

# Les modèles numériques

Regis Caloz\*      Dominique Schneuwly†

21 juin 2010

---

\*EPFL regis.caloz@epfl.ch, Overall

†UNIFR dominique.schneuwly@unifr.ch, Specials

## Table des matières

<b>Table des matières</b>	<b>2</b>
<b>Les modèles numériques</b>	<b>4</b>
Introduction à la modélisation numérique . . . . .	5
Résolution . . . . .	6
Les modèles spatiaux . . . . .	8
La notion de couche . . . . .	10
Les concepts du modèle vecteur . . . . .	11
La géométrie et ses propriétés . . . . .	12
Principes de numérisation de la géométrie . . . . .	14
Les primitives géométriques . . . . .	16
Objets complexes . . . . .	17
La topologie . . . . .	18
Structure non topologique (spaghetti) . . . . .	19
Structure topologique . . . . .	20
Comparaison . . . . .	21
Numérisation . . . . .	22
Comment nettoyer un jeu de données et constituer des objets spatiaux ? . . . . .	23
Les systèmes à topologie . . . . .	25
Les concepts du modèle raster . . . . .	26
Signification de l'information affectée à une maille . . . . .	27
Affectation . . . . .	28
Structure numérique . . . . .	30
Formats de structure numérique . . . . .	31
Quelques aspects propres à la structure image . . . . .	33
Propriétés comparatives des modèles vecteur et raster . . . . .	35
Tableau des aptitudes des modèles objet et image . . . . .	36
Combinaison des diverses technologies dans un SIG . . . . .	37
Modélisation du relief . . . . .	38

Le relief . . . . .	39
Les modèles numériques . . . . .	40
Exemple d'élaboration d'un modèle numérique . . . . .	43

## Les modèles numériques

La modélisation de l'espace géographique a été abordée jusqu'à présent sur un plan théorique, indépendant d'applications particulières. Le résultat est un découpage de l'espace en unités spatiales régulières ou irrégulières auxquelles sont rattachés des attributs décrivant leurs diverses propriétés. La modélisation s'est voulue également indépendante de la technologie, guidée seulement par la préoccupation de concevoir un modèle reflétant le plus fidèlement la réalité à observer et de disposer de moyens pour l'exploiter. Cette indépendance est cependant limitée par le contexte actuel des développements de l'informatique. En arrière pensée, celle-ci est présente et oriente la modélisation dans le sens de rendre compatible le modèle avec la logique informatique. Le risque serait de faire plier un modèle sous des contraintes qui n'ont rien à voir avec la réalité et qui le détourneraient de son but. L'outil prendrait l'ascendant sur le contenu ! Pour les SIG, compte tenu de la modélisation qui a été proposée, l'informatique n'a pas cette influence restrictive ; elle oriente simplement les choix de manière à trouver un compromis optimal entre la configuration du modèle et les moyens de le traiter.

## Introduction à la modélisation numérique

La première condition que doit remplir la phase de numérisation est la fidélité au modèle de la réalité. Cette contrainte posée, l'objectif est de disposer de moyens d'analyse et de traitement beaucoup plus puissants que ceux offerts par la cartographie conventionnelle. Ces moyens sont à la portée de l'informatique. En effet, elle met à disposition diverses fonctionnalités indispensables pour couvrir l'ensemble des besoins, notamment la structuration et le stockage des informations, le traitement et l'analyse numérique, la reproduction graphique, la mise en réseau pour la diffusion et l'échange d'informations. En définitive, la numérisation en modèle objet concerne, d'une part, la géométrie des unités spatiales et les attributs qui leur sont attachés ou les attributs d'entités non-spatiales. L'opération n'est pas indépendante du point de vue du sens accordé au modèle de la réalité, mais relève pour les 2 aspects de logique et de technique différentes. L'information géométrique couvre la représentation graphique, caractéristique de la forme et de la taille et, dans un système plus performant, les relations spatiales. Ensuite, les attributs décrivant les propriétés thématiques sont numérisés sous forme de tables. En modèle image, la géométrie se réduisant à la résolution et, le cas échéant, à la forme de la maille, la numérisation concerne seulement les attributs thématiques. Ainsi, une base de données géographique se caractérise par un double contenu : géométrique et thématique.

## Résolution

Tout comme une carte a une échelle et des spécifications, l'information dans une BD géographique a une résolution pour chacune de ses dimensions : spatiale, thématique et temporelle. L'office fédéral de topographie fournit des produits numériques à différentes résolutions : par exemple, VECTOR25 fournit la couverture nationale correspondant à une échelle au 1 :25'000 . VECTOR200 est le modèle du paysage de la Suisse équivalent à l'échelle 1 :200'000. Vous trouverez un aperçu de leurs résolutions dans l'animation qui suit.

### Résolution spatiale

La résolution spatiale est la résolution d'acquisition des données (ou la taille du plus petit objet représentable dans une surface). Dans la dimension spatiale, le processus de modélisation repose sur l'analyse de la géométrie des phénomènes spatiaux. En mode objet, la résolution spatiale peut être définie par 3 seuils de variables spatiales : la taille minimale que doit voir un objet pour être représenté (superficie pour un objet zonal, longueur pour un objet linéaire), la distance minimale de séparabilité entre 2 objets, la taille minimale que doit avoir un détail d'un objet pour être perçu (ou granularité). En mode image, la résolution spatiale est définie par le pas de la grille.

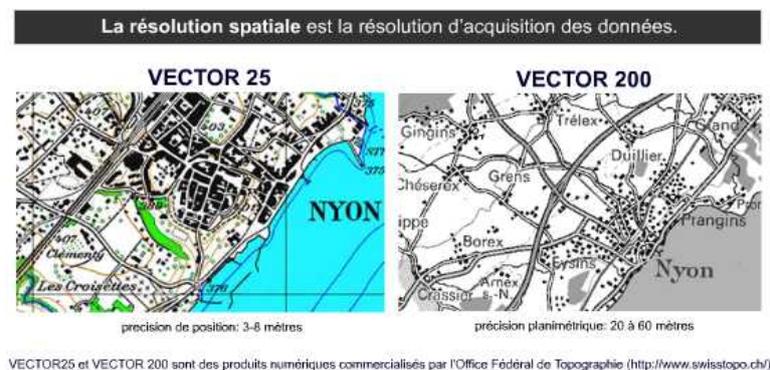


FIG. 1: resolution-spatiale.jpg

### Résolution thématique

La résolution thématique (aussi appelée résolution sémantique) correspond à la précision de la classification (hiérarchies de classes et domaines de valeurs hiérarchisés). Elle est inhérente au schéma conceptuel des données (Vangenot 2001). Il s'agit de définir le niveau de détail souhaité. Par exemple, une classe

d'arbres est elle suffisante dans le modèle ou l'information plus détaillée sur le type d'arbres (conifères, feuillus...) est-elle pertinente? Auquel cas plusieurs classes d'arbres sont nécessaires. La résolution définit l'information élémentaire.

