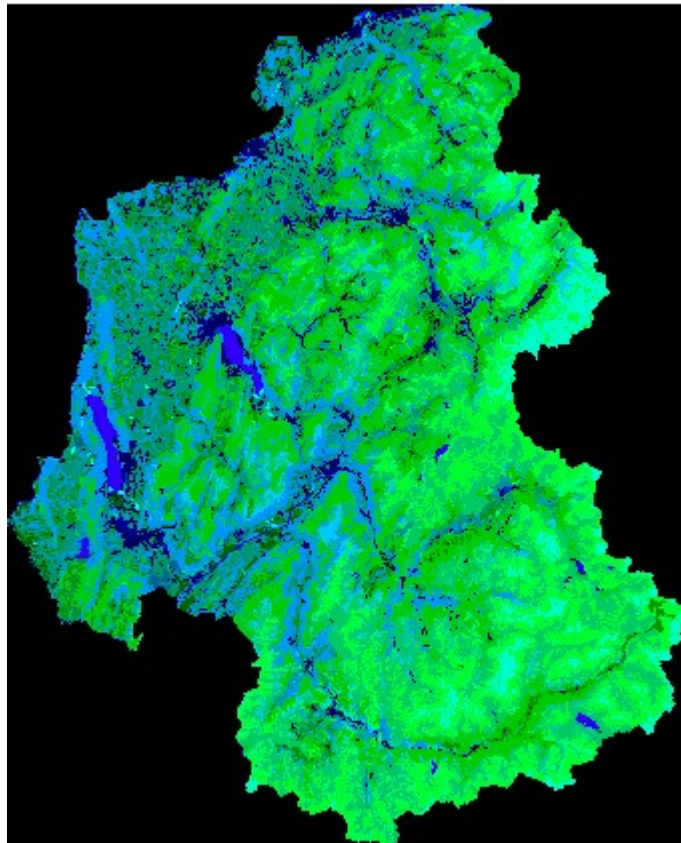


# L'OUTIL

# INFORMATIQUE

# EN GÉOGRAPHIE



*Pierre Dumolard*, Université Joseph Fourier, Grenoble  
*Françoise Allignol*, C.N.R.S. & Université de Nice  
*Emmanuel Paul*, Société « Géomap », Le Bourget du Lac  
*Erwan Quesseveur*, Université Technologique de Compiègne

# INTRODUCTION

## **POUR QUI CET OUVRAGE ?**

Le titre « *l'outil informatique en géographie* » destine clairement cet ouvrage à un public de géographes, d'aménageurs, d'environnementalistes, de spécialistes du « territorial » sous toutes formes et statuts : étudiants, enseignants, chercheurs, professionnels. Ce public n'a pas à avoir d'expertise particulière en informatique mais il est (au choix ou tout ensemble) :

- curieux de connaître certaines des possibilités qui lui sont maintenant ouvertes,
- confronté à de nouvelles pratiques professionnelles, de nouvelles façons de produire de la connaissance,
- désireux de maîtriser certains des concepts sous jacents à toute réalisation géo-informatique.

## **POUR QUOI CET OUVRAGE ?**

Pour diffuser auprès de ce public un panorama de méthodes aujourd'hui couramment utilisées et dont la pratique n'a été rendue possible que par l'apparition puis le développement de l'informatique sous ses formes matérielles, logicielles et ses nouvelles exigences en termes d'information (spatialisée, pour ce qui nous concerne).

## **POURQUOI MAINTENANT ?**

Parce qu'on n'en est plus aux temps pionniers où l'accès aux ordinateurs était rare et où la moindre application exigeait que l'on sache programmer. L'informatique d'aujourd'hui est un secteur d'activités offrant à l'utilisateur une assez large gamme de logiciels (commerciaux ou téléchargeables gratuitement sur Internet) dont l'ergonomie s'est largement améliorée. Les temps d'un usage « ordinaire » est maintenant venu.

Il est d'autant plus venu qu'est sorti des laboratoires un outil emblématique de la géo-informatique, les Systèmes d'Information Géographique. Cet outil s'est largement diffusé dans les collectivités territoriales, les services techniques de l'état, le secteur privé. Gérant des informations spatiales, il a généré de l'emploi pour géographes et favorisé des recherches pluri-disciplinaires.

Au delà de leur usage dans la gestion de territoires, les Systèmes d'Information Géographique sont aussi, en effet, un instrument commode de connaissance et de recherche, en ce qu'ils permettent certaines formes d'expérimentation par essais / erreurs, certains types de raisonnement visuel interactifs.

## **QUE NE CONTIENT IL PAS ?**

Tout d'abord, cet ouvrage ne contient aucune initiation à la bureautique, qui n'a rien de particulièrement géographique.

Il ne contient pas plus d'explications sur ce qu'est un ordinateur, quels sont ses composants matériels et logiciels, ce que réalise un système d'exploitation, quels sont les avantages de la mise en réseau d'un ensemble de machines : ce sont des connaissances aujourd'hui largement acquises, au moins empiriquement, par les jeunes et les moins jeunes (géographes ou pas).

Il ne contient pas non plus d'initiation aux pratiques statistiques en géographie, pour laquelle il existe d'excellents ouvrages (et de nettement moins bons !).

Enfin, n'étant pas principalement destiné à des informaticiens, du moins pour leur donner à voir de l'informatique technique, il ne contient ni algorithme, ni structures de données avancées, ni considérations savantes sur telle ou telle architecture, tel ou tel langage (« objets » ou autre), bien sûr sous jacents à tout logiciel.

## **QUE CONTIENT IL DONC ?**

Un panorama de méthodes, pas de toutes celles que l'on aurait pu décrire, seulement de celles devenues relativement courantes chez les géographes et/ou dont le potentiel d'usage nous a semblé le plus large et le moins difficile à appréhender. Les méthodes évoquées ont été, en effet, presque toutes mises en oeuvre avec des logiciels disponibles : elles sont donc à disposition du lecteur.

En fait, nous nous sommes attachés, le plus souvent possible, à montrer ce qu'il y a « entre l'écran et le clavier d'ordinateur » : des concepts, de l'information structurée, des possibilités d'expérimentation et d'innovation.

## **LA LOGIQUE DE L'EXPOSE ?**

Les deux premiers chapitres sont dédiés aux nouveaux médias et nouvelles formes de documentation géographique.

- **chapitre 1** : La documentation via Internet, source de données et de connaissances,
- **chapitre 2** : Le multimedia, permettant un contenu multiforme et une consultation conviviale.

Les trois chapitres suivants font le point sur des outils devenus classiques et posent des jalons vers des applications plus récemment apparues.

- **Chapitre 3** : Les bases de données et leur nécessaire structuration,
- **Chapitre 4** : L'Information Géographique, sous son double aspect sémantique et spatial,
- **Chapitre 5** : La Cartographie thématique, ses traitements statistiques et visuels.

Les derniers chapitres présentent des outils et connaissances maîtrisés par un nombre moindre de « spatialistes » mais qui sont en voie de diffusion rapide.

- **Chapitre 6** : Les Systèmes d'Information Géographique, incontournables et insuffisants,
- **Chapitre 7** : Le traitement d'images géographiques, analyse et synthèse,
- **Chapitre 8** : Les modèles d'analyse et de simulation, sous jacents à toute informatique géographique.

# **TABLE DES MATIERES**

## **Introduction**

### **Chapitre 1 : Documentation sur Internet**

#### **1.1 Généralités sur l'Internet**

- 1.1.1 définition et historique*
- 1.1.2 les composantes de l'internet*
- 1.1.3 avantages et inconvénients*
- 1.1.4 vérifier la validité des informations*
- 1.1.5 équipement nécessaire*

#### **1.2 Recherche d'information dans l'internet**

- 1.2.1 principes et stratégie de recherche*
- 1.2.2 outils de recherche d'information*

#### **1.3 Un exemple de recherche sur internet**

- 1.3.1 préparation de la recherche*
- 1.3.2 premières étapes de la recherche*
- 1.3.3 recherche de fond*
- 1.3.4 recherche avancée*
- 1.3.5 conserver l'information*

### **Chapitre 2 : Le multimedia**

#### **2.1 Qu'est ce que le multimedia ?**

- 2.1.1 cédéroms et organisation hypertexte*
- 2.1.2 principes de constitution et de navigation*

#### **2.2 Encarta, un exemple d'atlas interactif**

- 2.2.1 un produit « microsoft » pour « windows 9x »*
- 2.2.2 accès à l'information*
- 2.2.3 pourquoi un atlas électronique ?*

#### **2.3 Communiquer la révision d'un Plan d'Occupation des Sols**

- 2.3.1 multimédia et collectivités locales*
- 2.3.2 objectifs et enjeux*
- 2.3.3 dispositifs techniques*

### **Chapitre 3 : Bases de données**

#### **3.1 Bases de données mono-fichier**

- 3.1.1 types de tableaux*
- 3.1.2 implémentation informatique*
- 3.1.3 exemples de requêtes à une base de données mono-fichier*

#### **3.2 Bases de données relationnelles**

- 3.2.1 niveaux de modélisation*
- 3.2.2 implémentation informatique en SQL*

#### **3.3 La fouille de données (« data mining »)**

## **Chapitre 4 : L'information géographique**

### **4.1 Modélisations de la terre**

- 4.1.1 *coordonnées sphériques*
- 4.1.2 *projections : coordonnées planes*
- 4.1.3 *procédés d'acquisition de l'information spatiale*

### **4.2 Informatisation des données géographiques**

- 4.2.1 *le mode rasteur*
- 4.2.2 *le mode vecteur*
- 4.2.3 *transformations rasteur – vecteur*

## **Chapitre 5 : Cartographie thématique**

### **5.1 Connaissances de base en cartographie thématique**

- 5.1.1 *typologie pour la construction d'une légende*
- 5.1.2 *traduction visuelle de la typologie*
- 5.1.3 *plusieurs variables sur une même carte ?*
- 5.1.4 *un problème : la modification des unités spatiales*

### **5.2 Informatique et cartographie thématique avancée**

- 5.2.1 *capacités de l'informatique utilisables en cartographie thématique*
- 5.2.2 *bénéfices en termes de traitement des données géographiques*
- 5.2.3 *bénéfices en termes de visualisation des données géographiques*

## **Chapitre 6 : Systèmes d'Information Géographique**

### **6.1 Définitions des S.I.G.**

- 6.1.1 *définitions logicielles*
- 6.1.2 *définitions géomatiques*

### **6.2 Principales utilisations des S.I.G.**

- 6.2.1 *types de questions à un S.I.G.*
- 6.2.2 *domaines thématiques d'utilisation*
- 6.2.3 *secteurs d'activité géomatique*

### **6.3 S.I.G. vecteur**

- 6.3.1 *modèles de données*
- 6.3.2 *modélisation géométrique des objets cartographiques*
- 6.3.3 *modélisation de la géométrie et de la topologie*
- 6.3.4 *opérations possibles*

### **6.4 S.I.G. rasteur**

- 6.3.5 *modèle de données simple*
- 6.3.6 *améliorations possibles du modèle de données*

### **6.5 Rasteur ou vecteur ? Rasteur et vecteur ?**

### **6.6 Principales sources françaises de données géographiques**

## **Chapitre 7 : Traitement d'images géographiques**

### **7.1 Analyse d'images télédétectées**

- 7.1.1 *nature de l'image satellitale*
- 7.1.2 *principaux satellites civils d'observation*
- 7.1.3 *traitement d'images satellitales*
- 7.1.4 *exemple : la croissance urbaine de Tunis (1987-1996)*
- 7.1.5 *principales applications civiles*
- 7.2 **Synthèse d'images pour l'aménagement**
- 7.2.1 *trois étapes dans la construction d'une image virtuelle*
- 7.2.2 *quelques exemples d'application géographique*

## **Chapitre 8 : Modèles, analyse et simulation**

- 8.1 **Trois grands types de modèle**
- 8.1.1 *modèles classiques de l'analyse spatiale*
- 8.1.2 *modèles opérationnels*
- 8.1.3 *modèles de modèles*
- 8.2 **Modèles pour l'analyse spatiale**
- 8.2.1 *analyse des données spatiales*
- 8.2.2 *analyse spatiale des données*
- 8.3 **Modèles, simulation et aide à la décision**
- 8.3.1 *vieillissement potentiel d'une région*
- 8.3.2 *anticiper l'évolution du trafic au tunnel du Mont Blanc*
- 8.3.3 *effet d'un Plan de Déplacement Urbain sur les nuisances sonores*
- 8.3.4 *l'I.A. : Intelligence Artificielle ou Informatique Avancée ?*
- 8.3.5 *Modélisation dynamique par Système Multi-Agents*

## **Conclusion**

# **CHAPITRE 1 : DOCUMENTATION SUR INTERNET**

**Erwan Quesseveur**

## **1.1. GENERALITÉS SUR L'INTERNET**

### **1.1.1. L'INTERNET : DÉFINITION ET HISTORIQUE**

L'Internet (le « net » ou la « toile ») est un réseau de communication entre ordinateurs sur lequel circulent des informations de natures multiples. Par ce moyen, des millions d'ordinateurs sont aujourd'hui interconnectés à travers le monde, donnant ainsi accès aux connaissances mises à disposition sur le réseau par leurs propriétaires tout en permettant d'y diffuser sa propre contribution. De façon plus concise, l'Internet permet d'échanger des connaissances avec d'autres personnes connectées n'importe où dans le monde, quelle que soit la distance.

Le réseau Internet est né en 1969 sous le nom d'**ARPANET**, réseau militaire d'échange d'information. L'objectif initial était de concevoir un réseau informatique décentralisé et redondant tel que la destruction d'une de ses parties n'empêche pas la transmission de données, par d'autres chemins. Il a été rendu public en 1983, essentiellement pour relier entre elles des universités et des centres de recherche, mais c'est l'arrivée du **World Wide Web** (www, w3 ou web) - principe de navigation mis au point en 1990 - ainsi que l'arrivée des ordinateurs personnels dans les foyers qui ont lancé la « démocratisation » réelle du réseau.

### **1.1.2. LES COMPOSANTES DE L'INTERNET**

#### **A. Le World Wide Web**

Le Web est, à ce jour, l'élément le plus récent et le plus développé de l'Internet. L'information y est organisée par pages. Les liaisons entre pages sont basées sur le principe des liens hypertexte (**HTTP** pour « Hyper Text Transfer Protocol »), c'est-à-dire que des mots, généralement soulignés (ou des images) placés dans les pages, renvoient à d'autres pages. Ces pages peuvent se trouver aussi bien sur le même ordinateur que celle actuellement consultée ou sur un autre situé à des milliers de kilomètres du précédent. Les adresses sont écrites selon le modèle suivant :

<http://www.nomdusite.domaine>.

Le nom de **domaine** permet parfois de retrouver la nationalité du site. Ainsi les serveurs français auront le domaine **.fr**. Mais il existe d'autres domaines qui ne font pas référence à leur nationalité (**.com** pour les sites commerciaux, **.gov** ou **.org** pour les sites institutionnels).

Le langage qui permet de mettre en forme la plupart des pages web actuelles est le **HTML** (« Hyper Text Markup Language »).

On classe généralement les pages web en deux catégories :

- les sites à contenu,

- les sites de recherche de contenu (cf 1.2).

## B. Le courrier électronique

Le courrier électronique (**e-mail**, mël ou encore courriel) est une application datant de 1978 qui permet de rédiger un message sur ordinateur et de l'expédier via le réseau à une ou plusieurs personnes elles aussi connectées à la toile. Il est possible d'attacher au message envoyé des données de nature multiple (texte, image, ...) qui transitent sous la forme de fichier. Ce média est destiné à compléter, voire à supplanter, le fax du fait de la qualité des transmissions de document et le courrier traditionnel pour sa rapidité.

## C. Les communautés virtuelles

Les **groupes de discussion** (usenet, news, newsgroups) permettent de regrouper des abonnés autour de débats concernant des domaines les plus divers. Les utilisateurs laissent des questions sur un espace virtuel de discussion et viennent consulter les réponses. Pour trouver des informations dans ces « newsgroups », on les consultera eux-mêmes, mais aussi des pages web telles que <http://groups.google.com/>. Actuellement plus de 100000 « newsgroups » existent, couvrant des domaines allant du général au plus précis.

On peut regrouper des personnes intéressées par un même sujet par le biais des listes de mail (**mailing lists**). Dans ce cas, les messages sont envoyés à l'ensemble des personnes abonnées, directement à leur adresse électronique. Ces listes de mail fonctionnent suivant le même principe d'abonnement que les groupes de discussion.

## D. FTP (« File Transfer Protocol »)

Le protocole **FTP** (« File Transfer Protocol ») permet d'échanger des fichiers d'un ordinateur à l'autre. Pour ce faire, il faut se connecter à la machine sur laquelle se trouvent les fichiers souhaités, puis s'y déplacer suivant l'arborescence du disque dur de l'ordinateur pour trouver le fichier désiré. De même que pour les « newsgroups », on peut accéder à FTP par un navigateur tel Netscape ou Internet Explorer.

### 1.1.3 AVANTAGES ET INCONVÉNIENTS

L'Internet présente l'avantage d'offrir une quantité incroyable d'informations sous des formats extrêmement diversifiés, puisque toutes les formes de médias sont présentes (image, son, vidéo, texte, etc.). Néanmoins, la qualité de l'information disponible sur la Toile doit être vérifiée avant utilisation car dans les faits, chaque utilisateur est libre de publier, sans aucun contrôle, ce qu'il désire. Les problèmes liés à l'application d'un contrôle de la légalité des contenus de l'Internet est symbolique de cette situation.

Par ailleurs, il est parfois long et difficile d'accéder à des informations de qualité, si l'on n'a pas pris soin de bien délimiter le sujet et la méthode de recherche.

Enfin, l'Internet est victime de son succès. Avec un nombre constamment croissant d'utilisateurs, les lignes sont de plus en plus encombrées et le transfert des données est parfois très lent. C'est cependant un problème qui devrait être résolu par le développement de réseaux de communication à haut débit.



#### 1.1.4. VERIFIER LA VALIDITE DES INFORMATIONS

L'Internet présente l'avantage de fournir une masse très importante d'information ; il est, cependant, très utile de vérifier la validité des renseignements récupérés. Contrairement aux publications scientifiques qui bénéficient généralement d'un comité de lecture, les informations contenues dans l'Internet ne sont pas forcément vérifiées et validées avant publication. Notons tout de même quelques sites qui regroupent des sources évaluées par des modérateurs scientifiques, tels que

<http://www.britannica.com>.

Il convient donc d'évaluer par soi-même la qualité de l'information disponible. Pour cela on se posera diverses questions concernant, tout d'abord, l'auteur de la page consultée.

- Est-ce une institution, un expert ? Les compétences de l'auteur sont-elles énoncées ? Le site comporte-t-il des informations permettant de contacter facilement l'auteur (adresse électronique, postale) ?
- L'information trouvée est-elle référencée sur plusieurs pages ? Y a-t-il des liens entre ces pages ? Quel est le type de public ciblé par le site (spécialistes, initiés, débutants) ?

On peut également se questionner sur la nationalité des renseignements obtenus.

- Où le site est-il hébergé, et par qui ?

On peut également estimer la qualité de la réalisation du site.

- Est-il bien organisé (pages cohérentes) ? Les illustrations enrichissent-elles l'information ? Les liens avec d'autres pages fonctionnent-ils ?
- Le site est-il mis à jour régulièrement ? L'orthographe est-elle correcte ?

Enfin, il conviendra de s'interroger sur le but du site (information ?, publicité ?).

Pour conclure, on privilégiera l'information des sites comprenant des liens à jour vers d'autres pages à contenu « sérieux », qui équivalent aux références bibliographiques des travaux universitaires. Si ce n'est pas le cas, le croisement des informations entre sites ou avec d'autres médias (articles, ouvrages) est indispensable.

#### 1.1.5. ÉQUIPEMENT NECESSAIRE

##### A. Le matériel (« hardware »)

Pour se connecter à l'Internet chez soi, il est nécessaire d'avoir un ordinateur, équipé d'un **modem** (modulateur/**dé**modulateur) connecté à une prise téléphonique ou au câble Tv et Internet. Cet appareil permet de transformer le signal numérique de l'ordinateur en signal analogique pour transmettre les informations par les lignes téléphoniques. Un abonnement auprès d'un fournisseur d'accès (« provider ») est également indispensable pour accéder à l'Internet par le biais du téléphone. Les universités fournissent généralement un accès gratuit à l'Internet aux professeurs, chercheurs et étudiants. Dans ce cas, les connexions sont relativement rapides puisqu'elles s'opèrent directement par le réseau informatique, qui a des débits plus élevés.

## B. Le logiciel (« software »)

Les logiciels nécessaires à la connexion sont :

- un navigateur (butineur ou « browser ») pour se déplacer sur le web,
- un gestionnaire de courrier électronique,
- des logiciels ou protocoles pour FTP et les « newsgroups ».

Des logiciels gratuits tels que « Netscape Communicator » ou « Internet Explorer » permettent d'accéder à l'ensemble de ces services depuis la même plateforme.

## **1.2 RECHERCHE D'INFORMATION DANS L'INTERNET**

### **1.2.1. PRINCIPES et STRATÉGIES DE RECHERCHE**

#### A. Eléments méthodologiques.

##### a. Définition précise de la recherche d'information

Quel que soit le type d'information recherchée, une bonne délimitation préalable de ce que l'on cherche est nécessaire, sous peine de rapidement s'égarer dans le volume des données disponibles (« trop d'information tue l'information »). Ce principe est valable pour toutes les situations de recherche d'information, quel qu'en soit le support. Il s'agit d'acquérir des connaissances de base sur le sujet, passant par la maîtrise des différentes terminologies qui s'y rapportent et de déterminer l'approche particulière du sujet que l'on souhaite aborder. Par exemple on peut chercher sur Internet :

- la vulgarisation d'un domaine,
- une approche très synthétique
- ou bien encore des informations très précises.

##### b. Les particularités de l'Internet

A la différence d'une bibliothèque, le monde de l'Internet ne dispose pas d'un mode unique de classement des informations. Il n'existe pas de centre dans le réseau, tous les individus connectés peuvent être à la fois producteurs et consommateurs de données sans qu'un contrôle soit possible sur le contenu et les modes d'organisation de l'information proposée. Par ailleurs la multiplicité des types de données disponibles (texte en plusieurs langues, image, son) complexifie encore la mise en place d'une structure de classement.

Pour la compréhension de la suite, on a adopté les conventions graphiques suivantes

- [lien hypertexte](#),
- *commande de logiciel*,
- *requête rédigée par l'utilisateur*.

#### B. Recherche généraliste

Dans une recherche généraliste, la définition du sujet utilise surtout une indexation thématique. A partir d'une liste de thèmes (actualités, enseignement, informatique, zones géographiques, institutions, ...), l'utilisateur va naviguer vers les thèmes de son choix (par exemple :

[sciences humaines - géographie - instituts - institut géographique national](http://www.ign.fr). Depuis le site Internet <http://fr.yahoo.com/>). La figure 1.1 présente ainsi le portail d'accès au site de l'I.G.N. qui est renvoyé comme réponse à la requête ci-dessus et permet d'aller à des contenus plus précis (par exemple, « Cartes numériques et bases de données »).



**Figure 1.1 : Portail d'accès au site web de l'IGN** (source: [www.ign.fr](http://www.ign.fr))

Il n'existe pas de normalisation entre les différents outils proposant ce mode de recherche. L'organisation des thèmes est donc choisie arbitrairement et les cheminements pour atteindre un sujet ne sont pas forcément intuitifs. En outre les résultats proposés ne sont pas exhaustifs, ils sont issus d'une sélection réalisée par les gestionnaires de chaque index. Les divers sites de recherche sont ainsi plus ou moins efficaces selon la thématique abordée.

## C.Utilisation de moteurs de recherche

### a. Les types de moteur de recherche : moteur et méta moteur

Les moteurs de recherche sont des robots qui scrutent automatiquement l'Internet en quête de nouveaux sites à référencer dans leur base de données.

Cependant, pour une requête identique, les résultats peuvent être très différents d'un moteur à l'autre, car ceux-ci ont leurs propres méthodes d'indexation et des puissances différentes. Par ailleurs on estime aujourd'hui qu'à peine 50% de l'ensemble des sites de l'Internet sont indexés dans ces moteurs. L'utilisateur accède à l'information de la base de données par la rédaction d'une requête.

Les méta moteurs de recherche cumulent les efforts de plusieurs moteurs de recherche en lançant une requête simultanément sur toutes les bases de données. Pour l'utilisateur, ce traitement est transparent. Les résultats sont présentés regroupés et nettoyés des doublons éventuels. Ce type d'outil permet de scruter un ensemble plus large de sites mais la traduction automatique de la requête par le robot n'est pas toujours contrôlable.

On distingue les méta moteurs en ligne (Metacrawler par exemple) qui permettent de dégrossir une question et les méta moteurs à installer sur un poste de travail (Webferret, Copernic,...) avec lesquels il est possible de réaliser une veille beaucoup plus fine.

### b. Le langage de requête

Une requête simple consiste à rechercher dans des pages Internet un **mot clé** (exemple : *géographie*) ou un ensemble de mots clés contigus (chaîne de caractères), que l'on signale alors par des guillemets (exemple : *"géographie urbaine"*). Il est possible d'affiner la requête en utilisant des **opérateurs logiques** pour combiner les mots clés.

Les opérateurs logiques de base :




O p é r a t i o n	R é s u l t a t	O p é r a t e u r
U n i o n		O u , O r
I n t e r s e c t i o n		E t , A n d , +
E x c l u s i o n		S a u f , N o t , -

Figure 1.2 : les opérateurs logiques de base

L'**union** permet de rechercher des synonymes ou des concepts proches (exemple : *"système d'information géographique" OU "geographical information system"*).

L'**intersection** impose la présence de tous les mots clés dans la réponse.

L'**exclusion** permet d'en éliminer certains du résultat.

### *Les opérateurs de proximité :*

Ces opérateurs permettent d'affiner le critère d'intersection en proposant des requêtes sur la proximité des mots clés au sein d'un texte : proximité à moins de n mots, notion d'ordre entre les mots clés de la requête.

### *Les niveaux de recherche :*

Il est possible de réaliser la requête, au choix, uniquement sur l'adresse Internet, sur le titre de la page, sur les méta données (nom de l'auteur, date de mise à jour, mots clés de l'auteur et résumé du contenu de la page) ou sur l'ensemble de la page (image, texte, son, fichier associé ...). Cependant les moteurs de recherche n'intègrent pas tous ces paramètres.

#### **c. Les stratégies de recherche**

Afin de mettre en œuvre une recherche documentaire efficace il est conseillé de choisir au préalable une stratégie adaptée au résultat souhaité. A titre d'exemple on propose ici trois démarches de recherche différentes.

- Trouver le plus de sites sur un sujet : pour cela il est intéressant d'utiliser les méta moteurs de recherche pour allier les efforts de plusieurs moteurs.

- Trouver rapidement le plus de sites sur un sujet : il est souvent possible de se baser sur une compilation d'adresses réalisée par un spécialiste du domaine. On appelle ces pages particulières des signets, des favoris (« bookmarks » en anglais). Soit par exemple la requête : +*"geographic information"* +*bookmarks*.

- Trouver un minimum de sites, mais les plus pointus possibles : dans ce cas on utilise les niveaux de recherche les plus restrictifs. Les mots clés seront recherchés dans le titre ou l'adresse Internet des pages. Les documents sont sélectionnés selon leur date de création ou de mise à jour.

## **1.2.2. LES OUTILS DE RECHERCHE D'INFORMATION SUR INTERNET**

### **A. Les outils les plus courants**

L'essentiel de la recherche d'information peut être satisfaite par ces outils gratuits mais il est souvent intéressant d'en tester plusieurs pour une même recherche. Il n'est pas question ici de réaliser une présentation exhaustive des outils existants, l'offre est bien trop riche et fluctuante. Les moteurs de recherche et les outils de recherche généraliste avec des indexés thématiques sont souvent associés dans une même page d'accueil.

Adresse	Pages indexées	Opérateurs booléens	niveaux de recherche	requête sur la date	Langues
<a href="http://www.voila.fr">http://www.voila.fr</a>	100 millions	oui	non	oui	
<a href="http://www.google.com">http://www.google.com</a>	730 millions	oui	oui	oui	15
<a href="http://www.altavista.com">http://www.altavista.com</a>	346 millions	oui	non	oui	25
<a href="http://www.alltheweb.com">http://www.alltheweb.com</a>	550 millions	oui	oui	oui	49

**tableau 1.1 : Quelques moteurs de recherche**

La figure 1.3 fournit un exemple de requête au moteur « Voilà ».



**Figure 1.3 : exemple de recherche avec le moteur « Voilà »**

## B. Les outils de recherche avancée

L'objectif des outils de recherche avancée est d'alléger pour l'utilisateur les tâches répétitives de recherche d'information.

Les méta moteurs de recherche en ligne se présentent sous la forme de pages Internet de requête semblables à celle des moteurs de recherche simple.

Adresse Internet	nombre de moteurs enchaînés	compilation des résultats
<a href="http://www.metacrawler.com">http://www.metacrawler.com</a>	13	oui
<a href="http://www.debriefing.com">http://www.debriefing.com</a>	11	oui
<a href="http://www.profusion.com">http://www.profusion.com</a>	13	oui
<a href="http://www.kartoo.com">http://www.kartoo.com</a>	14	Oui visualisation graphique

**tableau 1.2 : Quelques méta moteurs en ligne**

Les méta moteurs hors ligne sont des logiciels autonomes à installer sur un ordinateur. L'exploitation du résultat d'une requête peut être réalisée en local, sans la connexion au réseau.



Nom du moteur	nombre de moteurs enchaînés	compilation des résultats
Webferret	9	non
Hurricane websearch	12	non
Copernic 2001	80	oui

**tableau 1.3 : Quelques méta moteurs hors ligne**

A titre d'illustration, la figure 1.4 présente le début de la liste des sites renvoyées par Copernic 2001 à la requête « climat ET températures ET alpes »



**Figure 1.4 : exemple de requête avec Copernic 2001**

L'information à la demande (le « push technology ») permet une veille permanente des évolutions des contenus de l'Internet. Le principe est de prévenir l'utilisateur lorsqu'une information nouvelle, relative à son sujet, est mise à jour. Soit le robot relance régulièrement une même requête sur des moteurs de recherche, ou bien il scrute directement les sites que l'utilisateur lui a spécifiés.

C4U (<http://www.c-4-u.com/>) est un bon exemple d'outil de veille.

Les outils qui utilisent la technologie des agents intelligents tendent vers une autonomie grandissante dans la recherche d'information. Ils peuvent consulter plusieurs moteurs de recherche à la fois, scruter les liens hypertexte des pages sélectionnées, rechercher des mots clés ayant une sémantique proche, créer la synthèse des sites les plus pertinents en traduisant leur contenu. Les agents intelligents de la dernière génération sont capables de tenir compte des préférences de l'utilisateur et de faire évoluer la stratégie de recherche d'information selon le contexte. Aujourd'hui ces outils sont encore peu accessibles aux utilisateurs finaux de l'Internet.

### **1.3. UN EXEMPLE DE RECHERCHE D'INFORMATION INTERNET**

#### **1.3.1. LA PREPARATION DE LA RECHERCHE**

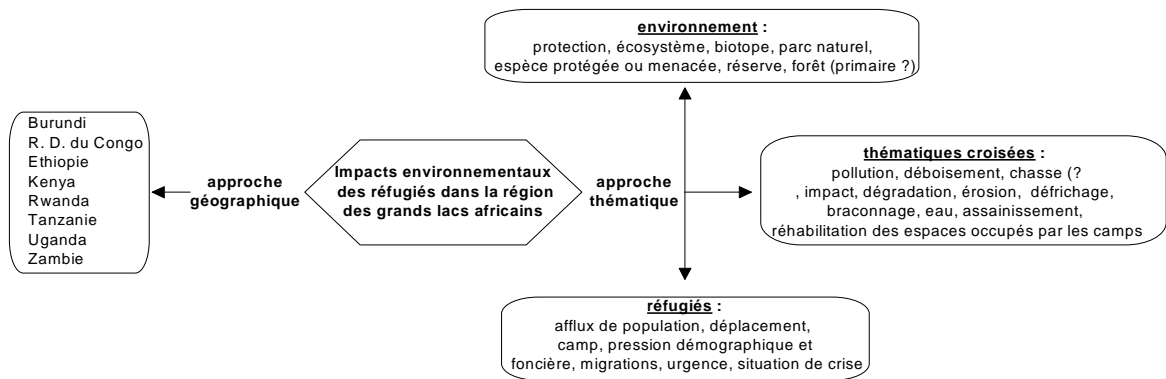
##### **A. Définition du sujet**

Pour donner un exemple, nous allons traiter ici le thème des impacts environnementaux dus aux flux de réfugiés dans la région des grands lacs africains.

A la suite des grands conflits ethniques qui ont bouleversé l'équilibre politique de la région des grands lacs africains, les flux massifs de réfugiés menacent l'environnement.

##### **B. Analyse du sujet**

Son « arbre sémantique » est :



**Figure 1.5 : arbre sémantique correspondant à l'exemple du § 1.3.1**

L'arbre sémantique ainsi construit permet de décomposer le sujet en mots clés simples. Au cours de la recherche, il est progressivement complété par des concepts issus des résultats.

*Trois étapes de recherche à enchaîner*

- se laisser guider dans des index thématiques à partir de mots clés simples. L'objectif est de mieux cerner le sujet, et son vocabulaire spécifique.
- exploiter les moteurs de recherche et les méta-moteurs. L'objectif de cette étape est de localiser l'essentiel des références accessibles traitant des deux thématiques croisées (« réfugié » et « environnement »).
- organiser une veille des mises à jour et engager des contacts avec les communautés virtuelles concernées par le sujet

#### **1.3.2. PREMIÈRES ÉTAPES DE LA RECHERCHE**

L'index thématique utilisé est "Yahoo" anglophone (<http://www.yahoo.com>)

Pour obtenir des informations générales sur la région des grands lacs, le chemin dans les index [Science](#) - [Geography](#) - [Country Profiles](#) mène à un site de description par pays : <http://www.un.org/Pubs/CyberSchoolBus/infonation/>



On en déduit le tableau suivant.

Pays	densité de population	Réfugié /habitants	Proportion de surface boisée	Nombre d'espèces menacées
Ethiopie	54	1/214	13.6%	60
Kenya	57	1/120	2.3%	107
Tanzanie	36	1/58	36.8%	132
Ouganda	85	1/102	30.6%	67

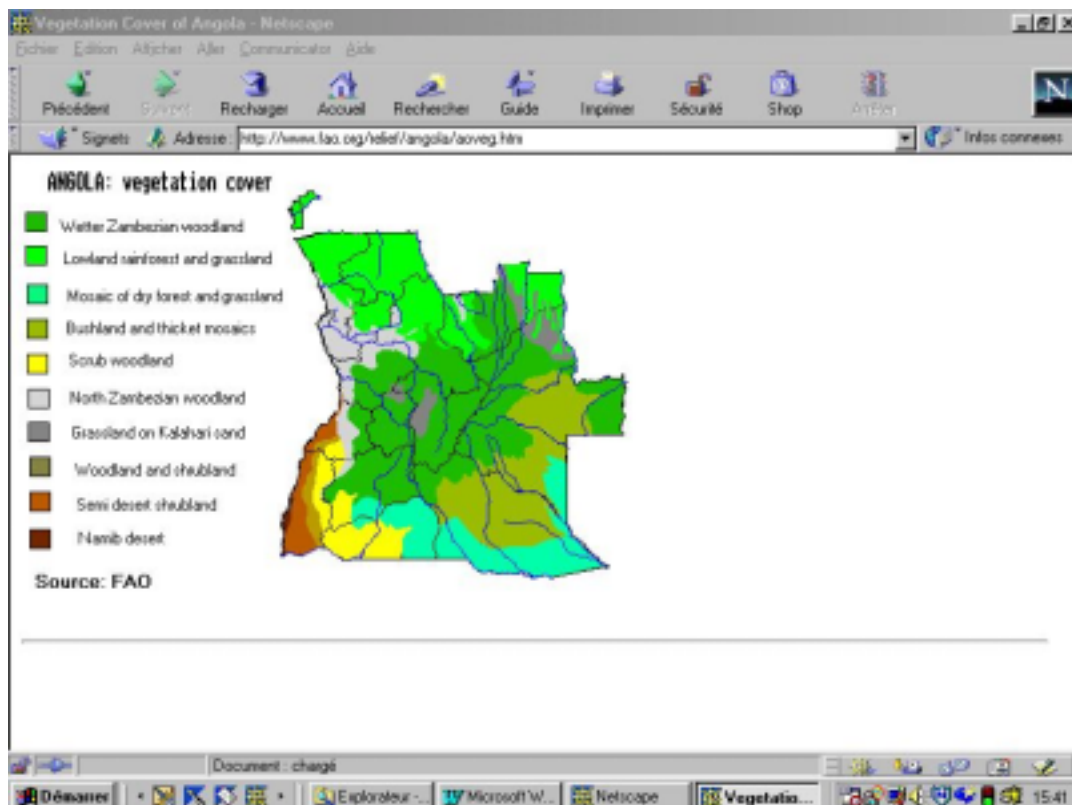
tableau 1.4 : exemple de résultats (Source publication des Nations Unies - 2000.)

Le chemin [Regional](#) - [Regions](#) - [Africa](#) - [News and media](#) > aboutit à un site africain d'information générale : <http://www.panapress.com/>.

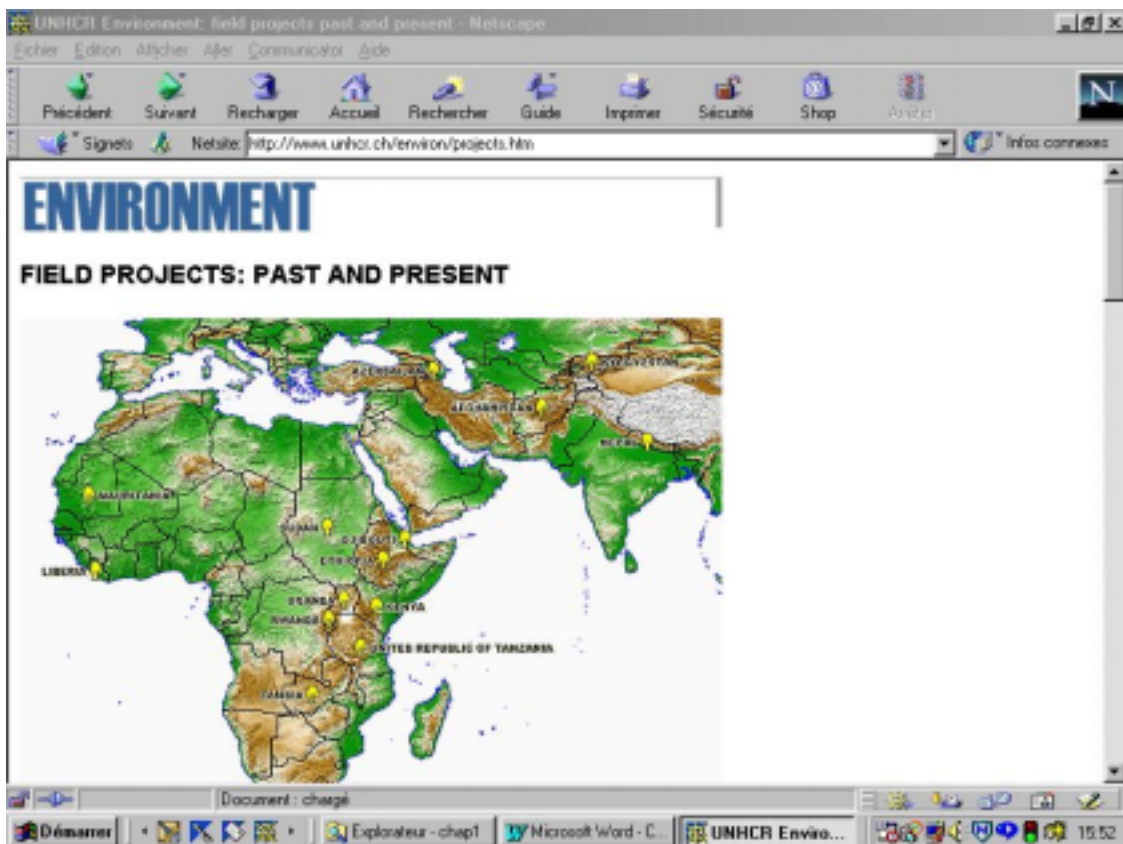
Pour trouver des informations sur les réfugiés, le chemin dans les index [Government](#) - [International Organizations](#) - [United Nations](#) - [Agencies](#) mène au site du Haut Commissariat des Nations Unies pour les Réfugiés : <http://www.unhcr.ch/>

Dans ce site, des pages particulières traitent du problème des réfugiés et de l'environnement et proposent des cartographies de zones d'intervention du HCR. La requête : *And maps And refugee And Africa And "Great Lakes"* dans Yahoo, donne accès à un site de cartographie sur les situations d'urgence : <http://www.reliefweb.int/w/map.nsf/home>.

Les figures 1.6 et 1.7 fournissent deux exemples de documents trouvés



**FIGURE. 1.6 : Exemple de document trouvé** (source : HDR)



**FIGURE. 1.7 : Exemple de document trouvé** (source : HDR)

### 1.3.3. RECHERCHE DE FOND

#### A. Avec méta-moteurs de recherche

L'objectif est de trouver des pages traitant du problème croisé des réfugiés et de l'environnement sans préciser la zone géographique.

- avec Kartoo, <http://www.kartoo.com>

l'expression : *réfugiés* +(dégradation Or pollution) donne 10 références dont 2 très pertinentes :

[http://www.msfc.ca/refugeecamp/francais/refugees\\_magazine.htm](http://www.msfc.ca/refugeecamp/francais/refugees_magazine.htm)

<http://gechs.org/avisofrench/twofr lg.shtml>

- avec Debriefing, <http://www.debriefing.com>

l'expression +*refugees* +*park* +*wildlife* donne 27 références dont 4 pertinentes :

<http://www.panda.org/resources/publications/species/igcp/page2.htm>

<http://ens.lycos.com/ens/oct2000/2000L-10-23-03.html>

<http://forests.org/recent/1996/sudanwar.htm>

<http://www.unesco.org/whc/sites/63.htm>

Les méta-moteurs sont adaptés aux recherches ne requérant pas de paramètres avancés.

#### B. Avec moteurs de recherche

L'utilisation des moteurs de recherche permet d'élaborer des requêtes affinées.

- avec Voila (<http://www.voila.fr>)

La requête +*refugiés* +*déboisement* + "*grands lacs*" +*afrique* donne 12 résultats dont trois très pertinents :

<http://nmainy.multimania.com/fichier6.htm>

<http://www.unhcr.ch/french/fdrs/ga99/ken.htm>

<http://roland.adjovi.free.fr/nguemb.htm>

- avec Google (<http://www.google.com>).

L'expression "environmental rehabilitation of refugee impacted areas" donne une seule référence très pertinente :

[http://www.worldbank.org/afr/afr\\_for/frmtheme.htm](http://www.worldbank.org/afr/afr_for/frmtheme.htm).

La requête "environmental rehabilitation" And refugee And "Great Lakes" And Africa en ne conservant que les pages mises à jour il y a moins d'un an, donne des résultats de très bonne qualité :

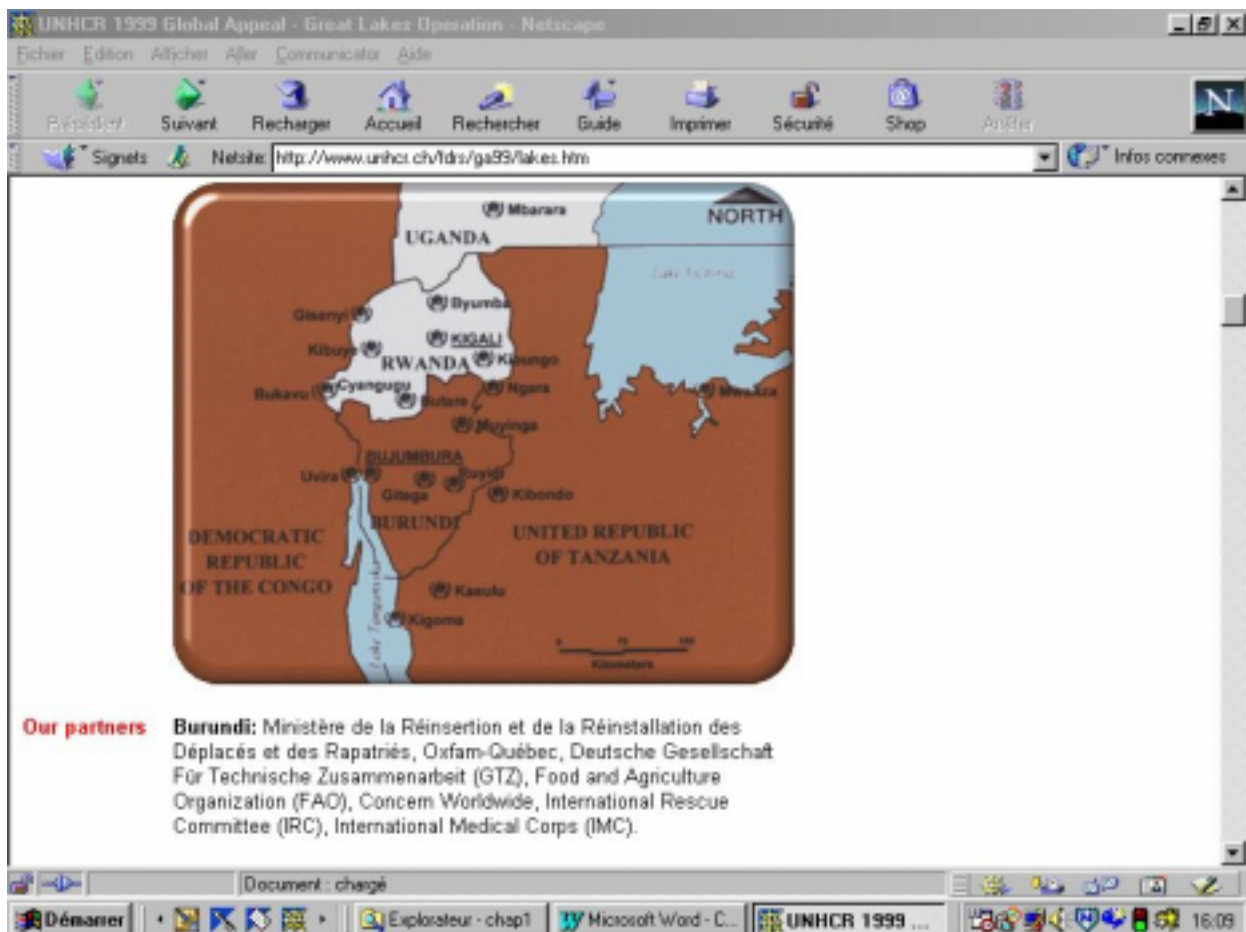
<http://www.international-alert.org/pdf/pubdev/tool.pdf>.

[http://www.capioli.cc/politica/poli\\_1en.htm](http://www.capioli.cc/politica/poli_1en.htm)

<http://www.frameweb.org/pdf/REDSO ETOA rev. 206-1-00.pdf>.

[www.odi.org.uk/hpn/newslet/acrobat/nl4e.pdf](http://www.odi.org.uk/hpn/newslet/acrobat/nl4e.pdf)

La figure 1.8 fournit un autre exemple de réponse à une requête au site du Haut Commissariat aux Réfugiés (<http://www.unhcr.ch/>).



**Figure 1.8 : Exemple de document répondant à la requête**

### 1.3.4. RECHERCHE AVANCÉE

#### A. Mise à jour de la recherche

L'utilisation d'un outil de type "Push" comme C4U (voir partie 1.2.) permet de surveiller automatiquement l'évolution des pages pertinentes. Par ailleurs, il est capable de renouveler une même requête et d'annoncer des mises à jour.

Les agents de recherche automatique permettent de scruter la toile en profondeur mais, du fait de leur rapide évolution, il est difficile de prévoir quelle en sera, pour un particulier, l'utilisation future. Le site Agentland (<http://www.agentland.fr/>) est spécialisé dans la présentation d'agents intelligents destinés au particulier.

