d'un service d'applications nouvelles, par la constitution d'une base de données importante sur l'Equateur. Ce service aurait pour objet la production à la demande de documents thématiques provenant de sources diverses : données d'inventaire, de télédétection, de recensements...etc.
Etant donné le niveau d'équipement informatique de l'IGM en 1985, les échéances des achats de matériel par l'IGM liés à ces deux projets sont absolument cruciaux pour la réalisation de l'APIQ.

Depuis 1985, l'objet des discussions de M. SOURIS avec les responsables de l'IGM a été d'essayer de définir une configuration informatique compatible avec les possibilités d'investissement de l'IGM, intégrant les systèmes de photogrammétrie envisagés et permettant une avancée technologique vers le développement d'applications thématiques, dont l'atlas informatisé de Quito constituerait un premier exemple.
A la demande des responsables de l'IGM, M. SOURIS a travaillé durant le second semestre 1985 à l'établissement d'une proposition d'un système interactif graphique pour l'IGM :

-proposition d'une configuration pour le matériel
-proposition d'une configuration pour le software
-modalités d'acquisition et proposition d'un contrat commercial pour l'achat de l'équipement du système.

Cette proposition a été présentée par M. SOURIS aux responsables de l'IGM, en Décembre 1985 à Quito.

La configuration proposée est organisée autour de trois axes, devant permettre de satisfaire l'ensemble des besoins de l'IGM à moyen terme :

* *un miniordinateur DIGITAL permettant la connexion des divers matériels de photogrammétrie et d'impression graphique,* ainsi que l'utilisation de nombreux logiciels scientifiques et graphiques : VAX-11/750, unité de disque fixe de 456 Mbytes, dérouleur de bande 1600/6250 BPI, contrôleur Ethernet, contrôleur pour 16 lignes asynchrones, 4 terminaux VT 220, terminal d'impression, imprimante 20cps.

* *un réseau local de stations de travail graphiques APOLLO pour la digitalisation manuelle, la constitution et l'interrogation de bases de données, la télédétection* : 2 stations couleur DN560, 3 stations noir et blanc DN300, un serveur de périphérique DSP80A auquel sont connectés un dérouleur de bande, un disque fixe 500 Mbytes, une imprimante et un controleur Ethernet

* *des systèmes de photogrammétrie automatique autonomes* (2 restituteurs du type C140 Zeiss ou BC2 WILD)

Enfin, la configuration proposée comprend : *un traceur électrostatique VERSATEC 7236, 3 digitaliseurs grand format, et un traceur électromécanique à plume.*

Ces trois pôles sont reliés par différents types de *liaisons* qui permettent la cohérence de la configuration :
- entre le miniordinateur et le réseau de stations de travail : liaison Ethernet, permettant de considérer les ressources de l'ordinateur DIGITAL comme partie intégrante du réseau local des stations de travail, et donc de bénéficier à ce titre de l'importante bibliothèque de logiciels existant sur le miniordinateur.

- entre les systèmes de photogrammétrie et le miniordinateur DIGITAL : les restituteurs analytiques seront connectés au miniordinateur pour le partage des périphériques (stockage sur bande magnétique, tracé sur plotter électrostatique) ainsi que pour les liaisons au réseau de stations de travail graphiques.

Chaque pôle de la configuration possède néanmoins des ressources propres de stockage et d'archivage, de manière à réduire la dépendance du sysème global par rapport à telle ou telle partie vitale de la configuration.

Le software associé au système proposé est de plusieurs provenances : software ORSTOM (dont TIGER), software intégré dans les systèmes de photogrammétrie automatique, software à acquérir dans le commerce (SAS, SPAD, UNIRAS, SPIDER pour l'application APIQ).

3 . L'ENQUETE DEMOGRAPHIQUE

3.1. Plan de sondage

Dans la phase de première application de la méthode de sondage stratifié sur image satellite à une situation réelle de ville d'un pays en développement, on distingue quatre objectifs principaux :

- Premier objectif : *L'adaptation du modèle des relations entre morphologie et population aux conditions spécifiques à la ville.*

L'intégration des connaissances accumulées sur la morphologie de la ville, (notamment par les collègues de l'ORSTOM travaillant sur l'atlas de Quito), ainsi que l'analyse des données morphologiques disponibles doivent permettre de définir une stratification efficace à Quito et de reformuler les hypothèses bayésiennes définies à Marseille.