**La résolution sémantique correspond à la précision de la classification**

	<b>VECTOR 25</b>	<b>VECTOR 200</b>
<b>Catégories de surfaces primaires</b>	Pépinière Rocher Rivière Buissons Eboulis avec buissons Eboulis sur glacier Eboulis Eboulis en forêt Eboulis en forêt clairsemée Glacier Piste sur herbe Piste sur revêtement dur Gravière Glaisière	Verger Vignes Lac Zone d'habitation Digue de retenue Barrage Carrière Marais et buissons Marais Marais en forêt Marais en forêt clairsemée Forêt Forêt clairsemée
<b>Réseau routier</b>	Autoroute Autoroute, chaussées séparées Semi-autoroute Autoroute - Entrée / Sortie Accès d'autoroute Route de 1re classe Route de 2e classe Route de 3e classe Route de 4e classe Route de 5e classe Route de 6e classe Route de quartier Route / chemin historique Piste pour chars d'assaut Chemin de parc Pont isolé Pont couvert isolé Passerelle isolée	Sens de circulation sur l'autoroute Autoroute Bac pour voitures Autoroute 2ème classe Chemin de fer de montagne et téléphérique Route principale (voie de transit d'une largeur de 4m) Route principale (voie de transit d'une largeur de 6m) Petite route Chemin réservé aux piétons Voie musée Route secondaire d'une largeur de 3m Route secondaire d'une largeur de 6m Voie à écartement normal Voie étroite Route principale (route de liaison d'une largeur de 4m) Route principale (route de liaison d'une largeur de 6m) Accès

VECTOR25 et VECTOR 200 sont des produits numériques commercialisés par l'Office Fédéral de Topographie (<http://www.swisstopo.ch/>)

FIG. 2: resolution-thematique.jpg

### Résolution temporelle

Tout comme une carte a une fréquence de mise à jour, une base de données spatiales a une résolution temporelle qui correspond à la périodicité de mise à jour de ces données. Par exemple, la périodicité des mises à jour de la base de données VECTOR25 de l'OFT est de 6 ans.

## Les modèles spatiaux

La modélisation numérique est guidée par les 2 formes de partition de l'espace en unités irrégulières et régulières. Les 2 modèles sont : le modèle à structure objet (ou vecteur) et le modèle à structure image (ou raster).

### Le modèle objet

La structure objet traduit la discrétisation de l'espace en éléments géométriques conformes aux modèles de l'espace ainsi que leurs relations de voisinage. Les attributs thématiques sont numérisés sous forme de table en base de données.

### Le modèle image

La structure image reproduit la composante géométrique par une grille à maille régulière et regroupe les informations de même catégorie thématique en une seule couche, sous forme de matrice. À chaque maille de la grille est donc affecté un seul attribut. Il y a autant de grilles ou de couches qu'il y a de thématiques.

#### Attention :

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 3: animation-vecteur-raster.swf

La dénomination des 2 structures n'est pas homogène dans la littérature scientifique. A l'origine, le modèle de discrétisation irrégulière a été nommé structure vecteur en référence à la manière de décrire la géométrie de l'objet. Aujourd'hui, de nombreux auteurs inclinent vers la désignation de structure objet plus fidèle aux processus de modélisation et à son contenu. Cette dénomination conduit malheureusement à quelques ambiguïtés avec le langage des informaticiens qui ont développé, de leur côté, le concept de base de données orientée objet. Les concepts ne s'opposent pas fondamentalement ; chacun, cependant, repose sur des notions propres aux disciplines qui les ont conçus. Il en est de même pour la dénomination de la structure image. La littérature fait état également de mode maille ou matriciel et surtout de mode "raster. Ces appellations ne soulèvent aucune querelle linguistique. En ce qui nous concerne, nous avons adopté la dénomination structure image parce que l'organisation des informations est identique à celle des images numériques produites notamment par les radiomètres imageurs à balayage.

De surcroît, les traitements numériques associés sont très similaires du point de vue des algorithmes mis en oeuvre.

### La notion de couche

Pour la cartographie conventionnelle, la couche est le plan graphique représentant les lignes et symboles devant être imprimés avec la même couleur indépendamment de leur signification thématique. Les cartes topographiques sont constituées de plusieurs couches, autant qu'il y a de couleurs imprimées. Par exemple, la couche bleue contient le réseau hydrographique, les courbes de niveau pour les glaciers et les milieux lacustres ainsi que les lignes électriques à haute tension. Pour les SIG, la couche n'est plus liée à la reproduction graphique mais représente les unités d'observation d'une même thématique censées être reproduites sur un même plan graphique. C'est l'assertion la plus courante. Elle fait partie de la structuration des informations par catégorie thématique. A noter que sous cette forme, la représentation des objets n'est pas individualisée. On ne peut représenter seulement une sélection d'objets faisant partie d'une même couche. Au contraire, les logiciels orientés objets considèrent individuellement chaque objet. La notion de couche n'est donc plus présente explicitement. Elle se réduit à la notion de catégories d'objets (les bâtiments, les tronçons de routes principales, les segments de rivière, etc.). D'un point de vue strictement géométrique, la couche correspond au plan graphique contenant l'ensemble des primitives géométriques partageant les mêmes propriétés topologiques de réseau ou de surface [?] . En mode image, la couche correspond à une thématique ou sous-thématique et, du point de vue informatique, à un fichier (type de sol, forêts, routes secondaires, etc.), c'est la grille accueillant les informations relatives à un thème. La couche est donc une primitive catégorielle qui offre de nombreux avantages pour la réalisation de cartes thématiques ou lors d'analyse spatiale ne faisant intervenir qu'une partie des informations contenues dans la base. Les logiciels exploitent ce concept dans la notion de Espace de travail réunissant les catégories d'objets, donc les couches, exploitées lors d'une analyse spatiale ou dans la production de cartes thématiques.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 4: Intro.layers.swf

**Les concepts du modèle vecteur**

Quelles sont les informations que nous devons numériser du point de vue géométrique ?