#### B. Prise de contact directe

Les groupes de discussion et les « mailing lists » (voir partie 1.1.) forment des communautés virtuelles qui permettent de communiquer directement avec des personnes partageant les mêmes centres d'intérêt. L'annuaire de listes (<http://www.francopholistes.com/>) s'attache à rassembler les organisations francophones. Comme l'anglais est aujourd'hui la langue la plus largement utilisée sur la toile, le nombre de thématiques traitées par des listes anglophones est naturellement plus important. Le mot clé "*mailing list*" associé aux mots clés du sujet permettent de localiser des listes à l'aide d'un moteur de recherche.

Pour affiner encore la recherche, il est conseillé de contacter directement les auteurs des pages pertinentes, ces derniers fournissent généralement une adresse électronique à cet effet.

### 1.3.5. CONSERVER L'INFORMATION

#### A. Imprimer une page web

Les navigateurs Internet actuels disposent de réglages de l'impression. Il est possible de choisir d'imprimer la totalité du document, le texte seul sans les images associées (afin d'économiser l'encre de l'imprimante) ou bien encore un cadre (ou frame) particulier du document Internet.

#### B. Sauvegarder en partie les informations d'une page

La première solution est de conserver l'adresse de la page dans les signets personnels (bookmarks ou favoris).

La procédure du "*copier / coller*" fonctionne très bien pour sauvegarder du texte. Après l'avoir sélectionné, il suffit de copier l'information depuis le navigateur *Internet* (*fenêtre Edition, onglet "copier"*) et de la coller dans un logiciel de traitement de texte (*fenêtre Edition, onglet "coller"*). Par cette procédure, la mise en page n'est pas conservée et la présence de cadres multiples peut compliquer l'opération. Il est possible de sauver le texte au format de lecture du navigateur (*fenêtre fichier, onglet "enregistrer sous"* du navigateur). Le fichier résultant est lisible en local.

Pour sauvegarder des documents associés tels que les images, les navigateurs les plus courants permettent l'enregistrement de l'objet en cliquant avec le bouton droit de la souris. Pour les fichiers disponibles au téléchargement (« download » en anglais), la

procédure est très simple. En cliquant sur l'objet à récupérer, le navigateur demande le répertoire de destination sur l'ordinateur.

### C. "Aspirer" un site Internet

Généralement, l'architecture des pages Internet n'est pas prévue pour faciliter la sauvegarde de l'information sur les ordinateurs locaux car le texte et les images qui constituent la page n'appartiennent pas forcément au même répertoire. Pour reconstituer la copie exacte d'une page sur votre ordinateur, il faut donc recréer complètement l'organisation des fichiers texte et image. Les "aspirateurs" de site Internet sont conçus pour réaliser ce travail.

**Pour vous exercer,** Voir, par exemple :

Web downloader : <http://www.pressibus.org/windows/internet/frwebdown.html>

Teleport Pro : <http://www.tenmax.com/>

Memoweb : <http://www.goto.fr/>

### Sites géographiques en français

<http://www.cybergegeo.presse.fr>

<http://www.libergeo.prd.fr>

<http://www.libercarto.prd.fr>

<http://geomatique.georezo.net/>

<http://www.ign.fr>

### Quelques sites pour obtenir des données en ligne

<http://www.insee.fr>

données statistiques

<http://www.ifen.fr>

données utilisation du sol (CORINE Land Cover)

<http://www.meteo.fr>

site de Météo France

<http://www.giscale.com>

données géographiques mondiales

<http://data.geocom.com/>

base de données cartographiques mondiales (DCW)

<http://apps.fao.org/>

données agricoles (monde)

<http://www.grid.unep.ch/data>

données O.N.U.

<http://www.odci.gov/cia/>

« World Fact Book »

<http://urfist.univ-lyon1.fr/gratuits/>

nombreuses adresses thématiques référencées

<http://www.crige-paca.org/>

données géographiques Provence-Cote d'Azur

# **CHAPITRE 2 : LE MULTIMEDIA**

*Emmanuel Paul et Françoise Allignol*

Tout un chacun a entendu ou entendra les mots « multimédia » et « hypertexte ». En toute première analyse, le multimédia désigne l'utilisation de textes, sons, images, images animées dans la communication d'un contenu tandis que l'hypertexte est une structuration de l'information qui en permet des parcours variés.

## **2.1 QU'EST CE QUE LE MULTIMEDIA ?**

tout comme l'internet, le cédérom est un vecteur important de contenu multimédia, aussi bien pour les jeux informatiques, pour les logiciels que pour les produits didactiques (grâce à son bas prix et son importante capacité de stockage).

### **2.1.1 CEDEROMS ET ORGANISATION HYPERTEXTE**

#### **A. Dans la jungle du multimédia**

Depuis quelques années, tout le monde parle de multimédia ou encore de "produits multimédia". Dans le même temps, un grand nombre de publications spécialisées sont apparues dans la presse ainsi que des ouvrages de vulgarisation. Comment s'y retrouver ? Au delà des définitions plurielles, le multimédia c'est quoi ?

#### **B. Petite histoire**

Par manque de normalisation, il n'existe pas de définition officielle du multimédia.

Certains auteurs parlent seulement de la possibilité d'utiliser simultanément des **informations de différentes natures** : texte, son, image, vidéo etc. D'autres auteurs introduisent en supplément la notion de support physique.

Un "Média", en France, est un support d'information (un journal, une chaîne de télévision) et on appelle "les médias" l'ensemble des activités qu'y s'y rapporte. Un "produit multimédia" est la réunion sur un même support de type disque compact du son, de l'image et du texte. On parle alors de **multimédia « off line »**, c'est à dire d'un média fonctionnant de manière autonome sur un ordinateur, sans connexion à l'Internet. On lui oppose le **multimédia « on line »**, contenu informationnel que livrent à distance des réseaux numériques appelés couramment autoroutes de l'information ou Internet.

Aujourd'hui, le formidable essor que connaissent ces domaines d'activité a conduit à une acceptation plus large du terme qui englobe la télévision interactive ou même le téléphone mobile, surtout celui de seconde génération (« WAP ») qui combine téléphonie mobile et Internet.

#### **C. Multimédia off line pour apprendre : les CD-Roms ou cédéroms**

Le cédérom (pour être en accord avec l'académie française) ou CD-Rom (Compact Disc-Read Only Memory c'est à dire disque compact en lecture seule) est un support de multimédia *off line* qui s'est banalisé dans les foyers et dans les entreprises. Il doit son succès initial aux jeux mais les programmes éducatifs sont en très forte augmentation. Leur interactivité est particulièrement adaptée à un enseignement personnalisé..



## 2.1.2. PRINCIPES DE CONSTITUTION ET DE NAVIGATION

### A. Eléments constitutifs d'un outil multimédia

#### - *Texte*

Tout comme dans un livre imprimé, le texte est un élément essentiel des produits multimédias. Il matérialise la pensée et permet, entre autres, d'accumuler la connaissance et de transmettre, diffuser et organiser l'information. Cependant il est généralement présent en moindre quantité que dans les livres imprimés car il cède plus facilement la place à des schémas explicatifs et surtout à des images.

#### - *Images*

Les images intégrées dans un produit multimédia peuvent être créées directement sur un micro-ordinateur équipé de logiciels de dessin et de peinture, ou importées sous la forme de photographies numérisées par un scanner. Les images de synthèse et les fractales ont fait récemment leur apparition. Ce sont de pures créations mathématiques qui permettent pour les premières de représenter tout ce qui n'est pas visible à l'œil nu (infiniment grand ou infiniment petit). Les fractales réduisent les formes en équation et sont particulièrement efficaces pour modéliser des reliefs ou des paysages accidentés virtuels. Mais les images fixes sont bien souvent insuffisantes pour représenter une information, donner une explication ou même pour faire rêver. Les techniques informatiques du dessin animé ou de l'animation cartographique permettent de créer des séquences graphiques. Ces séquences sont très utiles pour montrer, par exemple, l'évolution géomorphologique d'un petit bassin versant alpin soumis à une érosion importante ou la concentration séculaire de la population dans les zones urbaines.

#### - *Vidéo*

Au delà des animations graphiques, la vidéo avec sa bande son synchronisée est le support privilégié de l'image animée. La vidéo numérique tend à succéder à la vidéo analogique des caméras et des caméscopes. Elle présente l'avantage de fournir des séquences directement intégrables sur un support multimédia. De plus, l'imagerie numérique est modifiable et « montable » à volonté. Son inconvénient majeur est le volume important de données qu'elle produit. Ce volume rend tout d'abord les opérations de traitement délicates et rend ensuite nécessaire le stockage sous format compressé.

#### - *Son*

Les effets sonores revêtent une grande importance dans un produit multimédia. Ils accompagnent et rythment la progression de l'utilisateur dans le produit. Ils peuvent aussi avoir un rôle informationnel très précis, par exemple sur les CD-Roms encyclopédiques : musique folklorique d'un pays, texte parlé, ambiances sonores urbaines etc. Tout comme les images, le son dans les produits multimédias est intégré sous forme numérique et pose les mêmes problèmes de compromis entre qualité de restitution et volume de stockage.

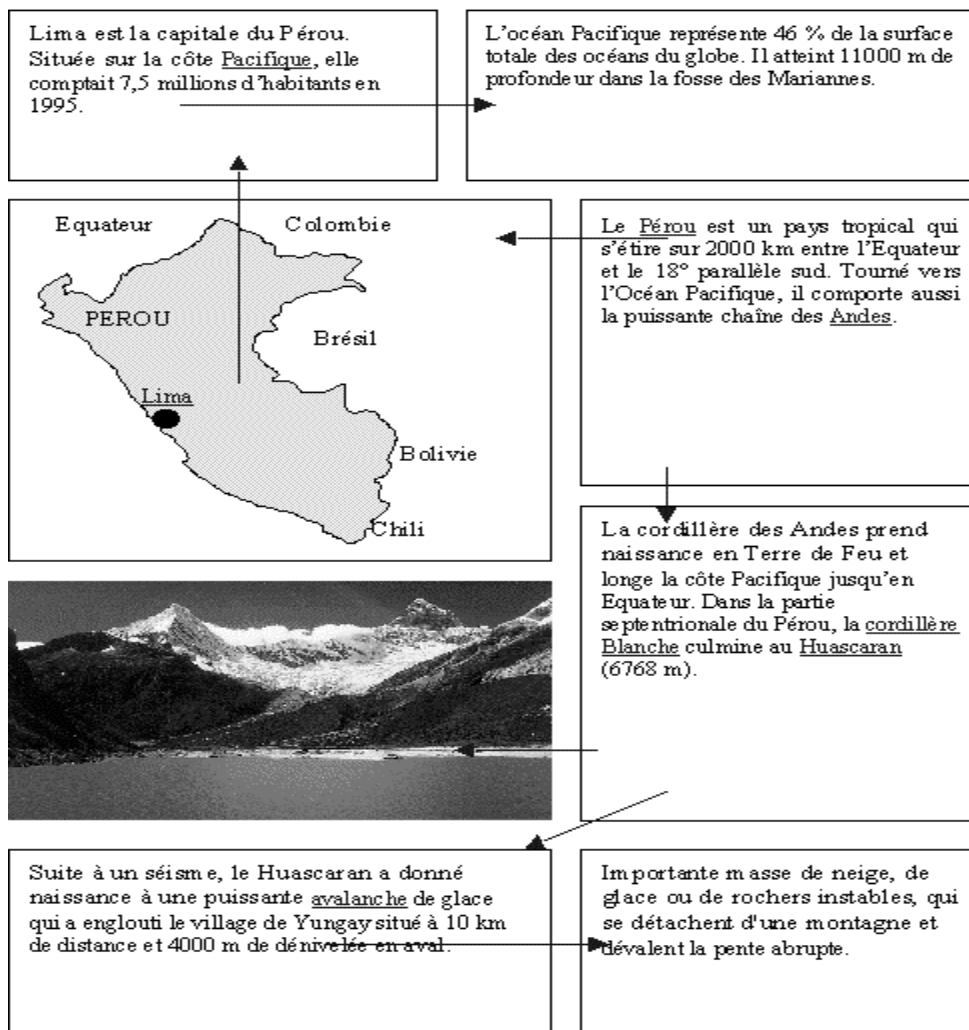
Sons, images, séquences vidéo représentent de grandes quantités de données. Par exemple, une image non compressée représente couramment quelques millions d'octets (caractères). Même avec les grandes capacités de stockage des supports actuels (disques, CD-Roms, ...), les images et les sons doivent être compressés (ramenés à quelques pour cent de leur volume initial sans perte notable de qualité) et très rapidement décompressés. C'est la mise au point de formats tels MPEG et MPEG3 qui a permis l'essor du Web et du

multimédia pour de grandes quantités de données et avec des temps de réponse acceptables par l'utilisateur.

## B. Navigation : Liens interactifs

Traditionnellement, un article ou un livre imprimé imposent une lecture linéaire. L'hypertexte ou hypermédia (lorsque l'image et le son sont aussi utilisés), invite par un accès libre à l'information, au vagabondage parmi les connaissances. Le terme d'interactivité souligne la participation active du bénéficiaire d'une transaction d'information. Il lui devient possible d'explorer les connaissances, en sautant par exemple d'un concept à son illustration concrète ou à un autre concept voisin etc.

l'hypertexte est utilisé aussi bien pour les CD-Roms que pour le Web. Il permet de passer d'un sujet à un autre par le biais de liens constitués lors de l'analyse du contenu à transmettre. Certains de ces liens sont hiérarchiques (dossiers, sous dossiers, ...), d'autres ne le sont pas, permettant de passer, par exemple, d'un dossier à un autre. L'ensemble de ces liens prend la forme d'un **graphe sémantique**. Du grand nombre de liens et de la bonne organisation de ce graphe découle, pour l'utilisateur, une impression de cheminement « libre » dans le corpus informationnel. Concrètement, comme l'illustre la figure 2.1, un clic de souris sur un mot souligné ou une image « active » déclenche une action. Généralement, le curseur classique de la souris se transforme en main lorsque l'utilisateur passe sur l'élément actif, cela lui indique qu'il peut cliquer dessus et atteindre d'autres données. Les mots sur lesquels un hyperlien a été créé, ont une police de caractères ou une couleur spéciale qui permet de les individualiser immédiatement.





Deux exemples vont nous permettre d'illustrer l'usage possible, en géographie, du multimédia.

## **2.2. ENCARTA, UN EXEMPLE D'ATLAS INTERACTIF**

### **2.2.1 UN PRODUIT « MICROSOFT » POUR « WINDOWS 9x »**

Encarta est un atlas mondial interactif, édité sur cédérom par la société américaine Microsoft pour système d'exploitation Windows 9x. Le contenu de cet Atlas a été construit à partir d'une bibliographie regroupant des ouvrages sur les Etats et Régions du monde, des Atlas, des dictionnaires thématiques (par exemple, un dictionnaire de géopolitique). Des ouvrages de référence en géographie physique, dynamique du globe, géographie humaine et environnement ont été utilisés comme support du contenu des concepts fondamentaux de la géographie. Une dizaine de « conseillers », généralement des enseignants chercheurs en Université, de diverses nationalités (américains, canadiens, anglais, japonais, français, espagnols, allemands etc.) ont été sollicités pour valider le contenu informationnel de cet Atlas.

### **2.2.2. ACCES A L'INFORMATION**

La page d'accueil d'Encarta (cf. Figure 2.2) propose d'effectuer un premier choix au niveau du type d'information recherché :

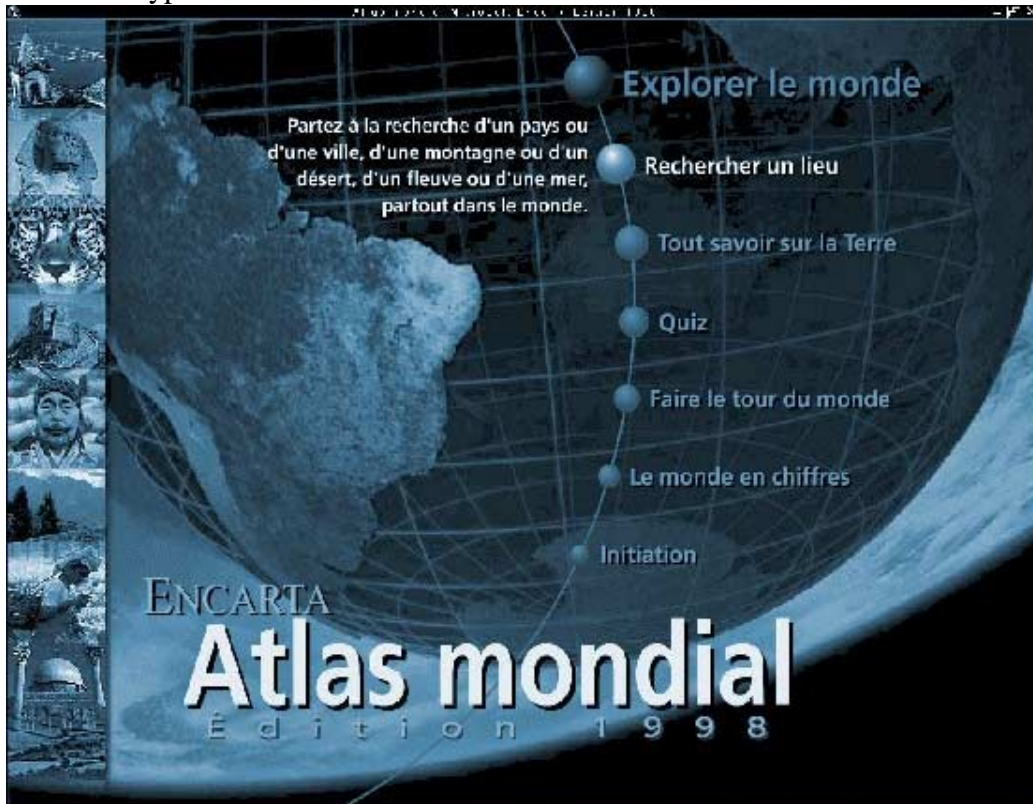


Figure 2.2 : page d'accueil d'Encarta présentant ses fonctionnalités (source: Microsoft)

- La fonction « Explorer le monde » permet d'accéder directement à toutes les fonctionnalités de l'Atlas présentées par un menu principal, un menu de gestion d'affichage (zoom, déplacement etc.) et une liste contextuelle d'onglets qui informe sur la nature des informations disponibles sur un lieu (cf Fig. 2.1).
- La fonction « Rechercher un lieu » place l'utilisateur dans le même contexte que précédemment avec un utilitaire de « recherche d'un lieu ». Elle est aussi accessible par le menu principal.
- La fonction « Tout savoir sur la terre » (sic !) ouvre un menu qui ressemble au sommaire d'un ouvrage de Géographie Générale. En cliquant sur les lignes, on accède aux articles correspondants. A gauche du texte, sont alignés des photos d'illustration qui peuvent être agrandies, des onglets qui listent les notions complémentaires pouvant compléter le sujet traité et éventuellement une vidéo. Le texte est en noir mais il comporte des mots écrits en rouge et d'autres en bleu. Ce sont des mots sur lesquels un lien hypertexte existe. Les mots en bleu sont des noms de lieu et en cliquant dessus, le texte est remplacé par une carte centrée sur le lieu en question. Les mots en rouge correspondent à des notions très précises de la Géographie. Ils donnent accès à une définition et le plus souvent à un schéma explicatif
- « Quiz » est un test ludique de connaissances en QCM à quatre niveaux sur des noms de villes et de pays.
- « Faire le tour du monde » propose de visiter sept régions du monde grâce à un simulateur de vol où l'on peut définir sa trajectoire, son altitude et sa vitesse (cf figure 2.3).
- « Le monde en chiffres » est un module de statistiques où l'on peut consulter, sur fond de carte mondiale, différents types de données portant sur l'agriculture, la population etc.. En déplaçant la souris, les chiffres du pays sur lequel on se trouve apparaissent à l'écran. Des définitions (telle que celle du produit intérieur brut) sont également accessibles. Les chiffres peuvent être visualisés sous la forme de courbes ou de tableaux.
- « Initiation » est une aide à l'utilisation d'Encarta qui permet d'apprendre à réduire, agrandir ou déplacer une carte, à rechercher un renseignement précis, ou à trouver toutes les informations disponibles sur un lieu.

### **2.2.3. POURQUOI UN ATLAS ELECTRONIQUE ?**

Par rapport à un atlas papier, un atlas électronique présente l'avantage d'offrir un accès à l'information plus facile, plus rapide et apparemment personnalisé. D'autre part, le multimédia donne la possibilité de visualiser des vidéos commentées sur des thèmes variés, comme la tectonique des plaques par exemple. Mais le contenu, commerce oblige !, est généralement « grand public » : les images sur les pays, la faune ou la flore sont assorties de bandes sonores (musiques du monde, cris d'animaux, ambiances sonores urbaines,...). A l'inverse de son équivalent papier, l'atlas électronique est davantage « anecdotique » que scientifique. En outre, les atlas électroniques sont limités par le volume maximum d'informations que peut contenir le support, ici le cédérom. Les sons

sont généralement de durée très courte et l'espace d'affichage des vidéos est réduit, ce qui restreint encore grandement leur capacité éducative. Il est cependant indéniable que ce type de produit est beaucoup plus ludique que son homologue papier, comme le montre l'exemple de la figure 2.3



Figure 2.3 : simulateur de vol Amérique du Sud (source: Microsoft)

On trouvera dans un second exemple, celui du projet **Ohrage** (responsable : M.Gaio) la description d'un programme de recherche inter-disciplinaire (informaticiens et géographes de l'université de Caen, géographes de l'université de Portsmouth) visant à construire et à maintenir (nombreuses et fréquentes mises à jour) un atlas électronique. Cet atlas **Trans-Manche** sera de type hyperdocument, c'est à dire qu'il contiendra textes et cartes régulièrement enrichis et qu'il sera de consultation aisée. On trouvera plus de détails sur le site <http://www.unicaen.fr/mrsh/modesco/Programmes.htm#OHRAGE>

## **2.3. COMMUNIQUER LA REVISION D'UN P.O.S.**

Ce cédérom a été créé à la demande de la municipalité grenobloise pour alimenter des bornes interactives présentant au public ses projets de modifications du Plan d'Occupation des Sols de la partie Nord Ouest de la commune.

### **2.3.1. MULTIMEDIA ET COLLECTIVITES LOCALES**

Les collectivités locales, comme les associations et les administrations nationales, prennent peu à peu conscience de l'intérêt à utiliser le multimédia pour communiquer avec les citoyens. Les projets et réalisations de cédéroms ou de sites internet se sont multipliés essentiellement à partir de 1996 aussi bien en milieu urbain qu'en milieu rural et quelle que soit la taille de la collectivité. Les thèmes moteurs les plus fréquents de ces réalisations sont la promotion touristique et culturelle et le développement économique.

C'est aux mairies que revient la responsabilité de donner la possibilité à tout un chacun d'avoir accès à des ordinateurs et à l'information, quelque soit le niveau social ou d'éducation. Cela passe en général par la mise à disposition d'ordinateurs en libre service dans les lieux publics comme les bibliothèques municipales.

## **2.3.2 OBJECTIFS ET ENJEUX**

### **A. L'exemple grenoblois**

Grenoble a réalisé un cédérom d'information sur la révision du Plan d'Occupation des Sols (P.O.S.) du secteur nord-ouest de la ville. C'est un exemple concret d'effort de communication, via le multimédia, de la part des élus et des techniciens. Le sujet est particulièrement difficile pour tous ceux qui ne sont pas spécialistes en matière d'urbanisme et de droit des sols. Pourtant, il était nécessaire d'impliquer les habitants du secteur dans cette révision, car leur avis allait être sollicité dans le cadre d'une « enquête publique ». Comment donner un avis positif ou négatif sur un document technique complexe si l'on n'a pas compris à quoi il sert et quel en est l'enjeu ? D'ailleurs, la loi impose que les personnes concernées puissent s'informer sur les raisons d'une telle procédure et obtenir des renseignements précis sur les projets urbains à venir. Il s'agit de garantir les droits des propriétaires et des usagers. Information du public ne signifie pas seulement mise à disposition des documents d'urbanisme qui sont particulièrement hermétiques pour un non spécialiste. Sans compréhension de la procédure, ni justification des décisions et sans exposé préalable de la politique de préservation du patrimoine urbain, les personnes concernées par la révision du P.O.S. ne sont pas susceptibles de donner un avis argumenté. C'est la raison pour laquelle la mairie de Grenoble a opté pour une présentation multimédia globale. Ce projet s'inscrit dans la politique de la ville concernant les nouvelles technologies. En effet, la réalisation de cette présentation multimédia coïncide avec la parution de la charte graphique et avec la présentation officielle du site web grenoblois.

### **B. Diffuser l'information**

Impliquer le citoyen dans l'aménagement de son territoire est nécessaire pour aboutir à une concertation effective. En conséquence, il faut lui proposer un outil actuel d'information qui lui rende accessible l'ensemble de la problématique et des enjeux du P.O.S. Ceci ne peut se faire qu'en élaborant un modèle qui sert de support à la communication. Le choix de la ville de Grenoble s'est porté sur une présentation multimédia conviviale qui permette la vulgarisation du dossier à plusieurs niveaux selon l'intérêt estimé de différents types de publics. Cette présentation a pour support un cédérom et elle a été mise à la disposition des habitants du quartier sur trois bornes multimédia placées en libre service dans des endroits publics, pendant toute la période de l'enquête publique. L'utilisateur est confronté, au cours de sa visite de la borne, à un schéma de la révision du P.O.S. qui correspond à une représentation abstraite et formalisée dégagée par les auteurs du contenu pour servir de support à son investigation. Cela doit l'aider à comprendre et à élaborer des réponses aux questions qu'il se pose ou, plus modestement, l'amener simplement à se poser des questions. Ce projet initié par la ville de Grenoble est novateur en ce sens que c'est la première commune à engager une action multimédia pour intéresser le citoyen à l'urbanisme. La ville souhaite informer ses habitants et les amener à jouer un rôle actif dans la production du P.O.S. révisé.

## 2.3.3 DISPOSITIFS TECHNIQUES

### A. Borne multimédia ?

Une borne se présente sous la forme d'un écran d'ordinateur sur lequel on peut agir en appuyant sur des images de boutons, par exemple (c'est ce que l'on appelle un écran tactile). Si l'écran n'est pas tactile, alors on agit sur l'application multimédia grâce à un « track pad », qui remplace la souris.

### B. Outil informatique de création multimédia

L'outil informatique multimédia choisi pour réaliser les bornes est « Toolbook ». L'application hypermédia (du type encyclopédie en ligne) résultante est exécutable sous système d'exploitation « Windows » qui équipe la plupart des ordinateurs personnels. « Toolbook » utilise la métaphore du livre comme base d'une application (chapitres, parties, sous parties, paragraphes,...). L'avantage de cette organisation est illustrée par la possibilité de créer et d'utiliser des techniques comme l'hyperliaison et l'animation. Une application comporte plusieurs livres. Comme un livre imprimé, un livre « Toolbook » est divisé en pages représentant les écrans de l'application. Les objets qui peuvent être intégrés à une page sont des champs de texte et d'hypertexte, des images, des boutons de navigation, des clips vidéo et du son.

### C. Cibles

Pour construire une application multimédia, il faut d'abord définir à qui elle va s'adresser. Trois types d'utilisateurs ont été définis selon leur niveau de compétences en droit, urbanisme, architecture : enfants des écoles du secteur, habitants et professionnels.

- Le premier type d'utilisateur n'a aucune compétence et il faut redéfinir chaque notion simplement en soignant l'interface graphique et en privilégiant les animations. Il s'agit du premier niveau d'information.
- D. Organigramme informatique Le deuxième type d'utilisateur possède une expérience pratique de la ville qui est son lieu de vie et qui se traduit par une perception, négative ou positive. L'habitant consulte la borne car il a besoin de renseignements précis qui touchent à son quotidien. Il s'agit d'un deuxième niveau d'information, optionnel par rapport au premier niveau.
- Le professionnel, ou l'étudiant futur professionnel, vient chercher sur la borne des informations très ciblées pour une étude précise. Il s'agit du troisième niveau d'information, optionnel par rapport au premier et au deuxième niveau.

L'organigramme informatique général des bornes P.O.S.(Tableau 2.1) a été créé de manière à respecter les impératifs de navigation de la charte graphique de la ville de Grenoble :

- 1 - Permettre à l'utilisateur de revenir à tout moment à l'index de l'application et d'accéder directement à la séquence de son choix sans être obligé de dérouler l'arborescence (toute la « table des matières »),
- 2 - Créer des séquences linéaires accessibles par des menus, sur une structure de type arborescent. Pour l'utilisateur, cette structure a l'avantage d'être simple au niveau de la représentation et donc du repérage.



P.O.S.	qu'est-ce que le P.O.S. ?	à quoi sert le P.O.S. ?
		de quoi est constitué le P.O.S. ?
		qui est concerné par le P.O.S. ?
		où se procurer le P.O.S. ?
	Elaboration du P.O.S.	Chronologie
	Révision du P.O.S.	Définitions à partir des mots soulignés
		Chronologie
Grenoble	Historique	Définitions à partir des mots soulignés
		Historique de la ville
	P.O.S. de Grenoble	Photo aérienne ➤ secteurs ➤ zonage
	Grenoblois en chiffres	démographie
		logement
		équipements
	Grenoble ville active	Les emplois
		Les secteurs
		Les entreprises
		L'implantation des emplois
Présentation du secteur nord-ouest	Le P.O.S. révisé	Plans ➤ P.O.S. ➤ secteur NO ➤ sous secteurs ➤ zonage ➤ photos ➤ grands projets
	Les grands projets	Lustucru
		Europole
		Bouchayer/Viallet
	Préservation du patrimoine	La naissance d'un quartier
		B4 protection du patrimoine et paysages
	Le quartier en chiffres	démographie
		logement
		équipements
	La vie économique du quartier	Les emplois
		Les secteurs
		Les entreprises
		L'implantation des emplois

Tab. 2.1 : Organisation générale de la borne « POS du N.O. Grenoblois » (source: M.Masson et F.Allignol)

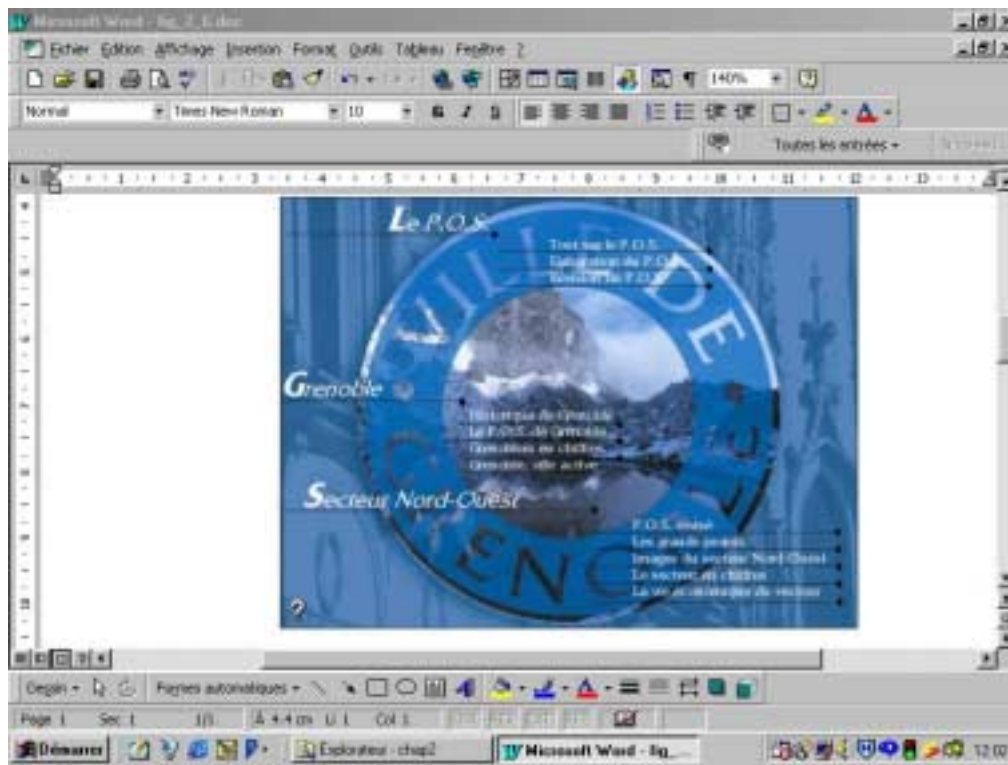


Figure 2.4 : Page d'introduction de la borne P.O.S. - Grenoble

Cette organisation par étages permet d'optimiser les chances de chacun d'arriver au bout de la présentation car il évite toutes les parties techniques rébarbatives.

## E. Contenu

La borne traite trois thèmes qui couvrent l'essentiel des connaissances nécessaires pour comprendre ce qu'est un P.O.S., les enjeux de la révision, la procédure, l'histoire de Grenoble (cf. Figure 2.5) et du quartier et leurs caractéristiques socio-économiques actuelles : le document P.O.S., la présentation de la ville de Grenoble et la présentation du secteur nord-ouest. Ces thèmes sont accessibles par l'index (cf. Figure 2.4) qui est, lui même, accessible de toutes les pages de la borne. C'est cet index qui permet à l'utilisateur de se resituer dans le contenu à tout instant.

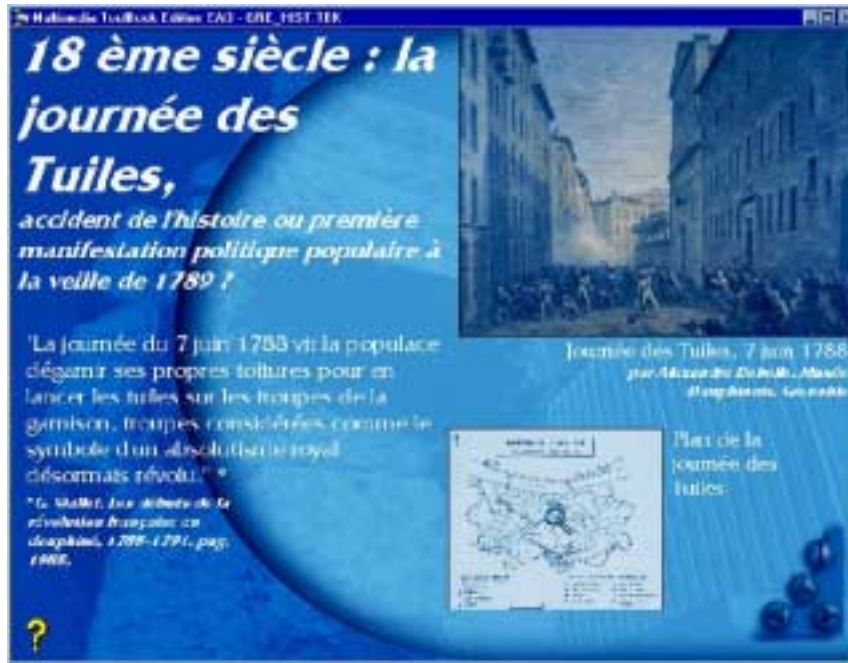


Figure 2.5. : exemple d'illustration du chapitre « histoire de Grenoble »

La navigation dans le document est à la fois thématique et géographique multiéchelles. La figure 2.6 montre ainsi qu'il est possible à partir d'un plan général de cliquer sur l'un des boutons pour obtenir une vue plus détaillée.

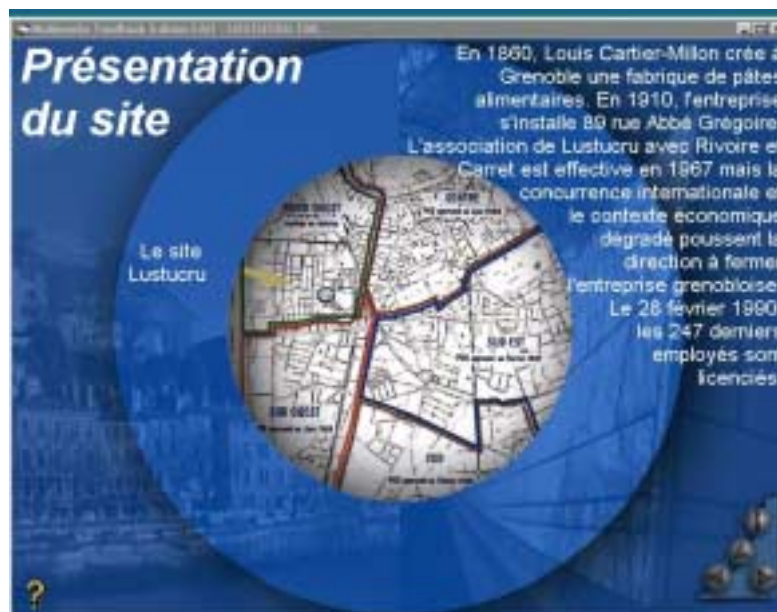


Figure 2.6: loupe menant à une vue plus détaillée

La figure 2.7 , quelques clics et quelques pages plus loin dans l'arborescence, montre à quel type d'information détaillée peut aller l'utilisateur sur tel ou tel projet.



Figure 2.7 : Aménagement retenu pour l'ilôt « Lustucru »

La figure 2.8 présente, quant à elle, la page d'introduction d'un aménagement de plus grande ampleur, celui du quartier Europole. Chacune des pastilles, numérotées de 1 à n , permet d'obtenir (comme dans le cas précédent), des détails supplémentaires .



Figure 2.8: page d'introduction du chapitre Europole



## Pour aller plus loin

[http://www.cyberformateur.com/biblio/biblio\\_ini.htm](http://www.cyberformateur.com/biblio/biblio_ini.htm)

## Référence citée

- Les figures 2.4 et suivantes sont extraites de la borne interactive « *Révision du P.O.S. du secteur Nord Ouest de la ville de Grenoble* » (F.Allignol et M.Masson-Vincent)
- M. Masson-Vincent : *Géographie et citoyenneté : :quels liens ?* (Faculté des Lettres d'Avignon, Actes du colloque Géopoint98, 1999)
- J.C. Muller : *Multimédia et aménagement du territoire* (Actes du Festival International de Géographie de St Dié, 1999)

## **CHAPITRE 3 : BASES DE DONNEES**

*Pierre Dumolard*

On appelle ici « **Base de Données** » (B.D. en abrégé) tout fichier ou ensemble de fichiers contenant des données relatives à une même application. Ces données sont arrangées sous une forme assimilable à des tableaux.

On distingue ici deux formes de B.D. :

- Des B.D. « élémentaires », composées d'un seul « tableau » relatif à un seul type d'entités,
- Des B.D. « relationnelles », décrivant les relations entre plusieurs « tableaux », chacun relatif à un type d'entités.

### **3.1. BASES DE DONNEES ELEMENTAIRES**

Les bases de données (B.D.) élémentaires sont composées **d'un seul fichier**, qu'on peut comparer à un **tableau** unique.

#### **3.1.1 TYPES DE TABLEAUX**

Un tableau peut être :

- symétrique (lignes et colonnes jouent le même rôle et peuvent donc être permutées),
- dissymétrique (lignes et colonnes jouent un rôle différent).

##### **A. tableaux dissymétriques**

Les lignes et les colonnes n'y ont pas même signification et jouent un *rôle différent*.

- Les *lignes* du tableau se réfèrent à des **entités** de même nature : entreprises, clients, tronçons routiers, rivières, régions,... Leur ensemble forme une **classe d'entités** : une branche d'activité, la clientèle d'une entreprise, un réseau routier ou hydrographique, un pays. Chaque ligne du tableau est précédée d'un **identificateur** (nom, numéro,...) unique, particulier à chacune. En géographie, les entités, souvent territoriales (on parle aussi d'**unités spatiales**) sont généralement identifiées par un nom et/ou un code. Par exemple, les départements français sont repérés par leur nom et/ou leur code minéralogique. Cet identificateur est unique : à chaque département correspond un numéro ou un nom qui lui est propre.

- Les colonnes correspondent aux divers **attributs** (ou descripteurs, ou **variables**) décrivant les entités : pour un fichier clients, par exemple, nom, adresse, produits et quantités commandées, dates, etc. Chaque colonne est « chapeautée » par un nom qui indique sa signification.
- Chaque case du tableau contient une **valeur** correspondant à une entité et à un attribut, numérique (ex. quantité commandée) ou alphabétique (ex. nom ou code du produit). Chaque colonne a même codage, chaque ligne pouvant juxtaposer des codages différents.

Exemple fréquent en géographie : des tableaux où les lignes concernent des lieux (départements, par exemple) et les colonnes des descripteurs, identiques pour tous les lieux. Un tel type de tableau est appelé « **matrice d'information spatiale** ». Dans le cas évoqué ci-dessus, le tableau aura 96 lignes et autant de colonnes que de descripteurs.

A titre d'exemple, le tableau (dissymétrique) ci-dessous est extrait d'un fichier relatif à une commune de Savoie. L'extrait concerne une partie de son territoire. Les 22 entités (en ligne) sont des parcelles d'occupation du sol, les 5 attributs (en colonne) qui les décrivent sont leur identificateur, leur mode d'occupation du sol (M.O.S., en 21 classes), leur niveau d'exposition au bruit (en décibels), leur surface (en ha) et leur population. La plupart des attributs sont numériques, un seul est alphabétique (M.O.S.).

Numéro	M.O.S.	BRUIT	Surface (ha)	Habitants
01	Z.I. ouverte	62	14,4	0
02	Z.I. fermée	56	8,1	0
04	Zone comm.	64	3,4	0
06	Z.I. ouverte	65	5,0	0
07	Habitat dense	70	2,2	269
08	Habitat diffus	68	4,2	143
09	Habitat dense	65	5,1	626
11	Loisirs ext.	66	6,8	0
12	Santé	58	1,7	0
14	Scolaire	55	1,0	0
17	Habitat dense	68	1,3	191
19	Loisirs int.	75	0,4	0
21	Religieux	60	0,6	0
22	Habitat dense	62	7,0	605
24	Bat. public	65	0,4	0
25	Bat. public	55	0,6	0
26	Habitat dense	58	1,0	92
28	Z.I. fermée	70	14,9	0
29	Loisirs ext.	55	0,9	0
30	Habitat dense	65	0,2	38
31	Terrain vague	62	0,6	0
36	Habitat diffus	58	4,2	106

Tableau 3.1 : extrait d'une B.D. élémentaire

Le modèle général des matrices d'information spatiale est :

## Entité → Valeur 1, ..., Valeur p

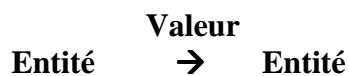
Ce schéma signifie que p valeurs correspondent à chaque entité : ce type de tableau décrit généralement des **états** (éventuellement successifs) d'où la dénomination de « variables d'état ».

### B. tableaux symétriques

Dans un tableau symétrique, lignes et colonnes jouent un *rôle équivalent*.

- Le premier cas de tableau symétrique est celui où figurent en lignes et en colonnes des **entités** et dans chaque case **une valeur d'attribut**, par exemple un nombre de migrants « alternants » entre communes de domicile et de travail (lignes et colonnes sont des unités spatiales). Ce sont des tableaux de **flux**.

Leur modèle général est :



Ce schéma résume un type de relation entre couple d'entités et valeur d'un attribut.

- Le second cas est celui de tableaux « croisés » où lignes et colonnes sont des **valeurs de deux attributs**, chaque case contenant généralement un effectif (**nombre d'entités**).

Leur modèle général est l'inverse du précédent :



Les tableaux de **contingence** de la statistique bivariée sont de ce type.

### 3.1.2 IMPLEMENTATION INFORMATIQUE

Un tableau dissymétrique comme le tableau 3.1 est une B.D. mono-fichier qui pourra être stockée en machine de diverses manières.

#### A. Types de fichiers

- Le tableau 3.1 peut être stocké sur disque comme une suite de 5 valeurs, d'abord pour l'entité 01, puis pour la 02, etc. Chaque ligne est un **enregistrement** d'un fichier, dit **séquentiel** parce que composé d'une suite d'enregistrements. L'inconvénient de cette structure est qu'il faut parcourir tous les enregistrements pour accéder au dernier. L'avantage, par contre, est que ce type de fichier est lisible, donc aisément vérifiable et modifiable par l'utilisateur (avec un éditeur de textes comme « Notepad » ou un

traitement de textes comme « Word »), parce que le codage des identificateurs et des valeurs est sous forme de caractères, d'où le nom de **fichier texte**.

- Il existe aussi des fichiers à **accès direct** où chaque valeur peut être accédée plus rapidement mais est codée en binaire (et n'est donc pas directement lisible sous éditeur ou traitement de textes). L'avantage de la rapidité d'accès est évident, comme est évident l'inconvénient d'absence de lisibilité directe du contenu du fichier (à cause du codage booléen).

## B. Opérations possibles dans une B.D. mono-fichier

Une B.D. n'a quelque utilité que si l'utilisateur peut lui adresser des « questions » (des **requêtes** en jargon B.D.) et en obtenir des réponses. Pour que réponse il y ait, il faut qu'existent des **opérateurs** combinant des données élémentaires.

Outre tri, lecture, modification, écriture, sélection d'attributs et d'entités de la B.D., trois types d'opérateurs permettent de combiner des valeurs.

- Les opérateurs **arithmétiques** ( +, -, \*, /, puissance) combinent des **valeurs numériques quantitatives**, soit d'un même attribut (exemple : addition des surfaces des entités 24 et 25 du tableau 3.1), soit de 2 attributs (exemple : calcul de densité en divisant population par surface, dans le même tableau). Le résultat de l'opération est **numérique quantitatif**.
- Les opérateurs de **comparaison** ( < , = , > ) comparent des **valeurs numériques ou alphabétiques** d'un même attribut ou d'attributs différents. Par exemple, la parcelle 04 a une surface inférieure à la parcelle 11, le M.O.S. « Scolaire » est réputé supérieur au M.O.S. « Santé » car alphabétiquement classé après. Le résultat d'une comparaison est **binaire** (vrai / faux).
- Comme dans le chapitre 1, les opérateurs **logiques** de base sont : ET, OU, NON. Ils mettent en relation des **valeurs binaires** (le plus souvent issues d'une comparaison) et permettent, entre autres, de sélectionner les entités ayant mêmes valeurs sur plusieurs attributs. Exemple : créer un sous tableau de celles dont le M.O.S. est « Habitat dense » **ET** la population est supérieure à 100 habitants. Le résultat d'un questionnement logique est **binaire** (vrai / faux).

Les divers types d'opérateurs sont, bien évidemment, combinables pour composer des requêtes complexes.

### 3.1.3. EXEMPLES DE REQUETE A LA B.D. MONO-FICHIER

Les exemples ci-dessous ne sont que quelques uns des possibles.

#### a. Requêtes avec opérateurs arithmétiques

- Calculer la densité (en Hb/Ha) des zones : Habitants / Surfaces  
Le résultat (0, 0, 0, 0, 122, 34, ...) peut être stocké comme attribut supplémentaire.

- Calculer le nombre moyen d'habitants par zone : somme des habitants / nombre de zones

Le résultat est ici un nombre et non un attribut (colonne) supplémentaire.

#### b. Requêtes avec opérateurs de comparaison

- Sélectionner les zones dont MOS = « Habitat dense »

Le sous tableau renvoyé aura 6 entités (lignes) correspondant aux zones 07, 09, 17, 22, 26, 30.

- Sélectionner les zones dont Habitants > 0

Le sous tableau comprendra les zones 07, 08, 09, 17, 22, 26, 30 et 36.

#### c. Requêtes avec opérateurs logique et de comparaison

- Sélectionner les zones dont (BRUIT > 65) ET (Habitants > 0)

(65 décibels est, en effet, la limite légale d'exposition de personnes au bruit).

Le sous tableau renvoyé correspondra aux 3 zones 07, 08, 17.

- Sélectionner les zones dont (BRUIT > 70) ET (Surface > 10)

(plus de 70 décibels définit un point noir sonore).

La sélection comprendra seulement la zone 28.

Les résultats de ces requêtes sont souvent stockés comme colonnes (attributs) supplémentaires dans le tableau et utilisés dans des requêtes ultérieures : une requête complexe est, en général, divisée en requêtes simples successives dont on peut conserver les résultats intermédiaires. Elle peut également être composée directement en utilisant l'ordre de priorité des opérateurs (logique > comparaison > arithmétique) et, éventuellement, des parenthèses.

Le Tableau 3.1 permet de répondre « de visu » à de nombreuses questions, parce que c'est un court extrait d'une vraie B.D. sur une petite commune ! Imaginez ce qu'il en serait avec de nombreuses zones et de nombreux attributs ! Dans ce cas, l'usage d'un gestionnaire de fichiers (comme dBase) ou d'un tableur (comme Excel) est le bienvenu !

