

Cartes Cellulaires

**Concepts, méthodes et techniques
dans le cadre d'analyses spatiales
avec MapInfo®**

Document d'exploration

Jacques Paris

Mars 2003

jacques@paris-pc-gis.com

Une définition

Une carte cellulaire est une table formée d'éléments géographiques constants qui servent à recueillir l'information à partir de sources diverses et à procéder aux manipulations nécessaires pour produire des résultats synthétiques.

Une carte cellulaire se situe entre les images raster (éléments fixes « pixels » référencés par leur position colonne-rangée avec calage géographique de l'image possible, une image par variable) et les cartes traditionnelles (éléments variables, référencés comme objets géographiques, une carte pouvant contenir plusieurs variables, plusieurs cartes avec des définitions géographiques différentes pouvant être nécessaires pour porter toute l'information utile).

Ce document

Nous allons explorer dans ce document des questions relatives à ce type de carte et à sa mise en œuvre dans le cadre du logiciel MapInfo. Il pourra ainsi servir de base pour élaborer un programme de développement de procédures d'exploitation et d'outils de support, ainsi que pour définir des expériences pratiques dans des projets concrets et dans des situations d'interaction avec d'autres logiciels.

Ce document est organisé autour de quatre pôles principaux :

- La conception de la carte cellulaire
couverture, forme des éléments, résolution

- L'assignation des attributs
conversion des données sur la base géographique unique

- La manipulation des données
re-classification, indices, variables composites intermédiaires et finales

- La production des résultats
présentation des résultats sous forme de cartes et de rapports

Une cinquième partie regroupe divers sujets assez spécifiques pour lesquels je n'ai pas trouvé une niche évidente ailleurs.

En forme de conclusion, je présente quelques éléments autour desquels un projet pourrait être élaboré.

Pourquoi avoir recours aux cartes cellulaires?

Les cartes cellulaires sont en utilisation depuis longtemps à deux fins principales : la collecte de l'information et la synthèse d'information spatiale de natures différentes.

Le découpage du territoire en grille formant des cellules est pratique courante dans bien des situations : décompte d'événements, inventaires, modélisation.

Il a été aussi à la base de beaucoup d'opérations de synthèse comme dans les analyses de potentiel (ex: développement urbain) ou de risques (ex: stabilité des sols), ou dans l'utilisation de modèles mathématiques pour l'étude et la prédiction de certains phénomènes (ex: pollution atmosphérique).

On pourrait imaginer recourir aux images raster pour atteindre ces objectifs. En fait, un raster est une carte cellulaire de dimension fixe et portant l'information relative à une variable donnée. Il est très possible de traiter plusieurs images raster pour en faire une synthèse et certains programmes le font très efficacement (IDRISI en est bon exemple www.clarklabs.org) mais MapInfo ne peut pas le faire. De plus, on ne peut pas limiter une image raster à la seule zone à étudier et sa résolution peut être bien trop fine pour les objectifs de l'étude.

Un tel type de synthèse pourrait aussi se faire facilement avec des cartes traditionnelles du moment que l'information est portée par des éléments géographiques de nature semblable (des régions, des lignes, des points). Ainsi, l'intersection de plusieurs couches de régions (on suppose plusieurs variables dont la définition des variations spatiales entraîne des découpages en régions différents) permettrait de définir un découpage en éléments communs « exact » (suivant les variations des variables originales) mais avec un risque de création d'éléments « résiduels » non significatifs (manque de coïncidence parfaite des limites qui devraient être communes entre plusieurs couches) ce qui requiert de nettoyer les cartes résultant de chaque opération d'intersection.

Une synthèse spatiale mettant en jeu des éléments de types différents requiert de toute façon une transformation/intégration des données sur des bases comparables. Il n'y a alors aucune différence entre cartes cellulaires et traditionnelles sur ce point.