Temps estimatif : 2 heures

### La géométrie et ses propriétés

Du point de vue géométrique, il s'agit de décrire en langage numérique des unités de forme irrégulière caractérisées par des relations spatiales. La géométrie est décrite par des points composant le pourtour de chaque unité, qu'elle soit ponctuelle, linéaire ou zonale. Les points (appelés points intermédiaires) sont reliés par des segments de droite ou par tout autre ligne défini mathématiquement (splines, courbes de Bezier, etc.). On distingue encore diverses nuances parmi les catégories d'unités. Une unité linéaire est dotée ou non de direction selon qu'elle fait partie d'un réseau modélisant un flux (eau, gaz, électricité, trafic) ou non. Une unité zonale est appelée simple ou complexe si elle est constituée de une ou de plusieurs unités spatiales (p.e. un pays possédant des enclaves). Pour numériser, on admet par hypothèse que l'espace de numérisation (plan graphique qui accueille les informations géométriques) est continu et conforme à une projection cartographique et géoréférencée. Les logiciels de numérisation offrent généralement un catalogue des projections connues, de sorte que l'utilisateur est à même de choisir celle qui lui convient. La marche à suivre pour procéder à la géoréférence de la base géométrique est examinée dans le module B-DC.

D'un point de vue sémantique, les attributs sont numérisés dans des tables liées aux unités géométriques. La numérisation des attributs se fait généralement par introduction au clavier. Parfois, les informations sont importées d'une autre base. Se posent dans ce cas des difficultés de compatibilité de format entre la source et la base réceptrice. Une table se définit comme la fiche type informatique correspondant à une catégorie d'entités sur laquelle figurent les attributs type qui la caractérisent. Pour chaque unité, on remplit une fiche où chaque attribut reçoit l'information propre à l'unité à laquelle il se rattache. La forme de présentation que nous adoptons intentionnellement ici est proche de celle des bases de données étudiées dans le module B-DM. Elle paraît, pour le propos actuel, suffisamment explicite. Il sera formulé plus tard le sens précis des symboles utilisés. La table est dite primaire si elle est directement liée par son identificateur à l'objet graphique. C'est elle qui contient les informations descriptives de l'entité. Elle est dite indirecte si elle complète l'information exprimée par un attribut de la table primaire. Dans la figure ci-dessous, une table indirecte pourrait par exemple mentionner les syndicats des communes figurant dans la première table. Il peut avoir plusieurs tables indirectes liées les unes aux autres. On comprend que si leur nombre augmente, une structuration des tables devient nécessaire. C'est le rôle des structures de bases de données, notamment relationnelles, présentées dans le module B-DM.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les  
flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en  
partie par des images d'écran.

FIG. 5: Graphic\_unit\_and\_table.swf

### Principes de numérisation de la géométrie

Quelles sont les informations que nous devons numériser du point de vue géométrique ?

- Les unités spatiales, le point, la ligne, la zone, de manière à pouvoir les reproduire graphiquement et à leur attacher des informations thématiques.
- Leurs relations spatiales tels que le voisinage, l'adjacence, l'inclusion, etc., c'est-à-dire la topologie, de manière à offrir la plus grande capacité pour les analyses spatiales sollicitées lors de l'exploitation de la base.

Ces souhaits, cependant, ne sont pas réalisés de manière satisfaisante sans un souci majeur vis-à-vis de la cohérence et de l'intégrité de la base de données géométrique. En effet, plusieurs situations sont susceptibles de se produire, notamment :

- Les objets adjacents possèdent par définition une frontière commune. Il est souhaitable de ne la numériser qu'une seule fois. Une double ligne risquerait de ne pas se superposer parfaitement. Elle est, d'une part, esthétiquement peu satisfaisante ; de surcroît, lors d'une modification éventuelle de la frontière deux interventions seraient nécessaires ; d'autre part, elle autorise la mise en relation de la topologie entre les différentes couches.
- Une ligne représente parfois une frontière commune pour plusieurs entités. Par exemple, lorsqu'une rivière fait frontière entre deux unités administratives et forme également une limite au même endroit pour une unité de végétation (forêt), quatre unités spatiales partagent ainsi une ligne commune. Comment organiser les informations pour ne numériser qu'une seule fois ce qui est partagé ?
- Une unité peut être totalement incluse dans une autre comme une île sur un lac ; ou encore, une zone est de type complexe et constituée par plusieurs unités géométriques sans frontières communes.
- Au cas où les entités spatiales de même catégorie sont censées couvrir tout l'espace, des frontières chevauchantes ne respecteraient pas l'intégrité des surfaces. La somme des surfaces des unités spatiales serait différente de la surface de l'espace géographique d'étude.

Toutes ces situations sont des risques encourus lors de la numérisation des unités géométriques. Ils sont réduits, voire supprimés, grâce à une structuration adéquate des informations et grâce à des modules de contrôle informatisés agissant lors de la saisie numérique.

En résumé, la géométrie de l'objet considère la forme de l'unité, sa taille et ses relations géométriques. Pour chacun de ces concepts, il est nécessaire d'insérer des informations qui leur sont propres. Selon le degré de perfectionnement, une base de données ne prend en compte que partiellement ces informations. On connaît ainsi deux catégories de bases de données se situant à deux niveaux de complexité : celle se limitant à la **structure**

**non-topologique** et l'autre incluant la **structure topologique** . Tout objet géométrique se laisse décomposer en quelques éléments qui constituent en définitive les briques avec lesquelles il est possible de le reconstruire : une ligne brisée, un polygone sont constitués de segments de droite, un segment se construit par une droite reliant deux points, etc. Si l'on maîtrise la numérisation des briques, celle des objets ne présente aucune difficulté ! Ces briques sont dénommées primitives géométriques. Elles sont évidemment différentes selon que la structure est topologique ou non. Les deux structures sont présentées parallèlement de manière à mettre bien en évidence leurs propriétés respectives.

### Les primitives géométriques

Les primitives géométriques se définissent comme les éléments géométriques à partir desquels les unités spatiales sont reconstituées dans leur forme, leur taille et dans leurs relations spatiales. Chacune des primitives est définie de manière rigoureuse et jouit de propriétés qui déterminent le contenu informatif de la BD-géométrie et, par voie de conséquence, la capacité d'analyse spatiale offerte.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashs etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 6: primitives.swf

### Objets complexes

L'agrégation de plusieurs primitives géométriques pour représenter une entité géographique complexe constitue un objet complexe. Par exemple, Le territoire français est une surface complexe : il est composé non seulement de la France métropolitaine, mais également de surfaces détachées que sont la Corse, la Guyane, la Guadeloupe, etc.