- Deuxième objectif : *La définition de la base de sondage.*

Quelles sont à Quito la taille et la forme des unités spatiales offrant le moindre risque d'erreur dans la collecte des données et la meilleur précision des estimations démographiques ?

- Troisième objectif : *La définition du niveau géographique de production des données démographiques localisées.*

Hormis l'estimation de l'effectif de population - ou de la densité moyenne - sur l'ensemble de la ville, quelles sont à une échelle plus grande, les unités spatiales où la connaissance des données démographiques est nécessaire et où la méthode mise au point peut fournir des estimations avec une précision suffisante?

- Quatrième objectif : *L'actualisation du plan de sondage.*

Quels sont les caractères morphologiques qui apportent une information sur l'évolution démographique de Quito et comment utiliser cette information pour actualiser la base de sondage, redéfinir les strates et reformuler les hypothèses bayésiennes?

3. 2. Objectifs démographiques de l'enquête par sondage stratifié sur image satellite

Après avoir analysé le contenu du questionnaire et pris en compte les remarques de D. COURET sur les modalités de collecte et traitement relatives à chacune des questions posées lors du recensement de 1982, il parait nécessaire de donner à l'enquête par sondage spatial stratifié sur image satellite les objectifs suivants : *dénombrement de population et analyse des modes de résidence.*

Ce second objectif devrait permettre :
- de mieux comprendre les différences entre population de droit et de fait recensées en 1982 à Quito
- de faire des propositions pour une définition opérationnelle de la résidence dans les opérations de collecte démographique telles que les recensements exhaustifs : la question posée en 1982 s'est révélée inopérante et comporte de trop nombreux biais (cf rapport de D COURET p. 14)
- d'apporter des éléments d'informations sur les systèmes migratoires : le recensement de 1982 ne saisit que les migrations par rapport au lieu de naissance et la dernière migration, sans distinction entre origine rurale et urbaine. Analyser les systèmes de résidence, ou de migration (il s'agit du même phénomène) est un élément essentiel de compréhension du fonctionnement de la ville.

DIAPOSITIVES SUR MARSEILLE

Réaliséesen Avril 1986

1- Visualisation sur écran Radiance. Typologie de la morphologie urbaine; 9 classes.

2- Visualisation sur écran Radiance. Densité du Bâti; 6 classes.

3- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 1.

4- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 2.

5- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 3.

6- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 4.

7- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 5.

8- Visualisation sur Péricolor 1000. Thematic Mapper; Canal 6.

10- Visualisation sur Péricolor 1000. Thematic Mapper; Composition colorée (1,4,7) avec une superposition des contours.

11- Visualisation sur Péricolor 1000. Thematic Mapper; Composition colorée (1,2,3) avec une superposition des contours.

12- Visualisation sur Péricolor 1000. Typologie de la morphologie urbaine; 9 classes avec une superposition des contours.

14- Visualisation sur Péricolor 1000. Densité du bâti; 6 classes; avec une superposition des contours.

16- Visualisation sur Péricolor 1000. Surface maximum du bâti recodée avec une superposition des contours.

18- Visualisation sur Péricolor 1000. Surface minimum du bâti recodée avec une superposition des contours.

20- Visualisation sur Péricolor 1000. Surface de l'élément majoritaire.

22- Visualisation sur Péricolor 1000. Surface de l'élément majoritaire; avec un grossiment de deux fois.

26- Visualisation sur Péricolor 1000. Surface maximum du bâti;

30- Visualisation sur Péricolor 1000. Typologie de la morphologie urbaine; grossissement x2.

33- Visualisation sur Péricolor 1000. Thematic Mapper; Composition colorée (1,4,7); grossissement x2.

35- Visualisation sur Péricolor 1000. Thematic Mapper; Composition colorée (1,2,3); grossissement x2.

36- Visualisation sur Péricolor 1000. Thematic Mapper; Composition colorée (1,4,7) avec un grossissement x8 montrant les défauts de calage entre canaux.

A- Visualisation sur écran radiance du canal 4 de thematic mapper avec la superposition des contours des zones de la relation morphologie.

B- Visualisation sur écran radiance de toute la zone de traitement. Typologie de la morphologie urbaine.