Les cartes cellulaires ont donc par rapport aux traditionnelles le gros avantage d'offrir une matrice unique de cueillette/synthèse des données bien contrôlée en ce qui concerne sa définition géographique et ouverte à toutes sortes d'interactions directes comme l'intégration de données raster (souvent simple agglomération de pixels), l'exportation vers des outils de traitements spécifiques (ex : modèles d'interpolation, de simulation) requérant des données sous forme de centroïdes documentés ou de grilles régulières, l'importation des résultats de programmes externes souvent disponibles sous l'une de ces formes.

Les désavantages dans cette comparaison tiennent à l'imprécision résultant de l'application d'une trame unique à toutes les variables dont les variations spatiales ne sont plus aussi finement et précisément reproduites, et au poids de ces cartes si on choisit de travailler avec une maille fine relativement à l'ensemble de la région étudiée.

Comparées aux images raster, elles permettent de limiter la zone à traiter, de la définir avec la précision voulue et de regrouper en une seule table toutes les données nécessaires à l'analyse ce qui permet entre autre de faire des synthèses sans avoir besoin d'outils externe.

Mais elles sont beaucoup moins efficaces pour porter l'information relative à un phénomène particulier et le visualiser.

1 - Conception de la carte cellulaire

Tout le processus de définition de la carte de base est imprégné d'une recherche de compromis entre la finesse et la précision de l'information contenue dans la carte tenant compte de sa complexité même. Il est bien entendu que l'utilisation de cartes cellulaires dans des logiciels comme MapInfo devient progressivement plus laborieuse avec l'augmentation du nombre de cellules; l'affichage peut devenir trop lent pour être productif, les fichiers trop gros pour être manipulables.

Le compromis se fera donc entre la définition de la carte, vue surtout comme le nombre de cellules, et la précision de l'information, estimée par rapport à deux aspects : la finesse du résultat désiré et la fidélité à représenter l'information de base.

Il ne peut donc pas y avoir de règle générale à utiliser dans cette phase. Ce ne peut être que la personne ayant une vue globale du projet qui peut savoir d'une part ce qu'il faut produire (finesse du produit final) et d'autre part comment les variations spatiales des variables de base seront le mieux prises en compte individuellement et globalement.

1- 1 Région couverte

Une carte cellulaire permet de limiter la région couverte à une approximation des limites réelles de la région à étudier. L'approximation provient de l'utilisation de cellules de taille finie. Nous avons donc une situation intermédiaire entre l'image (surface rectangulaire englobant la région d'étude) et la carte standard dont les limites sont celles de la région.

On devra tenir compte de cette approximation dans le choix de la taille des cellules de base de la carte cellulaire de façon à minimiser l'impact sur les données des cellules du pourtour; des cellules plus petites reproduiront mieux les limites et plus fidèlement les variables.

On peut aussi exclure d'une carte cellulaire les zones intérieures ne faisant pas partie de l'étude, comme les étendues d'eau. Il faut cependant ne le faire que si ces « trous » ne font pas réellement partie du phénomène à étudier. Un modèle atmosphérique est continu dans l'espace, une étude d'érosion s'arrête aux berges d'un lac.

1- 2 Trame de base

Une carte cellulaire peut être construite sur une trame de base, c'est-à-dire sur un graticule régulier sur lequel les cellules viennent se coller exactement. Ceci implique que la taille minimale des cellules est bien définie, celle d'une case du graticule. Cette approche est en contradiction avec toute subdivision éventuelle d'une cellule « minimale » mais ceci n'est qu'une considération théorique car la trame de base n'existe pas dans la définition géographique de la carte; ce n'est tout au plus qu'une référence potentiellement utile à la conception d'une carte cellulaire.

1- 3 Forme des cellules

Si on veut obtenir une mosaïque continue avec des éléments semblables et sans compliquer indûment la géographie des cellules, il faut s'en tenir aux formes les plus simples que sont le triangle équilatéral, le carré et l'hexagone régulier. Ces géométries ont des avantages et des limitations que nous allons identifier dans le cas de cartes faites de cellules uniformes.