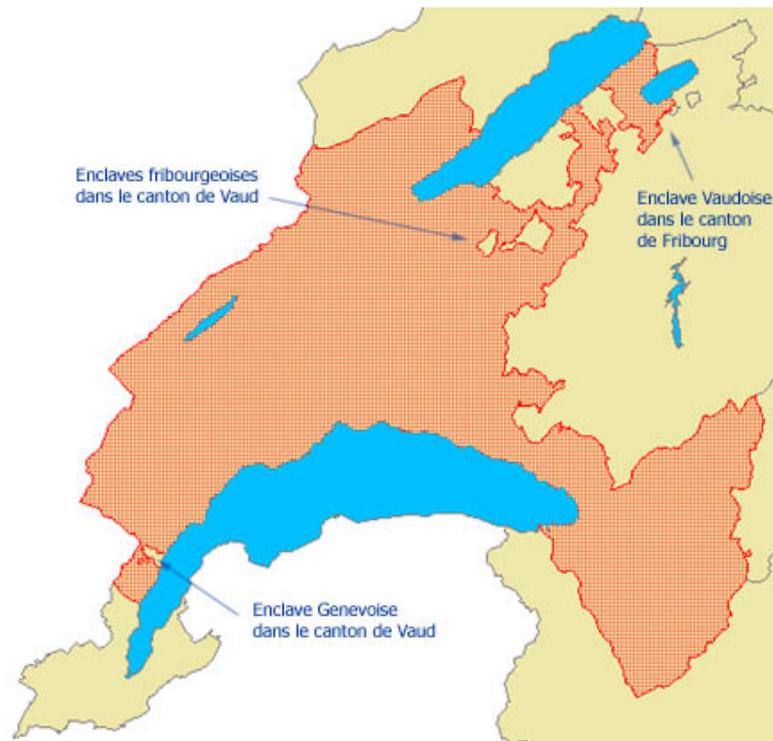


FIG. 7: surfaces-complex.jpg

## **La topologie**

Les données vecteur avec ou sans topologie ? Les systèmes à topologie ? Comment nettoyer un jeu de données ?

### Structure non topologique (spaghetti)

Un modèle non-topologique, dans sa version la plus simple, n'enregistre que les notions de points et de lignes fermées ou non qu'il stocke sous la forme d'un dictionnaire; il est communément appelé structure spaghetti. Seules la forme et la taille des unités spatiales sont identifiables. Les formes sont reproduites au moyen de points et de lignes de liaison droites ou curvilignes (arcs). Les informations à stocker sont réduites au minimum.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 8: structure-non-topo.swf

Les informations sont suffisantes pour produire une représentation graphique de l'unité spatiale. Les notions d'unité zonale ne sont pas représentées. Il est donc plus difficile d'identifier l'intérieur ou l'extérieur d'une zone délimitée par une ligne fermée, de la colorer comme il est souvent souhaité dans la confection de carte thématique. Notons cependant que cette opération peut tout de même être réalisée par calcul.

La structure spaghetti exige que toute ligne frontière soit dessinée autant de fois qu'elle forme une limite d'une unité géométrique. Il y a donc redondance dans le dessin. Cette structure, malgré ses faiblesses offre néanmoins des avantages de simplicité. Les informations de base étant réduites au minimum, elles se transfèrent facilement. Le format le plus répandu est le DXF. La topologie étant absente de cette structure, il n'est pas possible de travailler avec le concept de réseau. Cette situation est rédhitoire pour des applications traitant de distribution de flux dans ce réseau (d'eau, d'électricité ou de trafic routier). Cependant pour l'analyse spatiale, l'absence d'information topologique stockée peut être suppléée par des fonctions de calcul qui reconstituent les propriétés de voisinage. La version plus perfectionnée de cette structure ajoute la notion de polygones sous la forme d'un dictionnaire permettant de retirer des informations sur la taille des unités (surface et périmètre) et d'affecter des attributs graphiques tels qu'une trame ou une couleur. Outre les informations enregistrées dans la structure précédente, un fichier de polygones énumère les points caractéristiques de chacun d'eux. Un point pouvant appartenir à 2 polygones, cette structure autorise une seule ligne frontière entre 2 zones adjacentes. Cependant, lors de la reproduction graphique, les lignes frontières seront dessinées 2 fois. La qualité du dessin est parfois ainsi altérée si les 2 traits ne se superposent pas parfaitement.

## Structure topologique

Les bases de la structure topologique ont été posées par Peucker et Chrisman dans un article paru en 1975. Ces auteurs ont proposé un nouveau format qu'ils ont dénommés POLYVRT [?]

. Les structures exploitées actuellement n'ont pas été fondamentalement modifiées ; seule la manière de les présenter a gagné en simplicité. L'introduction tardive de la topologie dans les logiciels de traitements de données spatialisées est due à la faiblesse des premiers logiciels et des ordinateurs incapables de gérer la complexité des instructions séquentielles nécessaires à la description des notions de voisinage ou de le faire avec des temps de calcul acceptable. La structure topologique, en plus des informations répertoriées pour un modèle non-topologique, met en oeuvre les primitives de noeud et de chaîne.

### Attention :

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 9: structure-topo.swf

Notons que l'objet linéaire représenté dans l'animation n'intersecte pas le polygone 1 puisqu'aucun noeud n'apparaît sur les intersections de lignes. Ces 2 objets n'appartiennent probablement pas à la même catégorie thématique et ne sont pas situés sur la même couche au plan géométrique dans la réalité. Il sont dits non-planaires. Cette situation est fréquente, par exemple, dans la modélisation d'un réseau routier lorsqu'une route secondaire croise une autoroute par un pont. La notion de surface complexe ou de zone offre une grande liberté de description. Non seulement nous sommes en mesure de définir une zone formée de 2 ou plusieurs polygones, mais aussi il devient possible d'identifier une entité formée par l'enveloppe de plusieurs entités adjacentes. Un pays, par exemple, du point de vue géométrique est l'enveloppe des départements ou des provinces qui le composent. (topologie généralisée cf.) [?]