## **3.2. BASES DE DONNÉES RELATIONNELLES**

On appelle **base de données relationnelle** une B.D. constituée de plusieurs tableaux (classes d'entités) dont on a modélisé les **relations**.

### **3.2.1 DEFINITIONS : NIVEAUX DE MODELISATION**

#### **A. Les spécialistes distinguent 4 niveaux de modélisation.**

- le **modèle externe** (à l'informaticien) représente la vision que se fait un géographe d'un problème (défini, précisé et épuré),
- le **modèle conceptuel**, sous forme de schéma, en est une traduction selon un formalisme (par exemple le formalisme « Entités / Association »),
- le **modèle logique** est l'implémentation informatique, utilisant un langage d'interrogation (par exemple, le langage « SQL »),
- le niveau **interne** concerne l'organisation des données en machine (il est du ressort du logiciel B.D. lui même).

Les trois premiers niveaux de modélisation vont de la perception d'une application par un géographe à son implémentation en machine sous forme de B.D. relationnelle (tableau 3.2).

<i>Niveau de modélisation</i>	<i>concepteur</i>
« externe »	géographe
	géographe
conceptuel	+
	informaticien
logique	informaticien
interne	système

Tableau 3.2 : Niveaux de modélisation d'une B.D. relationnelle

Le modèle conceptuel d'une B.D. relationnelle est donc le lieu d'interface et d'interaction entre géographe et informaticien. Il met en œuvre un formalisme (« Entités / Association » ou, plus récemment, « Objets »).

Soit, par exemple, dans la commune savoyarde déjà évoquée au § 3.1, une B.D. à créer pour la révision du Plan d'Occupation du Sol (P.O.S.). On pourrait, bien sûr, ajouter des colonnes supplémentaires au tableau 3.1 (zone d'appartenance P.O.S., Coefficient d'Occupation du Sol, etc.). Mais la commune comportant des milliers de parcelles (et pas des milliers de zones P.O.S.) chacune décrite par de nombreux attributs, stocker toutes les données dans un fichier unique aboutirait à un fichier gigantesque, plein de redondances (donc de possibilités d'erreurs) et mal maîtrisable.

Il faut donc passer d'une représentation « naïve » des données à une autre qui en dégage les seuls classes d'entités, attributs et relations entre classes qui soient strictement **nécessaires et suffisantes** pour répondre à l'ensemble des requêtes des utilisateurs. C'est, en quelque sorte, l'équivalent d'une « **mise en facteurs** » de l'information.

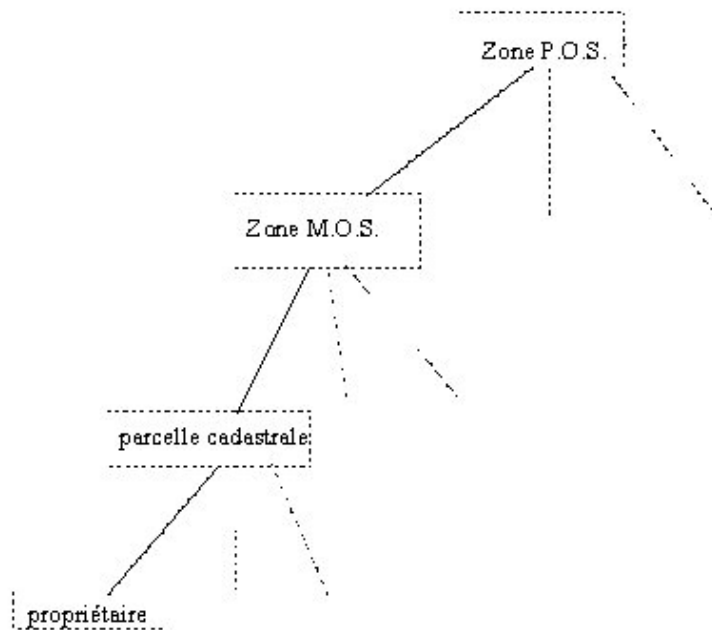
Dans notre exemple P.O.S., le modèle externe peut consister simplement à lister les classes d'entités et leurs attributs.:

- classe « zone P.O.S. » (attributs : Numéro, Affectation au P.O.S.).
- classe « zone M.O.S. » (attributs : Numéro, Mode d'Occupation du Sol, Bruit, Surface, Habitants, comme dans le tableau 3.1),

- classe « parcelle cadastrale » (attributs : Numéro, propriétaire(s), Surface, Utilisation),
- classe « propriétaire » (attributs : identificateur, Nom, adresse),

On peut également préciser la sémantique de leurs relations Une zone P.O.S. contient une ou des parcelles M.O.S., elles mêmes composées d'une ou plusieurs parcelles cadastrales (chacune ayant un ou plusieurs propriétaires).

Le modèle externe peut donc se présenter comme sur la **figure 3.1.**



**Figure 3.1 : Modèle externe de la B.D. « P.O.S. »**

Passer de ce type de représentation de la structure de la B.D. à un modèle conceptuel implique le recours à un formalisme admis par les informaticiens, par exemple ci-dessous le formalisme Entités / Association.

#### **B. Schéma conceptuel en formalisme Entités/Association**

Le formalisme **Entités/Association** (E/A) est le mode le plus courant de représentation graphique de modèles conceptuels. La figure 3.2 représente ainsi une relation entre les deux classes d'entités « personne » et « parcelle » d'une B.D. cadastrale.



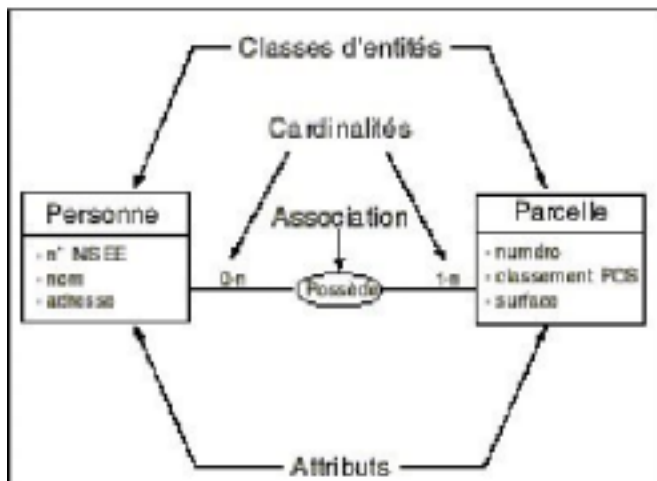


Figure 3.2 : exemple de modèle conceptuel d'une relation

Les différents composants de ce **schéma relationnel** sont ici:

- La classe « personne », munie de trois attributs (N° INSEE comme identificateur unique, nom et adresse),
  - la classe « parcelle », munie également de trois attributs (N° de parcelle, affectation P.O.S., surface),
  - la relation elle même, caractérisée par trois paramètres :
- 1) l'association qui donne la signification de la relation (« posséder » dans le sens n° 1 personne → parcelle, « être possédée » dans le sens n°2 parcelle → personne),
  - 2) la cardinalité dans le sens n°1 qui fournit les nombres **minimum** et **maximum** de parcelles que peut posséder une personne (de 0 à n, c'est à dire plusieurs),
  - 3) la cardinalité dans le sens n°2 indiquant qu'une parcelle est possédée au moins par une personne et au plus par plusieurs (cas d'indivision).

Reprenons maintenant l'exemple de la figure 3.1, toujours sur la même commune savoyarde mais avec quatre classes d'entités (Zone P.O.S., Zone M.O.S., parcelle cadastrale et propriétaire). Représentons les relations de cette B.D. sous forme de schéma Entités / Association (fig. 3.3).

A titre d'exemple, nous détaillerons l'une des relations : celle entre zones P.O.S. et zones MOS incluses dans les zones P.O.S.). Pour l'expliciter, nous devons fournir :

- le tableau des zones P.O.S. présentes dans l'extrait retenu (tab. 3.3),
- le tableau de correspondance entre zones M.O.S. et zones P.O.S. (tab. 3.4).

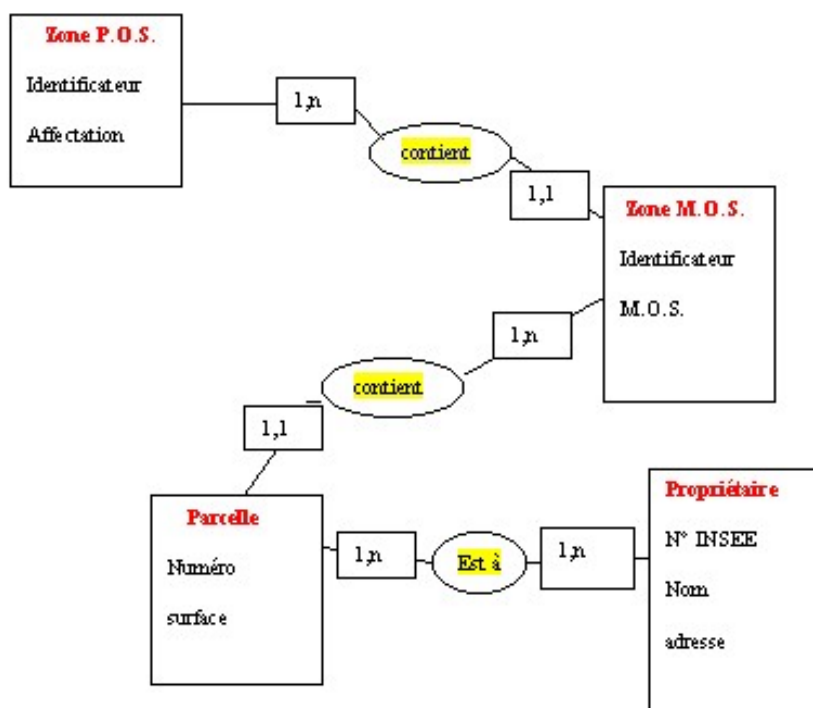


Figure 3.3 : schéma relationnel de l'exemple P.O.S. communal

C'est, en effet, une règle générale qu'une B.D. relationnelle comprend **autant de tableaux qu'il y a de classes d'entités et de relations « 1 à 1 » entre elles**. Aux deux tableaux «Zones P.O.S. » et « Zones M.O.S. » s'ajoute en effet le tableau de relation entre elles (Tab. 3.3). Il contient au moins une correspondance entre identificateurs des zones P.O.S. et des zones M.O.S. mais peut aussi contenir des attributs. Dans le cas de la fig. 3.1, la B.D. a 7 tableaux (4 classes d'entités + 3 relations « 1 à 1 » entre elles).

Identificateur P.O.S.	Affectation au P.O.S.
A18	UE ( <i>UE =zone d'activité</i> )
A05	UA ( <i>UA=bâti ancien dense</i> )
A16	UC ( <i>UC=extension, densité moyenne</i> )
A08	UF ( <i>UF=installation touristique</i> )
A21	ZAD ( <i>ZAD=Zone d'Aménagement Différé</i> )
B04	ND ( <i>ND=Espace naturel protégé</i> )
B12	NA ( <i>NA=urbanisation future</i> )
C24	UD ( <i>UD=zone constructible, faible densité</i> )

Tableau 3.3 : valeurs d'attributs pour les zones P.O.S. contenant les zones M.O.S. du tableau 3.1

Id. M.O.S.	Id. P.O.S.
01	A18
38	A18
04	B12
06	A18

07	A05
08	A16
09	A05
11	A08
12	A16
14	A05
17	A05
19	A08
43	A05
22	A05
47	A05
48	A05
26	A05
28	A21
29	A08
30	A05
31	B04
36	C24

**Tableau 3.4 : correspondance entre identificateurs de zones M.O.S. et P.O.S.**

NB : on peut noter que toutes les zones M.O.S. du tableau 3.1 ne sont pas référencées dans le tableau 3.4 (il manque les zones 02, 21, 24 et 25) et que des zones supplémentaires y figurent (zones 38, 43, 47 et 48).

### C. Modèle logique : opérateurs d'«algèbre relationnelle»

Les opérateurs (arithmétiques, de comparaison, logiques) décrits au § 3.1 pour traiter les données de **chaque tableau** (indépendamment les uns des autres) demeurent possibles. Mais il faut aussi, maintenant, définir des **opérateurs relationnels** pour combiner les données provenant de tableaux différents.

On indique ci-dessous les principaux d'entre eux, classés en deux catégories.

#### 1. Premier type , des opérateurs *ensemblistes entre tableaux*

l'**intersection** des 2 tableaux MOS et POS crée un tableau dont les entités sont dans le tableau MOS **et** dans le tableau POS (les zones MOS 02, 21, 24, 25, 38, 43, 47 et 48 n'y figureront pas),

l'**union** des tableaux MOS et POS crée un tableau dont les entités sont dans MOS **ou** dans POS, les éventuels doublons étant éliminés (toutes les zones MOS de l'un ou l'autre tableau y figureront),

la **différence**, MOS – POS, crée un tableau contenant toutes les parcelles du tableau MOS **sauf** celles non référencées dans le tableau POS (les parcelles 02, 21, 24 et 25 ne figureront pas dans le tableau résultant).

#### 2. Deuxième type : d'autres opérateurs ont été créés spécifiquement, dont le principal est la **jointure**. La jointure des tableaux MOS et POS crée un tableau résultat où :

- figurent les parcelles présentes dans les 2 tableaux, à l'aide du 3<sup>ème</sup> tableau de correspondance, associant numéro de zone MOS et identificateur de zone POS,

- figurent les attributs présents dans le tableau *MOS* ou dans le tableau *POS* ; on connaîtra ainsi la zone de POS des zones MOS référencées.

<b>Id. MOS</b>	<b>Id. POS</b>	<b>M.O.S.</b>	<b>dbA</b>	<b>Surf (ha)</b>	<b>Hab.</b>	<b>P.O.S.</b>
01	A18	Zi ouverte	62	14.4	0	UE
04	B12	Zone comm.	64	3.4	0	NA
06	A18	Zi ouverte	65	5.0	0	UE
07	A05	Hbtt dense	70	2.2	269	UA
08	A16	Hbtt diffus	68	4.2	143	UC
09	A05	Hbtt dense	65	5.1	626	UA
11	A08	Loisirs ext.	66	6.8	0	UF
12	A16	santé	58	1.7	0	UC
14	A05	scolaire	55	1.0	0	UA
17	A05	Hbtt dense	68	1.3	191	UA
19	A08	Loisirs int.	75	0.4	0	UF
22	A05	Hbtt dense	62	7.0	605	UA
26	A05	Hbtt dense	58	1.0	92	UA
28	A21	ZI fermée	70	14.9	0	ZAD
29	A08	Loisirs ext.	55	0.9	0	UF
30	A05	Hbtt dense	65	0.2	38	UA
31	B04	Terr. vague	62	0.6	0	ND
36	C24	Hbtt diffus	58	4.2	106	UD

Tableau 3.5 : Résultat de la jointure entre tableaux MOS et POS

Comme le montre le tableau 3.5, la jointure a retenu les entités présentes à la fois dans le tableau MOS et dans le tableau POS ; elle a, en outre, **joint** les attributs du tableau POS à ceux du tableau MOS. Elle a, en quelque sorte, « déplié » le tableau POS et effectué une déduction sur les données.

### 3.2.2. IMPLEMENTATION INFORMATIQUE EN « SQL »

Les opérateurs intra-tableaux (&3.1) et inter-tableaux (&3.2) sont exécutables dans un langage de programmation de requêtes, comme **QBE** (« Query By Example ») ou **SQL** (« Structured Query Language »). Nous présentons ici quelques rudiments de ce dernier, le plus courant des deux.

#### A. Langage d'interrogation SQL

Outre les commandes permettant de créer, corriger, mettre à jour, consulter, supprimer une B.D., le langage SQL permet des requêtes utilisant principalement l'instruction **SELECT**, dont la forme est :

```
SELECT attribut(s)
FROM tableau(x)
WHERE condition
```

- On peut, dans la clause SELECT, sélectionner un seul attribut (dans un seul tableau) ou plusieurs, provenant du même tableau ou de tableaux liés par un tableau d'association.
- Le ou les tableaux d'où proviennent les attributs sélectionnés doivent être précisés dans la clause FROM.
- La clause WHERE permet d'exprimer, à partir des attributs des tableaux sélectionnés, une condition logique ; celle-ci est formée à partir des valeurs d'attributs liés par opérateurs (numérique, de comparaison et/ou ensembliste). Seules les entités pour lesquelles la condition est vraie seront retenues.

L'instruction SELECT crée un nouveau tableau sur lequel on peut faire, à son tour, une requête : l'interrogation SQL d'une B.D. est souvent une suite des requêtes visant un résultat final.

## B. Exemples de requêtes SQL

Nous illustrons ci-dessous, à l'aide de requêtes SQL, les opérateurs relationnels décrits dans les paragraphes précédents.

- Sélectionner dans MOS l'attribut « BRUIT »

```
SELECT BRUIT
FROM MOS
```

Le tableau résultat de la requête sera un tableau avec toutes les entités et 2 colonnes (N° et BRUIT).

- Sélectionner dans MOS les attributs « surface » et « habitants » pour les points noirs acoustiques ( $BRUIT \geq 70$ )

```
SELECT surface, habitants
FROM MOS
WHERE BRUIT >=70
```

Le tableau résultat comportera 3 colonnes (N°, surface et habitants) pour 3 lignes : parcelles 07, 19 et 28 (du tableau 3.1) dont le niveau sonore est  $\geq 70$  décibels.

- Même type de sélection avec une clause WHERE composite : niveau sonore  $\geq 65$  db, parcelle habitée ou de plus de 5 ha.

```
SELECT surface, habitants
FROM MOS
WHERE BRUIT >= 65
AND (habitants > 0 OR surface > 5)
```

Le tableau résultat aura toujours 3 colonnes (N°, surface, habitants) et 6 lignes (parcelles de N° 07, 09, 11, 17, 28 et 30). Pour que la condition soit « vraie », il faut que la parcelle ait au moins 65 db et que, ou bien elle ait des habitants ou bien elle ait plus de 5 hectares. La parcelle 19, par exemple, ne sera pas retenue car elle n'a ni habitants ni taille suffisante, la parcelle 06 non plus (elle ne fait pas plus de 5 ha).

- Créer, par **jointure**, un tableau des parcelles MOS densément habitées ou incluses en zone d'activité et renseignées dans le tableau *POS*.

```

SELECT BRUIT, surface, habitants, MOS, POS
FROM   MOS, POS
WHERE M.O.S. = « Habitat dense »
OR     P.O.S. = « UE »

```

On a sélectionné dans la clause FROM les 2 tableaux dont il s'agit de joindre les entités ; la clause WHERE restreint à celles qui sont densément habitées ou appartiennent à une zone d'activité. La clause SELECT a retenu des attributs du 1<sup>er</sup> et du 2<sup>nd</sup> tableau . L'instruction ci-dessus va déduire la zone de Plan d'Occupation du Sol de chacune des zones MOS densément habitée ou incluses dans une zone d'activité.

Le tableau résultat sera :

<b>Id. MOS</b>	<b>Id. POS</b>	<b>dbA</b>	<b>Surf (ha)</b>	<b>Hab</b>	<b>M.O.S.</b>	<b>P.O.S</b>
01	A18	62	14.4	0	ZI ouverte	UE
06	A18	65	5.0	0	ZI ouverte	UE
07	A05	70	2.2	269	Hbtt dense	UA
09	A05	65	5.1	626	Hbtt dense	UA
17	A05	68	1.3	191	Hbtt dense	UA
22	A05	62	7.0	605	Hbtt dense	UA
26	A05	58	1.0	92	Hbtt dense	UA
30	A05	65	0.2	38	Hbtt dense	UA

Tableau 3.6 : tableau résultant de la requête impliquant une jointure

On remarque immédiatement que la zone MOS n° 38, bien qu'elle respecte la condition ( MOS = « Habitat dense » **OU** POS = « UE »), ne figure pas dans le tableau résultant de la jointure car elle n'est pas présente dans le tableau MOS. Bien que présente dans les 2 tableaux MOS et POS, la zone n°28 n'appartient pas au tableau 3.6 car elle ne respecte aucune des deux conditions posées ( MOS = « Habitat dense » **OU** POS = « UE »).

Les opérateurs relationnels permettent de prendre en compte l'association logique entre entités du monde réel, d'en déduire (notamment par jointure) des données nouvelles et, donc, d'exprimer l'information initiale sous une forme minimale. Le modèle relationnel représente donc bien l'équivalent, pour l'information non numérique, d'une **mise en facteurs**, ce qui a plusieurs avantages. Pour un même contenu, la décomposition en classes d'entités logiquement associées mène à un volume minimal d'information, assure la cohérence des données lors de mises à jour de la B.D., permet d'affecter des droits d'accès différents à différents types d'utilisateurs (chacun d'entre eux en « **voit** » alors une partie plus ou moins importante).

### **3.3 LA FOUILLE DE DONNEES (« DATA MINING »)**



Les Bases de Données des organisations deviennent, au fur et à mesure de leur existence, gigantesques (incluant des historiques), hétérogènes (appartenant à différents services), éventuellement éclatées sur différentes machines en réseau. On les qualifie d'**entrepôts de données** (« data warehouses »).

Le but de la **fouille de données** (« data mining ») est d'extraire de la façon la plus automatique possible, pour des utilisateurs non informaticiens, de la connaissance de cet amas de données. Cet objectif peut être atteint en combinant des techniques de différents types :

- De type Système de Gestion de Bases de Données (ou Système d'Information Géographique si les données sont spatialisées),
- De type statistique (analyse des données, classification),
- De type « Intelligence Artificielle » (Systèmes à base de connaissances, Systèmes d'apprentissage, Réseaux Neuronaux, algorithmes génétiques),

La fouille de données visant la combinaison de différentes méthodes et l'automatisme en est encore au stade de la recherche plus que de l'application courante. Des exemples de prototypage pour usage géographique peuvent être trouvés dans un **numéro spécial de la Revue Internationale de Géomatique** (Volume 9, N°4, Année 2000).

**N.B.** On trouvera :

- à la fin du chapitre 1 quelques adresses Internet pour télécharger gratuitement des bases de données géographiques,
- dans le chapitre 6 (& 6.6) les principaux fournisseurs de bases de données géographiques payantes.

### **POUR EN SAVOIR PLUS**

G.Gardarin : *Bases de Données, Objet et relationnel* (Ed. Eyrolles, 1999)

Revue Internationale de Géomatique, 2000, Vol. 9, N°4 : Numéro spécial sur le « data mining » sous la responsabilité de K.Zeitouni

## **CHAPITRE 4 : L'INFORMATION GEOGRAPHIQUE**

*Pierre Dumolard*

Toute information géographique est **d'abord de l'information** : en tant que telle, elle est comme toute information composée de descripteurs (attributs dans le langage des B.D., variables dans celui de la statistique) permettant de caractériser des objets (entités dans le langage des B.D., individus en jargon statistique). Ces descripteurs, communs à tous les objets, tentent de les caractériser, d'exprimer leur signification, d'où le terme de partie **sémantique** de l'information géographique.

Mais, les objets géographiques (à quelque échelle que ce soit) ont pour particularité d'être des objets localisés. Ils doivent donc être caractérisés par leurs **localisation, forme et dimension**. C'est là la partie **géométrique** de leur description, qui permet leur représentation cartographique. Cela suppose de pouvoir disposer d'un **référentiel spatial**, fondé sur une modélisation de la terre.

C'est à la compréhension de cette modélisation que nous voulons d'abord, brièvement introduire le lecteur avant d'aborder l'informatisation de ce type de données.

### **4.1 MODELISATIONS DE LA TERRE**

Forme et position « absolues » des objets géographiques s'expriment par des coordonnées (en 2 ou 3 dimensions), soit sphériques soit planes dans un **référentiel géographique** tandis que leur position relative s'exprime par leur contiguïté, inclusion ou chevauchement par rapport à d'autres objets. La position « absolue » est une notion géométrique, la position relative une notion topologique.

#### **4.1.1 COORDONNEES SPHERIQUES**

##### **A. Ellipsoïde, modèle géométrique de la terre**

Déjà, au 6<sup>ième</sup> siècle avant J. C., Pythagore concevait la terre comme une **sphère** et Eratosthène (3<sup>ième</sup> siècle av. JC) fit une première estimation de son rayon. Cette assimilation de la terre à une sphère est un modèle (géométrique) simple qui néglige les différences de relief. Ce modèle permet de définir des **parallèles** (dont l'équateur, les tropiques), des **méridiens** et les coordonnées en latitude et longitude de tout point de la surface terrestre (*coordonnées sphériques*, exprimées en degrés ou grades). Tout point de la surface terrestre peut donc être localisé par sa position en latitude et longitude sur cette sphère virtuelle mais cela suppose des mesures (terrestres ou par rapport aux étoiles).

Le moyen âge ne concevait pas la terre comme une sphère mais comme un disque aplati : les grandes découvertes démontrèrent que la représentation sphérique était plus pertinente.

Au 17<sup>ème</sup> siècle, débuta la cartographie « scientifique » de vastes territoires (Cassini, en France par exemple). Leur triangulation (par mesure d'angles et de distances entre sommets de triangles contigus successifs) permit de préciser positions et formes,. A partir de ces mesures, Newton démontra que la terre n'est pas exactement une sphère parfaite mais car légèrement aplatie aux pôles à cause de la rotation de la terre. La forme de la terre pouvait donc être représentée par un **ellipsoïde**, modèle mathématique tri-dimensionnel qui l'ajuste mieux qu'une sphère.

Au 19<sup>ème</sup> siècle, les grands pays se sont dotés, à partir d'ellipsoïdes centrés sur leur méridien origine, de **réseaux géodésiques** nationaux,. Ces réseaux sont constitués d'un ensemble de points, de coordonnées connues et matérialisés au sol par des bornes ; ils sont hiérarchisés en plusieurs niveaux (5 pour 80000 points dans le système français). C'est par mesure de distance aux points d'un réseau géodésique que toutes les coordonnées peuvent être déterminées. Les ellipsoïdes des différents pays étant différents, a été mis en place en 1984 un système géodésique mondial à coordonnées sphériques (WGS84 pour « World Geodetic System »).

Les coordonnées « sphériques » en latitude et longitude sont la projection sur une ellipse (particulière à un pays ou universelle) : elles sont donc le résultat d'une modélisation géométrique.

## **B. Géoïde, modèle physique de la terre**

La terre, cependant, n'est pas exactement une ellipse, elle est « bosselée », les montagnes étant composées de matériaux plus légers que ceux, plus lourds, du fond des océans. Relier les points de même pesanteur dessine une surface équipotentielle: le niveau moyen des océans a été choisi comme surface équipotentielle de base, permettant de mesurer des **altitudes**. Cette surface, le **géoïde**, est un modèle physique de représentation de la surface terrestre. Pour la France, le niveau zéro est, arbitrairement, celui défini par le marégraphe de Marseille. Pour d'autres pays, le niveau zéro est choisi différent : les niveaux zéro de leurs cartes sont donc différents.

Ellipsoïde (modèle géométrique) et géoïde (modèle physique) ne coïncident pas : il y a entre les 2 surfaces un écart (*Fig. 4.1*). La perpendiculaire au géoïde (ou normale, **N**) fait avec la perpendiculaire à l'ellipsoïde (**V**) un angle, appelé déviation de la verticale. Chaque pays choisit un **point fondamental** (**F**) où les 2 surfaces sont tangentes et où les 2 perpendiculaires sont confondues : pour la France, ce point est la croix du Panthéon à Paris.

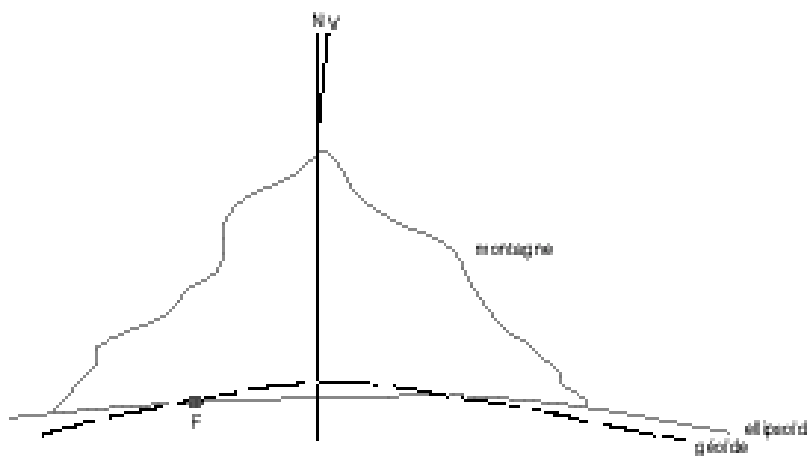


Figure 4.1 : différence entre Ellipsoïde et Géoïde

## 4.1.2 PROJECTIONS : COORDONNEES PLANES

### A. Types de projection plane

La cartographie transforme, par **projection** sur un plan, des coordonnées sphériques (longitude, latitude) en coordonnées planes (X, Y). Comme toute projection sur un plan déforme des surfaces courbes, on a adopté (selon la déformation à réduire et la position géographique du pays) des systèmes de projection différents.

On distingue classiquement :

- **projections cylindriques** (Fig. 4.2) sur un cylindre tangent à la terre (le long d'un parallèle, d'un méridien ou transversalement (par exemple les projections Mercator))

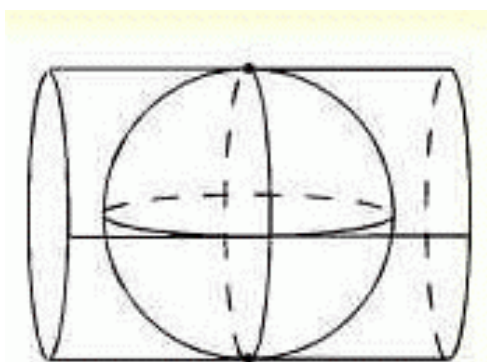
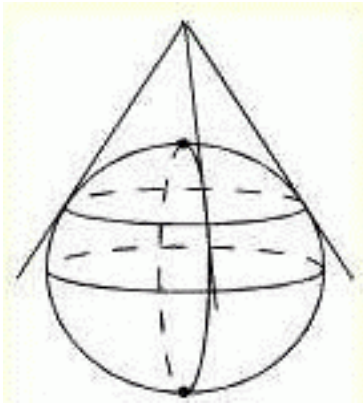


Figure 4.2 : Projection cylindrique (source: IGN)

- **projections coniques** (Fig. 4.3) , projections sur un cône dont le sommet est situé, par exemple, à la verticale d'un pôle (par exemple projections Lambert).



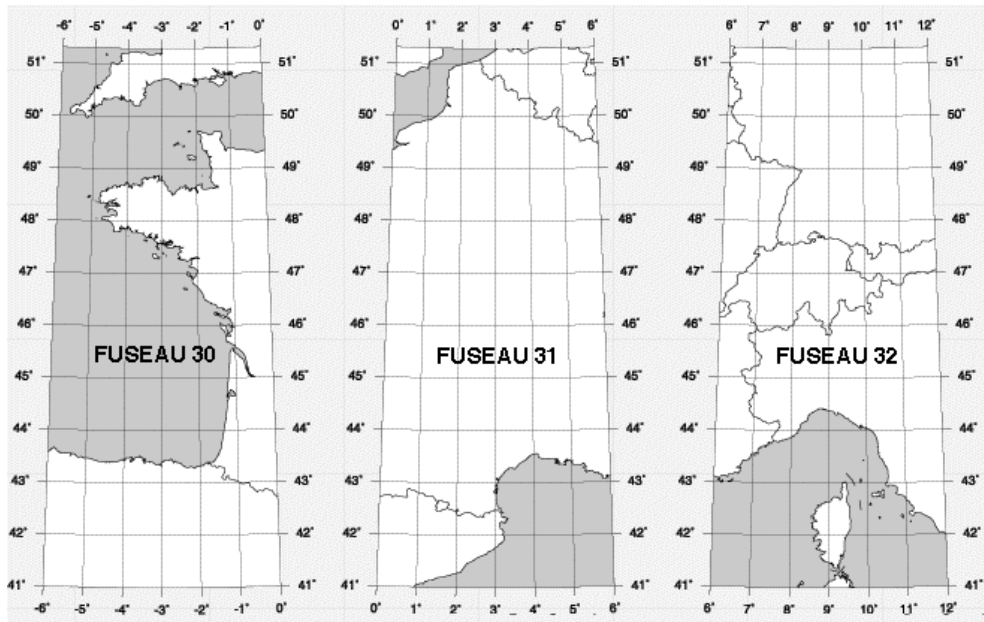
**Figure 4.3 : Projection conique** (source: IGN)

En outre, une projection peut être :

- **conforme** (déformant les longueurs, préservant les angles),
- **équivalente** (longueurs préservées, angles déformés)
- ou autre.

On détaille ci-dessous 2 projections, relatives au territoire de la France métropolitaine.

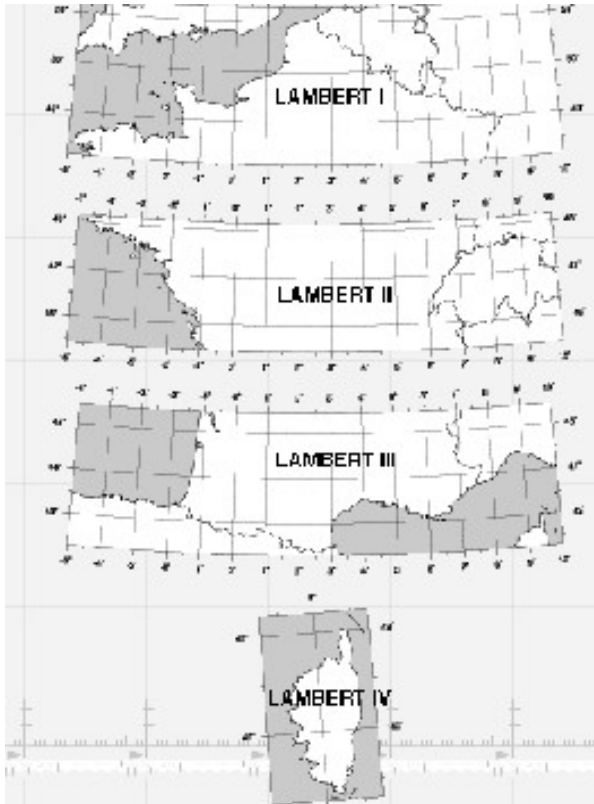
## B. Projections Mercator



**Figure 4.4 : La France métropolitaine en projection U.T.M** (source: IGN)

La projection de **Mercator** est souvent utilisée par des pays de forme allongée (N-S, ou E-O comme les USA). Une expression particulière en est la projection **UTM** (« Universal Transverse Mercator »). La terre y est divisée en 60 fuseaux de 6° de longitude : le cylindre de projection est tangent au méridien central de chaque fuseau. Ce type de représentation découpe, en quelque sorte, la terre en 60 « tranches d'orange ». La France, pour sa part, est située dans les fuseaux 30 à 32 (*Fig. 4.4*).

### C. Projections Lambert



**Figure 4.5 : Les quatre zones de la projection Lambert - France métropolitaine**  
(source IGN)

La projection la plus utilisée pour les cartes françaises est une projection *conique conforme de Lambert*. Le territoire métropolitain est divisé en 4 zones (*Fig. 4.5*), de Lambert 1 = Nord à Lambert 4 = Corse ; une projection France entière existe aussi (Lambert 2 « étendu »).

### B. Lecture des coordonnées sur une carte IGN au 1/50000

Sur les bords d'une carte topographique de l'IGN coexistent plusieurs systèmes de coordonnées :

- A *l'extérieur*,
  1. l'amorce du quadrillage UTM (en noir),
  2. latitudes – longitudes en degrés (en bleu)
- A *l'intérieur*,
  1. en noir, le quadrillage Lambert (I, II, III ou IV) : chaque coordonnée Y est faite du numéro de la zone Lambert et d'un nombre de kilomètres par rapport à la ligne origine,
  2. en bleu, l'amorce du quadrillage Lambert 2 étendu (valable pour toute la France métropolitaine), matérialisée sur la carte par des croisillons,

### 3. latitudes et longitudes en grades.

Les coordonnées de tous points de la carte (dans l'un des systèmes) peuvent être obtenues par interpolation.

La définition, la plus précise possible, d'un **référentiel spatial** (c'est à dire d'un système de mesure et de représentation de tout lieu d'un espace) est la tâche fondamentale des Instituts Géographiques Nationaux. C'est ce canevas qui est à la base de toute cartographie : le construire suppose la mise en œuvre de procédés d'acquisition.

#### 4.1.3 PROCEDES D'ACQUISITION DE L'INFORMATION SPATIALE

Jadis, les mesures de positionnement, d'angles, de distances se pratiquaient, en mer à l'aide de sextants, sur terre à l'aide de chaînes d'arpenteur, de lunettes de visée (modernisées en télémètres laser), le report sur plans ou cartes était manuel : la productivité était donc faible.

L'avion, le satellite, l'ordinateur sont, aujourd'hui, les outils de base d'une production cartographique plus automatisée.

- Pour les **plans** à « grande » échelle (du 1/50 au 1/5000), on pratique couramment des mesures à l'aide d'instruments de terrain (comme les télémètres).
- La **photographie aérienne** est encore la “ matière première ” de la plupart des cartes, notamment aux échelles comprises entre le 1/5000 et 1/50000 (selon l'altitude du vol). Les **orthophotos** sont des photos aériennes perpendiculaires à la surface terrestre : elles permettent un repérage en plan (coordonnées XY) d'objets et non la vision du relief. Mais, au cours de son vol, l'avion, muni d'un équipement de précision, réalise successivement des prises de vue de bandes de terrain se recouvrant sur 60% de leur largeur et 20% de leur longueur.

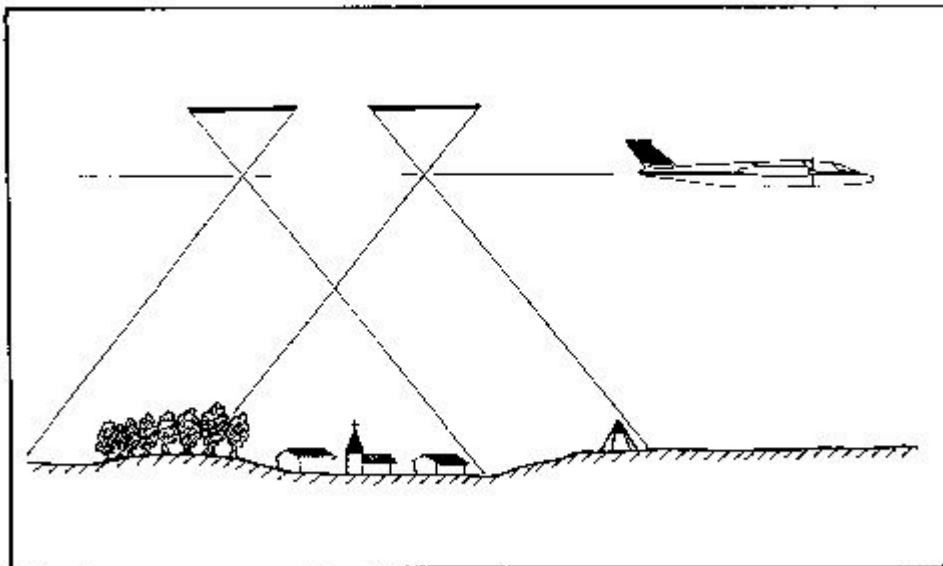


Figure 4.6 : Vision en coupe du trajet de l'avion lors d'une mission photo aérienne



L'observation, à l'aide de lunettes **stéréoscopiques**, des 60% communs à 2 photos successives (principe : figure 4.7) donne une vision en relief de l'utilisation du sol et permet le report de données sur une carte topographique (photo-interprétation). En intégrant lunette, appareils de mesure et de pointage, on peut faire de la photo-grammétrie (mesure des positions XYZ de tous les points cliqués). Le couplage à un ordinateur permet le stockage des coordonnées de ces points, en mode “**vecteur**” (ou dessin).

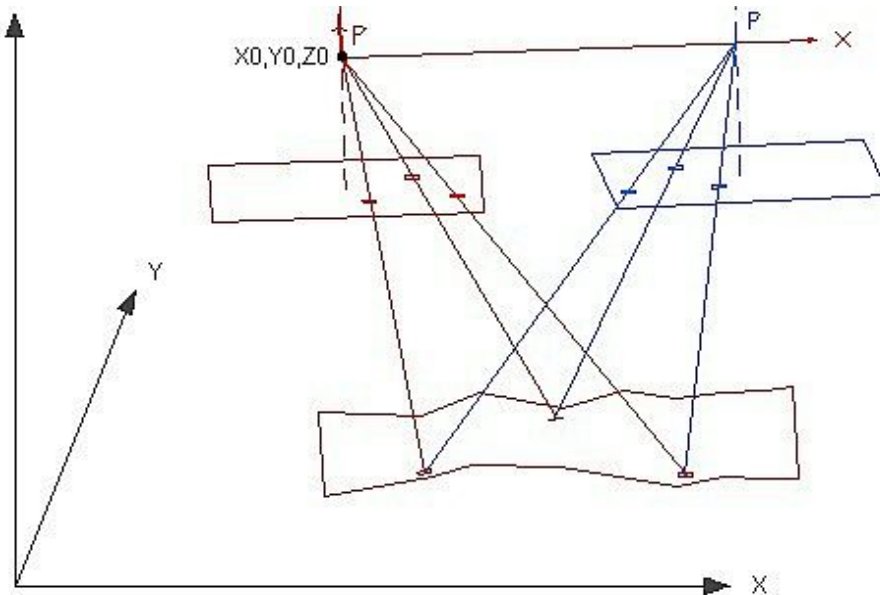
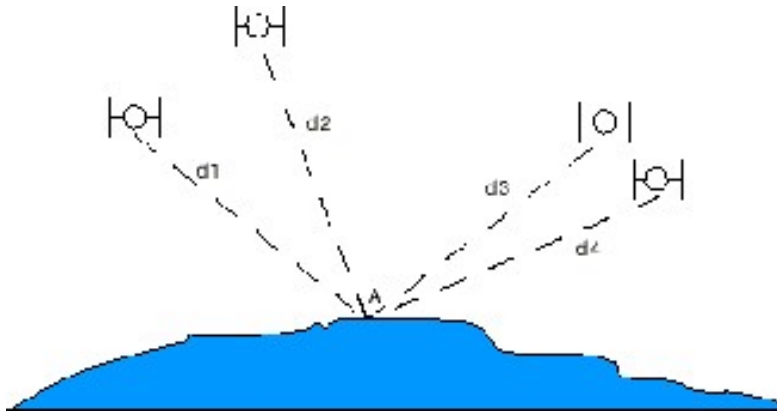


Figure 4.7 : Principe de la vision stéréoscopique

- Les **satellites** d'observation de la terre (par exemple SPOT) sont, eux, équipés de capteurs qui enregistrent le rayonnement électromagnétique réfléchi par la surface terrestre dans différentes longueurs d'onde. Les données, télé - transmises à des centres de stockage et de pré-traitement, sont relatives à des surfaces maillées, divisées en **pixels** (« **picture elements** »). Pour SPOT par exemple, les pixels sont des carrés de 10m de côté dans les images en niveaux de « gris ». Récemment lancés, de nouveaux satellites d'observation de la terre ont une bien meilleure « résolution » (pixels de 1m en niveaux de grisés pour IKONOS). Les images SPOT couvrant au sol 60km x 60km. par pixels de 10m se présentent donc comme des tableaux à 6000 lignes et 6000 colonnes, où chaque case contient un nombre, compris entre 0 et 255, codant un des 256 niveaux de grisé. Cette représentation est dite “**rasteur**”.  
L'interprétation d'images satellite consiste à transformer ces valeurs de réflectance électro-magnétique en utilisations du sol, permettant de renseigner des cartes ou des bases de données à intervalles de temps rapprochés.
- Dans les deux cas, photo aérienne et image satellite, on doit procéder à des « redressements » géométriques et au repérage de points de coordonnées connues. Par interpolation, l'ensemble des points ou des pixels auront une position estimée, sur un ellipsoïde de référence et dans un système de projection donné. Ce positionnement s'effectuait classiquement par visée astronomique (sextant, en mer) ou mesure de

distance par rapport à des bornes géodésiques. Il s'effectue de plus en plus à l'aide du **GPS** (Global Positioning System).

- Le système **GPS** (« Global Positioning System ») est fondé sur 24 satellites dont les orbites sont telles que tout point de la terre en ait toujours au moins 4 dans “ son ciel ”. L'appareil au sol **A** les localise (si son ciel est dégagé) et connaît donc leur éloignement  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  (*Fig.4.8*); un simple calcul permet alors de déterminer sa position, exprimée en coordonnées sphériques dans le système géodésique WGS84.



*Figure 4.8 : Principe de la localisation par G.P.S.*

La position de l'appareil **A** était, dans 95% des cas, connue avec une incertitude de 100 mètres en latitude – longitude et 150 mètres en altitude. Depuis le début de l'année 2000, suite à l'abandon par les USA du « brouillage » du signal, les ordres de grandeur de l'imprécision ont été divisés par dix. Pour atteindre une précision encore bien meilleure, on peut utiliser un GPS différentiel, composé de plusieurs GPS parfaitement synchrones, dont l'un au moins est situé en un point de coordonnées connues.

- Il existe aussi d'autres méthodes de positionnement qui ne se traduisent pas par des coordonnées géographiques précises (sphériques ou planes) mais par une adresse postale ou un code communal, fondés tous deux sur des référentiels spatiaux décrivant des **positions relatives**.

En mode vecteur, la **forme** des objets surfaciques est décrite par leurs limites tandis qu'en mode rasteur elle résulte de l'agrégation de pixels contigus de même valeur (ou couleur sur une image).

## **4.2. INFORMATISATION DES DONNEES GEOGRAPHIQUES**

On peut représenter informatiquement des objets localisés à la surface de la terre en mode **rasteur** ou en mode **vecteur**. Le choix de l'un ou l'autre type implique le choix d'un procédé de saisie informatique.

### **4.2.1. LE MODE RASTEUR**

L'exemple ci-dessous est extrait d'un fichier « texte » (cf chapitre 3.) en représentation rasteur : il est donc organisé en lignes et en colonnes et le nombre dans chaque case est ici

l'altitude moyenne d'un carré de 50m x 50m au sol ; l'extrait de 8 lignes et 10 colonnes représente donc une surface de 400m x 500m, située à peu près au centre de la cuvette grenobloise (fig. 4.9).

242	242	243	244	244	245	246	246	247	248
243	243	243	244	244	245	246	246	247	247
243	245	246	245	245	245	246	247	247	248
243	243	246	247	246	246	246	247	247	248
243	243	244	247	248	247	247	247	247	248
243	243	244	245	249	249	248	248	248	248
243	244	244	245	246	250	251	249	248	248
244	244	244	245	247	251	252	253	249	249

Tableau 4.1 : extrait du fichier rasteur de la fig. 4.9 (2 km² au centre de la cuvette grenobloise) source : IGN

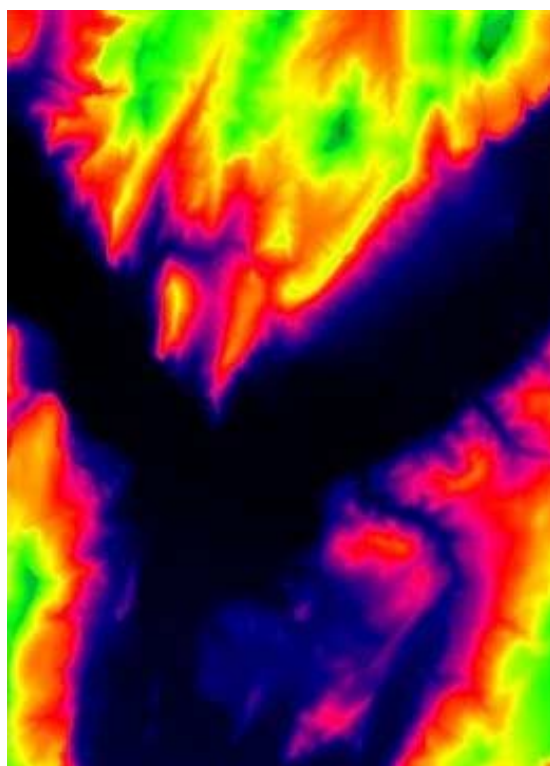


Figure 4.9 : image du relief de l'Y grenoblois

Chacun des pixels a même forme (carrée) et surface (2500 m²), sa position relative est donnée par ses n° de ligne et de colonne ; si l'on connaît les coordonnées (sphériques ou planes) des coins NW et SE de l'image, on peut donc calculer celles de chaque centre de pixel. La structure rasteur est simple du point de vue informatique : c'est un tableau, le plus souvent numérique, directement cartographiable à l'écran ou sur imprimante si l'on associe un grisé ou une couleur à chaque classe de valeurs. Le mode rasteur convient surtout à la description de surfaces, assez peu à celle d'objets ponctuels ou linéaires.