Frontières de région	On suppose une région rectangulaire. On décrit la possibilité de suivre les frontières définies donc comme rectilignes.
Déformation	Les formes de base sont « régulières », les déformations dans l'une des 2 directions et respectant la symétrie dans l'autre (ex. triangle équilatéral devenant isocèle)
Fragmentation	Découpage d'une forme de base en formes « semblables » recouvrant tout l'espace original, sans trou.
Alignement des centroïdes	Les centroïdes des objets sont placés le long de droites parallèles aux côtés de la région; il peut y avoir pour une même rangée ou colonne deux droites; les centroïdes d'objets voisins peuvent être sur la même ligne ou sur 2 lignes différentes, en quinconce..
Représentativité des centroïdes	Potentiel du centroïde à « représenter » l'espace autour de lui ce qui se passe dans la cellule. Plus les distances entre points les plus rapprochés et ceux les plus éloignés sont différentes, moins bonne est ce potentiel. Mesuré comme le rapport du plus proche au plus loin, soit le cosinus du demi angle au sommet.

	Triangles	Carrés	Hexagones
Frontières de région	2 (opposées) rectilignes, 2 en dents de scie	toutes rectilignes	2 (opposées) indentées, 2 en dents de scie
Déformation	triangles isocèles	rectangles	hexagones semi-symétriques
Fragmentation	découpage en triangles sous-multiples semblables	découpage en carrés sous-multiples	impossible
Alignement des centroïdes	2 droites parallèles au côté rectiligne et en quinconce, une simple pour l'autre direction	simple droite dans les 2 directions	simple droite dans les 2 directions et en quinconce
Représentativité	$\cos 60^\circ = 0.5$	$\cos 45^\circ = 0.71$	$\cos 30^\circ = 0.87$

Hormis la non-linéarité des frontières, les hexagones apparaissent comme les plus prometteurs d'un point de vue géométrique, du moment que l'on n'envisage pas de changer la taille de base des cellules. Nous aurions donc toujours ainsi une trame régulière simple et une bonne représentativité des cellules.

Les triangles offrent peu d'avantages par rapport aux carrés. La simplification de la géographie due à la disparition d'un nœud ne justifie pas la présence d'une double ligne de centroïdes dans une direction et la faible représentativité du centroïde.

Sans vouloir exclure systématiquement d'autres formes que le carré (il y a des applications spéciales conçues sur une base hexagonale, référence Spencer Simpson ¹ [ssimpson@baltometro.org]), le reste du document traitera essentiellement de cette forme, bien que plusieurs considérations puissent s'adapter aux autres.

1- 4 Fragmentation ou fusion

Une carte cellulaire peut ne pas être uniforme, c'est à dire peut contenir des objets de forme identique mais de taille différente. Pour créer une telle carte, on peut procéder par fragmentation ou par fusion locale d'une grille régulière.

Procéder par fragmentation indique que l'on a une bonne idée de la taille maximale des cellules; cette taille doit être acceptable vu la nature du phénomène dans son ensemble. La fragmentation portera sur les cellules pour lesquelles les variations spatiales d'une ou de plusieurs variables, donc du phénomène-problème, sont trop fortes pour être représentées correctement par de grosses cellules. Il n'y a pas alors théoriquement de limite inférieure de taille et la notion de trame sous-jacente est absente.

L'application CELLMAP.MBX de l'auteur procède par fragmentation sur les choix de l'utilisateur. Plusieurs exemples concrets devraient être montés pour montrer la gamme possible d'applications.

La fusion procède dans l'autre sens : une grille est d'abord constituée de cellules d'une taille minimale, ces cellules sont ensuite regroupées en plus grandes sur la base des variations spatiales du phénomène. On pourrait envisager d'explicitier des algorithmes optimisant les fusions pour minimiser certains paramètres statistiques tout en répondant à des critères géométriques comme le maintien de la forme d'une cellule en carré.

La préparation des données pour le programme CALPUFF a fait appel à la technique de fusion. Un résumé du travail serait très instructif.