## Comparaison

Structure non-topologique	Structure topologique
<ul style="list-style-type: none"><li>- Structure simple</li><li>- Facilité de transfert de données (format DXF)</li><li>- Suffisante pour cartographie numérique</li><li>- Analyse spatiale très réduite</li><li>- Doublement des limites de polygones adjacents (esthétique, manque de rigueur de gestion!!)</li></ul>	<ul style="list-style-type: none"><li>- Structure complexe</li><li>- Relations spatiales introduites lors de la saisie</li><li>- Analyse spatiale performante</li><li>- Indispensable pour la gestion de réseau</li><li>- Difficulté de transfert de l'information topologique (incompatibilité des formats)</li><li>- Aucun dédoublement, ni d'information, ni de dessin, même pour des couches différentes</li></ul>

FIG. 10: comparaison.jpg

**Numérisation**

La procédure de numérisation s'opère en 2 temps. Lors de la première phase, on ne numérise que le dessin sans se soucier des relations spatiales. Puis de manière interactive l'opérateur inscrit les relations topologiques par la détermination des noeuds, chaînes et zones ou celles-ci sont automatiquement détectées par le logiciel. La justification principale et essentielle des modèles topologiques, c'est leur aptitude à raisonner sur l'espace sans avoir à revenir chaque fois aux coordonnées  $x,y$ . (Drouet). En définitive, il s'agit d'un choix. Voulons-nous investir du temps lors de la numérisation pour en gagner lors de l'exploitation de la base de données, notamment pour les analyses spatiales, ou acceptons-nous de supporter les temps de calcul chaque fois qu'une étude requiert la reconstruction de la topologie? Une structure topologique assure, cependant, l'intégrité de changements opérés sur des frontières communes. Dans le cas du parcellaire par exemple, tout déplacement d'une frontière modifie celle de la parcelle adjacente. La frontière étant décrite par une seule chaîne, une seule transformation suffit pour les deux parcelles adjacentes.

### Comment nettoyer un jeu de données et constituer des objets spatiaux ?

Lors de la numérisation ou de la récupération d'un jeu de données, les erreurs peuvent être fréquentes. Elles rendent impossibles la définition d'objets numériques (lignes, polygones, zones, ...). On distingue principalement :

- les erreurs géométriques : point manquant ou mal placé, arc manquant ou mal placé, type ou forme de courbe inadéquat, etc.
- les erreurs topologiques : lignes “pendantes (arcs dépassés ou non atteints), polygones non fermés, arcs dupliqués, centroïdes dupliqués, noeuds superflus, etc.

Afin de permettre la création d'objets spatiaux, ces erreurs doivent être évitées ou corrigées : Dans la mesure du possible, il faut éviter les erreurs lors de la digitalisation. Pour faciliter la tâche des opérateurs, on peut recourir :

- à des fonctions d'accrochage (snapping) (disponibles sur la plupart des systèmes graphiques)
- à des fonctions de sélection de polyligne (sur systèmes spécialisés)
- à des systèmes géoinformatiques “à topologie, qui ne tolèrent pas les ambiguïtés géométriques lors de la saisie

A défaut, la topologie doit être “créée a posteriori en éditant les données “à la main, sur la base d'un examen visuel détaillé et en appliquant un programme de “création de topologie (topology builder), qui identifie les ambiguïtés et les résout conformément à des paramètres (tolérances métriques) définies par l'utilisateur.

#### Attention :

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 11: nettoyage.swf

### Démarche de création de la topologie

1. Nettoyage des données : raccordements sur points proches ou lignes proches, suppressions de dépassements, d'éclats et de lignes dupliquées
2. Création d'un graphe planaire : création des noeuds manquants, et agrégation de polygones décrivant les arcs
3. Identification et agrégation d'objets : points, lignes (avec ou sans orientation), surfaces (simples ou complexes)

**Remarques opérationnelles**

La création de la topologie peut se révéler une opération très fastidieuse et coûteuse dans le cas de données sources très “sales”. Une nouvelle digitalisation peut parfois s’avérer plus avantageuse. Les fonctionnalités et les performances de création de la topologie varient considérablement d’un système géoinformatique à l’autre :

- MapInfo : assistance au nettoyage et à l’agrégation des données (module externe) ; pas de structures de données topologiques
- MGE : “topology builder a posteriori (base pour l’analyse spatiale)
- Vision, Argis, SpatialWare, ArcInfo : contraintes topologiques à la saisie et “topology builder a posteriori

Dans les programmes constructeurs de topologie ( *topology builders* ), les algorithmes représentent des heuristiques discutables, s’appuyant sur des valeurs de paramètres empiriques. Les topology builders assurent la cohérence interne des données, mais pas leur cohérence externe.

### Les systèmes à topologie

On attend d'un SIG qu'il soit capable d'agréger automatiquement des informations telles que la connectivité de conduites industrielles, l'existence d'un mur commun à deux bâtiments, la desserte d'une parcelle par un chemin, etc. 2 solutions sont possibles : - Disposer des fonctionnalités permettant de déduire en tout temps ces relations par calcul : l'agrégation des informations est effectuée lors de chaque utilisation. - Stocker explicitement dans la base de données chaque relation géométrique pertinente pour l'application considérée : l'agrégation des informations est alors effectuée lors de leur saisie. On parle alors de systèmes à topologie. Dans les applications de la topologie, on s'intéresse essentiellement à la modélisation explicite des relations de connectivité et d'adjacence. Sur la figure, sont représentés deux exemples : la connectivité entre objets à 1 dimension : nodalité et l'adjacence de 2 surfaces le long d'une même ligne.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 12: relations-explicites.swf

Dans un système à topologie, la topologie est gérée à la fois alphanumériquement et graphiquement :

- La relation conduite CONNECTEE A conduite est explicite
- La connexion entre les 2 arcs est calculée métriquement.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 13: systemes-a-topologies.swf

## Les concepts du modèle raster

Le modèle image provient d'une partition de l'espace géographique en unités spatiales régulières réparties par couche thématique (occupation du sol, altitude, routes, réseau hydrographique, stations de mesure hydrométéorologiques, etc.). L'ensemble forme une sorte de grille virtuelle régulière, de mailles généralement carrées et définies par la résolution spatiale, c'est-à-dire, la taille du côté du carré exprimé en mètre. La structure image est du fait de la partition régulière indépendante de la géométrie (hormis la résolution de la maille). La notion d'objet est absente puisque celui-ci n'intervient en aucune manière pour la détermination des unités d'observation. Le mode image est indissociable de la notion de couche. A chaque thématique ou sous-thématique correspond une grille qui accueille, sur chaque maille, une seule information. Il y a donc autant de grilles que de thématiques.