La saisie informatique en mode image (rasteur) peut être faite de diverses manières :

- Saisie manuelle dans un éditeur de fichiers textes à partir d'une grille superposée à une carte,
- Saisie d'une carte simple ou d'une photo, en niveaux de gris ou en couleurs, sur un scanner,
- Les images satellite sont directement livrées comme un ou plusieurs fichiers rasteur.

L'information rasteur, étant numérique et de structure informatique simple, peut donner lieu assez facilement à de nombreux traitements, modélisations et simulations.

#### 4.2.2 LE MODE VECTEUR

Les données géographiques en mode vecteur ont une forme double :

- la description des propriétés des différentes entités se fait dans une B.D. (cf chapitre 3) où chaque objet a un identificateur unique,
- les objets spatiaux sont décrits par les coordonnées de primitives graphiques qui sont :
  - & le point,
  - & la polyligne, ligne formée d'un ou plusieurs segments,
  - & le polygone (polyligne fermée)

Le modèle le plus simple de leur description géométrique est :

1) Point	X, Y	
2) Polyligne	X1, Y1, X2, Y2, X3, Y3, ...	ligne brisée non fermée
3) Polygone	X1, Y1, X2, Y2, X3, Y3, ..., X1, Y1	ligne brisée fermée (frontière)

Dans cet exemple générique, les X désignent la longitude et les Y la latitude de points (dans un référentiel spatial ou un autre). Un seul couple XY suffit pour localiser un point. Pour décrire une polyligne (ex. un tronçon routier), le *crayon virtuel* (la souris ou le pointeur sur une tablette à digitaliser) se déplace de X1Y1 à X2Y2, de X2Y2 à X3Y3, etc. Un polygone est une polyligne fermée sur elle-même et désigne l'aire incluse, dont on fournit souvent le point moyen (de coordonnées XmYm). Le mode vecteur convient pour tout type d'objet spatial, quelle que soit sa dimension. Le zoom n'amène pas, comme en rasteur, une perte de qualité visuelle.

La reconnaissance et la description précise des formes, positions et dimensions (morphologie spatiale au sens propre du terme) sont bien souvent des préalables à toute quantification géographique (dessiner des frontières pour mieux caractériser un contenu), comme le démontre la délimitation morphologique des agglomérations de F.Moriconi (1).

#### 4.2.3 TRANSFORMATIONS RASTEUR - VECTEUR

La transformation d'un dessin vectoriel en image rasteur n'a de sens que pour des objets surfaciques ou, à la rigueur, linéaires et vaut surtout pour des dessins assez simples. Dans

le cas linéaire, rasteuriser automatiquement consiste à « noircir » les pixels traversés par un segment de droite. Des difficultés peuvent surgir s'ils ne sont pas suffisamment petits pour contenir un seul segment.

La transformation automatique d'une image rasteur en dessin vectoriel est, elle, relativement plus complexe. Pour des objets surfaciques, elle consiste à reconnaître dans l'image des zones homogènes et à les séparer (par une polyligne ou un polygone) d'autres zones homogènes. Pour des objets linéaires, l'idée de base est de les « suivre » jusqu'à leur terminaison. Des difficultés surgissent si la continuité du linéaire est mal rendue en rasteur.

L'idée est tentante de scanner des cartes (obtention de fichiers rasteur) et de les vectoriser automatiquement : en ce cas, les erreurs peuvent être importantes, surtout à cause de la toponymie et des éléments symboliques (l' « habillage » cartographique).

(1) F. Moriconi-Ebrard, 1994, *GEOPOLIS : pour comparer les villes du monde*, (Economica, Anthropos, coll. « Villes du Monde »)

### **Pour aller plus loin**

<http://www.ign.fr>

<http://www.cnig.fr>

<http://www.seig.ensg.ign.fr>

## **CHAPITRE 5 : CARTOGRAPHIE THEMATIQUE**

**Pierre Dumolard**

Les cartes thématiques se différencient des cartes topographiques par plusieurs caractéristiques :

- Elles représentent un seul thème alors que les cartes topographiques en visualisent plusieurs, fusionnés en un document “ composite ”,
- Ce thème est plus souvent abstrait que concret, alors que les cartes topographiques sont une simplification du paysage visible,
- Elles traitent le plus souvent de distributions statistiques spatiales.

On parle donc parfois à leur propos de “ *cartes statistiques* ” : la raison en est fort simple. Dès lors que l’on considère un certain nombre d’unités spatiales décrites de façon homogène par le même descripteur présentant les mêmes modalités (numériques ou pas), on fait de la statistique, comme Mr Jourdain faisait de la prose !

Une carte thématique se différencie également d’autres formes de représentation visuelle géographique (bloc diagramme, croquis régional, chorème, graphe, etc.).

Elle est fréquemment un document intermédiaire dans un **raisonnement spatial**, conjointement à d’autres (texte, graphique, tableau, calculs). A ce titre, elle est utilisée ou bien comme outil efficace de communication ou bien comme pièce essentielle (mais non unique) d’une analyse spatiale exploratoire.

### **5.1. CONNAISSANCES DE BASE EN CARTOGRAPHIE THEMATIQUE**

Toute carte, thématique ou autre, résulte d’un **traitement de données spatialisées**, d’une **méthode de visualisation** et peut faire l’objet, à son tour, de traitements. Si l’informatisation de la cartographie a grandement facilité les traitements de données, elle en a aussi changé les supports de visualisation (écran, imprimante) et a diffusé ce mode de représentation.

Au niveau le plus élémentaire, une carte thématique visualise la distribution spatiale simplifiée d’une variable statistique (ou descripteur ou attribut), par exemple, taux de fécondité, taille de villes, pluviométrie, classement de routes, utilisation du sol, etc..

Elle met en correspondance deux simplifications :

- spatiale (l’échelle donne le degré d’abstraction par rapport au territoire),
- statistique (l’œil ne perçoit pas aisément un grand nombre de figurés différents).

### 5.1.1. TYPOLOGIE POUR LA CONSTRUCTION D'UNE LEGENDE

Elle dépend, essentiellement, de cinq considérations, qui orientent le choix d'une représentation graphique des valeurs.

#### A. Nature des variables

On distinguera ici quatre types de variables:

- Quantitatives en valeur absolue qui expriment le résultat d'un dénombrement (nombres d'habitants par exemple) ou d'une mesure (surfaces, poids, quantités, etc.),
- Quantitatives en valeur relative qui traduisent aussi des quantités mais rapportées à une valeur de référence et souvent exprimées dans un intervalle borné connu (fréquences en % par exemple),
- Qualitatives ordonnées où existent entre les valeurs des différences non pas de quantité mais de niveau (exemple : classement de routes selon leur statut ou de vignobles selon la qualité de leur vin),
- Qualitatives nominales où il n'y a plus, entre les valeurs, ni différence de quantité ni différence de niveau mais seulement distinction de catégories (ainsi, l'utilisation du sol ne représente qu'une simple différenciation).

Les données statistiques spatiales sont le plus fréquemment **quantitatives**.

En effet, la géographie décrit des portions de surface terrestre et, dans celles ci, est souvent amenée à mesurer des phénomènes, à compter et à caractériser des personnes, des plantes, des animaux, des roches, etc. Ces " objets " sont des unités primaires qui ne nous intéressent pas en tant que tels mais dont nous importe la distribution dans des unités secondaires, territoriales (quartiers, villes, zones, régions, ....). Ces lieux sont des **agréats** et l'agrégation modifie le **nombre et la nature des variables** à représenter. Considérons, par exemple, la Catégorie Socio-Professionnelle (CSP) déclarée lors d'un recensement de population. Au niveau des personnes interrogées, c'est une variable qualitative nominale dont les 8 catégories sont codées de 0 à 7. S'intéresser au profil socio - professionnel des villes françaises amène à construire un tableau à autant de lignes que de villes et autant de colonnes que de CSP (8) : chaque colonne représente une des 8 CSP et chaque case contient (en nombres ou en %) une quantité d'habitants dans une ville et une CSP.

C'est seulement lorsque l'objet étudié est unité primaire (parcelle de culture, bâtiment, bassin versant, ...) qu'on peut le caractériser par des variables qualitatives. A chacune des natures de variable correspondent des possibilités différentes de traduction visuelle.

#### B. Variation unipolaire ou bipolaire ?

Le choix le plus fréquent est celui d'une variation unipolaire des valeurs, entre un minimum (zéro, souvent, pour des variables quantitatives) et un maximum. La représentation visuelle devra traduire cette variation unipolaire, par exemple par des grisés de plus en plus foncés. Mais il existe aussi des cas où s'impose une variation bipolaire des valeurs, de part et d'autre d'une valeur seuil. C'est souvent le cas lorsque la valeur zéro d'une variable quantitative est arbitraire et ne signifie pas " absence de ", par exemple pour des températures ou des altitudes. Autres exemples, des taux de croissance (démographique ou économique), des valeurs exprimées en écarts à la moyenne : il est ainsi judicieux, pour une carte de % de cadres supérieurs par département, de représenter



des écarts à la moyenne nationale pour bien souligner les surplus et déficits. Dans ce cas, il sera avantageux de différencier une classe centrale, conforme à la moyenne nationale et des classes de part et d'autre.

### C. Variables d'état ou de flux ?

La plupart des cartes thématiques représentent des variables d'état (dites de stock par les économistes) : chaque valeur traduit l'état d'un objet donné à un instant donné. Mais, chaque fois qu'il est question de flux, d'échanges, de mobilité, de transformation ou d'interaction spatiales, chaque valeur est relative à un couple de lieux (nombre de migrants entre communes A et B, pour des migrations de population) et la traduction graphique sera toute différente.

### D. Nombre de classes

Il est habituellement considéré comme impossible de représenter chacune des valeurs par un symbole ou un figuré différent : il faut donc les répartir en peu de classes, exclusives (pas de double appartenance) et exhaustives (pas d'unité non classée).

Deux considérations doivent guider le choix d'un nombre de classes :

- il doit être plus ou moins proportionnel au nombre N d'unités spatiales, deux ou trois quand N est petit, rarement plus de cinq ou six quand N est grand,
- en effet, l'œil peine à différencier aisément beaucoup plus de figurés et, pour cette raison de lisibilité, neuf classes doit être considéré comme un maximum absolu.

Pour des variables qualitatives à classes nombreuses, il faut regrouper celles-ci en quelques catégories ou niveaux, en fonction de leur proximité sémantique. Pour des variables quantitatives, il faut partitionner le continuum des valeurs en classes les plus homogènes possibles et les plus différentes possibles des autres. Elles deviennent ainsi, du point de vue de leur représentation visuelle, comparables à des variables ordinales.

### E. Etablir les limites des classes

Le choix des limites de classe doit être fonction de la forme de l'histogramme des valeurs.

Les techniques de partition les plus fréquentes sont :

& classes de même amplitude, système simple convenant assez bien à des histogrammes peu contrastés (proches d'une distribution uniforme),

& classes de même effectif, surtout pour des distributions assez symétriques,

& limites des classes en progression arithmétique ou géométrique, pour des distributions très dissymétriques,

& limites des classes aux discontinuités de la série, pour des distributions pluri-modales,

& limites de classes situées à 0.5, 1, 2 écarts type de part et d'autre de la moyenne, pour des distributions symétriques et des échelles de variation bipolaires.

En tout état de cause, il est recommandé, quand cela est possible, d'arrondir les valeurs des bornes de classes pour une meilleure lisibilité et mémorisation de la légende.

## 5.1.2 TRADUCTION VISUELLE DE LA TYPOLOGIE

Trois types de considérations sont ici fondamentales.

### A. Les variables visuelles de J. Bertin

J. Bertin et S. Bonin ont distingué 6 “ variables visuelles ” (fig.5.1), c’est à dire 6 types de figuré graphique, aptes à cartographier tout type de distribution

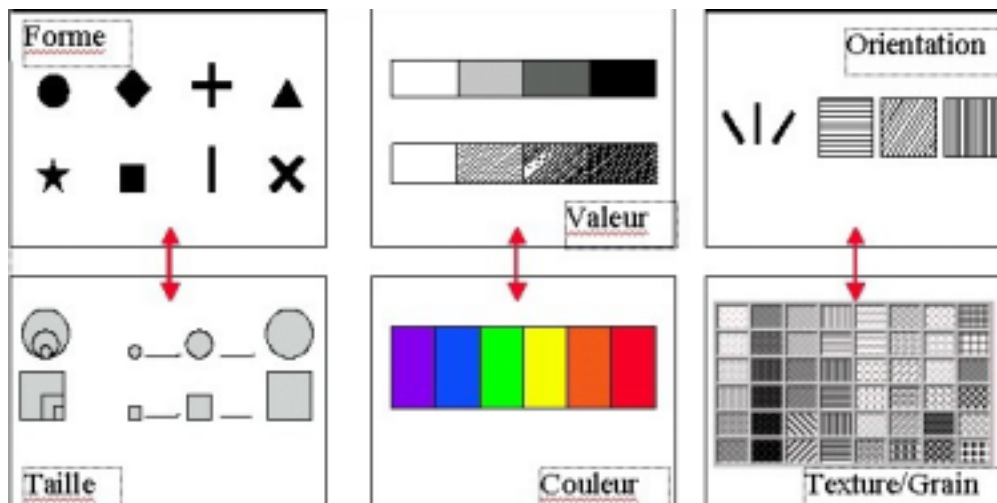


Figure 5.1 : Les 6 variables visuelles de J. Bertin et S. Bonin (source : S. Bonin)

- La variation de **forme** (carré, cercle, triangle, symboles divers,...) permet de représenter des variables *qualitatives nominales* mais, son pouvoir différenciateur étant assez faible, elle convient plus à des *cartes à lire* (avec attention) qu'à des *cartes à voir* (globalement).
- La variation d'**orientation** (traits de même épaisseur, horizontaux, verticaux, obliques) convient, elle aussi, à des variables *qualitatives nominales* pour des distributions à 4 ou 5 catégories (grand maximum 6) ; elle peut être combinée à d'autres variations, notamment en diversifiant l'épaisseur des traits.
- La variation de **couleur** est délicate à manipuler à bon escient. Il existe des palettes où le dégradé traduit :
  - & Soit, une simple différence de couleurs (de même intensité dans des tons différents) pour des variables *qualitatives nominales*,
  - & Soit, une gradation de couleurs de luminosité croissante pour des variables *qualitatives ordinales ou quantitatives relatives*.
  - & En cas d'échelle bi-polaire de valeurs, le cartographe emploie souvent des gradations de luminosité dans deux couleurs (par exemple du rouge clair au rouge foncé pour les taux de croissance démographique positifs, du gris clair au noir pour les taux négatifs).
- La variation de **texture** (ou de grain) est une variation de trame graphique (hachures, points,...) où la proportion de noir reste constante mais où les hachures ou les points sont ou bien gros et desserrés ou bien fins et serrés. Elle convient à des variables *qualitatives nominales*. En combinant variations de grain et de forme, on peut fort bien se passer de la couleur pour des échelles bi-polaire de valeurs.
- La variation de **valeur** d'une couleur est une variation d'intensité lumineuse par unité de surface, du clair au foncé. Elle convient pour représenter les valeurs croissantes de variables *ordinales ou quantitatives relatives*. Pour une échelle uni-

- polaire, un seul dégradé dans une seule couleur, pour une échelle bi-polaire, deux dégradés de valeur dans deux couleurs opposées.
- La variation de **taille** d'un symbole, cercle ou carré, est exclusivement réservé à des variables *quantitatives absolues*. La surface du symbole utilisé (cercle, par exemple) doit être proportionnelle à la valeur de la variable (nombre d'habitants de chaque ville, par exemple). La partition en classes, ici des tailles de ville, n'est d'ailleurs ni nécessaire ni recommandée car on peut faire correspondre une surface à chaque valeur.

	<i>Forme</i>	<i>orientation</i>	<i>couleur</i>	<i>grain</i>	<i>valeur</i>	<i>taille</i>
<b>Nominal</b>	X	X	X	X		
<b>Ordinal</b>			X			
<b>Qu. Relatif</b>			X		X	
<b>Qu . absolu</b>						X

Tableau 5.1 : correspondance entre nature statistique d'une série et variables visuelles

Si l'on considère le tableau 5.1 en ligne, du point de vue de la nature des données :

- les variables *nominales* peuvent être représentées par des figurés différant en forme, orientation, texture ou couleur,
- les variables *ordinales ou quantitatives relatives en classes* peuvent l'être par une variation soit de valeur soit d'intensité de couleur,
- les variables *quantitatives absolues* ne peuvent l'être que par une variation de taille.

Certaines variables visuelles peuvent être combinées, soit pour des échelles bi-polaires soit pour visualiser deux variables à la fois.

## **B. Primitives graphiques et variables visuelles**

Le type de symbolisation à utiliser doit aussi s'adapter aux types de localisation et forme des unités spatiales à représenter.

On distingue, en dessin assisté par ordinateur, 4 types de primitives graphiques (& 4.2):

- Le point (de dimension 0),
- La courbe (de dimension 1), souvent approximée par une polyligne (ligne brisée),
- L'aire (de dimension 2), délimitée par une courbe ou une polyligne fermée,
- Le volume (de dimension 3), utilisé en conception assistée par ordinateur ou en synthèse d'image mais peu en cartographie thématique.

La figure 5.2 présente une association possible entre :

- Primitive graphique (dimension et forme de l'unité spatiale),
- Nature statistique des données,
- Variable visuelle.

	POINT	LIGNE	SURFACE
NOMINAL	<i>Forme</i> 	<i>Forme</i> 	<i>Texture</i> 
ORDINAL / QUANTIT. RELATIF (CLASSES)	<i>Forme + autre</i> <div>Ville de niveau</div> 1 ○ 2 □ 3 ■	<i>Forme + autre</i> <div>50 %</div> <div>40%</div> <div>30%</div>	<i>Valeur</i> 
QTIT ABSOLU	<i>Taille</i> 	<i>Taille</i> 	<i>Taille</i> 

D'après M. Béguin et D. Pumain

*Figure 5.2 : Association entre variables visuelles, nature statistique et primitives graphiques*  
(source: M.Béguin, D.Pumain)

## C.L'habillage cartographique

De même qu'une carte thématique n'est pas interprétable sans sa légende, elle n'est pas "situable" sans un "habillage" qui doit, au minimum, comprendre :

- L'échelle, sous forme graphique de préférence (copiable en réduction),
- La direction du Nord géographique (pour permettre au lecteur de s'orienter),
- Quelques grands éléments de repérage (rivières, villes, réseau routier, sommets, noms, etc.).

### 5.1.3. PLUSIEURS VARIABLES SUR UNE MEME CARTE ?

La tentation est forte de surcharger la carte pour, d'un thème, "tout dire d'un seul coup". Tout dire n'est pas humainement possible mais dire plus qu'une chose l'est. Il existe trois façons de combiner plusieurs variables sur la même carte.

#### A.Combiner cartographiquement des variables

On ne peut (dans le meilleur des cas) combiner cartographiquement que deux variables :

- Soit en superposant des trames en bandes de largeurs proportionnelles (le résultat en est souvent illisible),
- Soit en définissant en légende des superpositions de couleurs (opération délicate),

- Soit encore en remplissant des cercles ou carrés proportionnels d'un figuré ou d'une couleur, mais cela ne vaut que pour une variable quantitative absolue et une variable qualitative à classes peu nombreuses.

Dans tous ces cas, la combinaison est limitée à 2 variables et la perception visuelle du message nécessite de gros efforts. La capacité combinatoire du procédé est faible et sa fiabilité également puisqu'il additionne deux fortes pertes d'information (deux discrétisations en deux ou trois classes au maximum pour chaque variable).

## B. Cartographier une combinaison obtenue par traitement graphique

La combinaison peut s'obtenir à l'aide :

- D'une collection de cartes c'est à dire de la juxtaposition de plusieurs cartes thématiques relatives au même territoire ; ces cartes juxtaposées ont, en général, une légende commune (en peu de classes, 2 ou 3, pour des raisons de lisibilité). Les collections de carte sont utilisées pour visualiser la distribution d'une variable à différents moments (exemple : croissances intercensitaires de population) ou celle, à un instant donné, de variables complémentaires (comme la distribution de plusieurs CSP sur un même territoire),
- Du graphique d'un nuage de points sur un plan XY orthonormé (croisant par exemple PIB/habitant et mortalité infantile); la partition en classes s'opère sur la distribution à 2 variables et la perte d'information est bien moindre qu'avec la superposition cartographique,
- Du graphique d'un nuage de points dans un diagramme triangulaire, à condition que la population ait été ventilée en 3 catégories sommant à 100% (comme la typologie des activités en secteurs primaire, secondaire et tertiaire) ; comme dans le cas précédent et pour les mêmes raisons, la perte d'information est bien moindre qu'en superposition cartographique,
- D'une matrice ordonnable qui est un tableau (lignes = unités spatiales, colonnes = variables) ; la valeur de chaque case est transformée en un grisé proportionnel. Permutations de lignes et / ou de colonnes permettent de faire des groupes de même profil visuel puis de les cartographier.

Les procédés 2, 3 et 4 sont des “ prémisses intuitifs ” de la statistique multivariée.

## C. Cartographier une combinaison obtenue par traitement statistique

Le traitement statistique à but typologique pour la cartographie généralise et codifie, en effet, des pratiques graphiques intuitives. L'analyse factorielle des correspondances, par exemple, est très proche du diagramme triangulaire mais en l'appliquant à un nombre de variables généralement supérieur à 3 . Comme les autres analyses factorielles, elle produit des projections sur des axes factoriels (résumés multi-variables). Les coordonnées sur les deux premiers axes factoriels permettent des regroupements graphiques en classes ; des méthodes de classification permettent de créer des groupes sur plus de 2 axes factoriels.

La carte devient, alors, une des aides à l'interprétation du traitement statistique.

### 5.1.4 UN PROBLEME: LA MODIFICATION DES UNITES SPATIALES

Vouloir combiner plusieurs variables sur la même carte peut se heurter à un problème connu sous le nom d' «unités spatiales modifiables » : Modifiable Areal Unit Problem en anglais, en abrégé MAUP (X). Les variables peuvent en effet être connues dans des zonages différents, d'extension différente.

Or, plus les zones sont de grande taille et différents, plus cela en perturbe l'analyse puisqu'on raisonne sur des agrégats importants et différents qu'on suppose, chacun, homogène. Cette hypothèse néglige bien évidemment les variances (hétérogénéités) internes. Le problème est en fait double : l'agrégation de données en zones pose un problème d'échelle (plus la zone est vaste, plus elle a de chances d'être hétérogène ce qui ôte sens aux sommes ou moyennes) et un problème de limites de ces zones. Pour une variable donnée, l'agrégation spatiale de valeurs individuelles en zones peut en effet donner lieu à une infinité de zonages différents, dont les sommes et les valeurs seront fort différentes. Si l'on traite, en outre, de zonages incompatibles d'une variable à l'autre, le problème est encore accru.

Il serait, par exemple, utile de pouvoir visualiser autour des grands axes de circulation le nombre d'habitants soumis à une nuisance sonore excessive. Mais, la population résidente n'est connue qu'à un niveau agrégé dans des limites administratives n'ayant rien à voir avec la diffusion du bruit en zone bâtie urbaine. Les deux zonages de connaissance sont incompatibles. D'autres cas d'incompatibilité peuvent se rencontrer s'il y a modification historique de limites.

Deux grands types de situations peuvent exister :

- La modification, par exemple d'une agglomération, peut se faire par ajout, au fil du temps, de communes entières. Comme la connaissance de base existe au niveau communal, gérer les données relatives à l'agglomération se fera par agrégation de données communales. Le géographe se trouvera face à une arborescence progressivement modifiée d'unités primaires. S'il veut établir les taux de croissance démographique successifs de l'agglomération « à géométrie fixe », il devra reconstituer les données du passé dans les limites actuelles.
- Un second cas, plus épineux, correspond aux zonages incompatibles : en témoigne l'exemple cité de « zones noires » sonores. Pour tenter d'estimer les populations résidentes gênées, il faudra désagréger les données démographiques :
  - & soit en affectant un nombre d'habitants proportionnel à la surface de chaque zone,
  - & soit en se servant d'une connaissance logiquement corrélée (le nombre et la hauteur des bâtiments observés par photo-interprétation, une base de données sur l'occupation du bâti, etc.)Il faudra ensuite agréger les nombres d'habitants estimés par immeuble dans les limites des zones noires sonores.

Les modifications de limites d'unités spatiales peuvent parfois même être volontaires, par exemple en matière de découpages électoraux. En jouer pour maximiser le nombre de députés d'une certaine « couleur » politique (« jerrymandering » en anglais) est une pratique occasionnelle (risquée en cas de modification de l'opinion publique).

## **5.2. INFORMATIQUE ET CARTOGRAPHIE THEMATIQUE AVANCEE**

### **5.2.1. Capacités de l'informatique utilisables en cartographie thématique**

La première, la plus évidente et la plus ancienne, est la capacité de **traitement** de données par des opérateurs logiques et / ou numériques, ce qui ouvre la voie à l'effectuation d'opérations. Vu la vitesse de fonctionnement des ordinateurs, les traitements peuvent être longs (quasi impossibles à la main) et complexes. La tâche fondamentale consiste alors à définir et conceptualiser précisément le problème à résoudre, à l'exprimer en une suite d'opérations élémentaires (ce que les spécialistes appellent *algorithme*) effectuable via un langage de programmation ou un logiciel existant. Vitesse et capacité opératoire se conjuguent pour permettre à l'utilisateur, une fois la démarche mise au point, de tester un certain nombre d'hypothèses pour en observer les résultats. On a là la base d'une **démarche quasi expérimentale** explicitement issue d'hypothèses et procédant par simulation et validation.

La deuxième caractéristique à mettre en avant est la capacité de **stockage** (donc de mémorisation) de grandes quantités d'informations, ouvrant la possibilité de grandes bases de données géographiques, aux multiples attributs, aux valeurs éventuellement étagées dans le temps. Cette grande capacité de stockage à moindre coût est un phénomène relativement récent.

La troisième caractéristique à considérer, spécialement pour la cartographie, est la baisse de coût et la grande amélioration récente des capacités **graphiques** (écrans, cartes d'affichage, imprimantes, dispositifs de saisie, etc.).

Les améliorations matérielles à coût décroissant ont permis la réalisation effective de méthodes restées jusque là virtuelles et théoriques, faute justement d'outil pour les réaliser. Tout le bénéfice n'en a pas encore été tiré par les cartographes.

### **5.2.2. Bénéfices en termes de traitement de l'information spatialisée**

Dans le contexte de la géo – informatique, une carte joue des rôles plus variés que jadis.

- Elle peut toujours être le résultat unique d'un traitement de données,
- Elle peut toujours être un instrument synthétique de communication,
- Elle est, de plus en plus, un document intermédiaire et transitoire dans une analyse,
- En tant que banque de données spatialisées, elle est le point de départ de traitements géographiques.

On négligera, dans ce chapitre, le traitement d'images et les Systèmes d'Information Géographique, détaillés ultérieurement.

## **A. La carte comme objet d'analyse**

On retiendra ici quelques grands types d'application relevant de cette problématique.

1. La carte thématique, papier ou informatique, peut être un excellent support d'**échantillonnage spatial**, aléatoire ou raisonné. On en trouve des exemples dans "l'analyse spatiale en géographie humaine" de P.Haggett. Une carte est en fait une base abondante de données spatialisées, à plusieurs attributs dans le cas d'une carte topographique, généralement à un seul dans le cas d'une carte thématique. Prenons pour exemple une image (carte rasteur) représentant l'utilisation du sol en quelques catégories : forêt, labours, prairies, eau, sol "artificialisé".



- Le premier mode envisageable de sondage consisterait à tirer au hasard (avec une calculatrice ou un ordinateur)  $n$  couples de coordonnées et d'y relever les utilisations du sol (mode de *tirage élémentaire sans remise*). La généralisation à toute la carte des pourcentages observés sur l'échantillon fournit des résultats relativement imprécis.
- On peut aussi envisager un *tirage systématique* consistant à superposer une grille à la carte et à relever l'utilisation du sol en tout croisillon de cette grille. On a là une méthode au moins aussi bonne que la précédente, voire meilleure car elle garantit le parcours de toutes les zones de la carte, sauf si le pas de sondage correspondait à une régularité paysagère (lignes de crête régulièrement espacées, par exemple).
- La généralisation à toute la carte à partir des résultats observés sur l'échantillon est, en fait, d'autant plus précise que la méthode de sondage intègre de la connaissance sur la carte, par exemple en *stratifiant l'échantillon*. Cela signifie, dans ce cas, d'y définir des zones de même utilisation du sol et de tirer au hasard un nombre de points proportionnel à leurs surfaces.

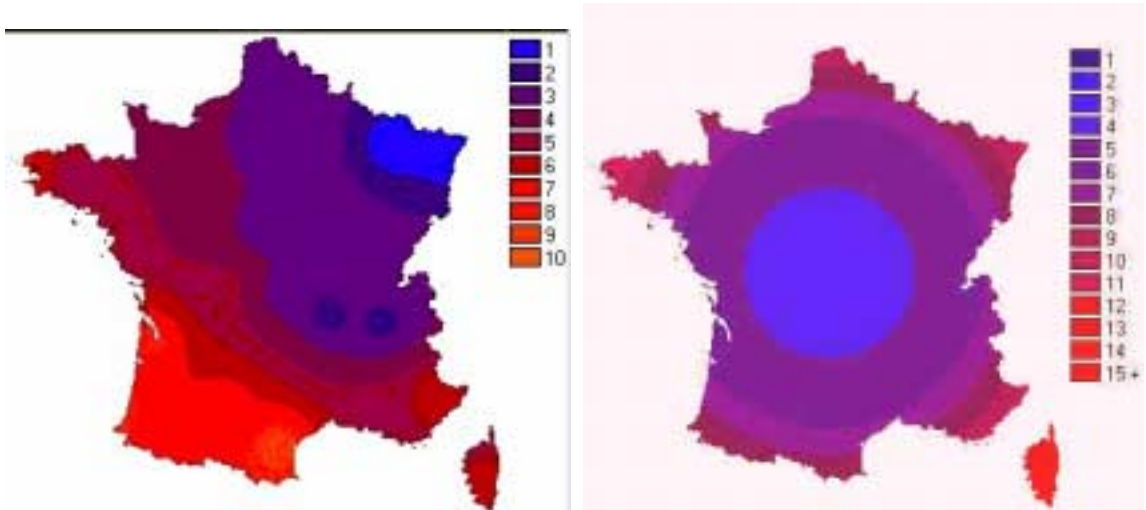
L'échantillon peut ensuite donner lieu à généralisation et traitements probabilistes, fournissant par exemple l'intervalle ayant  $x\%$  de chances de contenir la fréquence de telle utilisation du sol sur toute la carte. Bien sûr, l'une des exploitations consistera aussi à cartographier les points échantillon, pour y dépister une éventuelle mauvaise répartition.

2. A l'opposé, une carte ponctuelle (où les valeurs ne sont connues qu'en certains points) peut donner lieu à **interpolation** surfacique, manuellement ou à l'aide d'un algorithme (krigeage par exemple). La littérature en fournit maintes illustrations. Prenons pour exemple les températures minimales prévues pour le 2 janvier 2002 pour 25 villes de France métropolitaine (Tab 5.2).

VILLE	T° min	T° max	Longitude	Latitude
ajaccio	+1	10	7.11	46.183
biarritz	+6	12	-4.326	48.311
bordeaux	+5	10	-3.239	49.819
bourges	-5	3	0.065	52.315
brest	+3	9	-7.581	53.766
caen	-3	2	-3.007	54.646
cherbourg	-2	5	-4.391	55.154
clermont F.	-5	2	0.833	50.866
dijon	-6	2	3.005	52.581
grenoble	-7	3	3.776	50.207
lille	-5	-1	0.801	56.257
limoges	-1	5	-1.194	50.927
lyon	-4	3	2.782	50.843
marseille	-1	10	3.377	48.107
nancy	-10	-3	4.274	54.103
nantes	-1	7	-4.322	52.463
nice	+1	10	5.479	48.557
paris	-5	1	0.008	54.289
pau	+3	11	-3.007	48.111
perpignan	+8	13	0.62	47.441
rennes	-2	7	-4.463	53.46
st etienne	-8	1	2.281	50.481
strasbourg	-10	-1	6.012	53.981
toulouse	+3	11	-0.993	48.449
tours	-4	2	-1.831	52.658

Tableau 5.2: Températures mini et maxi prévues le 2/1/2002 (coordonnées en grades)  
(sources : Météo France et IGN)

Pour passer d'une carte ponctuelle (25 villes) à une carte couvrant tout le territoire, une interpolation est nécessaire. Nous allons utiliser, pour ce faire, la plus simple des techniques d'interpolation, dite "par inverse des distances". Pour tout point du territoire situé hors des 25 villes, la température minimale qui lui est affectée est la moyenne de celles des 25 villes, pondérées par l'inverse de leurs distances à ce point ( $1/d_j^2$ ). Cela revient à dire que la température de ce point doit être très comparable à celle des villes les plus voisines : ce type d'interpolation se justifie dans le cas d'une grandeur spatialement continue mais serait fallacieux dans le cas de fortes discontinuités spatiales.



Figures 5.3a et 5.3b: Températures minimales interpolées (2//1/2002) et fiabilité spatiale dde l'interpolation

La Figure 5.3.a visualise le résultat obtenu (10 classes d'amplitude  $2^\circ$  :  $[-10^\circ$  à  $-9^\circ$ ],  $[-8^\circ$  à  $-7^\circ$ ], etc.), tandis que la Figure 5.3.b indique la fiabilité spatiale des températures interpolées : pour chque pixel, sa somme de distances aux 25 villes ont été rangées en 15 classes de valeur, de la classe 1 (fiabilité maximale) à la classe 15 (fiabilité minimale). Sans surprise aucune, la fiabilité de l'interpolation est meilleure au centre que sur les marges : cet "effet frontière" est dû à l'absence de données en mer et à l'étranger.

Ces deux cartes comportent déjà un certain lissage puisque l'interpolation procède par moyenne pondérée des températures minimales prévues pour 25 villes.

3. Même dans les cas où les valeurs sont connues pour toutes les entités spatiales, un **lissage** cartographique peut s'avérer nécessaire,
  - pour des *données sensibles*, par exemple : entrées et sorties d'argent public pour chaque commune d'une région (l'objectif n'est pas alors de montrer du doigt certaines communes mais de dégager des zones floues, déficitaires ou bénéficiaires),
  - chaque fois que le cartographe veut suggérer une *variation continue* dans l'espace (un "champ" de forces) et s'affranchir de limites administratives arbitraires pour le phénomène étudié.
  - pour "adoucir" ("smooth" en anglais) une image.

Considérons un **Modèle Numérique de Terrain** sous forme d'image raster : à chaque pixel (représentant une certaine surface au sol) est associée son altitude moyenne. Diverses cartes peuvent en être déduites : représentations du relief en 2 ou 3 dimensions, cartes de pente, d'orientation, d'illumination qui, combinées à d'autres cartes thématiques, acquièrent un fort pouvoir explicatif.

Pour illustrer le fonctionnement d'un lissage, examinons l'image du relief de la plaine de Dirol en Mauritanie et son équivalente lissée par *moyenne mobile spatiale* d'ordre 7 (Figure 5.4). Ce procédé de lissage est un des plus simples qui soient : il consiste à définir une "fenêtre" de 7 lignes – 7 colonnes, d'y calculer l'altitude moyenne et d'affecter cette valeur au pixel central. La fenêtre est ensuite déplacée d'une colonne et, en fin de colonne, d'une ligne, opération répétée jusqu'à balayage complet de l'image.

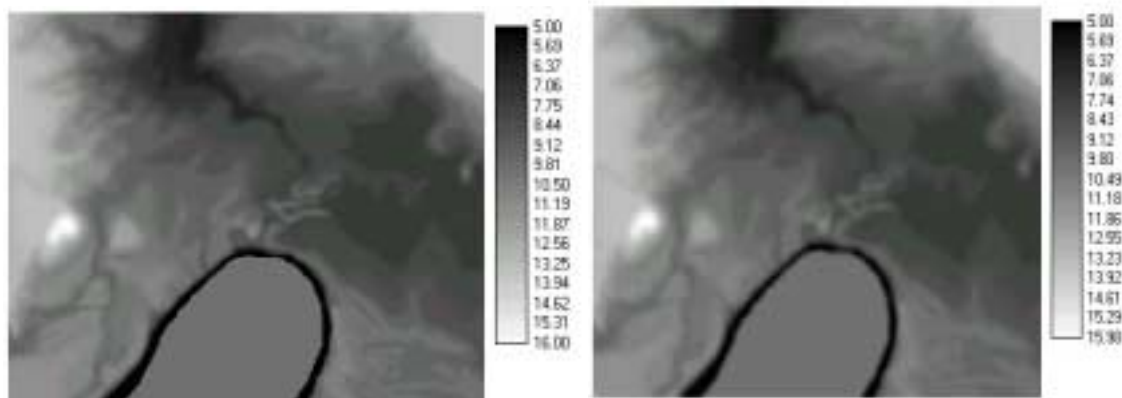


Figure 5.4: image originelle et lissée du relief de la plaine de Dirol (Mauritanie)

L'adoucissement des formes de ce micro - relief est manifeste mais l'apparence de sa structure demeure tout à fait lisible. Seules ont été gommées certaines « aspérités » secondaires. Le lissage d'une image en est, d'une certaine manière, une généralisation.

4. Tout au contraire, le cartographe peut vouloir souligner des **discontinuités spatiales** (fortes différences de valeur entre zones voisines ou franchissement d'une valeur-seuil).

La figure 5.5, due à C.Grasland, en fournit une illustration.

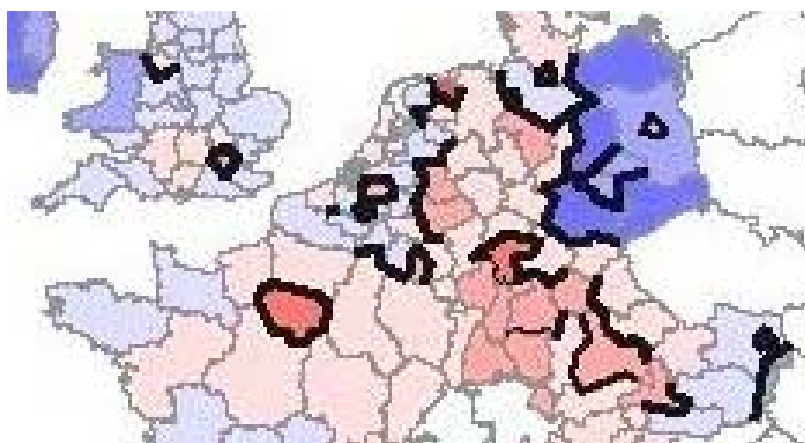


Figure 5.5 : Discontinuités de valeur entre régions contigües (source : C.Grasland)

5. La mesure de l'**autocorrélation spatiale** constitue un indicateur global de discontinuité spatiale. Si l'autocorrélation est faible, les unités voisines ne se ressemblent en général pas, ce qui justifie le marquage sur la carte des fortes différences entre voisins. Si elle est forte, il ne se justifie plus.

Pour faire saisir ce qu'est l'autocorrélation spatiale, prenons pour exemple le tableau 5.3. Il indique, pour 17 pays ouest européens (les 15 pays de la Communauté Européenne plus la Suisse et la Norvège) leur Produit Intérieur Brut / Habitant en 1999 (source : Eurostat).

Pays	PIB/h
Norvege	16028
Suisse	20874
Suede	17014
France	17405
Pays Bas	15695
Grande Bretagne	15804
Allemagne	18213
Danemark	16781
Finlande	16446
Autriche	16504
Belgique	16381
Luxembourg	19244
Eire	10589
Italie	15890
Espagne	11723
Grece	7366
Portugal	8770

Tableau 5.3 : PIB/Habitant de 17 pays ouest européens en 1999

La question posée par la mesure d'autocorrélation spatiale sur cet exemple est la suivante : pour ce qui est du PIB/habitant, la variance des unités spatiales voisines est elle plus faible que la variance globale ?

Si elle est plus faible, on parle d'autocorrélation positive : les lieux voisins sont plus semblables entre eux qu'ils ne le sont des autres lieux. Dans le cas présent, on a considéré comme voisins les pays contigus et on a mesuré l'autocorrélation spatiale avec l'indice C de Geary .

$$C = \frac{\sum_i \sum_j C_{ij} \{(X_i - X_j)^2 / C_{tot}\}}{\sigma^2_x}$$

avec  $C_{ij} = 1$  si unités spatiales contigües, 0 sinon

$C_{tot}$  = nombre total de couples d'unités spatiales contigües

$X$  = variable dont on calcule l'autocorrélation

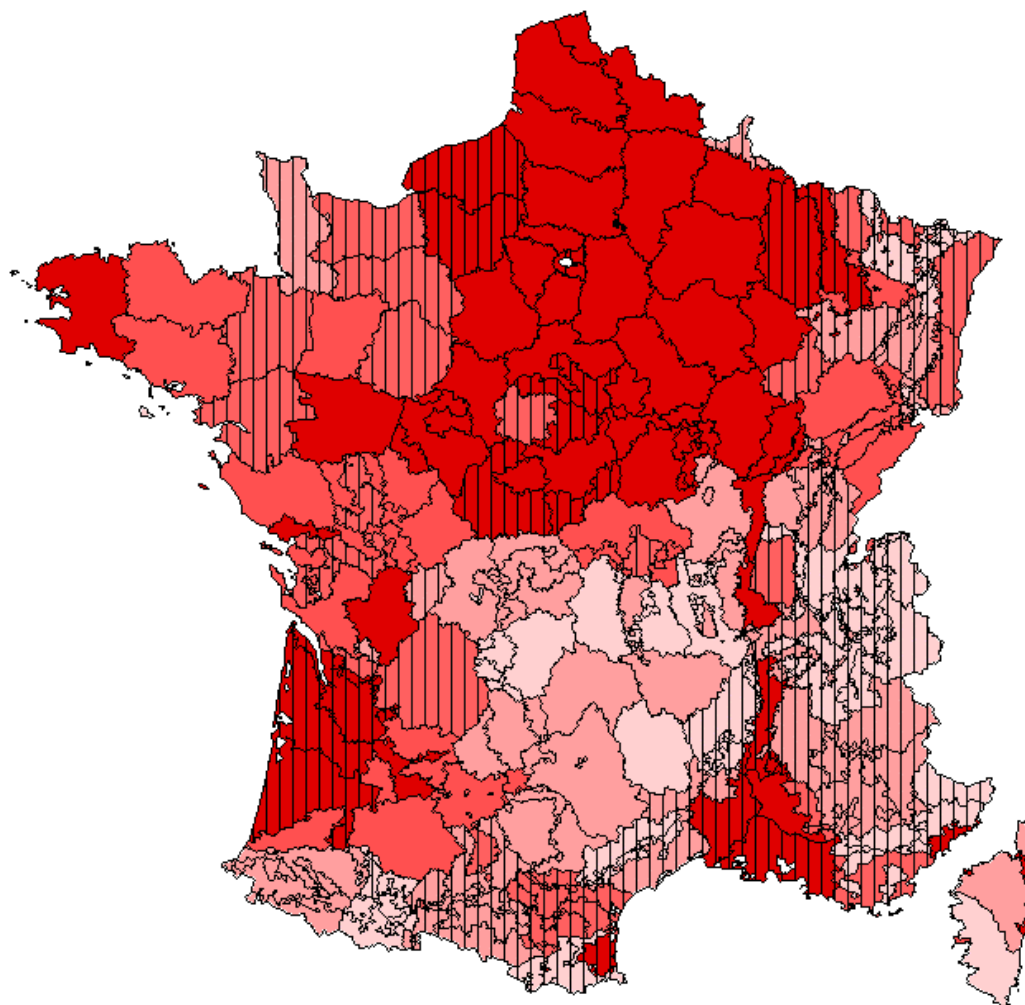
$\sigma^2_x$  = variance de  $X$

$C$  vaut ici 0,225, ce qui indique une forte autocorrélation positive : la variance de  $X$  entre pays contigus vaut moins du quart de la variance totale : la carte thématique du PIB / Habitant en Europe de l'Ouest est organisée par blocs de pays voisins présentant des valeurs comparables.

Un certain nombre de **paramètres** ont été proposés pour caractériser une carte thématique. Il semble qu'outre ceux là, l'un des paramètres les plus importants à déterminer et à visualiser concerne la **fiabilité**, globale et locale, de ce qu'affiche une carte après traitement de l'information.

La fiabilité *globale* d'une carte peut être estimée, par exemple par une analyse de variance. Mais il est plus important encore de faire nettement apparaître les différences *locales* de fiabilité, superposées ou juxtaposées à la carte des valeurs (cf Fig. 5.3). C'est plus fondamental encore si la carte doit être utilisée comme instrument d'aide à la décision.






## Mesure et variabilité de la dimension économique des exploitations







### Valeur de la dimension économique moyenne des exploitations :

- Avec indice de Gini  $< 0,518$   
(faible variabilité)

- Avec indice de Gini  $> 0,518$   
(forte variabilité)

-  Sans valeur
-   $6,24 \leq \text{mbs/ea} < 11,2$
-   $11,2 < \text{mbs/ea} < 16,3$
-   $16,3 < \text{mbs/ea} \leq 21,9$
-   $21,9 < \text{mbs/ea} \leq 69,52$

-   $5,95 \leq \text{mbs/ea} \leq 11,2$
-   $11,2 < \text{mbs/ea} \leq 16,3$
-   $16,3 < \text{mbs/ea} \leq 21,9$
-   $21,9 < \text{mbs/ea} \leq 43,5$

Source Eurofarm-Eurostat 1989 / Agrireg

Figure 5.6 : Superposition des valeurs d'une variable et de leur significativité locale (source: L.Merchez)

La figure 5.6 représente la dimension économique moyenne (mbs / ea = “ marge standard brute ” / nombre d’exploitations) des zones agricoles, défavorisées ou non (zonage départemental ou infra-départemental) et illustre l’idée de fiabilité spatiale différentielle. Connaissant, pour chacune des deux variables et de ces zones, leur distribution interne, il a été possible de calculer un indice d’hétérogénéité (indice de Gini, dans ce cas). Les zones à forte hétérogénéité interne ont été hachurées : leur valeur moyenne cache de fortes disparités et n’est donc pas “ fiable ” : une description correcte nécessiterait des données plus locales. En passant, on remarquera avec intérêt que certaines zones défavorisées (au sens de la politique agricole de la Communauté Européenne qui les aide) ont des exploitations de dimension économique fort diverse !

## B. La carte, objet informatique spécifique

A la différence de la carte papier, objet « définitif », la carte produite informatiquement est un objet « volatile » servant de test d’une idée ou d’étape dans un raisonnement géographique et la séquence des cartes est souvent tout aussi instructive que le document final. La cartographie informatique est ainsi source d’expérimentations visuelles.

Une première manifestation en est la pratique de l’**anamorphose cartographique**, qu’on pourrait définir comme une transformation graphique de la position des objets destinée à mieux traduire leurs valeurs sur une variable donnée (d’état ou de flux). La cartographie thématique classique accomplit cette traduction en affectant trames ou couleurs à des unités spatiales conservant leurs positions absolues dans l’espace euclidien alors qu’une anamorphose ne joue pour cela que sur des déformations spatiales.

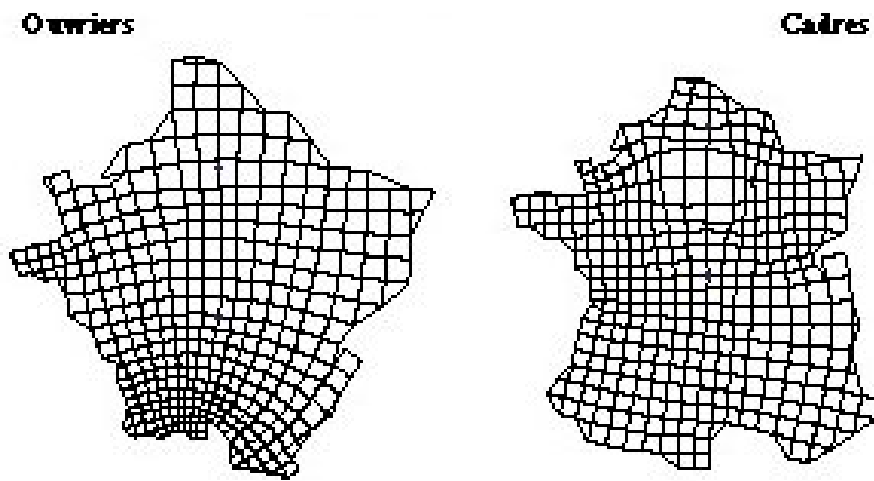
On en trouve une première illustration, simple, dans certaines mappemondes des manuels de classe Terminale, où la surface des états est proportionnelle à leur Produit National Brut, à leur poids démographique ou à toute autre variable significative de leur puissance.