La fragmentation maintient la trace implicite de la trame originale; les cellules plus grandes que la taille minimale s'inscrivent dans une trame qui est une fragmentation uniforme de la trame originale et toutes les cellules ont la même forme. Lors d'une fusion, le regroupement se fait sans contrainte particulière en ce qui concerne une trame de plus grande taille. Il faut même forcer la forme carrée des nouvelles cellules.

¹ "At our company we have turned the UK into a grid of 750m and 500m hexagons. This creates a stable form or geography that does not change, unlike our postcode sectors which are revised about twice a year! With this grid we attached spend information in those places where people actually live. This creates a demand grid which we use in our gravity modeling system which has been built in house and all in MI and Access! We really like the output and the ease of use." 21.02.03

Il serait intéressant de procéder à certaines comparaisons des résultats de ces deux approches sur un même terrain.

1- 5 Pré-définition des formes

Quand on fusionne 4 cellules carrées en un seul carré, les points qui correspondent aux sommets originaux ne disparaissent pas (avec MapInfo au moins) du périmètre du nouveau carré. Chaque côté du carré est en fait composé de 2 segments, et on peut dire que les sommets-intersections existant dans la trame originale sont conservés et donc que les formes sont prédéfinies.

Par contre, quand on procède par fragmentation, les nouveaux sommets ne sont d'abord définis que pour les portions fragmentées; les cellules voisines ne portent pas ces nouveaux sommets. Cette absence peut se révéler fâcheuse lors d'opérations subséquentes durant lesquelles des problèmes d'exactitude topologique peuvent surgir. Il faut se rappeler en effet que le point milieu d'un segment « rectiligne » c'est à dire défini par ses deux seules extrémités, n'est pas nécessairement exactement sur la ligne; cela provient d'une part du fait que l'on travaille sur des projections planes d'une surface sphérique et d'autre part de la façon dont les coordonnées sont enregistrées de façon interne par un logiciel avec la contrainte d'utiliser des nombres entiers.

Il faut donc pouvoir envisager la situation où le côté d'un grand carré fragmenté puisse ne pas « coller » exactement sur le côté d'un carré avoisinant non fragmenté, avec de possibles conséquences topologiques comme la présence de « splinters » internes lors de la fusion de ces deux carrés.

L'approche par fusion conserve tous les sommets au coût naturellement d'une définition des cellules conservées plus lourde par l'existence de sommets intermédiaires.

Pour éviter les problèmes soulignés plus haut quand on procède par fragmentation, deux approches sont possibles. La première utilisée dans CELMMAP applique les nouveaux sommets sur les cellules avoisinant la cellule fragmentée. Cette approche permet d'éviter tout problème topologique subséquent et minimise la « taille » de la carte en ne définissant que les sommets utiles.

L'autre approche consisterait à choisir une taille minimale envisageable de cellules, à définir les côtés des cellules comme des chapelets de segments de cette taille minimale et à contraindre toute fragmentation sur cette trame minimale virtuelle. Outre cette dernière contrainte qui a son propre poids, cette approche multiplie les sommets intermédiaires « inutiles » et donc le poids de la carte, mais il permet par contre de pouvoir envisager des superpositions « parfaites » de cartes cellulaires ayant des découpages différents du même espace.

Une expérience simple permettrait de mesurer objectivement les différences de performance (taille du fichier .map, vitesse d'affichage) avec ces 2 définitions (toutes les cellules sont définies avec des sommets intermédiaires correspondant à une taille minimale, seuls les sommets intermédiaires correspondant aux cellules fragmentées sont ajoutés)

2 - Assignation des attributs

L'assignation des attributs à une carte cellulaire consiste à inscrire dans les colonnes de son « tableau » les valeurs des variables correspondantes. Nous allons écarter de la discussion l'assignation manuelle et nous concentrer sur les principes de l'assignation automatique qui présuppose l'existence de ces variables sous la forme de carte électronique. Le problème essentiel est de transformer des valeurs de variables « applicables » à des objets de forme ou de type différents en des valeurs « applicables » à une cellule.