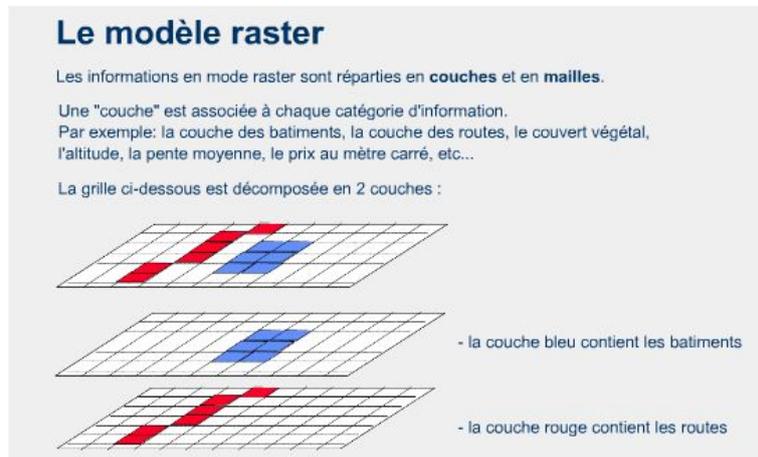


FIG. 14: raster.jpg

### Signification de l'information affectée à une maille

Chaque maille devient l'unité d'observation. La notion d'échelle a montré que la taille des mailles généralement appliquée est de l'ordre de 10 à 100 mètres. Une résolution trop fine engendre un volume et une diversité d'informations difficilement gérables. Une résolution trop grossière, au contraire, produit une généralisation qui perd très vite son sens opérationnel. Par ailleurs, on comprend très vite, sans analyse approfondie, qu'il est plus aisé de définir une altitude moyenne d'une maille que de décider sur la catégorie d'occupation du sol qu'il convient de lui affecter lorsque le milieu est hétérogène. Cette brève analyse met en lumière deux approches pour la description en modèle image de l'état d'une variable spatiale : la règle de dominance et l'échantillonnage statistique. Notons que la problématique est sensiblement similaire en modèle objet comme nous l'illustrerons plus loin par deux exemples.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 15: grid-affectation.swf

## **Affectation**

### **Selon la règle de dominance**

Selon la règle de dominance, on affecte à la maille soit une valeur moyenne si la variable est continue, soit la catégorie prédominante, la valeur nodale. La règle est logique ; elle affronte néanmoins des situations difficiles à gérer et fait disparaître des catégories qui, par l'espace qu'elles occupent, ne peuvent être dominantes, par exemple des routes, des chemins sous une résolution de 100m. Il est nécessaire parfois de subdiviser la thématique en sous-thématiques en créant ainsi une nouvelle couche. C'est la résolution thématique qui gouverne le nombre de couches ! La détermination de la catégorie dominante incombe à l'expert. Cette information est donc de nature interprétée, sujette de ce fait à la subjectivité de l'expert. Dans ce sens, le travail appliqué pour la maille est similaire à celui du botaniste ou du pédologue lorsque ceux-ci catégorisent un taxon. Ce procédé présente l'avantage que l'information produite est riche des connaissances et de l'expérience de l'interprète. Elle est localement au plus près de la réalité ... dominante ! Néanmoins, à moyen ou long termes, il présente un inconvénient souvent majeur. En définitive, il convient d'opérer les choix pertinents en référence à la finalité et à l'échelle considérée. Comment comparer objectivement des levés d'informations opérés par des experts différents et séparés d'une durée au cours de laquelle des sensibilités nouvelles ont pu se développer, comme c'est probablement le cas en ce qui concerne nos observations dans le domaine de l'environnement ? Lorsque la taille de la maille ou de l'unité spatiale est importante, l'hétérogénéité de l'occupation du sol contraint à des généralisations grossières en dépit de directives précises qui auraient pu être édictée. Il sera difficile, lors d'analyse de changements de faire la part des choses entre la subjectivité des experts et les modifications réellement survenues.

### **Selon un échantillonnage statistique**

L'inconvénient signalé est partiellement levé par un échantillonnage statistique. Au lieu de choisir la catégorie dominante, on affecte à la maille la catégorie (ou la valeur du paramètre) qui apparaît à un des sommets - toujours le même - de la maille. L'opérateur n'a pas interprété comme dans le cas précédent , il choisit dans la liste des catégories préalablement établies celle qui correspond. L'avantage est donc une forte diminution de l'intrusion de la subjectivité dans le levé. L'inconvénient réside dans le fait que localement l'information n'est pas toujours pertinente. Le sommet tombant sur un chemin de campagne, la maille entière, même si sa superficie est d'un hectare, est déclarée chemin. Une étude statistique montre, cependant, qu'à partir d'un certain nombre de mailles, la statistique de superficie tend à de-

venir correcte. La règle statistique paraît donc plus objective. Peu de liberté est laissée à l'interprète. Il n'a jugé que la classe apparaissant sous le point d'échantillonnage. La méthode suit ainsi des règles de normalisation plus rigoureuses et compatibles avec la logique numérique. Elle permet des mises à jour très ponctuelles. L'analyse des changements reposent donc sur des informations plus objectives.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 16: tableau-affect-OFT.swf

### Structure numérique

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les  
flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en  
partie par des images d'écran.

FIG. 17: numeric-structure.swf

## Formats de structure numérique

Du point de vue de l'utilisateur, la notion de format ne se révèle que lorsqu'il est nécessaire de transférer des informations d'un logiciel à un autre. Chacun s'est confronté au moins une fois à la fastidieuse expérience des changements de formats. Il est profitable d'appliquer quelques règles de conduite de manière à éviter les pièges principaux. Les quelques repères formulés ci-dessous permettent d'éviter des surprises. Les images numériques sont souvent multibandes. A chaque bande spectrale est attaché un fichier contenant les luminances. Il existe plusieurs manières d'enregistrer ces informations. Les formats les plus fréquemment utilisés sont : BSQ (Band sequential) et BIL (Band Interleaved by Line).



FIG. 18: BSQ-BIP.jpg

A cette diversité de formats, s'ajoute la présence ou non de la méta-information (en-tête) au début du fichier image ou encore entre chaque matrice spectrale pour les cas des images de télédétection. La structure image est gourmande en espace de stockage. Une résolution fine produit très vite un grand nombre de mailles. À chaque maille étant affectée une information de la taille d'un octet au minimum, on se trouve fréquemment confronté à un gros volume de données. Comment le réduire pour rendre la base de données plus manipulable? Deux solutions courantes sont évoquées dans le diaporama suivant : le compactage séquentiel et les structures en tétrarbre (QuadTrees).

Pour en savoir plus, voir les cahiers du NCGIA [?]

ou encore dans [?]