L’exemple permet déjà de fixer quelques caractéristiques : C’est l’*”espace thématique ”* qui est projeté sur un plan, plutôt que l’espace euclidien habituel,

- le procédé de construction consiste, dans ce cas, à juxtaposer un nombre de carrés élémentaires (“ pixels ”) proportionnel à la valeur de chaque état,
- la position exacte des états ne peut être respectée mais leurs positions relatives et leurs contiguités le sont,
- l’arrangement des carrés élémentaires schématise la forme de chaque état.

En un mot, la géométrie des états est chahutée mais leur topologie est préservée. Ce principe est essentiel pour toute anamorphose cartographique, procédé popularisé par W.Tobler et C.Cauvin.

L’exemple suivant, consultable « sous la plume de C.Cauvin sur le site Cybergeog, présente deux cartes « piézoplèthes » (piézo : pression apte à déformer).



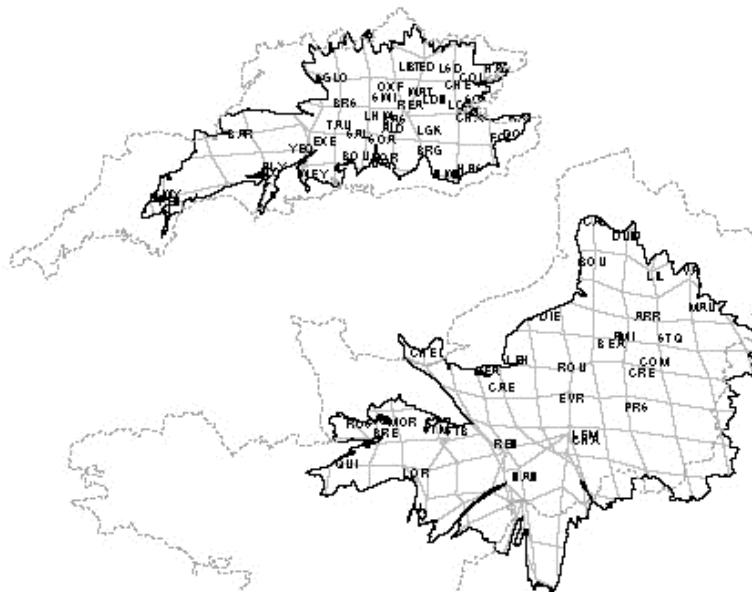
*Figure 5.7 : carte par anamorphose de la répartition des ouvriers et des cadres  
(source : C.Cauvin)*

Comme dans l'exemple des mappemondes de classe terminale, la surface d'une maille représente l'importance numérique de telle ou telle catégorie mais ici des forces leurs sont appliquées et réparties continûment sur toute la surface. Le procédé provient du calcul de résistance des matériaux employé en génie civil.

Ce sont bien les seules déformations géométriques de la grille qui révèlent l'importance des proportions d'ouvriers dans le Nord de la France et de cadres en régions parisienne et rhône-alpine.

L'anamorphose cartographique est fréquemment employée pour représenter un « espace d'accessibilité », routière, ferroviaire ou pédestre et sa déformation par rapport à l'espace euclidien de la cartographie classique. Dans ce cas, on définit les distances entre lieux comme des distances – temps de ou coût de parcours plutôt que comme des distances kilométriques sur un réseau. La figure 5.8 présente l'exemple, dû à P.Langlois, des espaces – temps d'accès de part et d'autre de la Manche depuis chaque capitale.

#### **DÉFORMATION À PARTIR DE CHAQUE CAPITALE NATIONALE (temps d'accès ferroviaire hiver 92-93)**



*Figure 5.8 Exemple d'anamorphose (source : P. Langlois, site Cybergeo)*



On peut noter des différences par rapport aux mappemondes des classes terminales :

- Les entités spatiales considérées sont encore surfaciques mais il s'agit cette fois des cases d'une grille et ce n'est plus le nombre de cases qui est proportionnel à la valeur de la variable mais leur surface (comme dans l'exemple de C.Cauvin),
- A la différence de l'exemple précédent, la variable n'est plus une variable d'état instantané mais de flux,
- Dans tous les cas, l'anamorphose cartographique suppose le calcul de déformations cohérentes de toutes les cases de la grille, de façon à maintenir les contigüités.

D'autres auteurs ont étendu l'anamorphose des pavages territoriaux aux réseaux pour la représentation d'espaces - temps ou coût de parcours.

A mon sens, deux cas se présentent :

- Si les utilisateurs du réseau peuvent entrer et sortir en tous points de celui ci, les anamorphoses linéaires (longueur de chaque tronçon proportionnel à son temps ou son coût de parcours) ne posent pas de problème supplémentaire par rapport au cas précédent, car elles changent la géométrie des lieux mais non leur topologie (leur disposition relative),
- Par contre, avec les moyens de transport à entrées / sorties éloignées (autoroute, train, avion), la géométrie ET la topologie sont changées. Considérant un espace - temps ferroviaire, Mâcon (sans arrêt TGV) devrait être situé au sud de Lyon par rapport à Paris ! Les relations d'ordre spatial sur le réseau ne sont plus préservées. Dans un tel exemple, il faudrait ou bien ne raisonner que sur l'espace des villes directement desservies par le TGV ou bien représenter les temps de parcours par un " relief virtuel ".

W. Tobler fournit aussi des exemples intéressants d'anamorphose cartographique appliquée à de tout autres thèmes.

### C. Transformations dimensionnelles d'objets cartographiques

On peut aussi évoquer brièvement d'autres transformations géométriques des objets que les anamorphoses, transformations que l'on peut qualifier de « **dimensionnelles** », en ce sens qu'elles consistent à transformer des objets de dimension 0 (points) ou 1 (lignes) en surfaces ou réciproquement. Le tableau 5.4 récapitule les transformations possibles.

De → A	POINT	LIGNE	SURFACE
POINT	-	1	2
LIGNE	3	-	4
SURFACE	5	6	-

Tableau 5.4 : transformations dimensionnelles des objets cartographiques

On omet ici les transformations non dimensionnelles, figurant sur la diagonale du tableau : point central d'un semis de points, tronçon stratégique d'un réseau, province centrale d'un pays par exemple.

- La transformation 1 consiste à relier en un trajet des points de passage obligés, comme dans le problème du voyageur de commerce (qui doit optimiser son parcours tout en visitant tous ses clients).

- La transformation 2 consiste, par exemple, à interpoler sur tout un territoire des valeurs connues ponctuellement (cf Fig. 5.3)
- La transformation 3 consisterait, entre autres, à résumer un réseau routier par ses carrefours
- La transformation 4 est habituelle dans tous les calculs d'accessibilité à une ville consistant à simuler le mouvement à partir d'un réseau de communication sur tous les espaces accessibles.
- La transformation 5 consisterait, par exemple, à déterminer le centre de la France.
- La transformation 6 résumerait, par exemple, les principales relations entre circonscriptions téléphoniques d'une région par un graphe hiérarchique.

### 5.2.3 Bénéfices en termes de visualisation

Un second type de traitements informatiques pour la cartographie thématique concerne des essais ou des idées pas encore complètement formalisés. On peut faire, à ce propos, quelques observations.

1) Ma première observation concerne l'emploi de la couleur sur les matériels actuels (16 Millions de couleurs, plus que l'œil ne peut en percevoir !).

- On peut avancer l'idée, contraire aux usages, qu'il n'est pas toujours nécessaire, en cartographie thématique, de réduire les valeurs d'une variable à peu de classes, surtout si l'on veut suggérer sa continuité spatiale. Les matériels informatiques actuels permettent, en effet, de varier assez finement la luminosité dans une couleur fondamentale.
- Par ailleurs, l'utilisation de la couleur en transparence permet de plus facilement lire la combinaison cartographique de deux variables.

Un élargissement de la sémiologie graphique et l'adaptation de ses règles aux supports informatiques sont activement recherchés.

2) Une autre pratique, celle du dessin animé cartographique, s'est déjà révélée féconde. Un exemple connu (cf Fig. 5.9) est celui de la diffusion du SIDA aux USA.

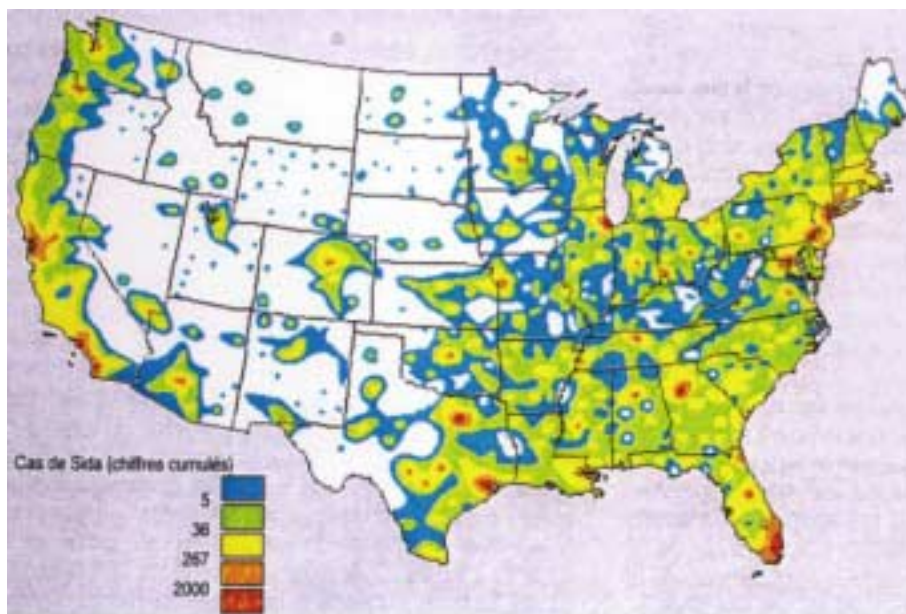


Fig. 5.9 : Une des planches du dessin animé cartographique de P.Gould

D'autres applications, moins médiatiques, pourraient être citées. Dans tous les cas, le procédé convient particulièrement bien pour représenter le spatio - temporel : évolutions territoriales (la concentration progressive de population dans les aires urbaines, par exemple), mobilité quotidienne (une sorte de " respiration " du territoire), traitement et cartographie d' " emplois du temps et de l'espace " de la Time Geography, etc.

3) Un autre mode de représentation a également été parfois pratiqué, le traitement de la carte thématique comme un paysage virtuel, avec son " relief " (fortes et faibles valeurs), son " utilisation du sol (une seconde variable, qualitative), la possibilité pour l'utilisateur de définir ses points de vue successifs (son itinéraire dans l'information)

4) On ne peut exactement ranger les chorèmes de R. Brunet dans la même catégorie d'adaptation de la cartographie thématique aux évolutions de l'informatique. Son ambition scientifique va bien au delà. Elle est de construire un **langage de synthèse cartographique** (alors que le " croquis régional " construit des synthèses cartographiques sans langage, c'est à dire sans règles explicites).

Par analogie aux phonèmes du langage parlé, les chorèmes sont les éléments du langage cartographique et leur assemblage, pour former des signifiés (des mots, des phrases, un texte), suit un certain nombre de règles, non encore complètement explicitées.

Du point de vue méthodologique, avant l'expression via un langage, il y a raisonnement spatial, " modélisation " du fonctionnement d'un territoire. Vu l'ampleur de l'ambition, la tâche est réalisée de façon artisanale, ce qui a pu amener des dérives par rapport à l'ambition première.

Si elle n'est pas informatisée, la chorématique est informatisable, à divers degrés, soit par un outil de conception assistée par ordinateur (contenant alphabet, règles de grammaire, opérateurs graphiques), soit comme système incluant aussi des possibilités de raisonnement spatial.

5) Sans doute, au delà de tous les essais isolés, le progrès le plus important, aux yeux d'un géographe utilisateur de technologie, serait qu'il existe un outil d'analyse spatiale exploratoire, intégrant et faisant interagir des méthodes d'interrogation, de traitement, de représentation graphique et de cartographie de l'information. C'est cette interactivité et cette intégration que proposent, partiellement, et l'analyse spatiale exploratoire et les Systèmes d'Information Géographique.

## Références citées

J. Bertin, *Sémiologie graphique* (Mouton & Gauthier-Villars, 1967)

J. Bertin,, *La graphique et le traitement graphique de l'information* (Flammarion, 1977)

S. Bonin, *Initiation à la graphique* (EPI, 1975)

S. Rimbert, *Carto-graphies* (Hermes-science, 1990)

R. Brunet, *La carte, mode d'emploi* (Fayard-Reclus, 1987)

M. Béguin, D. Pumain, *La représentation des données géographiques* (A.Colin, 1994)

- S. Openshaw, P.J. Taylor, *The modifiable Areal Unit Problem* in N.Wrigley & R.J. Bennett ed.  
« Quantitative Geography » (Routledge, 1981)
- D. Pumain, Th. Saint Julien, *Les interactions spatiales* (A.Colin, coll. U, 2001)
- P. Haggett, *L'analyse spatiale en Géographie Humaine* (A. Colin, 1973 )
- Chadule, *Initiation aux pratiques statistiques en géographie* (Masson, 1997)
- L.Sanders, *L'analyse statistique des données appliquée à la géographie* (Reclus, 1990)
- C.Grasland, in rubrique “ cartographie ” du site cybergeog
- L. Merchez, *Exploration of the Agrireg data base for a spatial analysis of the french agricultural space* (in  
C.Laurent, I.Bowler : CAP and the regions, InRA ed., 1997)
- C.Cauvin, *Cartographie théorique et anamorphoses*.(Comité Français de cartographie, Bulletin 146-147, 1996)  
Aussi <http://cybergeog.presse.fr/cartogr/texte/transcar.htm>
- P.Langlois, <http://cybergeog.presse.fr/>
- W.Tobler, *Pseudo cartograms* (The american cartographer, n°7, 1986)
- P. Gould, *Sources d'erreur dans une série de cartes* (Mappemonde, 1993-2)

# **CHAPITRE 6 : SYSTEMES D'INFORMATION GEOGRAPHIQUE**

*Pierre Dumolard*

En toute première analyse, les Systèmes d'Information Géographique (S.I.G. dans la suite de ce texte) sont des logiciels capables de gérer et de représenter des données spatialisées. Nés dans les années soixante, ils résultent de la confluence de plusieurs technologies :

- La cartographie et ses méthodes amont (cf chap. 5),
- Le dessin technique de type « génie civil »,
- Les bases de données, conçues pour la gestion des organisations (cf chap. 3),
- L'interprétation d'images satellitales (cf chap. 7).

Ils constituent aujourd'hui un secteur d'activité économique en plein essor et le noyau d'une convergence pluri-disciplinaire, la **géomatique**.

Encore faut-il préciser ce qu'elle est, ce qu'elle n'est pas, à quoi et à qui elle sert.

## **6.1. DEFINITIONS DES S.I.G.**

On peut donner du mot S.I.G. des définitions étroites, strictement logicielles ou des définitions élargies, géomatiques.

### **6.1.1. Définitions logicielles**

#### **A. Au sens le plus strict du terme,**

***S.I.G. = S.I. pour I.G.***

Traduisons cette petite charade : un S.I.G. est un **Système d'Information** adapté à une information particulière, l'**Information Géographique**.

Dans cette définition là, un S.I.G. comprend trois sous ensembles :

- un Système de Gestion de Bases de Données (cf chapitre 3),
- un module de cartographie vectorielle (cf chap. 5),
- un « pont », assurant la correspondance entre les deux.

Même d'un point de vue uniquement logiciel, cette définition est trop restrictive, car elle exclut tout système fondé sur la représentation image (rasteur) et sur d'autres opérateurs que ceux des bases de données.

#### **B. Si l'on étend la définition aux représentations « rasteur »,**

un S.I.G. assure **cinq grandes fonctionnalités** sur l'information spatialisée (les **5A**) :

1. L'**Acquérir**,
2. L'**Archiver**,

3. L'Accéder,
4. L'Analyser,
5. L'Afficher.

Les points 1 et 2 concernent essentiellement les producteurs (publics ou privés), les points 2 et 3 les informaticiens, les points 4 et 5 les thématiciens (dont cartographes et géographes).

Mais, on ne saurait réduire le domaine des S.I.G. à leur seul aspect logiciel.

### 6.1.2 Définitions géomatiques

La géomatique, outre ses facettes logicielles, met l'accent sur l'amont et l'aval des S.I.G. stricto sensu.

#### C. A l'amont des S.I.G.,

il y a des **données géoréférencées**, considérées :

& d'un point de vue technique (production, informatisation, cf chapitre 4),

& et, surtout, du point de vue de leur existence (sources de données), de leur accessibilité (aspects juridiques et institutionnels), de leur coût économique (environ les  $\frac{3}{4}$  du coût, hors personnel, d'un projet S.I.G. est imputable aux données).

#### D. A l'aval des S.I.G.,

Tout ce qui est relatif à l'aspect **organisationnel** de leur insertion dans des services. Tout système d'information est, en effet, un « système humain assisté par ordinateur » et, dans la pratique, les problèmes majeurs relèvent souvent davantage de la « science des organisations » que de considérations plus techniques.

Dans cet ouvrage, sans méconnaître les aspects économiques, institutionnels et organisationnels du domaine, nous nous limiterons à décrire des caractéristiques logicielles aptes à garantir un bon usage par des étudiants et chercheurs en géographie.

**On peut donc définir la géomatique comme un large ensemble de connaissances pluridisciplinaires destinées à gérer et aménager des territoires à différentes échelles.**

Elle inclut donc des aspects fort variés :

- Géoréférencement
- Bases de données
- Traitement d'image
- Données spatialisées
- Analyse spatiale
- Insertion dans les organisations .

### 6.1.3 Organisation de l'information en « couches » homogènes

Préalablement à toute application, les données doivent être **organisées en couches, homogènes** des points de vue thématique et géométrique. Cette

décomposition permet à l'utilisateur de les combiner à volonté à l'aide d'un S.I.G

Prenons pour exemple le stock de données nécessaires à l'agriculture de précision. C'est une agriculture visant à adapter les intrants et les pratiques culturales à la variabilité des conditions à l'intérieur des parcelles (et non plus à chaque parcelle considérée comme un tout). La figure 6.1 visualise une décomposition en couches d'information dans ce genre d'application.

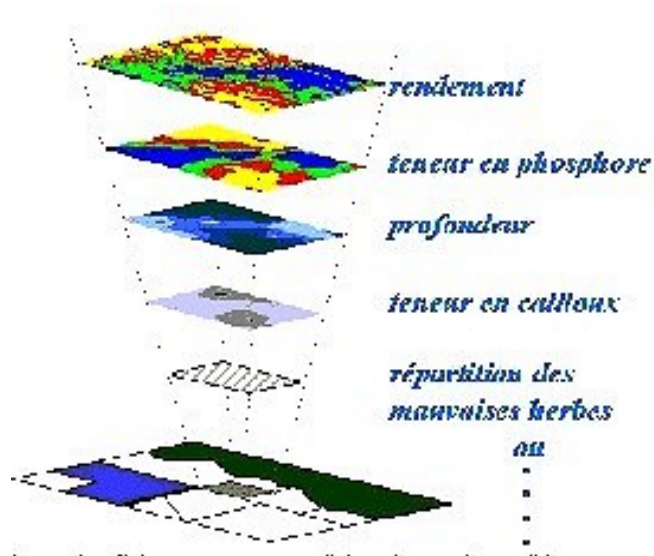


Figure 6.1 : exemple d'organisation en couches d'information homogènes (source : I.T.C.F.)

La provenance de ces informations est le plus souvent fort diverse. Le S.I.G. joue donc le rôle d'intégrateur, sur un même référentiel géographique, de données **multisources**.

La décomposition en couches doit s'opérer avant toute mise en œuvre, que le S.I.G. soit vectoriel ou rasteur. Les requêtes et opérations possibles n'y sont, évidemment, pas les mêmes.

## **6.2 PRINCIPALES UTILISATIONS DES S.I.G.**

On peut décrire ces utilisations :

- soit par rapport aux questions qu'on peut leur adresser,
- soit par leur domaine thématique d'application,
- soit encore par secteur d'activité.

### **6.2.1 Types de questions à un S.I.G.**

Un logiciel S.I.G. peut répondre à 5 grands types de questions.

1. ***Où*** trouve t'on cet objet ou ensemble d'objets ? La réponse à cette question est une carte de répartition des valeurs d'un attribut ou de plusieurs liés par des opérateurs (cf chap.3). Alors, la requête procède



- des attributs vers la carte*, qui en exprime le résultat. A la différence des cartes thématiques, une sélection seulement de toutes les entités y figure.
2. **Quoi** à cet endroit, désigné sur la carte ponctuellement ou par fenêtrage ? Lister les valeurs de tel et / ou tel attribut. Ici, *la requête procède de la carte vers les attributs* et les affiche.
  3. **Et si on modifiait** telle(s) ou telle(s) condition(s), que se passerait-il ? Quelles localisations changerait-on ? La possibilité d'introduire des scénarios spatiaux autorise la simulation, par exemple l'exploration des conséquences d'un projet d'aménagement.
  4. **Quelles relations spatiales** existe-t-il entre tout ou partie de la carte A et de la carte B ? On aborde évidemment là quelque chose d'essentiel pour l'analyse spatiale exploratoire.
  5. Quelles étaient les configurations territoriales des attributs C et D *au moment Ti* ou sur la *durée Ti à Ti+k* ? Une porte est, là, grande ouverte vers le spatio-temporel.

## 6.2.2 Domaines thématiques d'utilisation

Le tableau 6.1. résume les grands domaines thématiques d'utilisation.

<b><u>GESTION TERRITORIALE</u></b>	<b><u>PLANIFICATION / RECHERCHE</u></b>
S.I.G. urbains	S.I.G. recensement (population, activités)
S.I.G. réseaux techniques/infrastructures	S.I.G. environnemental
S.I.G. transports / communications	S.I.G. aménagement (P.O.S., S.D.A.U.,...)
S.I.G. cadastraux	S.I.G. relief (M.N.T.)
S.I.G. agricoles	S.I.G. paysager / patrimonial
S.I.G. études de marché	S.I.G. pour modélisation / simulation

Tableau 6.1. : Domaines thématiques d'utilisation des S.I.G.

## 6.2.3 Secteurs d'activité géomatique

Les principaux secteurs d'activité où peuvent intervenir des S.I.G. sont :

- *Cartographie* : Bases de Données Géographiques, Routières, cartes papier, atlas, presse, préparation et publication de recensements, etc.
- *Génie Civil / Urbanisme* : travaux sur réseaux techniques (eaux, électricité), projets de bâti ou de voie de communication, etc.
- *Collectivités Territoriales* : aménagement local ou régional,
- *Gestionnaires de réseaux* (ex : autoroutes) *et de flux* (ex : routage, gendarmerie)
- *Gestionnaires de Ressources naturelles* : énergie, mines, agriculture,
- *Environnement* : gestion de l'eau, des déchets, des nuisances (bruit, pollutions), des zones protégées et réserves naturelles, etc.
- *Santé, Gestion des Risques et des Secours* : protection civile, services d'urgence, pompiers, police, etc.
- *Etudes de marché* : publicité postale ciblée, enquêtes, localisation d'activités, etc.

- *Domaine militaire* : planification d'opération, visée de tir en temps réel, logistique, etc.

## 6.3 S.I.G. VECTEUR

### 6.3.1. Modèles de données

Un S.I.G. vectoriel **intègre** un système de gestion de bases de données (S.G.B.D.) et un cartographeur. Les opérations possibles dépendent des types de logiciel utilisé, qu'on peut affecter à l'une des quatre catégories du tableau 6.2. Ils se distinguent par leur « *modèle de données* » qui vont du plus simple et moins puissant (A) au moins simple et plus puissant (D).

MODELES DE DONNEES	<u>B.D. MONO-FICHIERS</u>	<u>B.D. MULTI-FICHIERS</u>
	(NON RELATIONNELLE)	(RELATIONNELLE)
<b><i>GEOMETRIQUE SEULEMENT</i></b>	<b><u>A</u></b>	<b><u>B</u></b>
<b><i>GEOMETRIQUE ET TOPOLOGIQUE</i></b>	<b><u>C</u></b>	<b><u>D</u></b>

Tableau 6.2 : Quatre types de S.I.G. vectoriels selon leur modèle de données, géométrie et sémantique.

Le modèle de données **A** est le plus fréquent dans les logiciels du marché : il permet un nombre restreint d'opérations (suffisantes pour des besoins simples) ; il peut, cependant, être amélioré par liaison avec un S.G.B.D. relationnel (type **B**). Le type **C** correspond à des logiciels spécialisés dans la gestion d'itinéraires. Le modèle de données **D** est le plus évolué et permet des requêtes plus complexes et plus efficaces mais le logiciel correspondant (Arc Info, par exemple) nécessite un apprentissage plus long. C'est donc le modèle de données qui va rendre possibles ou non certaines opérations. Nous avons détaillé (chap. 3) les modèles de « bases de données » ; il reste à détailler les modèles de données géométriques.

### 6.3.2 Modélisation géométrique des objets cartographiques

Dans les logiciels de type A, les plus élémentaires et les moins puissants, la géométrie des objets à cartographier est la même qu'en dessin assisté par ordinateur et en cartographie thématique. Les trois mêmes primitives graphiques y sont distinguées (cf chap. 5) : *point*, *polyligne*, *polygone*.

L'extrait ci-dessous montre le codage, dans le plus répandu des logiciels, d'un polygone (appelé ici « region »)

```

Region 1
9
864867.879 325466.061
864868.690 325473.889
864891.272 325475.310
864892.408 325463.388
864880.391 325462.668
864873.229 325462.202
864871.688 325462.112
864870.732 325462.978
864867.879 325466.061
Pen (1,2,0)
Brush (2,16777215,16777215)
Center 864886.678 325465.711

```

Ce polygone joint 9 points (le premier et le dernier étant le même). Pour chaque point sont indiquées ses coordonnées en X et en Y (ici en projection Lambert 3). Un segment va d'un point au suivant, jusqu'au dernier qui referme le polygone. Les instructions « Pen »(crayon) et « Brush » (brosse) , avec leurs paramètres, soulignent bien que l'on est dans une logique dessin. De même pour la spécification des coordonnées du centre (« Center ») du polygone, qui facilite son remplissage par une couleur ou un figuré.

Cette structuration élémentaire ( « **structure spaghetti** ») présente des inconvénients :

- Dans le cas du découpage d'un territoire en unités surfaciques jointives, tous les segments intérieurs contigus sont décrits deux fois, avec tous les risques d'erreur de saisie et de taille de fichiers que comporte cette façon de faire,
- Le modèle est trop près des opérations de saisie (pas assez abstrait) pour rendre compte de certaines configurations territoriales : réseau orienté (de transport ou hydrographique), polygones avec trous (lac dans une commune), objet complexe composé d'objets élémentaires différents, etc.
- Cette modélisation géométrique peut suffire dans les cas simples, sur de petites quantités de données mais devient inadaptée dès que les besoins de l'utilisateur deviennent plus complexes. Il faut, alors, coder séparément la géométrie et la topologie des objets spatiaux.

### 6.3.3 Modélisation de la géométrie et de la topologie

La topologie est une géométrie qualitative qui décrit la position relative des objets spatiaux (inclusion, chevauchement, contiguïté) et permet leur déformation pourvu que soient conservées la continuité et les positions relatives (on l'a parfois qualifiée de géométrie des objets déformables). La description topologique dans les S.I.G. est celle d'un graphe planaire (en 2 dimensions, pas en 3). Ce graphe peut être orienté (muni d'un sens de parcours) ou non.

La figure 6.2 nous permet de voir une façon de coder la topologie. Les primitives topologiques sont :

- **Point** (intermédiaire ou isolé)
- **Nœud** (point auquel aboutissent au moins 3 arcs, sauf pour une face incluse)
- **Arc** (polyligne allant d'un nœud à un autre, avec ou sans point intermédiaire)
- **Face** (polygone délimité par un ou plusieurs arcs).

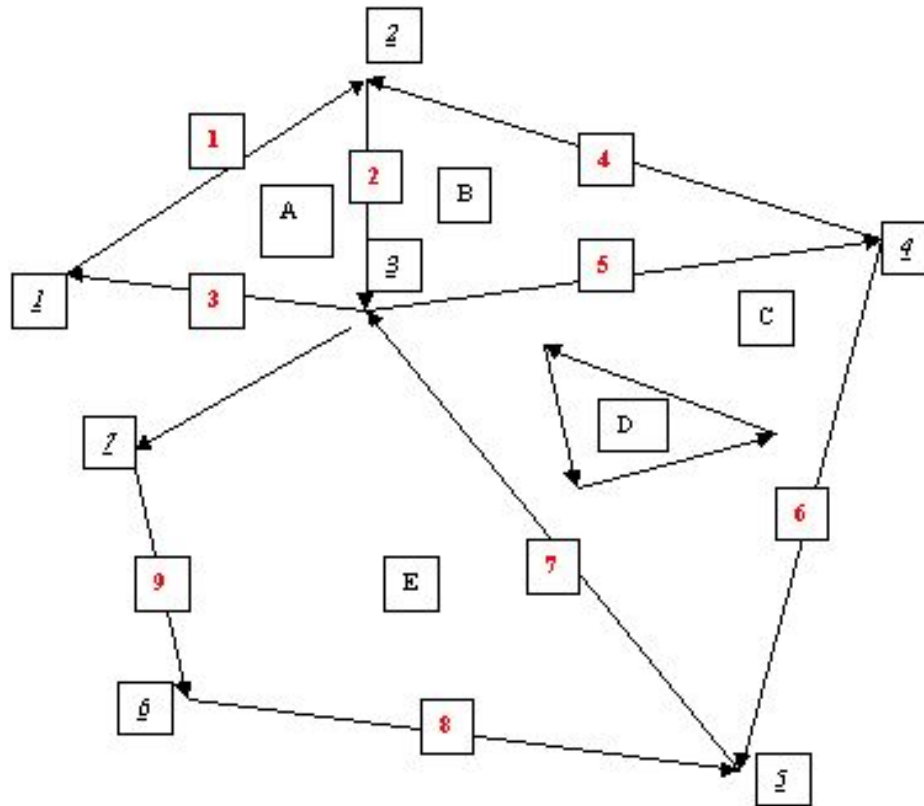


Figure 6.2 : schéma pour description topologique

Décrire la figure 6.2 peut prendre la forme suivante :

- Points (0, 1 ou 2 segments y aboutissent) : n° 1, 6, 7.
- Nœuds (plus de 2 arcs, sauf inclusion, y aboutissent) : n° 2, 3, 4, 5.
- Arcs (début en un nœud, finit en un nœud) ; par exemple, l'arc 9 débute au nœud 3, se poursuit par les points 7 et 6, finit au nœud 5 ; il a à gauche la face A et à droite l'extérieur,
- Faces (nœuds initial et final identiques) ; la face B a, ainsi, pour frontière les arcs 2, 4 et 5.

Si le graphe est orienté (comme sur la figure 6.2), des faces peuvent elles mêmes contenir des faces (le sens de parcours des arcs est alors inversé) : on peut représenter des polygones avec trou (inclus ou exclu), des réseaux avec sens uniques, ....

Que le graphe soit orienté ou non, la **description topologique rend compte de la structure qualitative du dessin**. Il reste à fournir au système les coordonnées XY précises des points qui composent les primitives.

L'avantage de séparer descriptions géométrique et topologique est multiple :

- minimisation des erreurs de saisie, corrections plus faciles, non redondance des données géométriques,
- topologie de toutes façons obligatoire à calculer pour toutes les requêtes portant sur des contiguités, des inclusions ou des chevauchements d'objets spatiaux. Si celle-ci est pré-calculée et stockée dans les données, les réponses sont très rapides.

#### 6.3.4 Opérations possibles

Un certain nombre d'entre elles sont communes à tous les S.I.G. vectoriels.

Il s'agit :

- d'introduire et de mettre à jour les données, géométriques et attributaires,
- de sélectionner des objets et / ou des valeurs d'attributs,
- de combiner celles-ci à l'aide d'opérateurs (cf chap. 3),
- de les cartographier, de zoomer et dé-zoomer,
- de pratiquer des mesures géométriques (longueurs, périmètres, surfaces...en projection 2D), de définir des zones tampon (« buffers » en anglais) autour des unités spatiales et d'agir sur leur contenu,
- etc.

Tous les S.I.G. vectoriels, même les plus simples, permettent donc une forte **interactivité** entre l'utilisateur et ses données spatialisées, ce qui a pu faire dire que les S.I.G. (vectoriels) étaient des « jeux vidéo géographiques ».

Cependant, certaines fonctions se comportent différemment en fonction du modèle de données propre à chaque logiciel. Il en va ainsi de la **superposition de couches** (« overlay » en anglais), si utile pour combiner des données multi-sources.

- Les logiciels de type **A** (un seul fichier B.D., structure géométrique spaghetti) sont davantage des cartographes interactifs que de vrais S.I.G.. Leurs possibilités d'overlay sont extrêmement limitées : les unités spatiales y sont toutes de même nature. Communes ou départements par exemple formant un fond de carte unique, on peut seulement combiner (arithmétiquement, logiquement) les valeurs d'attributs relatifs à ce fond de carte unique. La combinatoire est donc restreinte.
- Les logiciels de type **B**, les plus fréquents, admettent des objets spatiaux de types différents, donc une B.D. multi-fichiers reliés par des opérateurs relationnels (plus ou moins complets) et les couches graphiques afférentes aux différents types d'objet (parcelles cadastrales, îlots de recensement, rues, ...). L'organisation relationnelle fait gagner, par **jointure attributaire**, des possibilités d'enrichissement en données attributaires (création de tableaux joignant des données provenant de deux tableaux liés). La superposition de couches devient possible mais elle est seulement visuelle, sans stockage de son résultat pour utilisation ultérieure.

- Les logiciels de type **C** sont spécialisés en parcours de graphe (par exemple pour des estimations d'accessibilité sur un réseau). Les logiciels comme « Autoroute Express » de Microsoft appartiennent à cette catégorie.
- Les logiciels de type **D** offrent les possibilités maximales d'overlay de types d'objets différents, géométriquement et par leurs descripteurs. Géométrie résultant de la superposition et tableau produit par l'union des deux (**jointure spatiale et attributaire**) sont stockés, utilisables alors pour requêtes ultérieures. La solution d'un problème complexe revient alors à prévoir un enchaînement de requêtes (un algorithme). Les possibilités combinatoires sont grandes, les réponses aux requêtes topologiques rapides (puisque la topologie est stockée dans les données) : l'interactivité avec l'information géographique y gagne beaucoup et garantit de réelles possibilités d'analyse spatiale exploratoire.

Pour illustrer le propos, considérons la partie cadastrale d'un S.I.G. communal, contenant la géométrie des parcelles, les données relatives à chaque parcelle et à chaque propriétaire. Imaginons maintenant un projet de contournement routier (une autre couche géométrique, faite de tronçons contigus et des attributs les décrivant). On se pose la question d'identifier les **portions** de parcelles et les propriétaires touchés par le projet. Dans une modélisation de type B, la superposition plan cadastral - projet routier ne sera que visuelle et les **parties** de parcelles à exproprier ne seront pas stockées dans le S.I.G. alors qu'elles le seraient avec une modélisation de type D, ce qui permettrait ensuite d'en connaître les propriétaires, la surface touchée par le projet et donc le prix à proposer aux ayant droit. La puissance **combinatoire** est tout autre !

Fondamentalement, un S.I.G. vectoriel est donc un système de gestion de bases de données étendu à des données spatiales. Il est adapté à des types d'objets **préalablement identifiés** (spatialement et sémantiquement), correspondant le plus souvent à des éléments du paysage visible, à des zonages ayant valeur légale ou réglementaire, à des réseaux techniques,...

## 6.4. S.I.G. RASTEUR

Les S.I.G. rasteur présents sur le marché ont une structure de données très simple: une image par variable.

### 6.4.1. **Modèle de données simple**

Chaque image est un tableau où :

- numéros de ligne et de colonne indiquent la position des pixels, tous de mêmes forme (carrée) et surface,
- chaque case porte une valeur, très généralement numérique.

Ce modèle de données permet d'approximer des objets surfaciques et, plus difficilement, des éléments linéaires et ponctuels. La précision de la

représentation dépend de la « résolution » de l'image, c'est à dire de la taille du pixel au sol. L'améliorer coûte cher, puisque l'espace de stockage augmente très vite :

- au carré pour pixel de côté divisé par 2, au cube pour sa division par 3, etc.)
- et que chaque requête produit une nouvelle image, stockée sur disque.

Les opérations pertinentes sur les données raster dépendent de leur sémantique. Les valeurs d'un tableau raster peuvent, en effet, être de quatre natures. Elles peuvent :

- traduire une variable spatialement et mathématiquement continue (altitude, pente, pression atmosphérique, temps d'accès,...),
- représenter des états inconnus, à interpréter (cas des images satellitaires, cf chap. 7),
- représenter les états connus d'une nomenclature préalable (exemple des images d'utilisation du sol CORINE LAND COVER),
- être issues du scan de cartes ou plans préexistants (exemple : planches cadastrales scannées).

Vu la simplicité du modèle de données, la puissance d'un logiciel raster (combinaison entre valeurs d'une même couche ou de couches différentes) dépend alors uniquement des opérateurs mis à la disposition de l'utilisateur. La richesse est, ici, grande puisque le codage numérique de l'information ouvre de grandes possibilités. Selon la nature de la variable, l'utilisateur pourra effectuer des opérations arithmétiques, de comparaison, logiques (cf chap. 3).

La combinaison d'opérations élémentaires aboutit à des **fonctionnalités de second niveau** liées :

- à la contiguïté, à la connexité, à la distance, employées en analyse spatiale (par exemple, analyse de réseaux, de diffusion spatiale, d'intervisibilité paysagère, de cheminement optimal, ...),
- au lissage des valeurs d'une image,
- à l'interpolation de valeurs à partir des points de valeur connue,
- à la prise en compte, globalement, des contraintes de l'espace différencié.

D.Tomlin, sous le nom d' **«algèbre de cartes** », a rationalisé la combinaison de couches raster géo-référencées dans le même système de coordonnées. Il a défini quatre types de « fonctions cartographiques » :

- locales, travaillant sur chaque cellule d'une seule couche, sans référence à son voisinage (exemple : division des valeurs par une constante),
- focales, travaillant sur les valeurs d'une seule couche, avec référence au voisinage (exemple : lissage des valeurs de chaque case par moyenne pondérée avec celles des voisines),
- zonales, travaillant sur les valeurs de zones, contigües ou non, définies dans une autre couche (exemple : définir la densité de population dans



- chaque commune à partir d'une couche population et d'une couche fournissant les limites communales),
- *globales*, la valeur de chaque cellule résulte d'une fonction appliquée à toute la grille (exemples : calculs d'accessibilité, de cheminement, d'intervisibilité,...).

Ces fonctions, sous d'autres dénominations, sont aujourd'hui présentes (en tout ou en partie) dans la plupart des S.I.G. raster. Ceux ci se caractérisent donc par la **simplicité de leur modèle de données et la puissance de leurs opérateurs** de second niveau qui permettent analyse spatiale, modélisation et simulation.

Est ce à dire qu'ils ne sont pas améliorables et, si oui, quelles transformations de leur modèle de données cela suppose t'il ?

#### 6.4.2. Améliorations possibles du modèle de données raster

##### A. Quadtree et clés de Peano

Un progrès certain serait de pouvoir adapter la taille des cases d'une grille à la densité d'information qu'elles doivent contenir (par exemple sur une carte raster par départements, créer des pixels plus petits pour l'Ile de France). C'est possible à l'aide des « arbres quaternaires » (*quadtrees* en anglais).

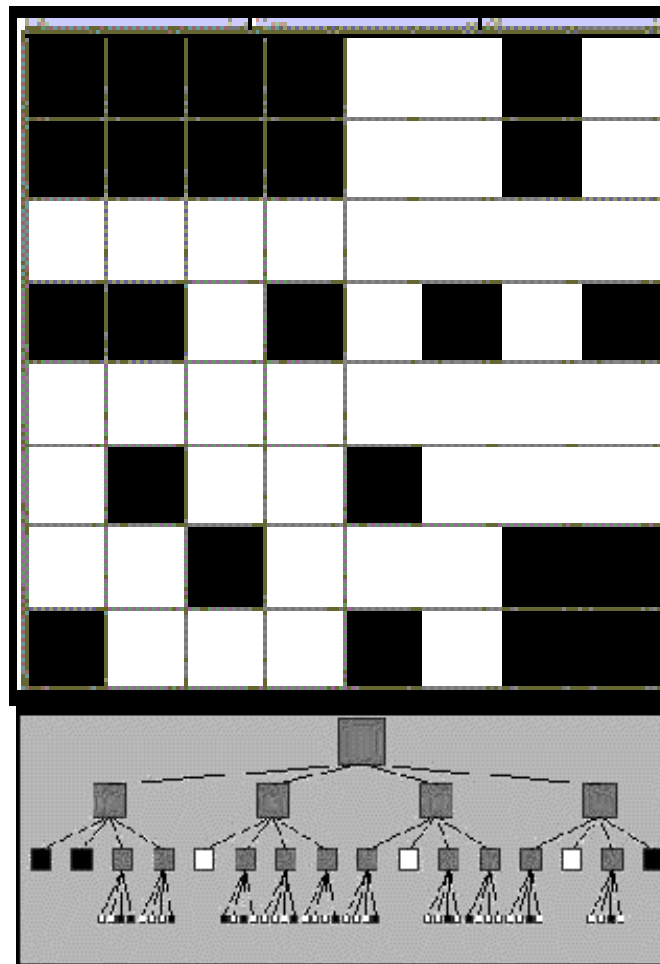


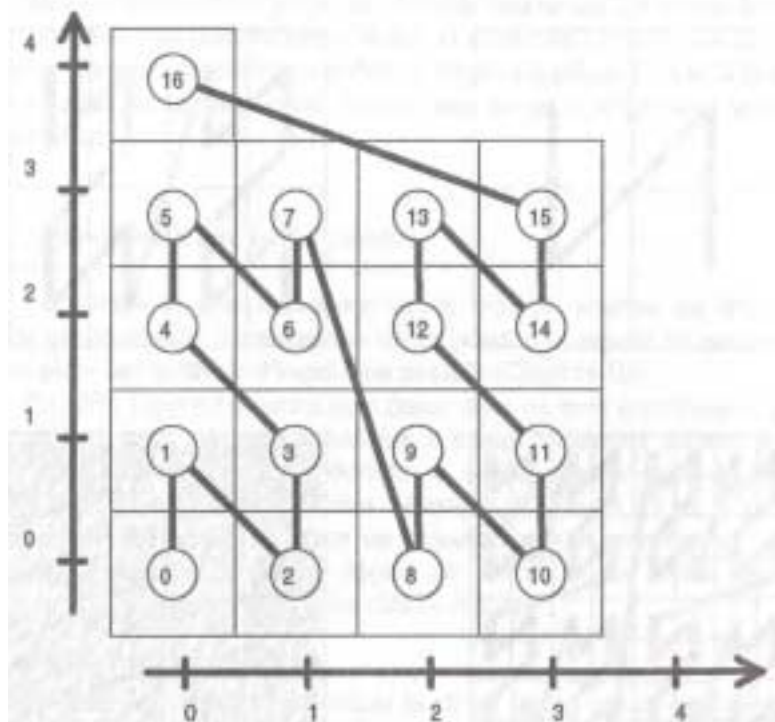
Fig. 6.3 : QUADTREE CORRESPONDANT AU DAMIER CI-DESSUS

Pour les expliquer de façon simple, considérons un cas d'école, présenté sur la figure 6.3, celui d'un damier carré de 8 lignes et 8 colonnes où une case noire signale la présence d'une valeur et une blanche son absence.

Si l'on partage le damier en quatre carrés égaux, on définit un *premier niveau* sur l'arborescence : les « branches » correspondantes y sont figurées en grisé car aucun des quatre sous damiers n'est de couleur homogène (tout blanc ou tout noir). Pour chacun des quatre, il faut donc itérer le même procédé de coupure en quatre, ce qui définit un deuxième niveau sur l'arborescence où trois des seize sous damiers (de quatre cases chacun) sont de couleur noire homogène (les deux du coin haut gauche, celui du coin bas droite) et trois des seize de couleur blanche homogène. Il n'est plus besoin pour ceux ci de continuer à les découper, c'est par contre nécessaire pour les dix autres. Au final, en itérant le procédé, l'arborescence a quatre niveaux (depuis le damier entier jusqu'aux cases individuelles) mais toutes ses branches n'ont pas été déployées, dans la mesure où des groupes de cases contigües étaient de même couleur. Le procédé a donc bien permis de définir des unités de taille différente en fonction de leur contenu.

Il reste maintenant à identifier chacune des feuilles de l'arbre que nous avons construit.

- Une première solution consiste à adopter une numérotation hiérarchique : 1 pour le damier entier, 1.1 pour le carré haut gauche de 16 cases, 1.1.1 pour le carré de 4 cases haut gauche, 1.1.2 pour son voisin de droite, etc. Cette numérotation arborescente est tout à fait semblable à celle de la table des matières d'un livre (partie, sous parties, etc.).
- Une seconde solution consiste à utiliser une *courbe fractale*.



*Fig. 6.4 : Courbe en N de Peano, numérotation et XY correspondants (source : R.Laurini et F.Milleret-Raffort)*

Une telle courbe (objet de dimension 1) est capable de passer par tous les points d'un plan (objet de dimension 2) ou d'un volume (objet de dimension 3) et donc d'identifier par un **numéro unique** toute unité spatiale, de dimension 0, 1, 2, ..., aussi grande ou petite soit elle. L'une de ces courbes fractales est la courbe « en N » de Peano. La figure 6.4 représente une telle courbe en N pour les cases du coin inférieur gauche du damier de la figure 6.3.

Quadtree (à unités spatiales de tailles différentes) et ordre de Peano peuvent être combinés.

Quels en sont les avantages ?

- l'organisation « quadtree » permet d'ajuster la taille des unités spatiales à la densité locale de contenu de l'image,
- la numérotation hiérarchique ou celle de Peano remplace, pour identifier une case, deux numéros (de ligne et de colonne) par un seul. Cela permet de trier les valeurs par numéro croissant de Peano et d'accéder à chacune de façon directe (et non séquentielle) : c'est bien plus efficace en termes de rapidité d'accès au contenu de chaque case ; cela fait donc progresser l'interactivité entre l'utilisateur et les informations de ses images.
- La combinaison Quadtree + identifiants hiérarchiques ou de Peano fait gagner sur les deux tableaux.

Malheureusement, ces améliorations n'existent pas, actuellement, dans les logiciels rasteur du commerce.

## B. Autre amélioration possible

L'amélioration que nous suggérons ici prolonge la précédente.

Elle consiste d'abord à coder, dans chacune des couches rasteur, non plus une valeur mais l'**appartenance** à une zone (d'un découpage administratif par exemple), à un tronçon de réseau, à un type d'objet spatial. Chaque case de chaque couche comporte ainsi un code d'inclusion dans une unité spatiale. Chacune de ces couches peut être ou non organisée selon le modèle Quadtree + Peano.

Les tableaux relatifs aux attributs décrivant les différents types d'objets spatiaux peuvent figurer dans une base de données relationnelle, ce qui permet des requêtes sémantiques combinatoires à l'aide d'un langage de type SQL (cf chap. 3). A chaque image peuvent correspondre plusieurs tableaux de sémantiques différentes.

Dans chacun, peuvent figurer deux types d'identificateurs : un numéro de Peano et un identificateur « composite », par exemple de la forme suivante pour repérer une commune française:

**||ZAT|RP|DPT|CT|COM||ZON1|...|ZONk||**

- où la partie 1, hiérarchique, décrit l'inclusion de la commune dans une Zone d'Aménagement du Territoire(ZAT, regroupant plusieurs régions), une

Région de Programme(RP), un département (DP), un canton (CT) et indique le numéro de la commune (COM) dans le département ou le canton, - où la partie 2, non hiérarchique, décrit l'appartenance de la commune à divers zonages (bassin d'emploi, parc naturel régional, communauté de communes, aire urbaine, etc.).