La nature des données initiales définit deux procédures totalement différentes; si les données décrivent les manifestations spatiales du phénomène de façon exhaustive (location d'hôpitaux, flux le long d'un réseau, utilisation du sol), aucun traitement préalable des données n'est requis avant l'assignation des attributs aux cellules. Par contre, si les données sont des mesures ponctuelles d'un phénomène dont la nature spatiale est continue (altitude mesurée à certains points, profondeur de la nappe phréatique pour divers sondages), un traitement préalable de généralisation des mesures est pré-requis.

2 – 1 Assignation directe

J'ai pu identifier deux approches principales pour assigner des attributs aux cellules, l'une "par sondage", l'autre "par intersection". "Par sondage" utilise un nombre fixe de points par cellule, récupère les valeurs pour la variable en question à ces positions et les synthétise avant d'introduire le résultat dans la cellule. "Par intersection" se sert de la surface d'intersection de chaque cellule avec la carte de la variable comme base pour pondérer les valeurs correspondantes dans le résultat.

Il est bien entendu que la technique par sondage ne peut s'appliquer qu'aux variables dont la manifestation dans l'espace est continue et qui sont donc représentées par des surfaces contiguës (il faut être sûr de trouver quelque chose aux points de sondage). CELLMAP offre une possibilité d'assignation par centroïde, c'est-à-dire que la valeur « trouvée » sous le centroïde de la cellule est celle qui est retenue. On pourrait imaginer des grilles de sondage plus fines, contenant 4 ou 9 points par exemple. La synthèse des valeurs lues doit ensuite être faite selon certaines règles ressemblant très fort aux celles que nous allons voir pour l'autre technique.

L'utilité de la technique par sondage avec points multiples reste à prouver comparée à celle par intersection. Elle est probablement plus lourde à programmer et à exécuter, mais elle reste celle qui est recommandée quand il s'agit d'assigner des valeurs à partir d'une autre carte cellulaire couvrant la même région (souvent dérivée d'une image raster) avec des cellules de même taille mais dans laquelle les limites de cellules risquent de ne pas coïncider exactement avec celles de la carte cellulaire de base. De légères imperfections d'alignement peuvent créer des difficultés parfois, des calculs inutiles souvent, alors que ce qui importe c'est la valeur « au centre ».

Je ne connais aucune expérience de mise en œuvre de la technique par sondage avec points multiples. Il serait peut-être intéressant de réaliser un outil

prototype pour évaluer sa performance potentielle comme outil d'assignation de valeurs.

Nous allons dans ce qui suit nous concentrer sur la techniques par intersection qui est applicable aux trois types d'objets, régions polylignes et points.

Variable décrite par des objets surfaciques

Géométrie prise en compte : surfaces interceptées par la cellule

Valeur à saisir

Numérique calculée² : éléments interceptés =

un seul: valeur de l'élément

plusieurs: sommation pondérée par la surface interceptée par rapport à la surface de la cellule des valeurs applicables aux régions en cause.

Numérique comptée³ :

Sommation des valeurs des éléments au pro rata de leur surface interceptée

Qualitative : éléments interceptés

un seul: valeur de l'élément

plusieurs: requiert des règles de synthèse comme la dominante (valeur ayant la plus grande surface totale interceptée) avec ou non un seuil minimum, un ensemble de conditions particulières propres à une variable particulière (si X occupe plus de N% de la cellule, automatiquement X; - si A et B présents, considérer que A=B; - ...), ...

Variable décrite par des objets linéaires

Géométrie prise en compte : longueurs interceptées par la cellule

Valeur à saisir

Numérique calculée : éléments interceptés =

un seul: valeur de l'élément

plusieurs: requiert l'accès aux termes ayant servi au calcul des valeurs.

Nécessite la sommation pondérée par la longueur interceptée des valeurs du numérateur et du dénominateur pour pouvoir calculer une valeur pour la cellule.