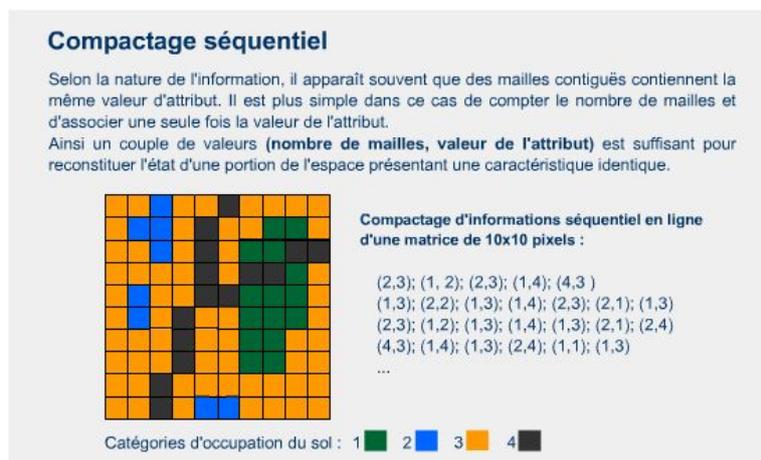


FIG. 19: compact-seq.jpg

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashs etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 20: quadtree.swf

## Quelques aspects propres à la structure image

### Représentation

La structure image est de nature discrète. L'élément de référence est la maille. Toute représentation des unités géométriques se base donc sur cette primitive. Un chemin de 3m de large est représenté par une succession de mailles adjacentes même si celles-ci sont définies avec une résolution de 30m. Il en est de même pour un objet ponctuel. Dans le même système, une borne d'hydrante occupe toute une maille. Le modèle image engendre donc des distorsions géométriques. En terme spatial, cette information est évidemment grossière. Il convient d'interpréter en terme de présence d'un thème dans la maille : Dans cette maille, il y a un chemin...

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 21: representation.gif

### Notion de région

La structure image se construit à partir de la maille. L'information est individualisée à ce niveau. Visuellement des mailles contiguës affectées d'une même valeurs d'attribut font apparaître des zones, des lignes et des points s'apparentant à la notion d'objet absente de la structure image. Les géographes les ont dénommées région. Certains logiciels, IDRISI en particulier, offrent une fonction de constitution de régions qui permet de l'identifier, de l'exploiter lors d'analyses spatiales. A l'instar des unités irrégulières, il est possible de calculer le centre de gravité de la région, sa taille, sa forme et de créer des zones tampons.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 22: region.gif

### Notions de distance

Dans un système euclidien, la distance se définit comme le plus court chemin entre 2 points. En structure image, le chemin le plus court dépend du choix qui est fait pour le définir. La distance entre 2 mailles est celle mesurée entre leurs centres de gravités respectifs ou en suivant les contours des mailles qui les séparent. En soi le mode de mesure est secondaire ; l'essentiel est qu'il soit clairement défini et appliqué de manière cohérente.

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 23: perimetre.gif

**Notion de topologie**

La topologie en structure image est implicite. La numérotation des mailles selon le mode matriciel permet d'identifier les cellules voisines tant au sens restreint qu'au sens étendu.

**Propriétés comparatives des modèles vecteur et raster**

La complémentarité est analysée en mettant en regard des aptitudes de chacun des modèles à remplir les fonctions principales attendues d'un SIG.

**Tableau des aptitudes des modèles objet et image****Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 24: table.swf

Cliquez ici pour ouvrir le tableau en PDF et l'imprimer au besoin... ([www.gitta.info/DigitModel/fr/multimedia/CompVectRast.pdf](http://www.gitta.info/DigitModel/fr/multimedia/CompVectRast.pdf))

**Combinaison des diverses technologies dans un SIG**

Les deux modèles offrent des avantages différents comme décrits dans le tableau ci-dessus. Combiner des informations dans les 2 modèles est intéressant pour de nombreuses applications SIG...

### Modélisation du relief

Le relief révèle les phénomènes qui se sont déroulés à la surface terrestre. Il en est le facteur explicatif prépondérant. Le relief conditionne l'exploitation, l'appropriation du territoire par l'homme. La caractérisation du relief est donc un besoin essentiel pour la compréhension et la gestion du territoire. Elle se fait par dérivation de nombreuses variables (pente, orientation, rugosité, etc.) Elle nécessite une modélisation numérique. Dans cette unité, nous abordons la modélisation. Les méthodes d'analyse sont décrites dans le module d'analyse spatiale BAN.

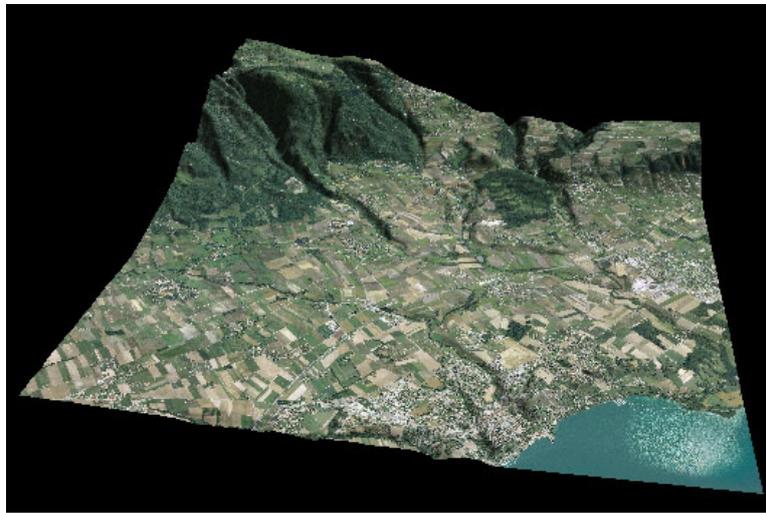


FIG. 25: Une vue 3D de la région de Nyon (source Swisstopo : orthophoto drappée sur le modèle numérique d'altitude au 1 :25'000)

### **Le relief**

Le relief a été perçu et étudié selon différentes approches :

- L'approche descriptive du relief (19<sup>ème</sup> siècle)
  - Description écrite intégrant la géomorphologie, le paysage
  - Description qualitative : montagneux, plat, vallonné, etc.
- L'approche quantitative basée sur des descripteurs statistiques
  - Indices statistiques du relief
  - Indice fractal

La géomorphologie est la science de l'étude de la forme et de l'évolution terrestre.

### Les modèles numériques

Le relief se modélise par l'altitude (élévation, hauteur d'un point) au-dessus du niveau de la mer. La cartographie le représente par les courbes de niveau.