L'avantage de ce type d'identificateur est de permettre de *gérer (voire de générer par agrégation numérique) des données spatiales multi-échelles*.

Comme le même type de codage hiérarchique peut s'appliquer aux valeurs d'attribut, une gestion multi-échelles des données peut fonctionner des unités spatiales vers les valeurs d'attribut et / ou des valeurs d'attribut vers les unités spatiales.

On peut, en outre, imaginer des systèmes « mixtes » où, aux mêmes tableaux de la base de données relationnelle, correspondraient une carte vectorielle et une image rasteur, ce qui augmenterait le spectre des opérations possibles.

### **6.5 RASTEUR OU VECTEUR ? RASTEUR ET VECTEUR ?**

Le tableau 6.3 récapitule les avantages et inconvénients de chacun des deux modes.

<b><u>RASTEUR</u></b>	<b><u>VECTEUR</u></b>
<b><u>AVANTAGES</u></b>	<b><u>AVANTAGES</u></b>
Modèle de données simple	Précision géométrique, cartographie
Puissance des opérations numériques	Compacité des données
Technologie classique	Forte interactivité utilisateur - données
<b><u>INCONVENIENTS</u></b>	<b><u>INCONVENIENTS</u></b>
Gros volume de données	Structure de données plus complexe
Qualité graphique irrégulière	Faiblesse sur opérations numériques
Difficile codage (points et polygones)	Coût important des données
<b><u>CONVIENT BIEN A</u></b>	<b><u>CONVIENT BIEN A</u></b>
Données relatives à des zones	Cartographie, précision géométrique
Données variant continuellement	Applications réclamant de l'interactivité
Applications « recherche »	Gestion de territoires

Tableau 6.3 : Avantages et inconvénients des modes rasteur et vecteur

Les deux modes sont complémentaires et tous les utilisateurs de S.I.G. voudraient, bien évidemment, pouvoir bénéficier, dans le même logiciel, des avantages de l'un et l'autre mode.

Cela n'est encore réalisé que bien imparfaitement et à des niveaux différents.

Comme les données sont l'élément le plus précieux, on devrait pouvoir les transférer automatiquement et sans difficulté. Le transfert dessin vectoriel

→ image raster est couramment assuré (mais avec risque de perte d'information). Par contre, le transfert inverse (vectorisation d'images) est plus problématique et conduit fréquemment à des erreurs ; par exemple, vectoriser une carte topographique scannée produit nombre d'erreurs liées à l'existence de l'habillage cartographique et de la toponymie. La plupart des logiciels S.I.G. peuvent aujourd'hui *superposer visuellement* carte et image et fournir des fonctions travaillant alternativement, ou sur du raster ou sur du vectoriel. Mais ils n'offrent pas d'opérateurs actifs indistinctement sur l'un et l'autre mode. Les combiner se pratique de façon visuelle et non de façon fonctionnelle. Les logiciels du commerce sont donc principalement vectoriels ou raster.

## **6.6.PRINCIPALES SOURCES FRANÇAISES DE DONNEES GEOGRAPHIQUES**

Nous n'évoquerons ici que pour mémoire les Bases de Données Géographiques (B.D.G.) à couverture mondiale (Digital Chart of the World) ou européenne (Eurostat, Gisco, Corine Land Cover, etc.).

### **6.6.1 Données publiques à couverture nationale**

Pour une description exhaustive des données géographiques à couverture nationale, on peut consulter le site web du C.N.I.G. (<http://www.cnig.fr>). Nous ne présentons ici qu'un bref aperçu.

#### **A. B.D.G. de l'Institut Géographique National**

Les deux principales sont des B.D. vectorielles : B.D. « Carto » et B.D. « Topo » (pour plus d'information : <http://www.ign.fr>).

- La B.D. « Carto » est, en contenu thématique et précision géométrique, comparable à la carte 1/50000. Elle convient donc à des institutions à assez grande compétence spatiale (collectivités territoriales et services techniques de l'état à compétence départementale ou régionale).
- La B.D. « Topo » est, en contenu thématique et précision géométrique, comparable à la carte au 1/25000 et convient donc à des usages localement plus détaillés (mais son coût est important, ce qui en limite l'usage).
- L'I.G.N. diffuse également d'autres produits : B.D. plus spécialisées comme la B.D. « routes » ou la B.D. « alti » (Modèle Numérique d'Altitudes, en format raster ou vectoriel), photos aériennes, scan des cartes 1/50000 ou 1/25000, etc.

#### **B. Autres B.D. nationales**

On peut citer à ce titre :

- La B.D. cadastrale de la Direction Générale des Impôts (spatialement très incomplète),
- Les B.D. environnementales (Corine Land Cover, B.D. des agences de bassin hydrographique, B.D. Météo France, etc.)

- Les B.D. localisées de l'INSEE (B.D. îlots, B.D. IRIS2000, Répertoire Informatisé des Logements, etc.)
- Les B.D. routières (comme celle de Michelin)
- Les B.D. d'images satellitales (cf <http://www.cnes.fr>)

De fait, toutes les activités administratives ou techniques de l'état central peuvent donner lieu à B.D.G. spécialisées. Pour être compatible spatialement les unes avec les autres, elles doivent s'appuyer sur les mêmes conventions géodésiques et de projection plane, définies par l'I.G.N.

### **6.6.2 Données privées à couverture locale**

Il existe d'autres moyens d'acquérir des données localisées, portant alors, généralement, sur des surfaces plus limitées :

- Acquisition auprès de compagnies spécialisées pratiquant la métrologie de terrain (géomètres experts) ou toutes les opérations aériennes (plan de vol, prises de vue, photogrammétrie).
- Il est probable également que ces activités, notamment la photogrammétrie, se pratiqueront dans d'autres institutions que l'I.G.N. et dans les services techniques des collectivités territoriales. Il faudra alors veiller au recalage des données ainsi produites par rapport au référentiel géographique national.

### **Références citées**

- Denègre J., Salgé F., *Les systèmes d'information géographique* (Presses Universitaires de France, coll. Que sais je n°3122, 1996)
- Tomlin D., *Geographic Information Systems and Cartographic Modeling*, (Prentice-Hall, 1990)
- Laurini R., Milleret-Raffort F., 1993, *Les bases de données en géomatique* (Editions Hermes)
- Charre J., Dumolard P., 1987, *Initiation aux pratiques informatiques en géographie, le logiciel INFOGEO* (Editions Masson, 1989)
- Dumolard P., 1997, *Multiscale statistical spatial databases* in C.Laurent & I.Bowler ed. (*CAP and the regions*, INRA éditions)
- Consultez le site du cours de MM. Brousseau et Thériault : <http://www.ggr.ulaval.ca/Cours/SIG/ggr-18708.htm>

# **CHAPITRE 7 : TRAITEMENT D'IMAGES GEOGRAPHIQUES**

*Pierre Dumolard*

Comme le texte, le schéma ou la carte, l'image joue un rôle important dans les travaux géographiques, non seulement pour la communication d'observations et d'idées mais aussi comme source d'analyse et comme résultat de synthèse.

Les « images » géographiques sont fort diverses : carte, image satellitale, photo (aérienne, terrestre), film, peinture, croquis, schéma, etc.

Dans ce chapitre, nous n'entendons pas détailler les pratiques classiques de l'image géographique, par exemple gloser sur le traditionnel commentaire de photos (oblique ou aérienne) ou sur le non moins habituel « croquis régional » mais insister sur deux techniques impliquant nécessairement le recours à l'informatique:

- l'analyse d'images de télédétection,
- la construction d'images de synthèse.

## **7.1 ANALYSE D'IMAGES TELEDETECTEES**

La télédétection est la détection, depuis un « capteur » éloigné, d'éléments présents à la surface de la terre ou dans la basse atmosphère. Le capteur peut être une matrice CCD, un appareil photo, un radar, ... et le vecteur un satellite, un avion, un ballon sonde, etc. Nous concentrons ici l'attention sur l'image satellitale, dont il faut bien comprendre qu'elle est plus et autre chose qu'une photo (à cause de la nature du signal capté et transmis à la terre).

### **7.1.1. Nature de l'image satellitale**

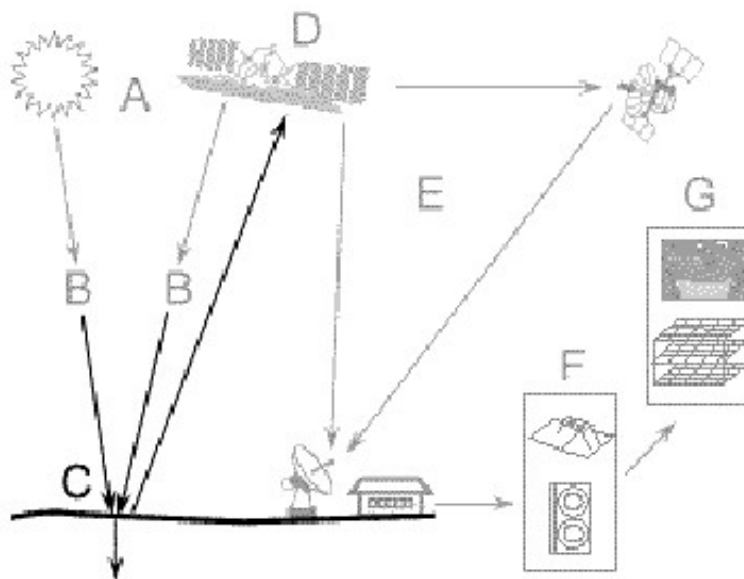


Figure 7.1 : les composants d'une prise d'image satellite (source : CCRS)



Les capteurs équipant les satellites captent l'**énergie** réfléchi ou émise par la surface terrestre, ce qui implique (cf Fig. 7.1):

(A), une source d'énergie : le soleil illuminant la terre, un rayon laser, un signal radar, ...

(B), l'interaction de cette énergie avec l'atmosphère traversée (du soleil vers la terre puis de la terre vers le satellite) qui perturbe le signal,

(C), l'interaction de cette énergie modifiée avec une portion plus ou moins rugueuse de surface terrestre

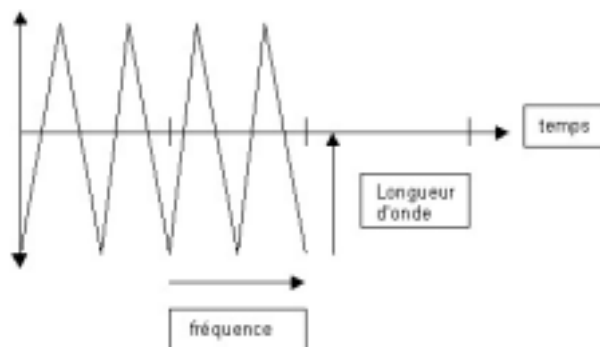
(D), la capture du signal par les appareils embarqués et (E), sa transmission à la terre, des redressements (pré-traitements) et sa transformation en fichier utilisable (F et G)

## A. Un peu de physique élémentaire

L'énergie captée est une **radiation électromagnétique** (composée d'un champ électrique et d'un champ magnétique) voyageant de façon ondulatoire à la vitesse de la lumière.

Les deux paramètres fondamentaux décrivant les mouvements ondulatoires sont (cf Fig. 7.2) :

□, la longueur d'onde (mesurant la « hauteur » d'un cycle et exprimée en micromètres),



□, la fréquence (exprimant, en Herz, le nombre de cycles par seconde).

Figure 7.2 : Longueur d'onde et fréquence d'un signal

La vitesse (constante) de la lumière ( $c = 300\,000\text{ km/s}$ ) est :  $c = \lambda * \nu$

Puisque  $c$  est constant, plus la longueur d'onde est petite plus la fréquence est grande (et vice versa). Ayant admis ces rudiments de physique, on peut maintenant examiner les longueurs d'onde des différents types de radiation électromagnétique (cf Fig. 7.3).

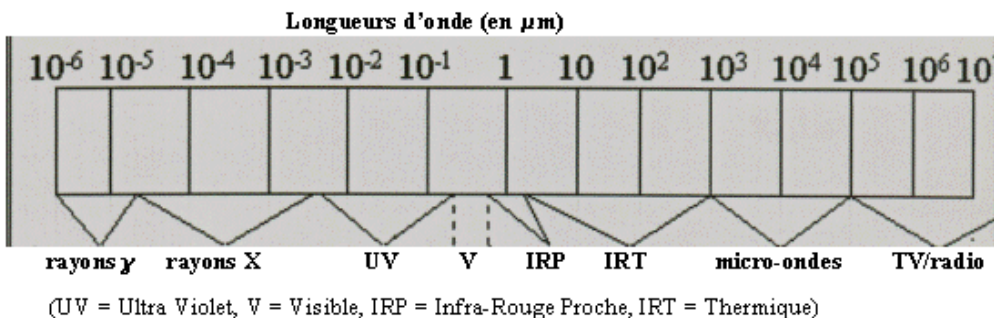


Figure 7.3 : types de radiations électromagnétiques selon leur longueur d'onde

Le spectre de la lumière visible (celui de l'oeil humain et de la photo) est limité aux longueurs d'onde comprises entre 0.4 et 0.7 micromètres. Il est lui même divisé en 3 parties, correspondant aux 3 couleurs fondamentales :

- le **Bleu** (longueur d'onde 0.4 à 0.5  $\mu\text{m}$ ),
- le **Vert** (longueur d'onde 0.5 à 0.6  $\mu\text{m}$ ),
- le **Rouge** (longueur d'onde 0.6 à 0.7  $\mu\text{m}$ ).

L'addition d'intensités données (plus ou moins claires ou foncées) dans chacune de ces 3 couleurs de base permet de fabriquer, sur les ordinateurs actuels, plus de 16 millions de couleurs (alors que l'oeil humain n'en différencie qu'un peu plus de 2 !).

Les capteurs des satellites ont un spectre plus large que le visible, débordant par exemple sur l'infra-rouge, ce qui explique qu'une image satellite n'est pas une photo, qu'elle est visualisée en fausses couleurs et qu'elle n'est pas d'interprétation immédiate.

## B. Capteurs embarqués

La plupart des satellites disposent de capteurs composés de barrettes ou de matrices CCD qui enregistrent électroniquement l'intensité du signal électromagnétique réfléchi par une portion de la surface terrestre.

Pour un ensemble de longueurs d'onde, chaque type d'objet a une **réponse spectrale**: réflectance plus ou moins forte dans différentes longueurs d'onde, ce qui se traduit par une courbe particulière (**Fig 7.4**) sur un graphique cartésien (illumination en Y, longueur d'onde en X). Par exemple, l'eau et la végétation se différencient par leur réponse sur un intervalle assez grand de longueurs d'onde (visible et infrarouge). Différencier des types de roches nécessite au contraire un intervalle de longueurs d'onde beaucoup plus petit.

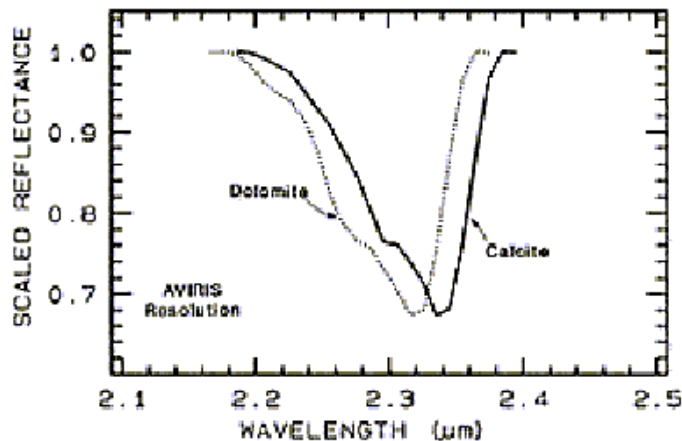


Figure 7.4 : Courbes spectrales de deux roches différentes, réflectance en ordonnée, longueurs d'onde en abscisse (Source : NASA)

On appelle **résolution spectrale** d'un capteur sa capacité à utiliser de petites plages de longueurs d'onde. Par exemple, une pellicule photo noir et blanc a une faible résolution spectrale car elle enregistre les longueurs d'onde dans toute la plage du visible, une pellicule couleurs en a une plus forte puisqu'elle enregistre, sur la même plage du visible, les longueurs d'onde dans 3 fenêtres, celles du bleu, du vert et du rouge : son pouvoir différenciateur est plus grand car l'ensemble des longueurs d'onde utilisées est segmenté en plus de fenêtres (ou bandes spectrales).

Les capteurs embarqués par les satellites sont donc:

- ou bien **panchromatiques**, ne différenciant pas de fenêtres dans l'ensemble des longueurs d'onde qu'ils couvrent
- ou bien **multispectraux** s'ils y différencient plusieurs bandes.

Par exemple, Landsat Thematic Mapper différencie 7 bandes de longueurs d'onde dans le visible et le proche infrarouge, SPOT-3 n'en distingue que 3 et a donc une résolution spectrale inférieure.

On appelle **résolution radiométrique** la sensibilité d'un capteur à enregistrer finement l'intensité de l'énergie électromagnétique dans les différentes fenêtres de longueurs d'onde. Une résolution radiométrique minimale serait celle d'un capteur qui n'enregistrerait dans chaque bande de longueurs d'onde que 2 intensités (forte codée 1 / faible codée 0). Les capteurs des satellites tels Landsat TM ou SPOT ont une résolution radiométrique nettement supérieure puisqu'ils distinguent 256 niveaux d'intensité dans chacune de leurs bandes de longueurs d'onde (7 pour Landsat, 3 pour SPOT).

Le « pouvoir » différenciateur d'un capteur résulte de ses résolutions spectrale et radiométrique.

### C. Itinéraire d'un satellite

Les satellites sont placés en **orbite** autour de la terre. Cette orbite est calculée en fonction des résolutions spectrale et radiométrique des capteurs embarqués, elles mêmes fonction de l'objectif assigné (météo, observation de la surface terrestre, GPS, ...). L'orbite d'un satellite se caractérise par son « altitude » (sa distance à la terre) et son itinéraire par rapport à la terre.

Les satellites en **orbite géostationnaire** sont situés à haute altitude et observent toujours la même portion de surface terrestre car leur vitesse de rotation est égale à celle de la terre (exemple : satellites météo). Couvrir toute la surface terrestre implique donc plusieurs dont les orbites sont « harmonisées ».

Les satellites en **orbite polaire** font le tour de la terre le long de l'axe des pôles, du nord au sud puis du sud au nord. Ils sont en général **héliosynchrones** (synchrones avec le soleil), passant ainsi toujours à la même heure solaire au dessus des mêmes régions. Les capteurs embarqués n'enregistrent donc l'énergie solaire réfléchie par la terre que sur une partie de leur trajet (le jour et pas la nuit). Comme la terre est en rotation d'ouest en est, le satellite « observe » lors de ses trajets successifs des bandes successives de surface terrestre.

A un instant donné, le satellite est perpendiculaire à un point de la surface terrestre qu'on appelle **nadir**. Son cycle de passage est le temps nécessaire pour repasser au dessus du même nadir, ce qui définit la **résolution temporelle** du satellite, temps nécessaire pour « observer » toute la surface terrestre et repasser sur le même point. Cette résolution temporelle est un paramètre important, notamment pour la surveillance d'incendies de forêt, d'éruptions volcaniques, d'inondations, d'échouage de navires pétroliers, etc. L'altitude d'un satellite définit, conjointement à la nature des capteurs, la dimension de la portion de terre observée lors d'un passage orbital (60 \* 60 km pour une scène SPOT) et la finesse du grain de l'image produite qu'on appelle **résolution spatiale**. Par exemple, les capteurs de SPOT-3 ont un pixel de 10 \* 10 m en mode panchromatique et 20 \* 20 m en mode multispectral (dans chacune de ses 3 bandes de longueurs d'onde), LANDSAT TM a un pixel de 15 \* 15 m dans chacune de ses 7 bandes. La résolution spatiale d'un capteur définit le « grain » de l'image satellitale.

En résumé, quatre paramètres qualifient un ensemble d'images satellitales :

- **résolution spectrale** (nombre de bandes dans un ensemble de longueurs d'onde),
- **résolution radiométrique** (finesse de la mesure de l'énergie dans chaque bande),
- **résolution temporelle** (temps de retour au dessus des mêmes régions),
- **résolution spatiale** (surface au sol d'un pixel de l'image).

### 7.1.2 Principaux satellites civils d'observation

Le premier satellite civil a été TIROS-1, satellite d'observation météorologique lancé en 1960 par les USA. Le premier satellite d'observation de la surface terrestre, LANDSAT-1, a été lancé par la NASA en 1972 mais la commercialisation de ses images ne date que de 1985. Le satellite français SPOT-1 a été lancé en 1986, le satellite indien IRS-1 en 1995. C'est donc seulement à partir de la fin des années 80 / début des années 90 que l'image satellitale est devenue réellement disponible comme source de données pour les géographes.

#### A. Satellites météorologiques

L'observation météorologique a été une des premières applications civiles des satellites : le premier satellite météorologique, TIROS 1, a été lancé en 1960. Il existe, aujourd'hui deux réseaux principaux de satellites météorologiques, de faible résolution spatiale mais de forte résolution temporelle, permettant un suivi quasiment en temps réel.

- GOES est un ensemble de satellites géostationnaires faisant partie du réseau mondial d'observation et de prévision météorologiques et océanographiques. Dotés de capteurs dans le visible et le proche infrarouge, ils permettent aussi l'observation nocturne. Leurs principales utilisations sont la mesure des températures, vents, humidité, couvert nuageux, la détection d'incendies, d'orages, d'éruptions volcaniques, de pollutions.
- NOAA AVHRR complète le dispositif. Composé de satellites en orbites polaires héliosynchrones doté d'une résolution temporelle de 6 heures maximum et de capteurs dans l'infrarouge proche et thermique, il permet, outre des observations météorologiques complémentaires, le suivi de la végétation, de l'agriculture, des étendues d'eau.

Dans ces deux systèmes, la taille du pixel des images varie, selon les capteurs, de 1 à 4 km mais les images couvrent une large portion de la surface terrestre.

#### B. Satellites d'observation de la terre

Leurs images sont, pour leur haute résolution spatiale, les plus utilisées par les géographes. Historiquement, Landsat 1, lancé en 1972, a été le premier satellite d'observation civile mais les images n'en ont été commercialisées et largement diffusées qu'à partir de 1985.

- LANDSAT7 fournit des images d'une résolution spatiale de 15 m dans 7 bandes spectrales (3 dans le visible, 2 dans le proche infrarouge, 2 dans l'infrarouge thermique). Son principal avantage est cette bonne résolution spectrale.
- SPOT4 un satellite héliosynchrone d'orbite polaire. Il est doté d'un capteur panchromatique (image blanc/noir, pixel de 10 m au sol) et de 3 capteurs dans le visible et le proche infrarouge (pixel de 20 m au sol). Un de ses avantages, outre sa bonne résolution spatiale, est sa possibilité d'acquérir des couples stéréoscopiques d'image permettant analyse et vision en 3 dimensions. Ces propriétés permettent des applications en zone urbaine et des représentations en relief de zones dépourvues de cartes récentes.



- IRS est un satellite indien combinant des caractéristiques des séries Landsat et Spot : il est en effet doté d'un capteur panchromatique dans le visible (pixel de 5 m au sol) et de 6 capteurs « couleurs » dans le visible et l'infrarouge (résolutions variables, de 23 m à 188 m). Comme Spot, il est capable de capter des couples stéréoscopiques d'images.
- IKONOS est un satellite commercial lancé en 1999 par la compagnie Space Imaging : il délivre des images noir/blanc à pixel de 1 m au sol et couleurs à pixel de 4 m au sol. L'image satellite « haute résolution » converge vers les utilisations traditionnelles de la photo aérienne, comme le montre la Figure 7.5.



Figure 7.5 : Image du centre de Rome prise par le satellite IKONOS (source : Space Imaging)

### C. Autres types de satellites et/ou de capteurs

- satellites d'observation marine spécialisés (dont la série des MOS japonais),
- satellites munis de RADAR (comme les ERS européens et le canadien RADARSAT), capables d'acquérir des images de jour et de nuit, par temps clair ou nuageux,
- capteurs LIDAR émettant des rayons laser, utilisés notamment pour l'acquisition d'images 3D de pollution atmosphérique.

#### 7.1.3 Traitement d'images satellitaires

Les réflectances enregistrées par les capteurs sont transmises à des stations de réception terrestres (Fig. 7.1, opération **E**), soit en temps réel (directement ou via un autre satellite) soit en temps différé (lorsque le satellite arrive « en vue » d'une station de réception). Dans tous les cas, les données transmises se présentent sous la forme brute d'un flux de données qui doit être traité (Fig. 7.1, opération **F**) avant d'être délivré à des utilisateurs.

#### A. Pré-traitements d'images

Très rapidement après réception, les images peuvent être visualisées sous forme de « vues rapides » (« *quick looks* »), notamment pour s'assurer que la scène n'est pas cachée par des nuages. Elle donne ensuite lieu à deux types de redressements : radiométriques et géométriques.

- Les redressements radiométriques sont rendus nécessaires par les variations d'illumination de la terre, de conditions atmosphériques, de perturbations des capteurs. Les corrections de variations d'illumination sont rendues nécessaires pour la comparaison d'une même scène prise à deux passages successifs dans le temps du satellite ou par la juxtaposition de scènes spatialement contigües, prises à des moments différents. La dispersion et l'atténuation de l'illumination lors des traversées de l'atmosphère (du soleil à la terre, de la terre au satellite, du satellite à la terre) peuvent être différentes selon les conditions du moment. Corriger l'effet de perturbations atmosphériques peut passer par des opérations simples (égaliser les réponses d'une même occupation du sol) ou complexes (modélisation des conditions atmosphériques lors de l'acquisition de l'image). Les irrégularités de fonctionnement des capteurs ou de la transmission sont aujourd'hui devenues plus rares qu'à l'époque « pionnière » des débuts (années 70). Elles pouvaient parfois être corrigées.
- Les redressements géométriques sont liées à des distorsions dues au mouvement du satellite, à son altitude, à sa vitesse comme à la courbure, à la rotation et au relief de la terre. La modélisation du mouvement du satellite par rapport à la terre en permet la correction. Il reste encore à positionner l'image dans un référentiel géographique (cf chapitre 4). Cela peut être fait par repérage de points caractéristiques
  - & soit dans deux images de la même scène (positionnement « relatif »),
  - & soit dans une image et sur une carte (ou au sol) où ces points ont des coordonnées connues en latitude et longitude (positionnement « absolu »). Dans ces deux cas, à partir de ces points d'amer, le positionnement de tous les pixels de l'image est calculé. Dans le second cas, l'image peut être exprimée dans un système de projection.

L'image est alors prête à être délivrée à ses utilisateurs. Elle se présente sous forme d'un fichier (image panchromatique) ou de plusieurs (images « fausses couleurs »). Chaque fichier est un tableau numérique, chacune des cases portant un nombre, indice de la réflectance du pixel dans une fenêtre de longueur d'onde. Ce nombre est généralement (Landsat, Spot,...) compris entre 0 et 255.

## B. Interprétation d'images satellitaires

Puisque l'image n'est pas une photographie mais l'enregistrement d'une réflectance dans une fenêtre de longueurs d'onde, elle doit être « interprétée » par l'utilisateur qui doit faire correspondre à chacun de ses pixels un code numérique pour une occupation du sol. Cette interprétation fait très massivement appel à l'informatique.

La première étape consiste souvent à **améliorer l'image** à l'aide de traitements graphiques.

- Si l'histogramme du nombre de pixels par valeur de réflectance est trop resserré sur une plage restreinte de valeurs, l'utilisateur peut « étirer » cet intervalle entre 0 et 255 afin de rendre les contrastes plus évidents. Il peut le faire sur toute l'image ou sur une partie seulement.
- Il peut aussi appliquer à l'image un « filtre ». Un filtre « passe bas » consiste à définir une fenêtre carrée de 3\*3, 5\*5 ou plus, de remplacer la valeur du pixel central par la moyenne (pondérée ou non) des valeurs des pixels de la fenêtre. Cette opération est répétée pour tous les pixels de l'image. L'effet est un lissage mettant en évidence les grandes zones homogènes d'occupation du sol. L'application d'un filtre

« directionnel » a, au contraire, pour but de mettre en évidence des discontinuités et des éléments linéaires (routes, par exemple).

Une première **interprétation visuelle** peut alors être esquissée. La reconnaissance d'objets différents s'appuie sur leurs différences de « couleur » (claire - foncée), de taille, de forme, sur la répétition d'un même « motif » (des blocs d'immeuble, par exemple), sur l'association qu'on sait fréquente de certains éléments (zones commerciales et échangeurs autoroutiers en périphérie urbaine, ...). Une première interprétation visuelle est, le plus souvent, incomplète (nombreux pixels incertains) et possible seulement sur une seule image (pas de combinaison de plusieurs images de la même scène) : on doit donc recourir à des traitements complémentaires.

Des **calculs arithmétiques simples** entre deux images peuvent être d'un grand secours.

- La soustraction de deux images de la même scène prise à deux moments différents crée une image des différences temporelles,
- La division de deux images de la même scène dans deux bandes de longueurs d'onde différentes (« ratio spectral ») permet de lever certaines ambiguïtés. Par exemple, des forêts de montagne de composition identique ont souvent des réflectances variables selon leur orientation mais la division de deux bandes spectrales les fait apparaître comme semblables. Le plus connu des ratios spectraux est le NDVI qui révèle l'état de la végétation.

Des **calculs plus sophistiqués** sont souvent nécessaires.

- L'analyse en Composantes Principales (ACP) se pratique sur les images correspondant aux différentes bandes de longueurs d'onde d'une même scène (7 pour Landsat, 3 pour Spot). L'information qu'elles contiennent est partiellement redondante : il faut donc calculer des « facteurs » non corrélés (pour pouvoir, ensuite, faire des opérations arithmétiques licites entre eux). En outre, l'analyse résume l'essentiel de l'information en peu de facteurs. Par exemple, 90% environ de l'information contenue dans les 7 bandes Landsat est généralement résumée par les 3 premiers facteurs d'une ACP.
- Diverses méthodes de classification multivariée sont très souvent utilisées pour identifier / discriminer des types d'utilisation du sol. Une première façon de faire consiste à employer un algorithme regroupant les pixels en classes auxquelles l'interpréteur affectera une occupation du sol. On parle dans ce cas de classification non supervisée. Il existe aussi des méthodes de classification supervisée dont le principe est le suivant. Par sa connaissance du terrain et/ou des opérations préalables, l'interpréteur a identifié des utilisations du sol dont les signatures spectrales sont donc connues. Celles ci vont ensuite servir à affecter les pixels non encore interprétés à l'un ou l'autre des types déjà identifiés. Une « matrice de confusion » permet d'évaluer la qualité d'une classification supervisée en comparant les pixels identifiés par l'interpréteur et ceux classés automatiquement.
- Sur une image panchromatique, les opérateurs de l'analyse d'image peuvent également se révéler utiles pour la recherche de structures et de textures.

La mise en oeuvre de traitement d'images satellitaires implique le recours à des logiciels, soit relativement simples comme Titus ou Idrisi (principalement SIG raster) soit plus complexes comme Erdas, PCI, ENVI, Geoimage, etc.



Une fois interprétées, les images satellitaires peuvent être intégrées à un SIG raster (cf chap. 6) et être combinées à des couches de données géographiques acquises par d'autres moyens. Leur principal avantage est, alors, évident : elles constituent des couches d'occupation du sol, éventuellement étagées dans le temps et, surtout, récentes pour les dernières acquises.

#### **7.1.4 Exemple : la croissance urbaine de Tunis (1987 – 1996)**

Cet exemple, emprunté à L. Bouassida, Ch. Maréchal et N. Ben Khatra, figure avec d'autres TP sur le site <http://www2.ac-toulouse.fr/msoe/sommaire/peda/tp3/>.

Le but de l'exemple est de mesurer, à partir d'images du satellite SPOT, la progression du front urbain de Tunis puis de préciser les évolutions de sa texture et de sa structure.

##### **A. Progression du front urbain de Tunis de 1987 à 1996**

Cette progression peut d'abord être estimée par des moyens visuels (sans traitement mathématique) en combinant à l'écran les trois canaux XS1 (Bleu en fausses couleurs, canal du visible « vert »), XS2 (Vert, canal du visible rouge) et XS3 (Rouge, canal du proche infrarouge). La végétation active apparaît, par exemple, en rouge vif à l'écran. L'évolution d'occupation du sol se lit par superposition des canaux des images de 1987 et de 1996 (Figure 7.6).

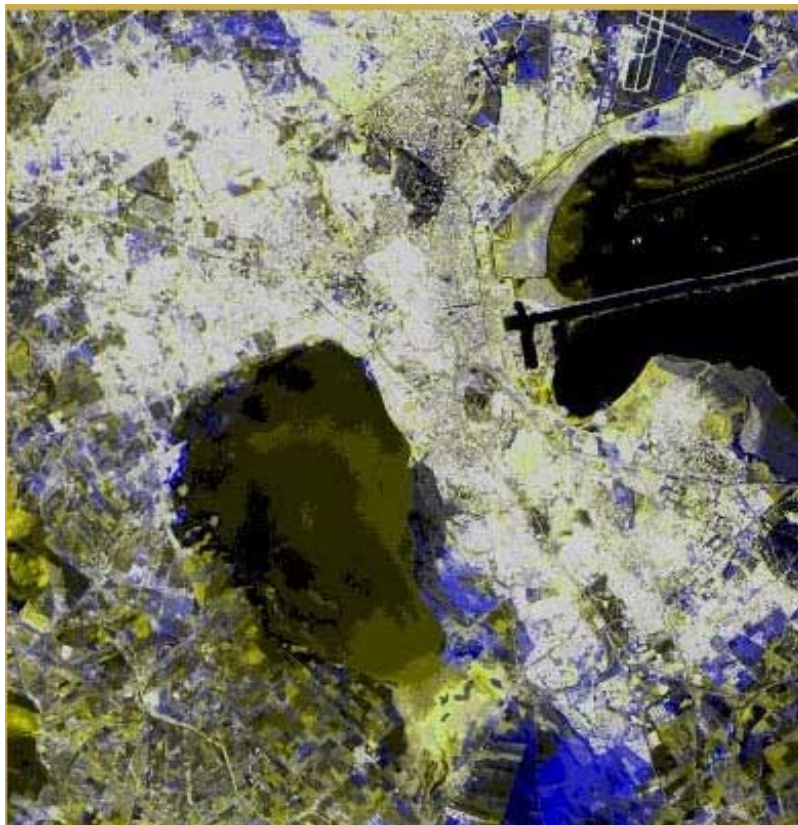


Figure 7.6: superposition des canaux des images SPOT XS de 1987 et 1996 (source: L.Bouassida, Ch.Maréchal, N.Ben Khatra)

Apparaissent en bleu sur la figure 7.6 les secteurs urbanisés entre les deux dates, en jaune les zones végétales et dans les autres couleurs celles qui n'ont pas changé. On distingue

clairement les secteurs en forte progression urbaine : périphérie sud principalement mais aussi croissance auréolaire quasi générale.

Cette première approche par interprétation visuelle peut être utilement complétée par des traitements mathématiques , notamment pour différencier le tissu urbain.

## B. Différenciation du tissu urbain par traitement d'image

L'indice de végétation (NDVI = « Normalized Difference Vegetation Index ») met bien en évidence les zones végétalisées et, par contre coup, les secteurs artificialisés, à bâti urbain dense. Cet indice s'obtient, pour chaque pixel, de la façon suivante :

$$\text{NDVI} = (\text{PIR} - \text{R}) / (\text{PIR} + \text{R})$$

Où PIR désigne la valeur du pixel dans le canal infrarouge et R dans le canal rouge. Cet indice varie entre -1 et +1, les valeurs voisines de 0 identifiant les sols nus.

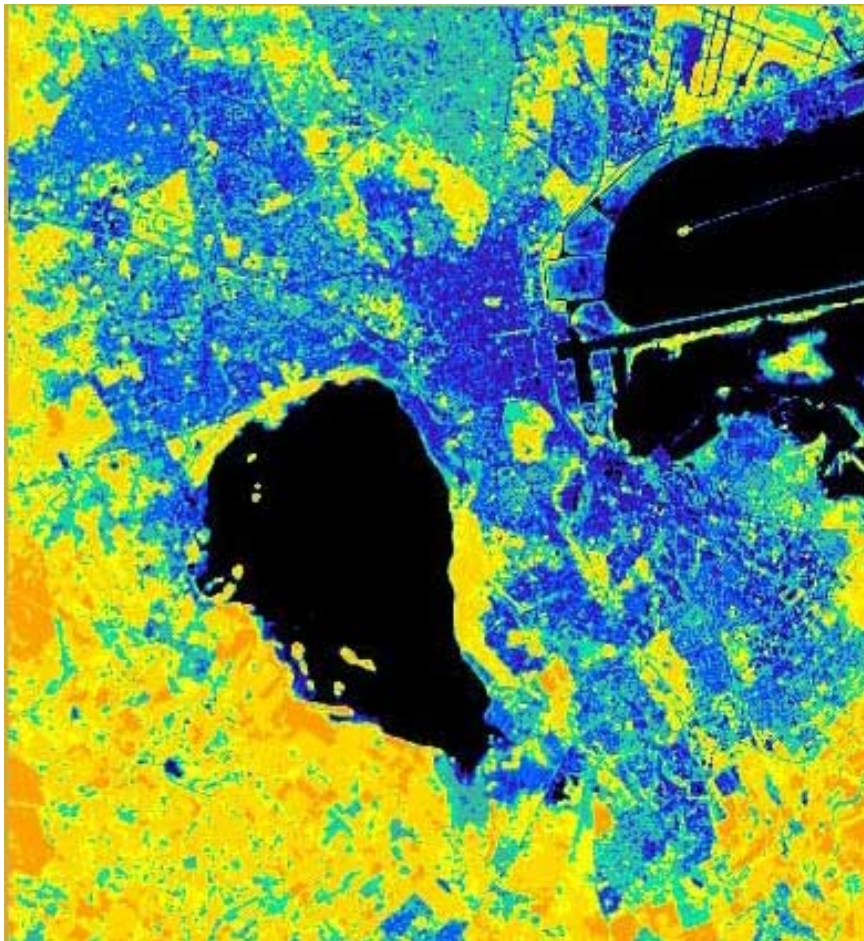


Figure 7.7: Mise en évidence de la densité du bâti par l'indice NDVI (source: L.Bouassida, Ch.Maréchal, N.Ben Khatra)

La figure 7.7 met bien en évidence les noyaux de forte densité du bâti : centre ville, quartier nord est, bord de mer.

On peut aussi vouloir délimiter et qualifier de façon plus précise le périmètre urbain. Il vaut mieux, pour cela, utiliser une image SPOT panchromatique (niveaux de gris) dont la résolution spatiale est meilleure (pixel de 10\*10 m au lieu de 20\*20m). Les

caractéristiques texturales peuvent s'y obtenir par application de filtres, de lissages, d'opérateurs de morphologie mathématique :

- Erosion qui amincit les formes,
- Dilatation qui les épaissit,
- Ouverture (dilatation puis érosion)
- Fermeture (érosion puis dilatation).

On trouvera d'autres exemples d'application dans l'ouvrage de M. Robin (*La télédétection*, Nathan Université, 1995). On peut tenter ici de dresser un panorama général des principales applications civiles de la télédétection par images satellitales.

### 7.1.5 Principales applications civiles

L'occupation du sol, les océans, la basse atmosphère, observés fréquemment, sur tout le globe terrestre et avec des capteurs à longueurs d'onde plus larges que le visible constituent des informations utiles dans de nombreux domaines, civils et militaires. Pour ce qui est des applications civiles, on en retiendra ici quatre grands types.

Nous avons déjà vu qu'une des plus anciennes applications civiles concernait la prévision météorologique et qu'il existe, pour cette activité, un réseau de satellites spécialisés. Cela permet, entre autres, le suivi de trajectoire des cyclones tropicaux et le déclenchement d'alerte. Il existe de même des satellites spécialisés dans l'observation océanographique. Les applications de type « sciences de la terre » (géologie, hydrologie), scientifiques ou industrielles, ont beaucoup bénéficié de l'image satellitale. En géologie, cela concerne la recherche pétrolière et minière, l'exploitation de matériaux, l'observation géomorphologique de surface (affleurements rocheux, dépôts sédimentaires et alluviaux, sols, ...), le suivi et le contrôle de risques naturels (glissements de terrain, éruptions volcaniques, ...). L'image satellitale permet aussi de suivre l'eau sous toutes ses formes (réseau hydrographique, zones humides, glace, neige, humidité des sols), la structuration du relief par les eaux (délimitation de bassins versants, avancée et retrait des littoraux), les étiages et les crues.

Un grand nombre d'applications concernent la gestion de l'agriculture et de la forêt. Pour ce qui est de l'agriculture elle-même, cela va de la cartographie des types de culture à l'estimation des récoltes en passant par la surveillance des conditions de sol, d'humidité, de croissance végétale ou bien encore de respect des règles de la Politique Agricole Commune européenne. On peut mettre ici en exergue les utilisations futures de l'image à haute résolution spatiale pour l'agriculture de précision (qui adapte les intrants à la variabilité des conditions de culture à l'intérieur des parcelles). Quant à la gestion forestière, elle fait à diverses échelles un important usage de l'image satellitale aussi bien en matière de déforestation tropicale, de reconnaissance des espèces, de planification de la production, d'estimation de la croissance végétale que d'incendies de forêt.

Le dernier grand type d'applications concerne la cartographie (en 2 ou 3 dimensions) dans les régions du monde dépourvues de cartes de base ou dont les cartes de base sont anciennes.

Le caractère multispectral et multitemporal de l'image satellitale, ses possibilités de combinaison avec d'autres données géoréférencées au sein d'un SIG en font une source d'information précieuse pour les géographes travaillant de façon analytique à petite ou moyenne résolution spatiale.

Certaines applications prospectives à grande échelle spatiale en aménagement (urbanisme, par exemple) font de plus en plus appel aux techniques de la « réalité virtuelle » .

## **7.2 SYNTHÈSE D'IMAGES POUR L'AMÉNAGEMENT**

L'idée fondamentale est de construire des images **virtuelles**, fixes ou animées, les plus « véristes » possibles de façon à simuler les modifications apportées à un paysage par un projet d'aménagement important (une autoroute, un viaduc, un groupe d'immeubles, etc.). En arrière plan de cette pratique, il y a, chez les décideurs, la nécessité de communiquer l'impact visuel de cet aménagement sur son environnement, voire même la possibilité d'en négocier des variantes. On a là un outil pouvant favoriser la « gouvernance territoriale ».

### **7.2.1 Trois étapes dans la construction d'une image virtuelle**

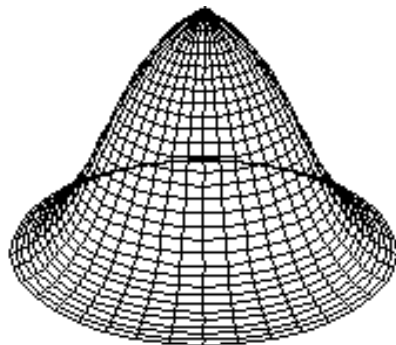
#### **A. Représentation géométrique en 3 Dimensions**

La première étape consiste à construire une représentation géométrique en 3 dimensions de la **structure topologique** de tous les objets d'une scène, en commençant par les plus élémentaires (une fenêtre, une cheminée, un mur, etc.) qui seront ensuite assemblés en objets complexes (une maison par exemple). La « construction » suit donc une procédure hiérarchique ascendante.

Pour les objets n'existant pas dans la scène réelle, ils seront composés au moyen de figures géométriques élémentaires (point, ligne, rectangle, sphère, ...) ensuite assemblées comme avec un meccano. Ces opérations s'effectuent à l'aide du « *modeleur* » présent dans les logiciels d'image de synthèse.

Pour les objets existant dans la scène réelle, il faut pouvoir disposer de données décrivant leurs topologie et géométrie tridimensionnelles. Selon l'échelle de la représentation, elles proviennent de bases de données géographiques, d'une campagne de photogrammétrie, de plans d'urbaniste ou d'architecte, de modèles de végétation, etc.

Le modeleur, comme celui de 3D Studio Max (de Autodesk), permet de stocker en machine une représentation 3D de chaque scène selon un modèle dit « fil de fer » illustré par la figure 7.8.



*Figure 7.8: représentation «fil de fer» d'un cône (source : [www.ensimag.fr](http://www.ensimag.fr))*



La représentation en 2 dimensions (écran, imprimante) d'une scène tridimensionnelle implique un observateur extérieur à la scène, situé (à un moment donné) en un point XYZ et, donc, le calcul des parties de la scène qu'il peut voir et une projection en perspective qui réduit la dimension des objets lointains.

Si l'on désire simuler un observateur en mouvement, ou bien chaque scène doit être précalculée ou bien les lointains doivent être des images préenregistrées.

## B. Habillage avec une texture

La structure géométrique d'un objet ne suffit pas à lui donner une apparence de vérisme, il faut l'habiller d'une **texture**, groupe de pixels raster simulants tel ou tel matériau, tel ou tel végétal, tel ou tel nuage, etc. Les logiciels d'image de synthèse contiennent, en général, des « banques de texture » dans lequel l'utilisateur pourra choisir la mieux adaptée. Un bon infographiste peut créer ses propres textures. Il doit alors jouer sur les propriétés de la surface de l'objet (plus ou moins lisse ou granuleuse), sur ses couleurs, ses motifs, son éclairage par des sources lumineuses. C'est en effet l'éclairage de la scène qui, via ombres, reflets, transparences, contribue (avec les calculs de perspective) à donner l'impression de relief.

La figure 7.9 fournit un extrait d'une « banque de textures ».

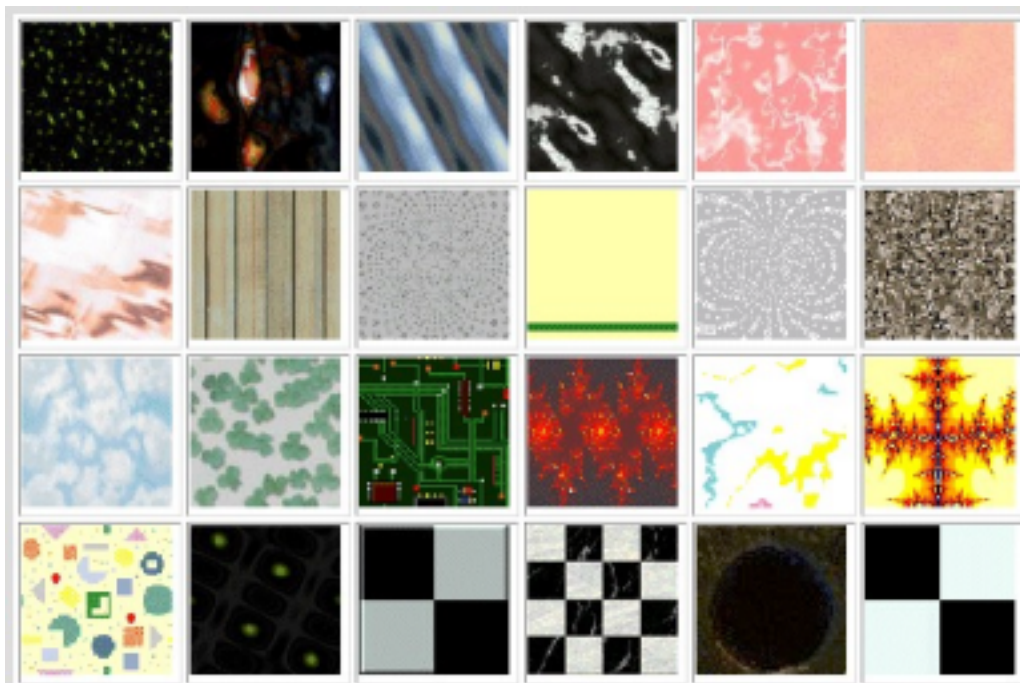


Figure 7.9 : exemples de texture (source : [www.nepthys.com/textures/](http://www.nepthys.com/textures/))

## C. Eclairage d'une scène

Il s'agit de reproduire :

- l'éclairage **diffus**, constant en tout point de la scène,
- et l'éclairage provenant de sources **ponctuelles et directionnelles**, variable en fonction de l'orientation des diverses surfaces par rapport à l'observateur,

Dans ce cas, jouent les phénomènes de :

- **réflexion** (la couleur d'une lumière perpendiculaire à une surface est celle de la source lumineuse et non celle de l'objet)
- et de **réfraction** (distorsion par traversée d'un objet translucide).