Numérique comptée :

Sommation des valeurs des éléments au pro rata de leur longueur interceptée

Qualitative : éléments interceptés

un seul: valeur de l'élément

plusieurs: requiert des règles de synthèse comme la dominante (valeur ayant la plus grande longueur totale interceptée) avec ou non un seuil minimum, un ensemble de conditions particulières propres à une variable

² Rapports, moyennes, taux ... qui s'appliquent uniformément à l'élément surfacique entier

³ Décomptes, inventaires, recensements des cas situés dans l'élément et supposés être uniformément répartis dans l'élément. On peut inclure dans ce type de données des caractéristiques géométriques de l'élément pas forcément dans le tableau (surface, longueur en particulier)

particulière (si X occupe plus de N% de la cellule, automatiquement X; - si A et B présents, considérer que A=B; - ...), ...

Variable décrite par des objets ponctuels

Géométrie prise en compte : points situés dans la cellule

Valeur à saisir

Numérique calculée : éléments présents =

un seul: valeur de l'élément

plusieurs: requiert l'accès aux termes ayant servi au calcul des valeurs.

Nécessite la sommation des valeurs du numérateur et du dénominateur pour pouvoir calculer une valeur pour la cellule.

Numérique comptée :

Sommation des valeurs des éléments présents; peut n'être qu'un simple décompte des points présents.

Qualitative : éléments interceptés

un seul: valeur de l'élément

plusieurs: requiert des règles de synthèse comme la dominante (valeur ayant la plus grande « présence ») avec ou non un seuil minimum, un ensemble de conditions particulières propres à une variable particulière (si X représente plus de N% des points, automatiquement X; - si A et B présents, considérer que A=B; - ...), ...

On peut aussi inclure comme valeurs synthétiques attribuables d'autres mesures de statistique descriptive autre que la moyenne tirées de la distribution des valeurs incluses dans une cellule. Ainsi on peut choisir parmi les valeurs du minimum, du maximum, du coefficient de variation ((maxi-mini) / moyenne), d'écart type, etc., celle(s) la(les) plus pertinente(s) au sujet traité.

2- 2 Traitement préalable

Dans le cas où les mesures n'existent que pour certains points précis dans l'espace, il faut d'abord générer des valeurs pour toutes les cellules. Ceci se fait en général de deux façons suivant la nature du phénomène. Si les variations de celui-ci sont quantitatives et continues, on fait appel à une procédure d'interpolation; si elles sont qualitatives et donc discontinues, on construit des zones d'influence autour des points de relevé (des polygones de Voronoï par exemple). Nous ne rentrerons pas dans le détail de ces diverses techniques, nous contentant de remarques pratiques en ce qui concerne l'attribution subséquente des attributs aux cellules.

Une fois les polygones d'influence construits, l'attribution des attributs peut suivre la procédure esquissée plus haut. C'est dans le cas de l'interpolation que nous avons le plus à dire. En effet bien des outils d'interpolation fonctionnent sur la base d'une grille de cellules auxquelles des valeurs calculées (interpolées) sont attribuées selon les résultats du modèle mathématique. Il est donc très important de pouvoir définir cette grille de base de telle sorte que l'intégration des résultats qu'elle portera dans la carte cellulaire soit facilitée et ne donne pas lieu à d'autres calculs qui pourrait invalider partiellement les résultats de l'interpolation.

On doit prendre en compte deux points en particulier : le calage de la grille et la taille respective des trames d'interpolation et cellulaire. On ne peut pas assez mettre l'emphasis sur la nécessité d'avoir une excellente coïncidence des trames de la carte cellulaire et de la carte des données à y transférer. Il faut donc que la définition de la région à soumettre à l'interpolation le point origine des coordonnées de la trame qui sera construite durant l'interpolation soit le même que celui utilisé pour la trame de base.

Il faudrait vérifier avant tout appel à une interpolation si l'outil utilisé va livrer des données qui seraient applicables aux centroïdes de la cellule, ou à un coin (le coin supérieur gauche le plus souvent). Une autre façon d'exprimer le problème serait de savoir si le point origine de la trame d'interpolation est utilisé comme coin supérieur droit de la première cellule ou comme son centroïde. Il faudra donc dans certains cas envisager de donner au point d'origine un décalage d'une demi-cellule pour que la région couverte par l'interpolation et par la carte cellulaire soient identiques.