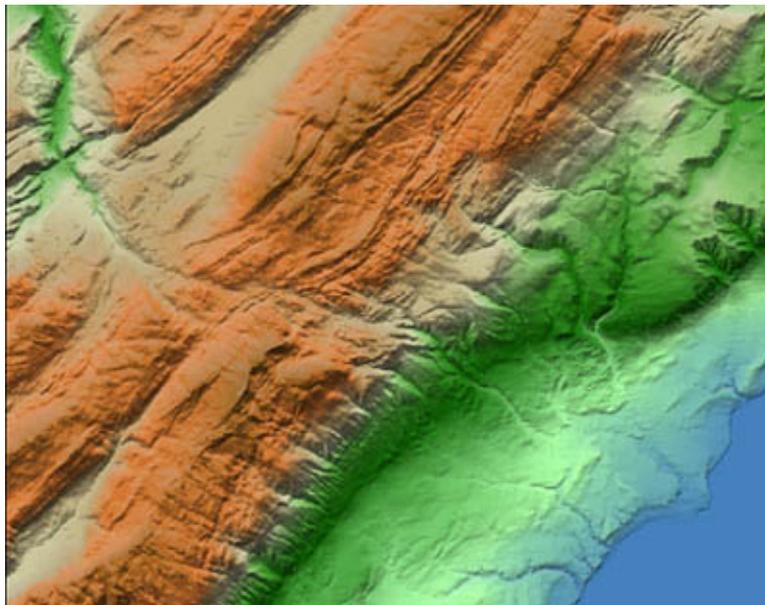


FIG. 26: Extrait du modèle numérique d'altitude sur la région de Nyon (source Swisstopo MNA2 - maille de 25 mètres). Les couleurs représentent des intervalles d'altitude du moins élevé en bleu au plus élevé en brun.

### Les modèles numériques d'altitudes (MNA)

Depuis l'avènement du numérique, le relief est modélisé le plus souvent par une grille à maille régulière d'altitudes mesurées ou estimées. Un modèle numérique d'altitude (MNA) est un modèle du relief. D'autres termes sont souvent employés : MNT (modèle numérique de terrain), ou DEM (Digital elevation model) en anglais, ou encore DTM (Digital Terrain Model). Un MNA est un tableau numérique exprimant les altitudes estimées ou mesurées aux noeuds d'une grille régulière virtuelle appliquée sur la carte topographique du territoire.

Un MNA se caractérise par son mode d'acquisition, sa résolution et sa précision. Par exemple, on peut trouver les MNA suivants :

- MNA à faible résolution : Digital Chart of the World - Résolution 30" d'arc (environ 800m) - Mode d'acquisition et précision inconnus, couverture totale de la Terre - Disponibilité : Web, gratuit

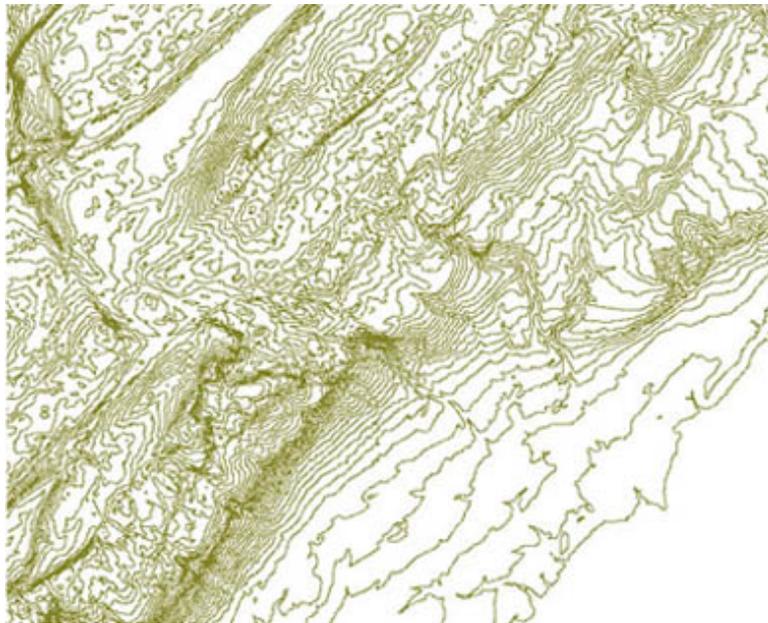


FIG. 27: Les courbes de niveau à 30 mètres sur la même région.

- MNA à résolution moyenne : RIMINI - couverture de la Suisse - Résolution 250m, acquisition par carte topographique au 1 :25'000 - Précision : équivalente à celle de la carte topographique
- SRTM (Département de la défense US) : Résolution d'acquisition : 30 m - Précision de l'ordre de 2 m - Mode d'acquisition : radar sur navette spatiale US - Pas encore disponible
- MNA25 de l'Office fédéral de topographie - Mode d'acquisition et propriétés décrites dans le paragraphe suivant
- MNA à très haute résolution : Résolution : 1 m - Acquisition par balayage laser (laser scanning), par radar aéroporté ou par photogrammétrie

### **Les modèles par triangulation (TIN)**

Le relief est aussi parfois modélisé par un échantillonnage irrégulier semi-dirigé intégrant les lignes de structure. La représentation prend la forme de TIN (Triangular Irregular Network)

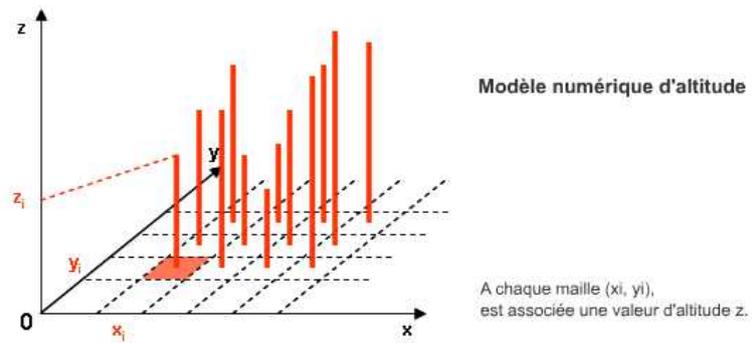


FIG. 28: MNA.jpg

**Attention :**

Seulement les images peuvent être représentées dans la version PDF ! Les flashes etc. ne sont visibles que dans la version en ligne et sont remplacés en partie par des images d'écran.

FIG. 29: TIN.gif

**Exemple d'élaboration d'un modèle numérique**

Les illustrations de cette unité proviennent d'extractions du MNA 25 de l'Office Fédérale de Topographie. La procédure d'élaboration de ce modèle peut être résumée comme suit :

- Scannage des planches graphiques contenant les courbes de niveau
- Jonction des différents fichiers, "nettoyage des textes, rajout des portions manquantes
- Érosion de l'image
- Vectorisation des lignes formées par les pixels contigus
- Intersection avec la grille de référence
- Interpolation
  - Bicubique avec pondération inverse de la distance
  - Le nouveau modèle (2000) - niveau de qualité Level 2 - comprenant les lignes de crêtes principales des Alpes, la mise à jour des courbes de niveau sur les glaciers, une topologie du modèle de base améliorée ainsi qu'une nouvelle interpolation du modèle matriciel est disponible depuis la mi juin 2001. Ce jeu de données répond aux exigences demandées pour des applications d'une très haute précision