Dans tous les cas, doivent aussi être calculés les ombrages, si importants pour donner l'impression de relief et de profondeur.

L'éclairage d'une scène est réalisé à l'aide de modèles d'optique géométrique reproduisant la propagation des ondes lumineuses dans différents milieux telle qu'elle est perceptible par un œil humain. La figure 7.10 en fournit un exemple.



*Figure 7.10 : exemple d'éclairage d'une scène (source : <http://perso.easynet.fr/~philay/>)*

## 7.2.2 Quelques exemples d'application géographique

Nous n'abordons ici que quelques uns des exemples possibles.

### A. Simuler des changements du paysage agricole

L'étude tente d'anticiper, dans une petite région rurale située au Sud de Grenoble (Trièves), les possibles effets paysagers de l'extensification de l'agriculture dans un contexte de déprise agricole et d'essor du tourisme vert. Les simulations devaient servir, pour la Chambre d'Agriculture de l'Isère, de base pour négocier avec les agriculteurs le maintien subventionné des parcelles les plus visibles. L'essentiel de l'étude préalable a pu être menée avec des logiciels courants : un SIG raster (Idrisi), un « retoucheur » d'images (Adobe Photoshop) et des données simples : modèle numérique d'altitude de terrain et de végétation et photo aérienne. La figure 7.11 présente le « bloc diagramme » ayant servi de support aux simulations (paysage visible depuis la petite ville de Mens, poursuite de l'enforestement, modifications volontaires des systèmes de culture).



*Figure 7.11 : simulation paysagère à l'aide d'un SIG (source : J.Guilloux)*

A l'échelle de cette étude (le 1/25000), l'image de synthèse n'était pas nécessaire pour simuler des modifications paysagères. Elle le devient quand la précision désirée augmente et l'échelle de représentation devient plus grande (1/2000 ou plus).

## **B. Modélisation procédurale de grands paysages (CIRAD / IAURIF)**

Le projet « Imago Metropolis », dont les promoteurs sont le CIRAD (<http://www.cirad.fr>) et l'IAURIF (<http://www.iaurif.fr>), s'appuie sur les mêmes principes qu'un logiciel (AMAP) simulant la croissance de 400 espèces végétales. Imago Metropolis est conçu pour transformer les cartes d'occupation du sol en maquettes de paysage en 3D. Les types d'objet élémentaire (toit, pan de mur, trottoir, segment de rue...) y sont construits par une procédure de modélisation particulière (qui admet des paramètres variables comme dimensions XYZ, texture, ...) et ces objets peuvent se combiner en objets complexes (une maison, un immeuble, une rue, ...). On a là un logiciel adapté à la modélisation 3D rapide de paysages urbains & périurbains vus d'une certaine distance (paysage d'urbaniste, non d'architecte). La Figure 7.12 illustre les possibilités d'Imago Metropolis.



*Figure 7.12 : exemple de documents produits par Imago Metropolis (source : IAURIF)*

## **C. Paysages « véristes » précalculés : un exemple par l'association Ideralpe (Lausanne – Grenoble)**

Ideralpe était l'association de membres de l'Ecole d'Arts de Lausanne et de l'équipe SIG de l'Université J. Fourier de Grenoble. L'idée de base (1992) était de créer un outil de communication / négociation d'aménagement couplant un SIG (pour les aspects réglementaires et le stockage d'information géographique), un imageur de synthèse et des

modules de simulation de l'impact de telle ou telle conséquence de l'acceptation du projet d'aménagement. Par rapport à Imago Metropolis, les paysages produits sont beaucoup plus véristes mais requièrent des temps de préparation et de calcul beaucoup plus longs.

La Figure 7.13 représente ainsi, reconstituée en image de synthèse, la poste du village de St Pierre de Chartreuse.



Figure 7.13 : en image de synthèse, la poste de St Pierre de Chartreuse ( source: IDERALPE et S.Aubel )

#### **D. De grands parcours urbains calculés en temps réel : le projet « IMAGIS »**

Le projet, commun au laboratoire GRAVIR (IMA Grenoble) et au Massachusetts Inst. of Technology (<http://www-imagis.imag.fr/VILLE/>), a pour objectif de représenter en image de synthèse le parcours d'un observateur dans un quartier urbain de grande taille. La particularité du projet est que la vision du paysage qu'a l'observateur au long de son parcours doit être calculée « en temps réel » (au fur et à mesure de son déplacement) avec une qualité visuelle acceptable pour un urbaniste. Cela implique nombre d'optimisations, sans doute matérielles (probablement un calcul « en réseau » sur des machines puissantes) mais aussi logicielles.

L'une de celles ci concerne les « imposteurs multimaillés ». Le principe est le suivant : à un instant donné, les objets proches de l'observateur sont représentés (en mode vectoriel 3D) avec une géométrie précise et une texture tandis que les objets des arrière plans le sont en mode rasteur avec déformations de perspective. La transition entre les divers plans est graduelle car le deuxième arrière plan est une image à pixels nombreux tandis que les suivants, plus lointains, sont faits de pixels plus grands et moins nombreux. Ces diverses images, correspondant aux différents plans de la scène, sont qualifiées d'imposteurs, multimaillés car constitués de maillages emboîtés. La figure 7.14 illustre ce principe.





Figure 7.14 : principe d'un imposteur maillé (source : équipe IMAGIS, IMA Grenoble)

Malgré les progrès informatiques déjà accomplis, il faut choisir entre :

- l'image de grande qualité visuelle, coûteuse car longuement précalculée (structures « fil de fer », textures, éclairages), nécessitant l'intervention d'infographistes qualifiés et dont l'utilisateur serait plutôt un architecte ou un paysagiste travaillant à l'échelle du 1/50 à 1/500,
- l'image calculée en temps réel, dont la qualité visuelle est nécessairement moindre mais qui peut représenter un observateur mobile dans des espaces importants considérés à plus petite échelle (1/1000 à 1/5000) et dont l'utilisateur serait plutôt l'urbaniste ou l'aménageur local,
- l'image paysagère construite à partir d'un SIG raster et dont l'utilisateur serait plutôt l'aménageur régional ou le géographe (échelles du 1/10000 au 1/50000).

### Quelques sites Internet sur l'image satellitale

- CCRS : cours de télédétection à <http://www.ccrs.nrcan.gc.ca/ccrs/eduref/>
- SPOT IMAGE : <http://www.spotimage.fr/>
- NASA : <http://www.nasa.gov/>
- SPACE IMAGING : <http://www.spaceimaging.com/>

### Quelques applications de l'image de synthèse

- J. Guilloux : *Activité agricole et paysage visible, application au canton de Mens (sud isère)* (Revue Internationale de Géomatique, 1995 n° 3-4)
- IAURIF : *Imago Metropolis* (Note rapide n°3, mars 2000)
- S. Aubel, P.Dumolard : *Contribution à l'outil Ideralpe, l'application St Pierre de Chartreuse* (Revue Internationale de Géomatique, 1995 n° 3-4)
- M.Bécholey et al. : *Aménagements paysagers et aide à la décision, apport des nouvelles technologies* (Revue Internationale de Géomatique, 1998 n°3)
- <http://www-imagis.imag.fr/> (exemples de développements informatiques)

## Chapitre 8 : Modèles, analyse et simulation

Le mot « **modèle** » est un des mots clé de l'informatique mais aussi, en géographie, de ce que l'on appelle habituellement « **analyse spatiale\*** ». Celle-ci est souvent bien autre chose que de l'analyse et est temporelle autant que spatiale. En effet, à la différence d'autres manières de faire, l'analyse spatiale s'efforce de formaliser, souvent par un langage mathématique ou statistique, les ressorts de la localisation d'activités ou de l'organisation de l'espace géographique. Pour ce faire, on utilise ou aboutit à un modèle. On définit ici un modèle comme une représentation générale, abstraite d'un phénomène, dans son essentiel, son essence. Pour reprendre une image fréquemment utilisée, modéliser c'est garder le « bébé dans la baignoire et jeter l'eau du bain ». Cela signifie que le « bruit », les phénomènes considérés comme non essentiels dans la compréhension de la complexité du réel sont éliminés au profit d'une représentation jugée suffisante et valide, bref adéquate.

Au-delà de cette définition commune, il y a une grande diversité dans la pratique.

- Certains modèles sont essentiellement faits pour l'**analyse** de données empiriques : ce sont en général des modèles formels (statistiques par exemple).
- D'autres modèles, produits d'analyses précédentes, sont des modèles dédiés à un problème précis (par exemple l'estimation de production d'un polluant atmosphérique par la circulation automobile). Suffisamment validés, ils servent à **simuler** des futurs possibles sous diverses hypothèses.

### I. Grands types de modèles

En géographie, on associe fréquemment le mot « modèle » aux modèles classiques de l'analyse spatiale (datant du 19<sup>ème</sup> siècle et de la première moitié du 20<sup>ème</sup>). Les plus utilisés sont :

- Le modèle de Von Thünen décrivant la répartition des cultures, plus ou moins intensives, selon leur distance au siège d'exploitation et aux moyens de communications
- Le modèle gravitaire (Ravenstein, Reilly, Zipf) assimilant l'interaction entre deux centres urbains à celle existant entre deux planètes du système solaire (loi de Newton)
- Le modèle de Weber prédisant la localisation d'une industrie en fonction de ses approvisionnements en énergie, matières premières et de son marché
- Le modèle de Christaller décrivant la hiérarchie et l'espacement de centres de service selon leur niveau d'équipements ainsi que la forme hexagonale de leurs aires d'influence. Le modèle de Lösch complexifie ce modèle initial en s'intéressant à l'ensemble de l'économie et à l'économie urbaine.
- Les modèles intra-urbains (Burgess, Hoyt, Harris/Ullman) décrivant la forme auréolaire, et parfois sectorielle, de la distribution interne de leurs fonctions et densités.

Ces modèles ont une grande ambition puisqu'ils visent à expliquer et prédire une structuration répétitive dans le temps et l'espace (cf. Haggett, 1973 ; Pumain, Saint-Julien, 1997). On peut les qualifier de **modèles généraux** (et presque de lois). Ils présentent un certain nombre de points communs. Tous partent du constat de la concentration spatiale, à diverses échelles (éventuellement emboîtées) et, le plus souvent, d'un « appauvrissement fonctionnel » avec l'accroissement de la distance au(x) centre(s). Cet appauvrissement implique une hiérarchie des aptitudes à payer la rente de localisation, dans la pure lignée de l'économie libérale néo-classique. Les distributions spatiales sont alors la projection au sol de l'économie de marché. Elles aboutissent aux formes géométriques simples (des cercles, des triangles, des hexagones) car l'espace, dans ces modèles, est considéré comme un espace euclidien, homogène (comme une « feuille de papier ») et isotrope où les distances sont calculées « à vol d'oiseau ». On sait qu'on calcule ainsi des distances minimales, ce qui peut être acceptable à petite échelle, pour de vastes territoires (mais se pose alors le problème de différentes projections planes !).

L'intérêt principal de ces modèles est **didactique** (ils illustrent de façon simple l'effet tendanciel d'un mécanisme). Il peut être **heuristique**, si l'on observe les écarts entre valeurs observées et celles prédites par le modèle (qui sert à annihiler une évidence).

Un deuxième type de modèles est d'ambition plus modeste, visant à tenter de prédire la distribution spatiale d'un phénomène limité : le cheminement d'une avalanche ou d'un glissement de terrain, l'évolution d'un flux automobile, un risque localisé, la diffusion d'un polluant, la fréquentation d'un espace naturel, etc. On peut qualifier ces modèles de **modèles particuliers**, intermédiaires entre lois et réalités empiriques. Ils sont nombreux et variés, dans la mesure même où chacun représente un phénomène particulier. Ils possèdent cependant un certain nombre de caractéristiques qui les différencient des modèles généraux. Ils tentent, le plus fréquemment, de mimer des processus plutôt que des distributions invariantes, ils sont donc souvent utilisés pour tenter d'anticiper une évolution ou l'effet d'un aménagement, ce qui en fait des instruments opérationnels de **simulation** et d'**aide à la décision**. Une autre de leur utilisation possible est de **pallier l'absence de données**, ce qui est fréquent en géographie physique où les grandeurs varient de façon continue dans l'espace mais ne sont mesurées que ponctuellement. Par exemple, les précipitations affectant toute une région ne sont mesurées que là où existent des stations météorologiques. Une interpolation, créant entre ces points de mesure des valeurs « vraisemblables » respectant une logique explicite, généralise spatialement l'information.

Un troisième type est constitué de ce qu'on peut appeler des **modèles de modèles**, en fait des langages formels (numériques ou non) utilisables pour modéliser. Il s'agit surtout des mathématiques (science pour **construire** des modèles) et de l'informatique (science pour **opérationnaliser** des modèles). La statistique, avec ses innombrables lois de probabilités (Laplace-Gauss, Poisson, etc....) et son recours aux méthodes algébriques fournit au géographe nombre d'instruments pour créer tel ou tel modèle particulier. C'est également le cas d'autres branches des mathématiques, comme la théorie des graphes, la topologie, etc. On remarquera, en passant, l'utilisation grandissante de modèles informatiques qualitatifs.

On signale ici pour mémoire le terme « **modèle de données** » qui signifie « structuration de l'information » destinée à faciliter la résolution informatique d'une classe de problèmes. Ainsi, pour ce qui concerne l'information géographique « image » (cf. chap. 6), le modèle de données matriciel est il le plus simple mais d'autres sont envisageables : par exemple, le modèle de données quadtree permettrait d'ajuster la

taille des pixels au contenu informatif local tandis que leur indexation par clés de Peano permettrait une connexion efficace à un système de gestion de bases de données.

Un *modèle* est une épure représentant l'essentiel d'un phénomène, d'un processus ou d'un système. Il peut prendre diverses formes : textuelle, graphique, mathématique. L'avantage qu'offre une formulation mathématique est d'être opératoire et de permettre de simuler des hypothèses et des scénarios différents. Quelle que soit sa forme, un modèle doit être *validé* par confrontation au réel.

## II. Modèles pour l'analyse spatiale

L'analyse spatiale procède de deux façons différentes pour mettre en évidence l'organisation de l'espace. Il s'agit de deux approches essentiellement analytiques et numériques mais différentes dans leur conception :

- L'une, que l'on qualifiera d'**analyse des données spatiales**, traite des données de façon non spatiale, cartographie les résultats et juge, *a posteriori*, des inégalités ou différences spatiales,
- L'autre, que l'on qualifiera d'**analyse spatiale des données**, introduit une ou plusieurs contraintes spatiales (la distance souvent) dans la méthode de traitement des données.

### 1. Analyse des données spatiales : quelques exemples

Une application peut en être trouvée dans l'utilisation conjointe, maintenant classique, de l'analyse des données et de la cartographie thématique. L'exemple ci-dessous en est une illustration simple.

Le tableau 8.1 est un tableau relatif à 17 pays ouest européens et 7 des indicateurs retenus par l'ONU pour construire son indice de développement humain, mieux à même que le seul produit intérieur brut par habitant de caractériser le développement d'un pays ou d'une région. Les 7 indicateurs ici retenus (année 1995) sont :

- L'espérance de vie à la naissance,
- La mortalité infantile,

- Le nombre d'habitants pour un médecin,
- Le nombre de journaux lus pour 1000 habitants,
- Le nombre de télévisions pour 1000 habitants,
- Le produit intérieur brut par habitant
- Le nombre moyen d'années d'études des habitants.

	<i>esp.</i>	<i>mort.</i>	<i>Habtts</i>	<i>jrnaux</i>	<i>TV</i>	<i>PIB/hb</i>	<i>années</i>
	<i>vie</i>	<i>enft</i>	<i>/docteur</i>	<i>/1000h</i>	<i>/1000h</i>		<i>etude</i>
Norvege*	77.10	4.00	503.00	614.00	425.00	16028.00	98.00
Suisse *	77.40	6.00	696.00	463.00	407.00	20874.00	93.00
Suede *	77.40	7.00	387.00	533.00	474.00	17014.00	100.00
France *	76.40	13.00	320.00	210.00	406.00	17405.00	102.00
Holland*	77.20	14.00	450.00	311.00	495.00	15695.00	104.00
G.B. *	75.70	11.00	700.00	395.00	435.00	15804.00	102.00
All. *	75.20	8.00	380.00	390.00	570.00	18213.00	90.00
Dk *	75.80	4.00	399.00	352.00	535.00	16781.00	98.00
Finl. *	75.50	15.00	443.00	559.00	497.00	16446.00	98.00
Autr. *	74.80	11.00	388.00	357.00	481.00	16504.00	90.00
Belg. *	75.20	4.00	331.00	305.00	452.00	16381.00	100.00
Luxbg *	74.90	2.00	554.00	389.00	255.00	19244.00	95.00
Eire *	74.60	3.00	681.00	159.00	276.00	10589.00	102.00
Italie *	76.00	6.00	234.00	107.00	424.00	15890.00	99.00
Espagne*	77.00	7.00	360.00	82.00	396.00	11723.00	92.00
Grece *	76.10	7.00	351.00	140.00	196.00	7366.00	89.00
Portug.*	74.00	14.00	412.00	38.00	177.00	8770.00	76.00

Tableau 8.1 : données de l'exemple (Source Eurostat, 1995)

L'objectif, à partir de ces données, est de déterminer les principaux profils de pays en Europe de l'Ouest en matière de développement économique, social et culturel. Une analyse en Composantes Principales (cf. Chadule, 1997) permet de résumer les 7 indicateurs par 2 « facteurs » qui expriment 60% de l'information du tableau 8.1. Le premier facteur (41% de l'information) est un axe de développement économique, fortement corrélé avec PIB/hab, TV et journaux pour 1000 habitants. Il oppose les pays de l'Europe « riche » (Pays Bas, Finlande, Norvège, Suède, Suisse) aux pays de l'Europe « périphérique » (Espagne, Irlande, Grèce, Portugal). Le deuxième facteur (19% de l'information) est un axe de « structure médicale » opposant les pays à

mortalité infantile relativement élevée (France, Pays Bas, Finlande, Italie) à ceux où elle est faible (Suisse, Luxembourg, Irlande) malgré un nombre de médecins relativement faible. Les 17 pays peuvent donc être représentés, en résumé, sur un graphique cartésien (en X les scores sur le facteur 1, en Y ceux sur le facteur 2) cf figure 8.1.

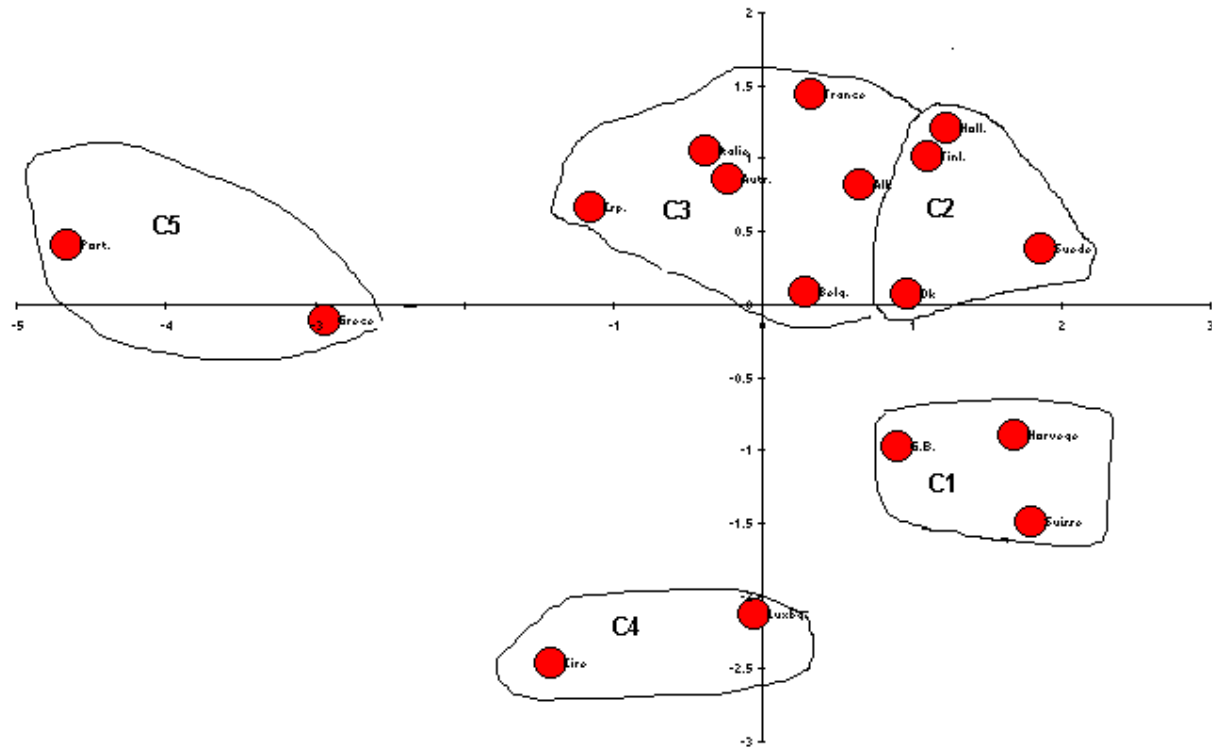


Figure 8.1: projection des pays du tab. 8.1 sur le 1<sup>er</sup> plan de l'AFC

Une typologie peut être effectuée directement sur ce graphique. Elle peut aussi être créée par une méthode de classification, par exemple celle des « nuées dynamiques » (Chadule, 1997), comme nous l'avons fait ici avec pour résultat 5 profils de pays.

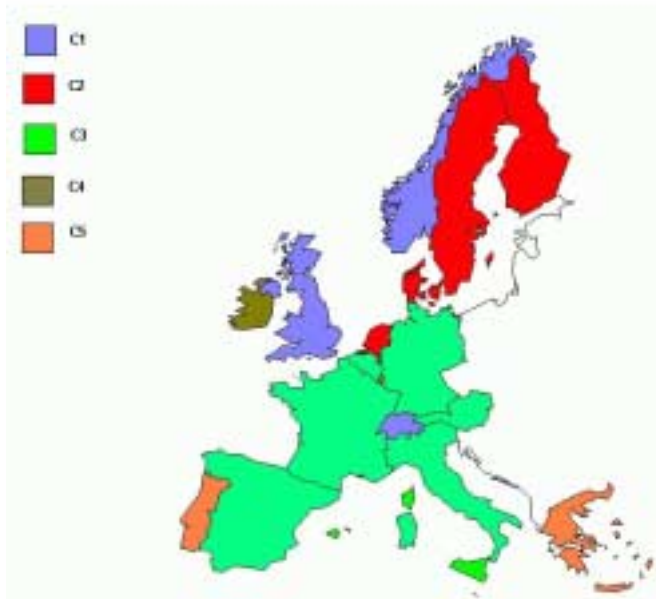


Figure 8.2: représentation cartographique de la typologie

La classe C1 (Norvège, Suisse, Grande Bretagne) est caractérisée par un fort PIB/hab et un nombre de médecins relativement faible (mais malgré tout, en moyenne, une mortalité infantile réduite)

La classe C2 (Suède, Danemark, Finlande, Pays Bas) est également un groupe de pays riches mais à structure sanitaire différente (mortalité infantile assez importante)

La classe C3 comporte la plupart des grands pays d'Europe de l'ouest, avec un PIB/habitant moyen et une mortalité infantile relativement importante,

La classe C4 est, paradoxalement, composée du Luxembourg et de l'Irlande : ce qui les rapproche est essentiellement leur très faible mortalité infantile,

La classe C5 (Grèce, Portugal) est composée des deux pays ouest européens les moins développés.

Une typologie simplifiée pourrait en être dérivée en joignant classes 1 et 2 d'une part, 4 et 5 de l'autre (chaque classe composite ayant alors deux sous types). Elle confirme assez bien les connaissances empiriques qu'a tout un chacun des inégalités de développement en Europe de l'Ouest.

Bien sûr, l'exemple ci-dessus proposé n'est qu'un exemple d'école et la méthodologie « ACP + classification + carte thématique » est d'autant plus puissante qu'il y a beaucoup



d'unités spatiales et beaucoup d'indicateurs. C'est alors qu'elle acquiert toute sa valeur heuristique.

On remarquera que cette méthodologie, séquence d'opérations algébriques et cartographiques, ne fait sur les variables aucune hypothèse préalable de nature sémantique et qu'elle utilise du quantitatif pour produire du structurel (c'est-à-dire du qualitatif). Elle est caractéristique d'une attitude d'exploration des données, pour laquelle existent bien d'autres façons de faire (*Exploratory Data Analysis*, par exemple).

## **2. Analyse spatiale des données : quelques illustrations**

À la différence des techniques précédentes, l'analyse spatiale des données fait explicitement intervenir une contrainte spatiale dans le traitement de l'information. Elle recouvre un assez grand nombre de techniques, relevant de mots clé tels que :

- Caractérisation d'un semis de points (centre, dispersion, configuration),
- Caractérisation d'un réseau, d'un découpage spatial,
- Interpolation (et son dual, l'échantillonnage spatial),
- Lissage d'une distribution spatiale,
- Interactions et hiérarchisation de lieux centraux,
- Localisations et allocations « optimales »,
- Diffusion spatio-temporelle,
- Agrégation / désagrégation de données spatiales
- Résidus d'un modèle d'analyse spatiale,
- Etc.

Nous n'en présentons ici que deux exemples simples, renvoyant le lecteur à l'abondante bibliographie pour de plus amples connaissances.

*a. Exemple d'utilisation d'un modèle : le modèle gravitaire*

À titre démonstratif, nous nous proposons de détailler quelque peu un modèle simple, le modèle gravitaire, et les différentes manières de l'utiliser dans une application géographique.

Au 17<sup>ième</sup> siècle, Isaac Newton a découvert la loi de l'attraction universelle qui décrit le mouvement des planètes du système solaire dû à l'attraction mutuelle qu'elles exercent les unes sur les autres. La forme de cette loi est :  $A_{ij} = g (M_i M_j) / d_{ij}^2$  où  $A_{ij}$  est l'attraction mutuelle des planètes  $i$  et  $j$ ,  $d_{ij}^2$  est (le carré de) leur distance euclidienne,  $g$  est la force de gravité,  $M_i$  et  $M_j$  les masses des 2 planètes en interaction ;  $g$  et  $M$  sont des grandeurs physiques universelles et cette loi est parfaitement symétrique ( $A_{ij} = A_{ji}$ ).

À la fin du 19<sup>ième</sup> siècle, un certain nombre d'auteurs du domaine des sciences sociales (Ravenstein par exemple), frappés par l'intensité et la relative symétrie des échanges entre lieux en fonction directe de leur « importance » et en fonction inverse de leur distance. Ils ont donc développé, métaphoriquement, le modèle gravitaire d'interaction entre lieux, de forme générale :  $I_{ij} = k (M_i M_j) / d_{ij}^a$  où  $I_{ij}$  est l'interaction (symétrique) entre lieux  $i$  et  $j$ ,  $k$  est une constante d'ajustement numérique,  $M_i$  et  $M_j$  la « masse » (à définir) des lieux  $i$  et  $j$ ,  $d_{ij}$  leur distance et l'exposant  $a$  estime la difficulté du déplacement (la rugosité de l'espace) entre  $i$  et  $j$ .

Les difficultés de mise en œuvre de ce modèle sont de différents ordres :

- d'ordre notionnel si, dans une application donnée, les échanges (de population, de biens, de services, d'informations) sont davantage hiérarchiques que symétriques,
- d'ordre « calculatoire » car on doit estimer deux inconnues (les masses et la distance) à l'aide d'une seule équation.

Face à cette dernière difficulté, diverses stratégies d'application du modèle sont envisageables. Le premier cas correspond aux situations où l'on a besoin d'information sans disposer de données empiriques. Le modèle va donc être utilisé pour créer des données « vraisemblables ». Par exemple, on dispose difficilement d'informations fiables sur les déplacements domicile-travail intra-urbains entre îlots de petite taille qu'on pourra estimer en considérant  $M_i$  comme les actifs à leur lieu de résidence,  $M_j$

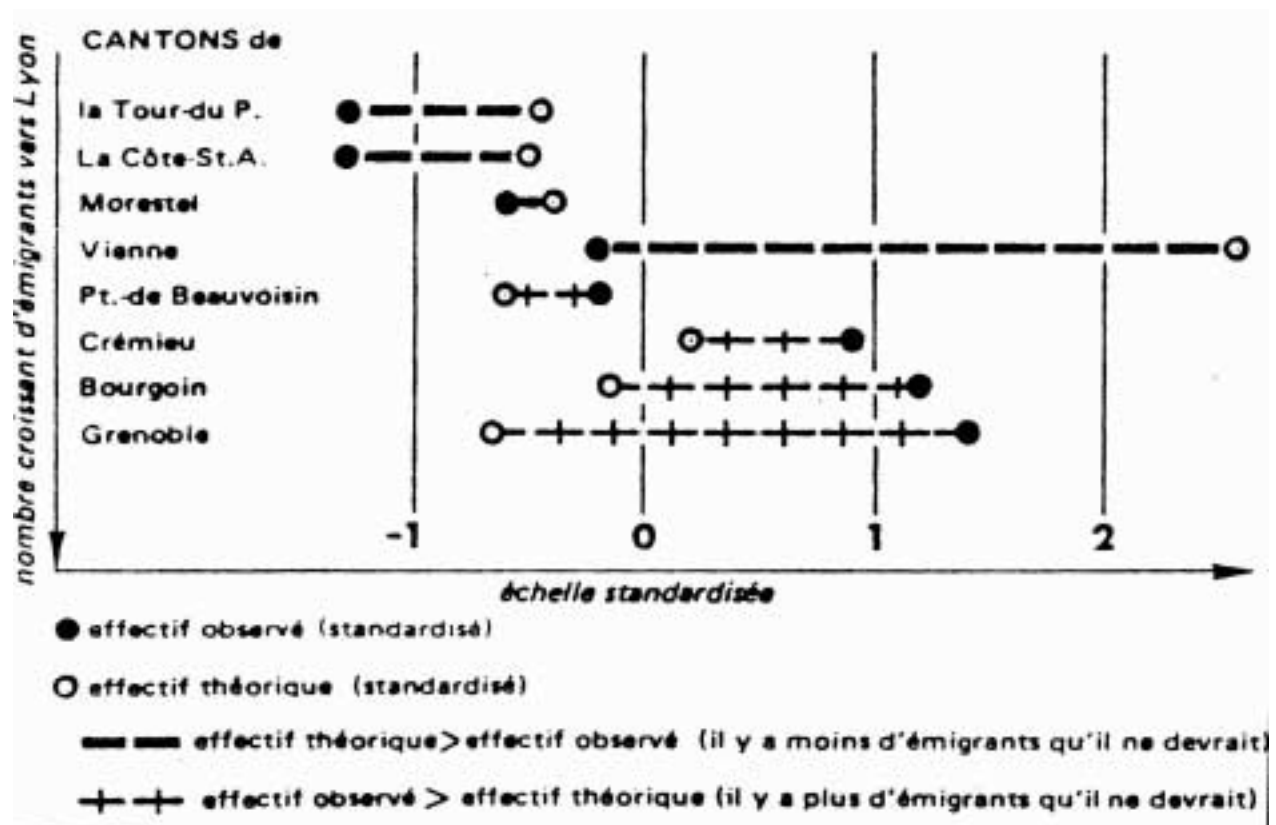
comme les actifs à leur lieu de travail et la distance comme le plus court trajet (en temps ou km) via le réseau entre îlots  $i$  et  $j$ .

Le deuxième cas correspond aux situations où l'on dispose de données empiriques et où l'on fixe l'une des deux inconnues pour pouvoir **ajuster** l'autre aux données. On peut, par exemple, définir la distance comme euclidienne (à « vol d'oiseau », avec le décimètre sur une carte), ce qui est fort commode. On peut aussi, pour tous les problèmes concernant la mobilité, fixer la distance comme une mesure d'accessibilité utilisant le réseau de communication. Dans les deux cas, cela permet de tester différentes grandeurs pour définir les masses à considérer<sup>1</sup>. On peut, au contraire, définir ce que sont les masses (nombres d'habitants ou d'actifs par exemple) pour tester différentes distances.

Le troisième cas correspond aux situations où l'on dispose de données empiriques, où l'on utilise le modèle gravitaire avec des paramètres simples (distance euclidienne, masses démographiques par exemple) et où l'accent est mis sur les résidus (ce que le modèle ne prend pas en compte). L'application présentée ici consiste à comparer, pour la seconde moitié du 18<sup>ème</sup> siècle, les nombres annuels moyens d'émigrants de cantons du bas Dauphiné vers Lyon à ce qu'ils auraient été si n'avaient joué que la distance à Lyon et la masse démographique de ces cantons. En un mot, le modèle a ici servi à effacer leur effet pour faire apparaître un (ou des) bassin(s) migratoire(s) préférentiel(s), un (ou des) types d'espace plus particulièrement liés à Lyon. La figure 8.3. indique les forts surplus et déficits pour les cantons lui envoyant au moins 10 émigrants en moyenne annuelle.

---

<sup>1</sup> D. Pumain et Th. Saint-Julien (2001) en fournissent un bon exemple dans leur ouvrage sur les interactions spatiales.



*Figure 8.3: Écarts « émigrants observés – modèle gravitaire », bas Dauphiné, fin 18<sup>ème</sup> siècle, (source P.Dumolard)*

Les cantons envoyant peu d'émigrants sont tous déficitaires tandis que ceux en envoyant davantage en envoient plus que si ne jouaient que leur masse démographique et leur distance à Lyon. Lors de cette phase précoce de concentration de la population, ce sont surtout les cantons les plus urbains.

Enfin, le quatrième cas d'utilisation évoqué ici fait référence aux raffinements successifs du modèle gravitaire, notamment celui dû à A.Wilson. On dispose d'une matrice origine – destination, indiquant les flux des lieux  $i$  (en ligne) vers les lieux  $j$  (en colonne). Le flux entre deux lieux peut s'écrire :

$$F_{ij} = A_i O_i B_j D_j e^{-bc}$$

Où  $c$  est l'accessibilité entre  $i$  et  $j$ ,  $O_i$  et  $D_j$  sont le flux émis par le lieu origine ( $i$ ) à destination du lieu  $j$ ,  $A_i$  est le total des flux émis par  $i$  et  $B_j$  le total des flux reçus par  $j$ .

Dans cette équation exponentielle négative,  $b$  est le paramètre à estimer<sup>2</sup>. Cette formulation du modèle gravitaire permet une meilleure prise en compte de certains types de données que la formulation simple.

#### *b. Centre et dispersion d'un semis de points*

Le résumé d'un semis spatial de points peut consister en l'ajustement d'une loi de probabilité bi-dimensionnelle (par exemple une loi de Gauss). Elle peut aussi passer par une carte qui en présente des résumés : centre et courbes d'iso-dépassement par rapport à celui-ci. Le centre d'un semis de points (par exemple des villes) munies d'une masse (par exemple leur population) se définit comme le point tel que sa distance aux autres points (pondérés par leur masse) soit minimale. Si la distance utilisée est la distance euclidienne, ce point central en sera le centre de gravité (tel que  $\sum_j (W_j D_j)$  soit minimale avec  $W$  les poids et  $D$  les distances). En revanche, si l'on utilise la distance de Manhattan, le point central en sera le point médian (tel que  $\sum_j W_j |D_j|$  soit minimal). Enfin, dans le cas d'un réseau urbain, il sera le lieu minimisant les déplacements des habitants de toutes les villes du réseau.

On peut aussi déterminer l'éloignement, absolu et relatif, de toute ville par rapport à ce point central (Charre, Dumolard, 1989) et tracer des courbes de même éloignement. C'est ce que vont illustrer les deux applications suivantes.

La première considère l'ensemble des agglomérations françaises de plus de 100 000 habitants en 1990. Le procédé de calcul procède, en effet, de la façon suivante : une grille est superposée au territoire de la France métropolitaine, les agglomérations munies de leur poids démographique y sont localisées. De toutes les cases de la grille à toutes ces agglomérations, on calcule la somme des distances pondérées par leur poids démographique. Le point médian est la case présentant la distance pondérée minimale

---

<sup>2</sup> On trouvera de plus amples explications sur le modèle gravitaire dans l'ouvrage de D. Pumain et Th. Saint-Julien (2001) sur les interactions spatiales.

et les courbes d'iso-dépassement sont déterminées en considérant de quel pourcentage chaque case a une distance supérieure au point médian.

Le point médian de ce réseau urbain français en 1990 est très voisin de Paris, ce qui signifie que, même sans tenir compte des réseaux ferré et routier, Paris est le lieu central de la France urbaine, probablement surtout à cause de son poids démographique relativement exorbitant.

Une courbe d'iso-dépassement traduit un dépassement, par rapport au point médian, de  $x\%$  du coût de déplacement des urbains français. Les courbes d'iso-dépassement en Ile de France (2%, 5%, 10%) sont plus resserrées vers le Nord (à cause de la masse démographique des villes de cette région) tandis que la courbe de 50% sur la figure 8.4 limite à peu près une France centrale excentrée et dessine une large « périphérie ».



Figure 8.4: Centre et dispersion du réseau de villes françaises de plus de 100 000 habitants, (source J.Charre et P.Dumolard)

L'expérience peut être renouvelée en considérant les villes de plus de 500 000 habitants dans l'Europe des 12 de 1990 plus la Suisse (Figure 8.5) : le centre de L'EUROPE OCCIDENTALE est situé en France du Nord Est. Il s'agit très précisément du centre de la mégapole européenne, mise en évidence par R. Brunet (1989) bien que les auréoles successives n'en dessinent pas la forme, popularisée sous le nom de « banane bleue ».



Figure 8.5: Centre et dispersion du semis de villes (> 500 000 habitants) de 13 pays ouest européens, (source J.Charre et P.Dumolard)

### 3. Un problème : la modification des unités spatiales

L'analyse spatiale rencontre plusieurs problèmes intrinsèques au procédé de l'utilisation de données statistiques d'une part et du recours aux données géoréférencées d'autre part. Un problème particulier plus récemment mis en évidence est connu sous le nom d'« unités spatiales modifiables », *Modifiable Areal Unit Problem* en anglais, en abrégé MAUP (Fotheringham, Openshaw 1981, Grasland, 1996, Openshaw, 1983). Les variables peuvent en effet être connues dans des zonages différents, d'extension différente, ce qui interfère avec les calculs et la cartographie et peut ensuite conduire à des interprétations erronées.

Or, plus les zones sont de grande taille et différents, plus cela perturbe l'analyse puisqu'on raisonne sur des **agréats** importants et différents qu'on suppose, chacun, homogène. Cette hypothèse néglige bien évidemment les variances (hétérogénéités) internes. Le problème est en fait multiple : l'agrégation de données en zones pose un problème d'**échelle** (plus la zone est vaste, plus elle a de chances d'être hétérogène ce qui ôte sens aux sommes ou moyennes) et un problème de **limites** de ces zones. Pour une variable donnée, l'agrégation spatiale de valeurs individuelles en zones peut en



effet donner lieu à une infinité de zonages différents, dont les sommes et les valeurs seront fort différentes. Il y a donc une *variation* de solutions possibles (et convenables) qui fait qu'on appelle les unités « modifiables ». « The areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating » qu'on peut traduire ainsi : « Les unités surfaciques (objets aréaux) utilisés dans nombre d'études géographiques sont arbitraires, modifiables et sujettes aux lubies ou aux fantaisies de qui fait, ou a jadis fait, le zonage » (Openshaw, 1984, 3).

Une solution possible à ce problème peut, par exemple, être trouvée chez C. Grasland:  
<http://www.parisgeo.cnrs.fr/cg/hyperc/wp3/wp3.htm>

Si l'on traite, en outre, de **zonages incompatibles** d'une variable à l'autre, le problème est encore accru. Deux grands types de situations peuvent exister : l'incompatibilité spatiale et l'incompatibilité temporelle des zones étudiées. Il serait, par exemple, utile de pouvoir visualiser autour des grands axes de circulation le nombre d'habitants soumis à une nuisance sonore excessive. Mais, la population résidente n'est connue qu'à un niveau agrégé dans des limites administratives n'ayant rien à voir avec la diffusion du bruit en zone bâtie urbaine. Les deux zonages de connaissance sont incompatibles. Pour tenter d'estimer les populations résidentes gênées, il faudra **désagréger** les données démographiques :

- soit en affectant un nombre d'habitants proportionnel à la surface de chaque zone,
- soit en se servant d'une connaissance logiquement corrélée (le nombre et la hauteur des bâtiments observés par photo-interprétation, une base de données sur l'occupation du bâti, etc.).

Il faudra ensuite **agréger** les nombres d'habitants estimés par immeuble dans les limites des zones noires sonores.

Un autre cas d'incompatibilité constitue la modification de limites au cours du temps. La modification, par exemple d'une agglomération, peut se faire par ajout, au fil du temps, de communes entières. Comme la connaissance de base existe au niveau communal, gérer les données relatives à l'agglomération se fera par agrégation de données

communales. Le géographe se trouvera face à une arborescence progressivement modifiée d'unités primaires. S'il veut établir les taux de croissance démographique successifs de l'agglomération « à géométrie fixe », il devra reconstituer les données du passé dans les limites actuelles.

Les modifications de limites d'unités spatiales peuvent parfois même être volontaires, par exemple en matière de découpages électoraux. En jouer pour maximiser le nombre de députés d'une certaine « couleur » politique (« *gerrymandering* » en anglais) est une pratique occasionnelle (risquée en cas de modification de l'opinion publique).

### III. Quelques modèles pour la simulation et l'aide à la décision

Le but de la modélisation est ici tout différent : plutôt que de décrire sous forme générale une distribution spatiale, il s'agit maintenant de tenter d'anticiper une évolution pour éclairer des décisions territoriales. Vu la complexité des processus spatio-temporels, il ne peut s'agir que de phénomènes limités, à condition encore qu'ils montrent une tendance affirmée et des variations point trop aléatoires. Si cette condition est remplie, le modèle joue alors le rôle d'instrument d'**expérimentation**.

Expérimenter consiste ici à **simuler** les conséquences logiques de différentes hypothèses. Par exemple, pour tenter d'évaluer l'évolution de la pyramide des âges d'une région donnée, il faut faire des hypothèses sur l'évolution de la fécondité, de la mortalité et du bilan migratoire de sa population au cours des x prochaines années. Simuler consiste alors à calculer **itérativement** (jour par jour, semaine par semaine, ...) le nombre de naissances, de décès et le nombre de personnes passant dans la tranche d'âge supérieure.

#### 1. Vieillissement potentiel d'une région

La démographie, du moins pour ce qui est de la croissance naturelle des populations, se prête assez bien à la prospective car une partie de celle-ci dépend des structures d'âge, qui évoluent lentement et continûment (hors catastrophe). Mais, comme les migrations varient de façon beaucoup moins prévisible dans l'espace et le temps, on se cantonne

généralement à faire évoluer une structure d'âge en fonction d'hypothèses sur l'évolution de la mortalité par âge et sexe et de la fécondité par âge des femmes entre 15 et 49 ans. C'est ce que nous allons illustrer à travers l'exemple du « vieillissement potentiel » du département du Doubs (Pepe, 1987). Le *vieillissement potentiel* d'une population peut être défini comme l'évolution du taux ( $>65\text{ans}/<20\text{ans}$ ), comme si ne jouait que l'évolution de sa pyramide des âges sous l'effet des seuls paramètres de son comportement démographique (fécondité par âge des femmes de 15 à 49 ans, mortalité par âge et sexe). On ne prévoit pas ainsi ce que sera une population dans x années (car cela dépend aussi du bilan migratoire par âge et sexe) mais quel potentiel de vieillissement de la population contient une structure d'âge actuelle.

Au recensement de population de 1982, le Doubs apparaissait comme un département stable en effectif et jeune (relativement à la France), mais avec des disparités internes notables et une pyramide des âges laissant présager un certain vieillissement. Il a donc été décidé de simuler l'évolution sur 20 ans des structures d'âge de ses zones internes à partir des hypothèses suivantes :

- pas d'apport migratoire (ou balance migratoire par âge et sexe nulle),
- maintien sur 20 ans de la fécondité par âge et de la mortalité par âge et sexe.

L'objectif était non pas de prévoir le nombre d'habitants en 2002, mais de montrer le potentiel de vieillissement présent dans les années 80 dans les structures d'âge de ses différentes zones. Le premier pas a donc consisté à définir des zones, homogènes par leur structure d'âge, spatialement continues et d'effectif suffisant. À partir de 17 unités urbaines et 26 parties rurales de canton (renseignées par leur pyramide d'âge quinquennal), une analyse factorielle des correspondances puis une classification automatique a permis de dégager 6 grands types de zones. Leur cartographie a défini 12 micro-régions démographiques dans le département. La simulation a été effectuée pour chacune d'entre elles, pour le Doubs, pour la Franche Comté et la France (afin d'avoir des éléments de comparaison). Sur les 20 ans de 1982 à 2002 ont été calculés, semaine par semaine :

- Le nombre de survivants,
- Le nombre de naissance,
- Le nombre de personnes passant dans la tranche d'âge supérieure.

Figure 8.6: Vieillissement potentiel du Doubs (1982-2002), (source: E. Pepe, 1987)

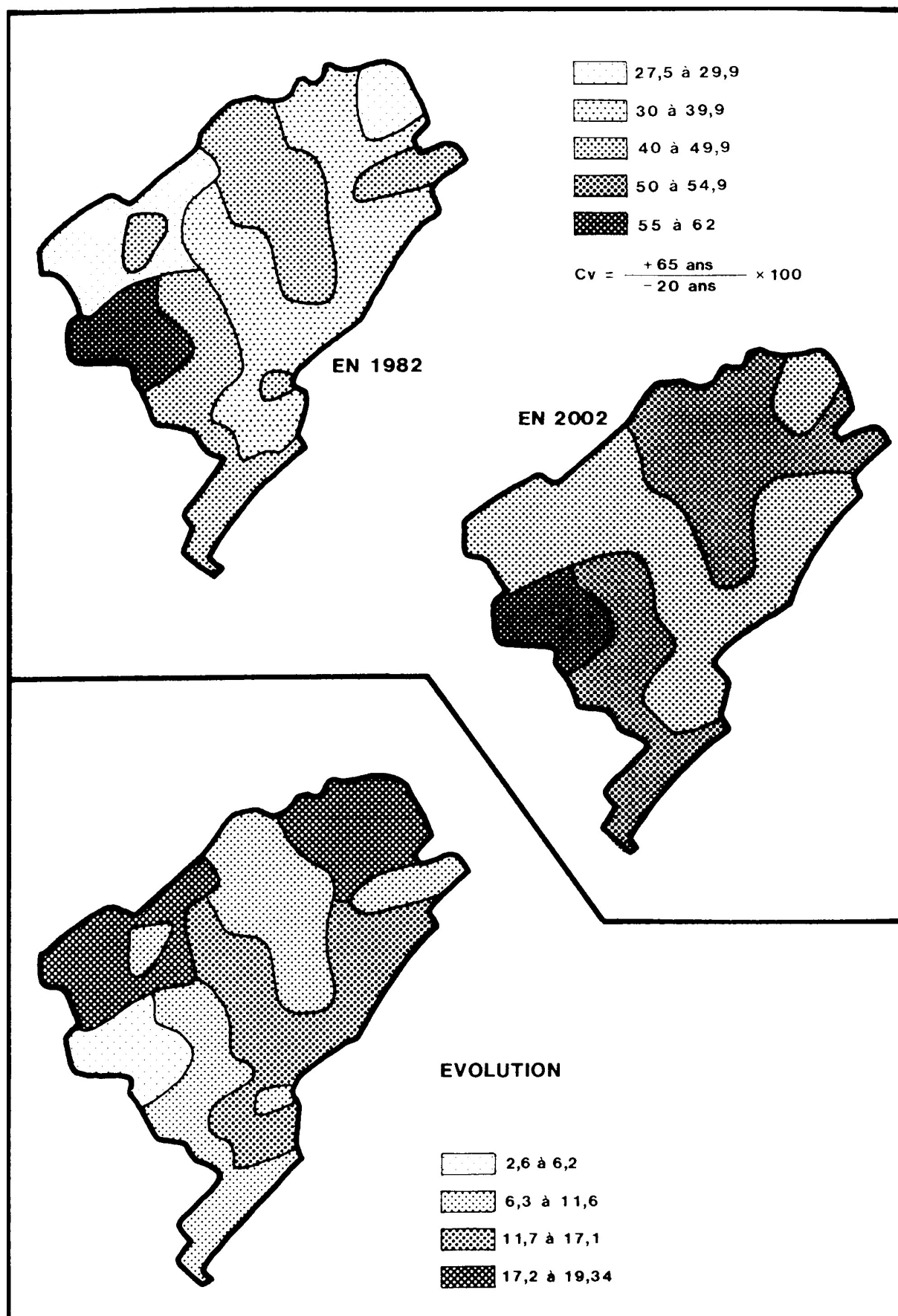


Figure 8.6 : Vieillissement dans le département du Doubs, 1982-2002 (source : E.Pepe, 1987)

Des récapitulations permettaient de calculer les taux annuels (natalité, mortalité, vieillissement, âge moyen des habitants) et de représenter les pyramides d'âge comparativement aux pyramides initiales. *In fine*, des récapitulatifs 1982-2002 ont montré le vieillissement potentiel des 12 micro-régions (*cf.* fig. 8.6). On peut montrer trois processus : d'abord, on assiste à une certaine homogénéisation de leurs structures d'âge dans le département. Ensuite, le vieillissement du Doubs est plus accusé que pour d'autres départements franc-comtois. Enfin, la natalité reste potentiellement supérieure à celle du reste de la France mais diminuée de 2 points par le seul effet de la structure d'âge (hors diminution de fécondité).