Pour alléger la procédure d'assignation, il est aussi recommandé de choisir un rapport simple entre les tailles des cellules des deux trames. L'idéal serait d'utiliser la même taille ce qui par la suite simplifie de beaucoup les opérations. L'utilisation de valeurs interpolées sur la même base n'entacherait pas la qualité des résultats de l'interpolation puisqu'il n'y aurait alors aucun calcul de pondération.

MapInfo a un interpolateur intégré avec ses cartes GRID et une petite application MBX comme InfoExMig (par l'auteur) permet en outre d'extraire sous forme de carte cellulaire le contenu d'une table MIG dans son entier ou pour une fenêtre rectangulaire donnée. Si le type d'interpolation offert par MI est jugé trop limitatif, d'autres interpolateurs existent sur le marché, y compris des versions shareware comme Visual Explorer; une assez ancienne application MBX/VB (XYZ4MI, par l'auteur) permet de passer d'un format x, y, z de carte interpolée sur VE à un format MI (cellules) et de MI (carte de points ou de lignes de contour) à un format lisible par VE.

Il conviendrait de vérifier le calage exact d'une grille mig générée comme carte thématique par MapInfo dans différentes circonstances pour pouvoir utiliser sans problèmes ce type de fichier comme source de données extrapolées.

3 – Manipulation des données

Il serait des plus surprenants de voir une carte cellulaire utilisée uniquement pour recueillir des données sans vouloir les utiliser en conjonction avec d'autres variables. Ceci implique d'avoir recours à des manipulations pour lesquelles il faut avoir des règles et des outils. Même si l'on se contentait de produire une carte des données originales, il faudrait établir certaines règles de représentation qui sont implicitement une manipulation car les données originales doivent être transformées en « plages » de valeurs (le cas de valeurs individuelles ne peut pas être exclus a priori car des regroupements pourraient être effectués avant représentation).

Pour entreprendre des analyses spatiales multi-variées, il faut avoir un **modèle** de ce que l'on veut faire, c'est-à-dire une procédure d'analyse ou de synthèse des données indiquant les diverses opérations à effectuer. Ces opérations font appel à certaines **règles** de transformation et de synthèse et pour pouvoir les faire, il faut des **outils**. Mais il faut souvent faire un pré-traitement des données brutes pour changer les valeurs d'une variable (transformation) ou créer de nouvelles variables (transgénération).

3 – 1 PRÉ-TRAITEMENT

TRANSFORMATION

Rappelons qu'il s'agit de transformer une seule variable à la fois.

Dans le cas d'une variable quantitative, il s'agit de **changement d'échelle et d'unité** par rapport à celles utilisées pour l'enregistrer. Les raisons de ces transformations sont en général de ramener plusieurs variables dans un espace où les comparaisons sont plus faciles et pour cela utiliser une échelle commune ou prédéterminée (normalisation par rapport à la moyenne relativement à l'écart-type, gamme de variation 0-maximum ramenée à 0-100) ou d'accentuer ou de diminuer la grandeur des variations dans certaines portions du domaine de la variable en lui faisant subir une transformation non-linéaire (logarithmique, exponentielle, sinusoïdale,...)

Dans le cas d'une variable qualitative, il est surtout question de ramener l'espace des catégories initiales à une liste prédéfinie de catégories pour en avoir une expression plus universelle ou plus concise. Cette **recodification** requiert simplement des listes de correspondances entre « anciennes » et « nouvelles » catégories.

On ne peut pas oublier non plus la transformation d'une variable quantitative en variable qualitative en utilisant un certain nombre de catégories pour couvrir la gamme de variation de la variable. Après une telle **codification**, la variable originale perd certains de ces attributs associés à une variable numérique et ce n'est pas l'utilisation d'un code numérique qui peut les lui rendre.