## **2. Anticiper l'évolution du trafic au tunnel du Mont Blanc**

Une étude sur le trafic de transit au tunnel du Mont Blanc a été entreprise au début des années 90, bien avant l'accident de mars 1999, la fermeture du tunnel et sa réouverture en 2002 (*cf.* Dumolard, Colomb, 1993). La raison en était l'augmentation rapide du trafic, qui a plus que triplé en moins de 30 ans : il est passé de 600 000 véhicules en 1966 à presque 2 millions en 1992. Ce trafic, de véhicules légers et de tourisme dans un premier temps, était devenu un transit massif de poids lourds entre l'Europe du Nord Ouest et l'Italie (avec les retombées imaginables en termes d'environnement, de qualité de vie et de tourisme). Ce transit massif et régulier tout au long de l'année se superposait à un trafic local important et aux pointes de l'activité touristique (hiver et surtout été). Cette évolution faisait envisager le percement d'un second tunnel. Il était donc pertinent de se poser deux types de questions :

À quoi attribuer cette rapide évolution ?

Pouvait-on anticiper l'évolution future ?

Tout d'abord, on peut rapporter cette rapide évolution à l'explosion des échanges (effectués principalement par la route) à l'intérieur de l'Union Européenne et à sa politique économique ultra-libérale de « laisser faire, laisser passer ». Cette explosion avait été amoindrie par l'ouverture du tunnel du Fréjus (quasi exclusivement passage de

transit poids lourds). Dans le même temps, elle a été augmentée par le **différentiel** de législation entre l'Union Européenne et la Suisse, où la réglementation sur le transit de poids lourds est sévère. En effet, les camions autorisés ne doivent pas y excéder 28 tonnes (au lieu de 42) ou emprunter le ferroutage, le gazole y est plus cher (comme l'ensemble des services annexes). La législation protectrice de la Suisse (excellente dans son principe si l'on veut réglementer davantage le transport routier) reporte une partie du transit poids lourds sur les passages autrichiens et français. Comment estimer ce report dans l'angle franco-italo-suisse des Alpes ?

Un premier modèle, purement économétrique, consiste à calculer les coûts de transit poids lourds sur un certain nombre d'itinéraires : le tableau 8.2 en fournit quelques exemples.

Itinéraires	Nb km	Durée	Coût	Coût/tonne
<b>Lyon-Milan</b>				
• via Mont Blanc	468	6h40	3052	73
• via Fréjus	463	6h00	2810	67
<b>Paris-Milan</b>				
• via Mont Blanc	888	11h00	4855	116
• via Fréjus	836	11h20	5006	119
<b>Anvers-Milan</b>				
• via Mont Blanc	1199	14h40	5942	141
• via Fréjus	1304	14h40	6072	145
• via St Bernard	1129	14h00	5115	183
• via Gothard route	1005	12h20	3991	143
• via Gothard ferroutage	993	15h00	6445	230
<b>Cologne-Milan</b>				
• via Mont Blanc	1095	14h00	5545	132
• via Fréjus	1185	15h00	5935	141
• via St Bernard	963	12h30	4525	162
• via Gothard route	879	11h00	3400	121
• via Gothard ferroutage	864	13h30	5830	208

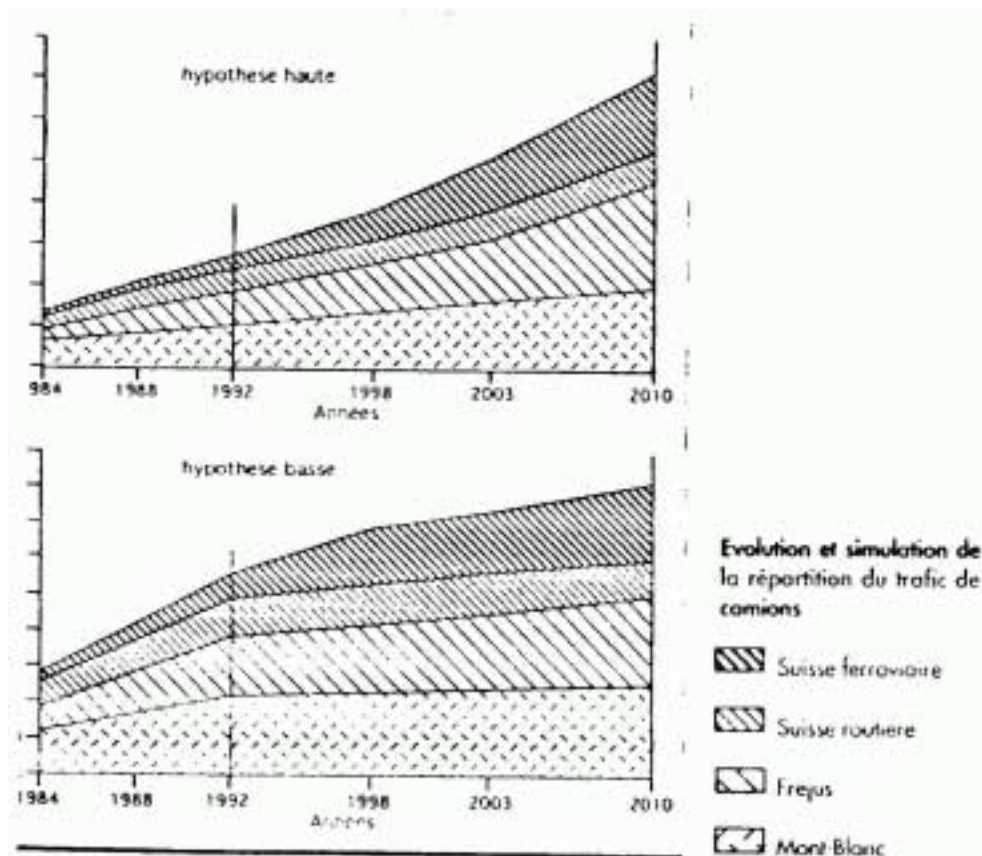
tableau 8.2: Coût à la tonne de quelques itinéraires trans-alpins (en 1992)(source : Dumolard, Colomb, 1993)

Les estimations du tableau 8.2 ont été calculées à l'aide du logiciel Autoroute Express et d'une base de données routières européenne : tous les coûts de transport routier de marchandises ont été comptabilisés et traduits en coûts à la tonne (divisés par 28 pour les trajets suisses, par 42 pour les autres puisque la législation suisse n'autorise pas le

transit routier des camions de plus de 28T au lieu de 42T pour l'Union Européenne). Ces coûts à la tonne ont été ensuite convertis en francs français.

En termes purement économétriques (sans le différentiel réglementaire), la Suisse devrait capter 53% du trafic de transit poids lourds (32% en réalité) et la France 47% (68% en réalité). À cause de ce différentiel, la Suisse s'en épargne presque 40% et la France en récupère 45% de plus. Il y a là détournement de nuisances !

Une seconde question pertinente portait sur l'évolution possible, à l'horizon 2010, des flux de poids lourds en transit au tunnel du Mont Blanc. Les simulations ont consisté à construire des scénarios pour les flux locaux et régionaux d'une part, internationaux d'autre part. Pour chacun, des facteurs clés d'évolution ont été déterminés, ont donné lieu à des hypothèses haute et basse et ont été combinés. La figure 8.7 présente les deux évolutions extrêmes, qui constituent une fourchette de prévision.



*Figure 8.7: simulation d'évolution du trafic P.L. au tunnel du Mont Blanc(source: Dumolard, Colomb, 1993)*



La prospective pour ce qui concerne le trafic local (vallée de Chamonix) et régional (vallée de l'Arve, Genevois) s'appuie essentiellement sur des évolutions démographiques probables (toutes inférieures à celles du passé) et fournit des évolutions de trafic comprises entre 15% et 33%. Les scénarios pour le transit international sont plus difficiles à établir puisqu'ils dépendent du maintien (partiel ou total) du différentiel réglementaire, de l'évolution des échanges (en masse et origine/destination) entre Europe du N.O. et Italie, de la montée en puissance des autres passages alpins (Fréjus, par exemple). La fourchette va de 2600 à 4200 camions/jour, contre 2100 en 1992. Une simulation effectuée par le CETE de Lyon donnait un résultat intermédiaire (CETE Lyon, 1991, document de travail).

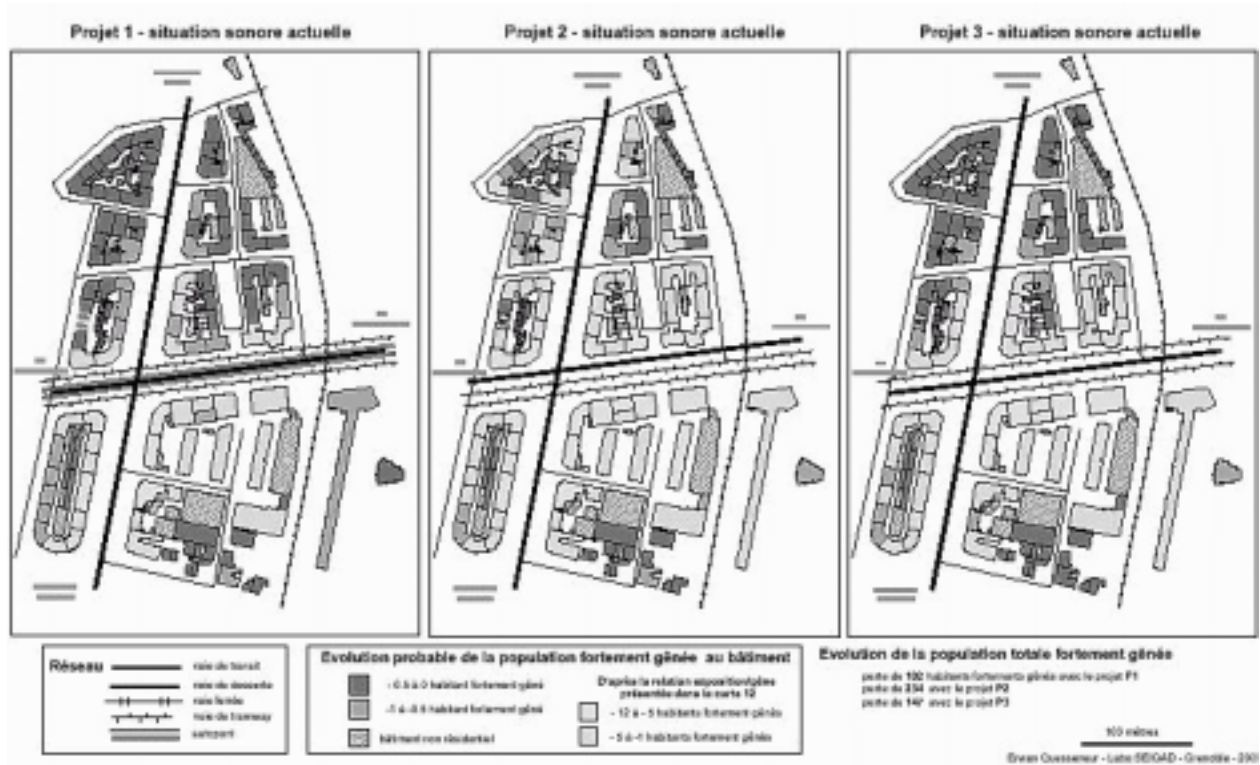
L'intérêt de ces simulations est de montrer la difficulté de la prospective quand jouent des variables pesant lourd et dont le comportement peut changer du tout au tout en fonction de décisions politiques (comme le différentiel de législation).

### **3. Effet d'un P.D.U. sur les nuisances sonores**

Les nuisances sonores constituent la principale source de gêne citée par les interviewés lors des enquêtes d'opinion. Les Plans de Déplacements Urbains (P.D.U.) ont pour objectif de favoriser des modes de communication plus respectueux de l'environnement que la circulation en automobile individuelle. Limiteront-ils les nuisances sonores ? Dans quelle proportion ?

Répondre, même approximativement à cette question, implique de disposer d'un modèle de diffusion spatiale du bruit, phénomène complexe résultant de facteurs nombreux tels la caractérisation des sources (usine, rail, route, trafic) et les conditions de diffusion (topographie, bâti en 3D, utilisation du sol, mesures anti-bruit, météo). Le logiciel MITHRA, créé au Centre Scientifique et Technique du Bâtiment, simule cette diffusion spatiale (à condition de disposer des données nécessaires). Il a permis de tester trois options d'aménagement du secteur le plus bruyant (et le plus pollué) de la ville de Grenoble (« carrefour de l'estacade »), pour dégager la meilleure du point de vue acoustique (Quesseveur, 2001). Il s'agissait de l'installation de voies de tramway. La première option installe des voies de tramway latéralement, la deuxième les installe

d'un seul côté, la troisième supprime, en outre, l'autopont passant sur l'avenue. La figure 8.8 montre que toutes les trois options améliorent la situation sonore, mais avec une répartition spatiale différente des diminutions du nombre d'habitants gênés.



*Figure 8.8: Baisse des gênes acoustiques (3 options testées)(source: E. Quesseveur, 2001)*

#### 4. L'I.A. : Intelligence Artificielle ou Informatique Avancée ?

Depuis une vingtaine d'années, se développent les simulations par ce que l'on nomme habituellement « Intelligence Artificielle ». À la différence d'autres modélisations de type statistique ou mathématique, il s'agit de modélisations informatiques de nature fondamentalement **qualitative** puisqu'elles tentent de doter les logiciels de capacités de déduction ou d'induction (de raisonnement, donc).

##### *a. Petit panorama de l'I.A.*

Nous n'avons présenté ici que des modèles simples (ou des formes simples de modèles), qui puissent inspirer des étudiants de géographie sans expertise particulière

en mathématique et/ou en informatique. Ce parti pris nous a prémuni de développer des applications fondées sur l'« Intelligence Artificielle » (Intelligence est à prendre ici au sens anglais du terme c'est-à-dire « acquisition d'information », « adaptation à une situation »). L'I.A. recouvre pourtant un ensemble de méthodes convenant particulièrement bien pour l'aide à la décision territoriale.

Sans entrer dans les détails, nous allons cependant en citer quelques-unes :

- L'utilisation, en géographie, des **systèmes à base de connaissances** (dits aussi systèmes « experts »), illustrée par les travaux de l'équipe de M. Guigo (1995),
- L'usage géographique de **réseaux neuronaux** par D. Josselin (1995),
- L'emploi, dans notre discipline, de modèles de type **automates cellulaires** par le modèle SIMURB d'A. Langlois et M. Phipps (1997). Ce ne sont là que quelques références permettant aux plus curieux de pénétrer, via des applications spatiales, le monde de l'I.A..

L'« Intelligence Artificielle » est une branche récente de l'informatique qui dote les logiciels de capacités de raisonnement, soit par déduction (systèmes expert) soit par induction (automates cellulaires, systèmes multi-agents) soit encore en mimant la sélection naturelle de type darwinien (algorithmes génétiques).

#### *b. Modélisation dynamique par système multi-agents*

Nous développerons ici une application géographique des systèmes multi-agents. Comme les réseaux neuronaux et les automates cellulaires, les systèmes multi-agents (S.M.A.) participent de l'**Intelligence Artificielle Distribuée** où un comportement « intelligent » du système émerge progressivement des interactions entre ses parties. Un **agent**, au sens informatique du terme, est un automate capable d'agir et d'interagir (avec d'autres agents, avec un utilisateur). Chaque agent, simple ou complexe, réalise une tâche parcellaire dans un **système** où la prise en compte de la complexité et de la dynamique provient des **interactions**. L'émergence d'un comportement collectif adapté

en est le produit. Les S.M.A. sont donc un des moyens de mettre en œuvre une conception systémique du monde.

Les composants d'un modèle S.M.A. sont :

- Un « environnement », dans notre cas un territoire différencié par les facteurs influençant le comportement des agents,
- Des agents, « objets » dotés d'une certaine autonomie et de capacités (réflexion ou réflexe) et, souvent, d'une « démographie » métaphorique (apparition, multiplication, disparition),
- Des actions des agents dans et sur l'environnement,
- Des communications et interactions entre agents.

Les spécialistes distinguent souvent S.M.A. cognitifs (peu d'agents, doués de raisonnement), assez proches des systèmes experts et S.M.A. réactifs (nombreux agents aux capacités plus élémentaires). Ces deux types ont été utilisés en agronomie et géographie, lors du colloque SMAGET (Cemagref, *Actes du colloque SMAGET*, LISC Clermont Ferrand, 1995 et site <http://www.lisc.clermont.cemagref>).

On peut, par exemple, citer AGRIPA (Touret, 1995), modèle de détermination de courbes isochrones à partir d'un lieu central.

Il comporte les éléments suivants :

Un espace différencié décrit, en mode rasteur, par son réseau routier (routes, autoroute, échangeurs, ponts), ses zones bâties (plus ou moins densément), l'occupation naturelle ou agricole du sol, les obstacles au mouvement (réseau hydrographique, falaises). Chaque type d'objet spatial est doté d'une vitesse de transit.

Des agents dotés de quelques capacités simples.

- de type « démographique » : ils naissent, procréent et meurent.
- de type « cinématique » : ils cherchent à se déplacer « au moindre coût ».
- de type « interaction » avec les autres agents (qu'ils peuvent tuer) et avec l'environnement (ils inscrivent sur chaque pixel traversé le coût cumulé de leur déplacement depuis le lieu central).

L'algorithme procède de la manière suivante :

- un agent naît au centre du lieu central et commence à se déplacer en suivant un chemin de moindre coût (celui autorisant la plus grande vitesse de transit)
- et inscrit sur les pixels empruntés le coût cumulé de son déplacement.
- A chaque déplacement, il donne naissance à un fils qui hérite les capacités de son père et son coût cumulé de déplacement.
- En phase intermédiaire, les agents se multiplient par scissiparité et de plus en plus de pixels sont marqués d'un coût d'accès (« optimal » car l'agent qui arrive sur un pixel déjà marqué meurt si le coût cumulé inscrit est inférieur au sien).
- L'algorithme s'arrête lorsque tous les pixels sont marqués et qu'il ne reste plus qu'un seul agent (tous les autres sont morts).

Ce modèle peut être amélioré (Dumolard, 1999), au moins conceptuellement, de diverses manières : en distinguant déplacements à courte et longue portée spatiale (leurs logiques ne sont pas les mêmes), en considérant plusieurs lieux centraux, en intégrant alors les idées fondatrices du modèle gravitaire, en dotant les agents de capacités évoluant au long de leur « cycle de vie », etc.

La figure 8.9 illustre, à partir du modèle AGRIPA, quelques simulations pour des configurations particulièrement simples : un seul centre et espace homogène, un seul centre au carrefour de 2 routes, un seul centre avec 2 percées dans un obstacle, concurrence d'une route et d'une autoroute reliant deux centres.

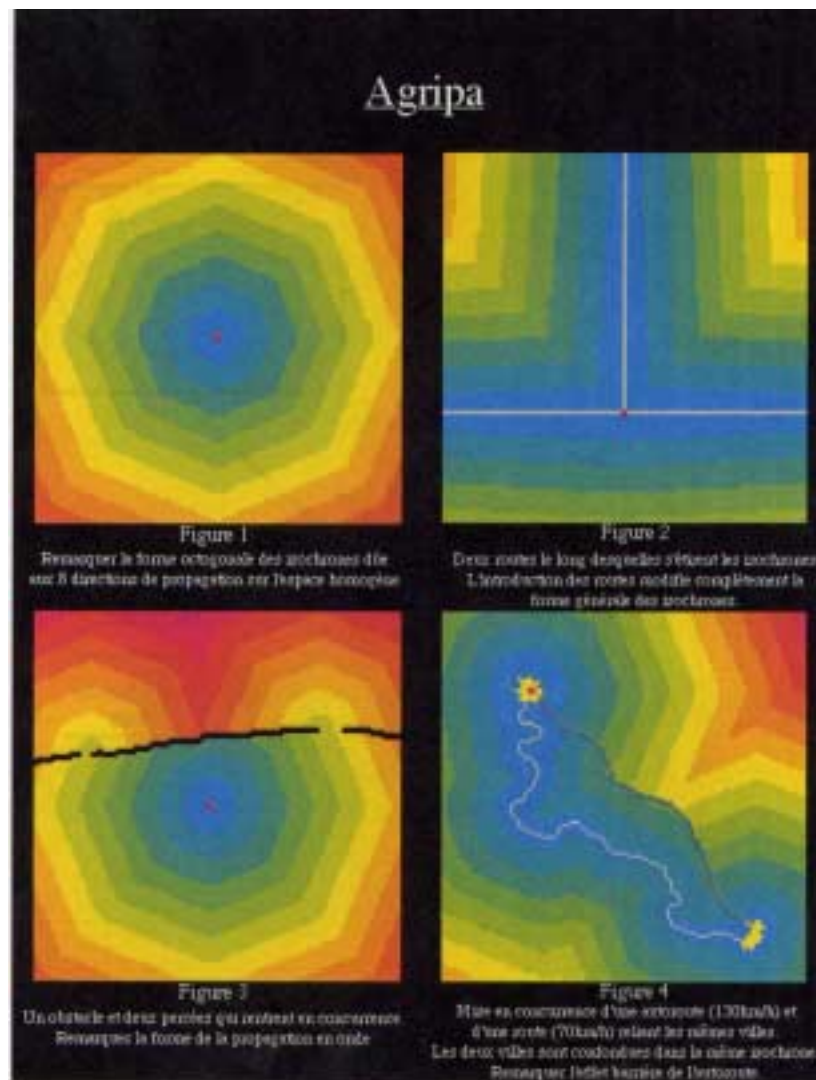


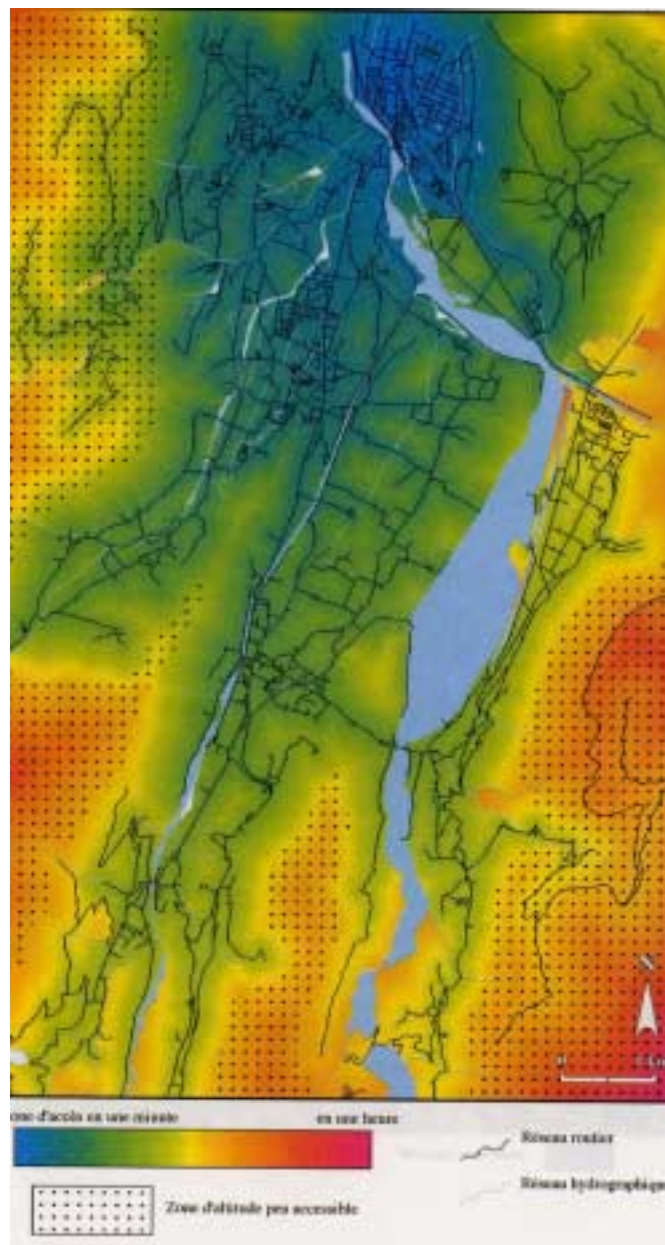
Figure 8.9. Modèle AGRIPA appliqué à des cas d'école (source: A.Touret, 1995)

On peut aussi présenter ici une application du modèle AGRIPA à une situation réelle (Vigneron, 1995).

Au début des années 90, existait le projet de compléter l'autoroute Marseille – Dijon (doublant à l'est l'autoroute Rhône – Saône, saturée) par le tronçon manquant (A51, Sisteron – Grenoble), le plus difficile du point de vue relief, géologie et faiblesse des densités démographique et économique. Un projet concurrent à l'autoroute envisageait de transformer en 2x2 voies la Nationale 75 ou la Nationale 85 (comme cela existe en Bretagne ou dans le sud du Massif Central). La zone la plus délicate à franchir par chacune des 3 options envisageables (A51, doublement de la RN75 ou de la RN85) était le sud immédiat de l'agglomération grenobloise sur une quinzaine de km. On a pu

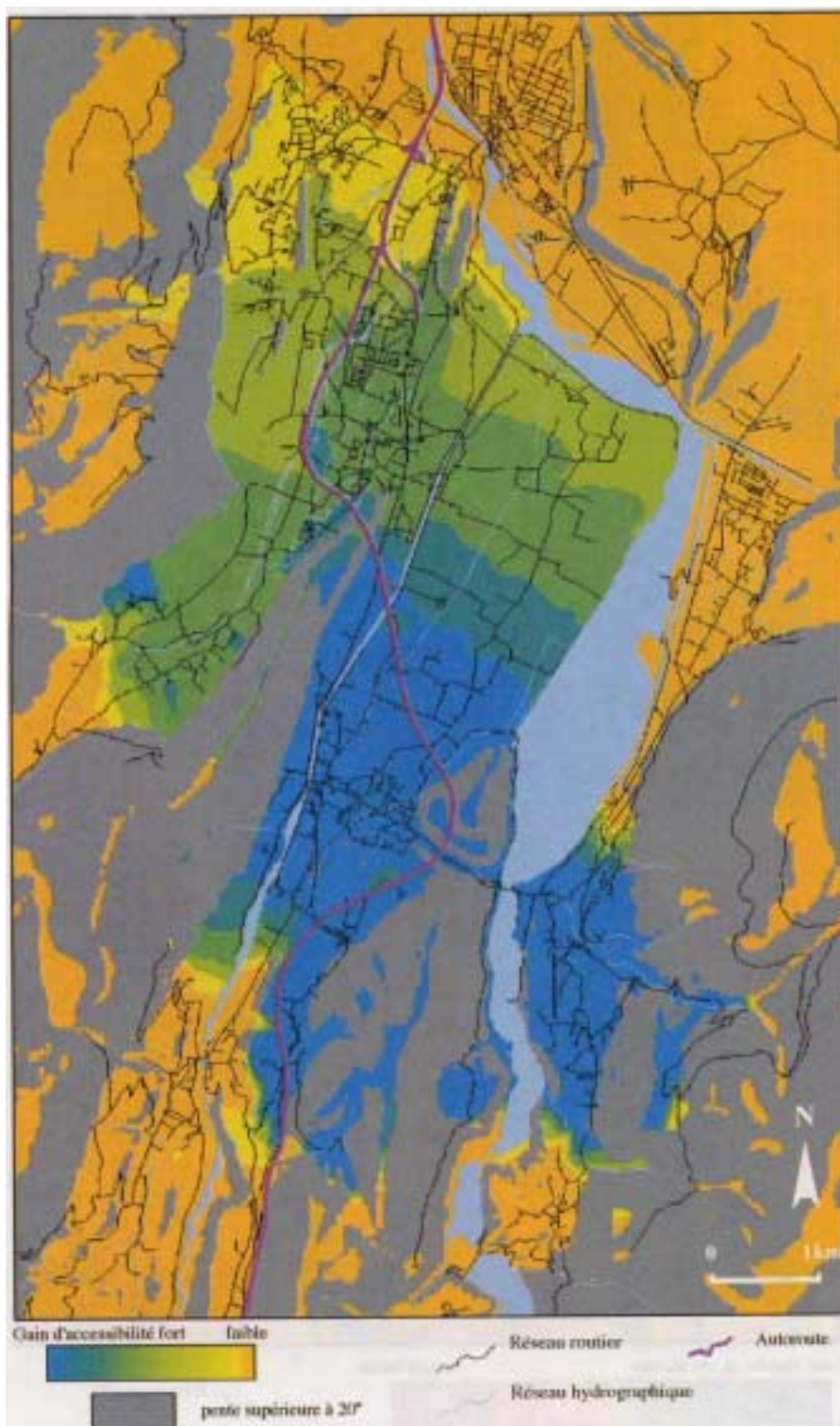


notamment tester, selon chacun des 3 scénarios, la localisation des gains d'accessibilité et le « risque » induit de pression foncière et de péri-urbanisation. La Figure 8.10 présente l'accessibilité calculée par AGRIPA pour l'année de début de simulation .



*Figure 8.10 : Accessibilité au sud de l'agglomération grenobloise en 1976 (calculée par le modèle AGRIPA), Source : Vigneron, 1995*

La Figure 8.11 présente les différences d'accessibilité routière entre 1976 et 2005 dans l'hypothèse de construction du premier tronçon (aujourd'hui existant) de l'autoroute A51.



*Figure 8.11 : simulation de la différence d'accessibilité au sud de l'agglomération grenobloise entre 1976 et 2005, (source : D.Josselin, V.Vigneron, 1995)*

Pour aller plus loin : (<http://www.lisc.clermont.cemagref/>).



## Références citées :

- (1) P.Haggett : *L'analyse spatiale en géographie humaine*  
D.Pumain, Th. Saint Julien : *L'analyse spatiale*  
Tome 1 : *Localisations dans l'espace* (A.Colin, 1997)  
Tome 2 : *Interactions spatiales* (A.Colin, 2001)
- (2) Chadule : *Pratiques statistiques en Géographie* (Masson, 1997, 4<sup>ième</sup> édition)
- (3) J.Charre, P.Dumolard : *Pratiques Informatiques en Géographie, le logiciel Infogéo* (Masson, 1989)
- (4) E.Pepe : *Le vieillissement potentiel dans le département du Doubs* (mémoire de maîtrise, Université de Franche Comté, 1987)
- (5) P.Dumolard, L.Colomb : *L'évolution du trafic au tunnel du Mont Blanc* (Revue de Géographie Alpine, 1993 n°4)
- (6) E.Quesseveur : *Traitement spatial des impacts du bruit des transports terrestres* (Thèse, Université Grenoble 1, 2001)
- (7) M.Guigo et al : *Gestion de l'environnement et systèmes experts* (Masson, 1995)
- (8) A.Langlois, M.Phipps : *Automates cellulaires, application à la simulation urbaine* (Hermes-Scence, 1997)
- (9) D.Josselin : *La déprise agricole en zone de montagne, vers un outil d'aide à la modélisation spatiale* (Thèse, Université Grenoble 1, 1995)
- (10) A.Touret : *AGRIPA, un modèle de calcul de courbes isochrones fondé sur un système multi- agents* (Revue Internationale de Géomatique, 1995 n°3-4)  
N.Ferrand : *Modèles multi-agents pour l'aide à la décision et à la négociation en aménagement du territoire* (Thèse, Université Grenoble 1, 1997)  
*Actes du Colloque SMAGET* (Cemagref Clermont Ferrand, 1998)
- ( 11) P.Dumolard : *Accessibilité et diffusion spatiale* (l'Espace Géographique, 1999 n°3)
- ( 12) V.Vigneron : *Accessibilité routière au sud de l'agglomération grenobloise en liaison avec la construction de l'autoroute A51* (Revue Internationale de Géomatique, 1995 n°3-4)

## **Pour conclure**

Les applications géo-informatiques que nous avons présentées ne recouvrent pas l'ensemble des pratiques de l'informatique par des géographes / aménageurs, pour plusieurs raisons.

- Nous avons volontairement exclu ce qui ne leur est pas propre, comme les outils bureautiques (traitements de textes, tableurs, logiciels de présentation, ...) dont l'usage (bien que devenu universel) est utile mais non nécessaire.
- Bien qu'elles soient devenues de pratique assez courante, nous avons fort peu insisté sur des applications statistiques. Leur mise en oeuvre informatique n'est qu'une simple traduction, sans grande plus value informatique : l'essentiel des difficultés à affronter et de la connaissance à mobiliser est statistique et il existe, pour ce faire, de bons logiciels, de manipulation abordable.
- Nous avons fort peu insisté sur les méthodes ressortant de l' »Intelligence Artificielle » parce qu'il en existe encore assez peu d'applications dans nos disciplines et qu'elles sont généralement de mise en oeuvre assez difficile, impliquant le plus souvent une connaissance informatique avancée et la programmation dans des langages assez complexes.
- Il n'est cependant peut être pas trop tard pour présenter aux esprits les plus curieux un bref panorama de ces méthodes et leur donner envie « d'aller y voir d'un peu plus près ».
  - Les systèmes à base de connaissances (ou systèmes expert) sont parmi les plus anciennement utilisées par certaines équipes de chercheurs. Pour plus de détails, on se reportera à l'ouvrage de M.Guigo et al (Masson, 1995).
  - Les réseaux neuronaux ont une utilité voisine de celle de l'analyse statistique des données mais en s'affranchissant d'un certain nombre de limitations (ajustements linéaires, par exemple). Ils sont de manipulation assez complexe et n'ont encore été que peu employés.
  - La famille des techniques de l' »Intelligence Artificielle » distribuée l'a déjà été davantage car elle correspond assez bien à une approche systémique dynamique des faits spatiaux. Les automates cellulaires ont donné lieu à quelques applications emblématiques, explicitées par exemple dans l'ouvrage de A. Langlois et M. Phipps (Hermès, 1998). Les systèmes multi-agents bénéficient déjà, parmi les chercheurs du territorial, d'une certaine renommée, raison pour laquelle nous y avons fait allusion dans le chapitre 8. En outre, leur mise en oeuvre peut être amorcée à l'aide d'un logiciel téléchargeable gratuitement sur Internet (« STARLOGO »).
- Nous n'avons pas davantage évoqué ce que peuvent apporter les algorithmes génétiques, les méthodes de la géostatistique (variographie, krigeage) ni le traitement d'image à l'aide des opérateurs de la morphologie mathématique.

La liste de ce que nous n'avons pas présenté pourrait encore être allongée, tant l'évolution de l'informatique est rapide. En somme, nous avons concentré l'attention sur ce qui est devenu courant ... ou est en train de le devenir.

Il reste donc encore beaucoup à explorer pour les esprits curieux et inventifs !

## **Glossaire**

### **Algèbre de cartes**

Tentative, due à Dana Tomlin (1990), pour définir des opérateurs cartographiques (en mode rasteur), regroupés en 4 types :

- local : fonctions mathématiques appliquées à un pixel d'une ou plusieurs couches
- focal : fonctions sur le voisinage de chaque pixel d'une seule couche
- zonal : fonctions sur des ensembles de pixels de couches différentes
- incrémental : pour dériver par exemple le plus court chemin entre 2 pixels.

### **Algorithme**

Séquence d'opérations visant à la résolution d'un problème.

### **Analyse spatiale**

Ensemble de modèles permettant de créer, résumer, analyser des distributions et processus spatiaux.

### **Anamorphose cartographique**

Transformation géométrique visant à cartographier des positions topographiques et des formes non pas réelles mais proportionnelles aux valeurs localisées d'une variable. On pourrait qualifier l'anamorphose cartographique de « projection plane » sémantique.

### **Attribut**

Employé dans le domaine des bases de données, ce terme équivaut à ceux de descripteur ou de variable. Le codage d'un attribut peut être numérique ou alphabétique.

### **Autocorrélation spatiale**

Corrélation des valeurs des unités spatiales avec celles de leurs voisines.

### **Base de données (BD)**

Ensemble d'informations structurées dans une table (BD élémentaire) ou plusieurs (BD relationnelle). Une BD doit être conçue pour être aisément mise à jour et consultée.

### **Chorèmes**

Éléments de base d'un langage graphique (comme les phonèmes sont les éléments de base d'un langage parlé), surtout adaptés à la représentation de dynamiques spatiales.

## **Courbe de Peano**

Courbe fractale, donc apte à parcourir tous les points d'un plan ou d'un volume. Pourrait être utilisée dans les S.I.G. raster pour accélérer l'accès à l'information.

## **Échantillonnage spatial**

Ensemble de méthodes pour construire des échantillons spatialement représentatifs, permettant par exemple de généraliser des observations à tout un espace.

## **Entité**

Objet élémentaire (nommé « individu » en statistique). Dans une BD, l'ensemble des entités de même nature forme une classe d'entités et donne lieu à une table.

## **Fiabilité cartographique**

Probabilité qu'à chaque localisation d'une carte d'être affectée de sa vraie valeur (joue notamment en cas d'interpolation spatiale).

## **Fichier**

Ensemble plus ou moins structuré de données, stocké sur un support informatique (disque, bande magnétique, etc.). L'arrangement des données à l'intérieur du fichier répond à un format imposé par le programme qui les utilise. Ces données peuvent être de nature quelconque (numérique, texte, image, son) mais doivent être activables par un type de fichier particulier, un programme, qui a pour caractéristique d'être exécutable sur ces données.

## **Fouille de données (*Data mining*)**

Elle a pour but d'extraire automatiquement de la connaissance de gigantesques entrepôts de données (*data warehouses*). Elle combine pour cela différentes méthodes. Le data mining spatial fait, quant à lui, un fort usage de techniques géomatiques, de l'analyse exploratoire des données et de l'analyse spatiale.

## **Géodésie, réseau géodésique**

La géodésie est la science qui a pour but de modéliser la forme de la terre (ellipsoïde) et de décrire le champ de la pesanteur terrestre (géoïde). Elle fournit donc le fondement des systèmes de positionnement. Un réseau géodésique est ainsi constitué d'un ensemble de bornes de coordonnées connues.

## **Géo-informatique**

Ensemble des connaissances et moyens de mise en oeuvre informatiques utiles à la géographie et à l'aménagement du territoire. La géomatique n'en est qu'une partie.

## **Géomatique**

Ensemble pluridisciplinaire de connaissances pour la mise en oeuvre de Systèmes d'Information Géographique dans des organisations (collectivités territoriales, entreprises, services techniques de l'état). La géomatique comprend :

- la production de données localisées (photogrammétrie, GPS, image satellitale, géodésie, projections planes),
- leur implémentation au sein de systèmes d'information,
- les types de combinaison de données qu'ils permettent,
- leur insertion dans les services et la circulation de données qu'ils induisent, modifiant leur fonctionnement.

## **GPS (*Global Positioning System*)**

Système de positionnement fondé sur l'existence de 24 satellites. Leurs orbites sont telles que chaque portion de la terre en aient toujours au moins 3 dans son « ciel ». Le GPS sert à localiser chaque position terrestre avec une précision de l'ordre de 30m.

## **Hypertexte**

Navigation simple dans l'information à l'aide de liens (individualisés par une couleur ou un figuré différent) sur lesquels on clique pour atteindre le contenu désiré. Si outre du texte, l'information contient des images et/ou du son, on ne parle plus de liens hypertexte mais hypermedia.

## **Image de synthèse**

Image informatique artificiellement construite et non enregistrée à partir d'une scène réelle. La construction de telles images met en jeu 3 stades :

- la description géométrique des objets d'une scène,
- leur habillage par une texture,
- l'éclairement de la scène.

La synthèse d'images est une des techniques de base de la « réalité virtuelle ».

## **Image satellitale**

Image (et non photographie) prise depuis un satellite et envoyée à une station terrestre sous forme raster.

## **Intelligence artificielle**

De l'anglais mal traduit (intelligence y signifie information) et compris différemment :

- soit simulation par ordinateur de processus de la pensée, exemple : systèmes expert,
- soit résolution distribuée (entre ordinateurs, entre agents logiciels) de problèmes, exemples : réseaux neuronaux, automates cellulaires, systèmes multi-agents

## **Internet**

Réseau informatique interconnectant des millions d'ordinateurs dans le monde qui peuvent ainsi échanger des informations en temps réel.

## **Logiciel**

Terme désignant un ensemble cohérent et fonctionnel de programmes. Employé de façon générique, désigne (par opposition au matériel) l'ensemble de tous les programmes.

## **Modèle**

Représentation épurée et généralisée d'une réalité. On a distingué dans ce manuel :

- les modèles généraux, visant à expliciter un mécanisme d'organisation de l'espace (exemple : modèle gravitaire),
- les modèles opérationnels, visant à simuler (pour expérimenter) sur une représentation d'une réalité limitée (exemple : modèle de diffusion d'un incendie de forêt),
- les modèles de modèles, méthodes et techniques permettant de construire des modèles.

## **Modèle conceptuel de données**

Représentation graphique codifiée de la structuration des tables dans une base de données (exemple : modèle Entités / Associations).

## **Modèle numérique de terrain**

Un modèle numérique de terrain (MNT en abrégé) est la couche d'information d'un S.I.G. représentant l'altitude du sol d'une zone donnée. Il peut se présenter sous forme :

- raster : altitude moyenne de chaque pixel (de taille 50\*50 mètres au sol dans la BD Alti de l'IGN),

- vecteur : courbes de niveau et points cotés (comme sur une carte topographique) ou réseau des triangles joignant les points cotés les plus voisins créés par triangulation de Delaunay. On dispose alors de ce qu'en anglais on appelle « *Triangular Irregular Network* », TIN en abrégé).

### **Modeleur**

Logiciel utilisé en image de synthèse pour représenter informatiquement des objets tridimensionnels.

### **Modification des unités spatiales**

Cette modification peut provenir de l'ajout, au fil du temps, de zones de niveau inférieur (ex. : une agglomération). Elle peut aussi exister si l'on veut combiner des variables connues dans des zonages incompatibles. Dans les deux cas, il faut désagréger l'information pour la ré-agréger ensuite. En anglais, MAUP, il désigne le problème qu'il n'y a pas une seule solution, mais plusieurs voire une infinité de solutions et de cartes possibles lors des processus d'agrégation en zones et de cartographie de données.

### **Moyenne mobile spatiale**

La valeur de chaque unité spatiale est remplacée par une moyenne des valeurs d'elle-même et de ses voisines : c'est donc un procédé de lissage des valeurs.

### **Multimédia**

Ensemble des matériels et des techniques permettant de gérer tout à la fois du texte, du son, de l'image, de la vidéo.

### **Opérateur**

Symbole qui, dans une expression mathématique, permettent d'effectuer des opérations (arithmétiques, logiques, autre) sur des données. Un exemple peut en être les opérateurs relationnels des bases de données (union, intersection, jointure, etc. de tables).

### **Paramètre**

Option d'un programme offrant à l'utilisateur le choix entre différents modes de traitement.

### **Photogrammetrie**

Procédé de mesure en 3 dimensions à partir de couples de photographies (aériennes, terrestres) ou d'images satellitaires. Ce procédé utilise, comme la photo-interprétation, la

vision stéréoscopique (en relief) d'images d'une même zone prises avec deux angles de vue différents.

### **Pixel (« PICTure Element »)**

Un des carrés élémentaires (de surface constante) constituant le « grain » d'une couche d'un S.I.G. raster. La surface au sol des pixels d'une image satellitale, par exemple, fournit sa résolution spatiale.

### **Projection plane**

Opération consistant à projeter sur un plan (de dimension 2) des coordonnées sphériques (de dimension 3). Cette opération génère nécessairement des déformations que les différents types de projection tentent de minimiser (préservant soit les angles soit les surfaces).

### **Quadtree (arbre « quartique »)**

Arborescence composée de 4 branches, elles mêmes divisées en 4, etc... itérativement jusqu'à une condition d'arrêt du découpage. Appliqué à des distributions spatiales, ce type de partitionnements successifs permet d'adapter la taille des pixels d'une image à la densité locale d'information.

### **Redressements (d'image satellitale)**

Avant de pouvoir être utilisée, une image satellitale doit subir 2 types de redressement :

- des redressements radiométriques pour corriger des différences d'illumination de la surface terrestre,
- des redressements géométriques pour rendre l'image compatible avec un référentiel cartographique.

### **Réflectance**

Radiation électro-magnétique émise ou réfléchie par la terre.

### **Requête**

Question adressée à une base de données à l'aide d'un langage (comme SQL) ou d'une interface graphique.

### **Résolutions d'un satellite**

Un satellite d'observation de la terre est caractérisé par 4 types de résolutions :

- résolution spatiale : surface au sol d'un pixel de l'image,



- résolution temporelle : intervalle de temps entre 2 passages successifs du satellite au dessus de la même zone,
- résolution spectrale : nombre de fenêtres de longueurs d'onde différentes générées par les capteurs du satellite (3 pour SPOT, 7 pour Landsat),
- résolution radiométrique : capacité des capteurs à différencier finement les intensités lumineuses à l'intérieur de chaque fenêtre (256 intensités pour SPOT et Landsat).

### **Sémiologie graphique**

Ensemble de règles (dues à J. Bertin) de traduction visuelle de variables statistiques en fonction de leur nature mathématique et de la dimension géométrique (ponctuelle, linéaire, surfacique) des objets à représenter.

### **Signature spectrale**

Chaque objet homogène de la surface terrestre a une signature spectrale propre (son « empreinte digitale »), due à sa capacité à réfléchir l'énergie solaire dans les différentes longueurs d'onde.

### **Simulation**

Procédé consistant à faire fonctionner un modèle avec différentes hypothèses.

### **SQL (*Structured Query Language*)**

Langage de requêtes à une base de données relationnelle.

### **Système d'information géographique (S.I.G.)**

Au sens strict du terme, un S.I.G. est un logiciel pour gérer des données cartographiques, ce qui réfère plutôt aux S.I.G. VECTEUR (intégrant système de gestion de bases de données et cartographeur). Les S.I.G. RASTEUR, quant à eux, logiciels de combinaison d'images, sont davantage utilisés pour l'analyse et la modélisation spatiales.

Par extension, les S.I.G. incluent les données nécessaires au fonctionnement du système et, pour certains auteurs, l'analyse des organisations dans lesquels ils fonctionnent.

### **Système multi-agents**

Système logiciel composé d'objets et d'agents. Le fonctionnement du système est essentiellement dû aux interactions entre agents et entre agents et objets. Un

algorithme a, dans cette optique, pour but de converger vers un état stable. On différencie généralement :

- SMA cognitifs à peu d'agents mais dotés de capacités importantes,
- SMA réactifs à nombreux agents dotés de capacités limitées, de type réflexe pavlovien.

### **Topologie**

Géométrie qualitative décrivant les propriétés structurelles d'objets déformables : contiguïté, voisinage, inclusion, chevauchement, etc.

### **Variables visuelles** (de J.Bertin)

Trames graphiques destinées à traduire visuellement la nature statistique d'une variable et la forme (ponctuelle, linéaire, surfacique) des unités cartographiques.

### **En complément, on pourra consulter :**

- |                                      |  |
|--------------------------------------|--|
| - pour les définitions informatiques | <a href="http://www.linux-France.org/prj/jargonf/">http://www.linux-France.org/prj/jargonf/</a>  |
| - pour les définitions géomatiques   | <a href="http://www.seig.ensg.ign.fr/">http://www.seig.ensg.ign.fr/</a><br><a href="http://www.geomatique.georezo.net/">http://www.geomatique.georezo.net/</a> |
| - pour l'analyse spatiale            | <a href="http://www.cybergeopresse.fr/">http://www.cybergeopresse.fr/</a>  |