TRANSGÉNÉRATION

Ce terme recouvre toutes les opérations de création de nouvelles variables faisant appel à deux ou plusieurs variables originales. Parmi les types de résultat, on peut identifier les combinaisons (sommations pondérées ou non), les rapports (numérateur divisé par

dénominateur le tout calibré, en % par exemple, ou non) pouvant représenter des performances moyennes (personnes par km², voitures par ménage), les distributions (répartition proportionnelle d'une population par catégorie), les changements (taux de changement sur une période donnée), toutes sortes d'indices ou d'indicateurs, et des variables qualitatives provenant de combinaisons spécifiées par des règles particulières.

Une petite note pour attirer l'attention sur la possibilité d'introduire ou d'utiliser dans ces opérations des caractéristiques géographiques des objets impliqués (surface, longueur et aussi proximité ou distance). En cela MI est une plate-forme des plus intéressantes.

La diversité des possibilités laisse entrevoir la complexité des commandes à prévoir. D'un autre côté, ces types de manipulations sont essentiellement les mêmes que celles faites sous le couvert des règles de la phase de traitement analytique qui suit. De même les outils sont en grande majorité utilisables en phase de pré-traitement tout autant que durant l'analyse même.

3 – 2 ANALYSE

MODÈLE

L'expression d'un modèle particulier est la responsabilité particulière de l'analyste et ne rentre pas dans le cadre de ce document. Nous ne nous permettrons ici que quelques réflexions générales. Tout d'abord, il faut reconnaître deux courants différents par la nature du sujet même, l'un étant orienté vers la compréhension du phénomène et sa description spatiale (analyse, description) et l'autre vers l'élaboration de contexte et la formulation de réponse (synthèse, prescription). En dépit de ces orientations que certains voient comme incompatibles, bien des éléments de construction de modèles sont identiques; ainsi des règles et des outils.

Nous avons pu observer deux approches différentes à la conception de modèle. Une façon de procéder est de laisser les chiffres faire leur travail; toutes les variables sont mises ensemble et quelque modèle mathématique (composantes principales, analyse de correspondance,...) produit des résultats qui « révèlent et mesurent » des dimensions sous-jacentes du phénomène. Cette approche est souvent utilisée dans l'analyse de situation, rarement dans une synthèse quoique certaines dimensions ainsi repérées puissent être parfois introduites dans une synthèse.

Mais cela ne répond pas aux situations dans lesquelles le nombre ou la nature des variables ne se prête pas à ce type d'analyse ou lorsqu'il faut parvenir à une synthèse des composantes selon des règles prédéterminées. Que la synthèse se fasse en une seule étape ou qu'elle soit progressive (souvent représentée sous la forme d'une arborescence d'opérations impliquant variables et résultats intermédiaires), elle requiert certaines règles de synthèse.

RÈGLES

Les règles peuvent impliquer deux ou plusieurs variables. Certaines approches entraînent des règles exclusivement binaires (c'est le cas souvent des manipulations d'images quand l'image A combinée à l'image B donne C). Il est en effet beaucoup facile d'imaginer et de spécifier des règles de combinaison impliquant 2 variables que

plusieurs. De telles règles doivent en effet prévoir le résultat à donner à la valeur synthèse sur l'ensemble des gammes de variation des variables de base.

Si une règle peut s'exprimer par une expression mathématique quelque simple ou complexe qu'elle soit, il suffit de construire l'expression dans le langage du programme pour obtenir le résultat.

Si la règle fait appel à des conditions particulières (A n'entre en jeu que dans la condition c1 de B, si condition c2 de A et condition c5 de B alors résultat = r2, si condition c2 de A alors résultat = 0 pour toutes conditions de B et de C), la diversité des formes et opérateurs devient incontrôlable; je ne connais pas de syntaxe « générale » qui permette d'exprimer ces conditions sous une forme aussi universelle que celle d'expressions numériques ne mettant en jeu que des opérateurs « arithmétiques ».

Dans le cas de deux variables qualitatives ou numériques codifiées en plages, on peut facilement représenter une règle par un tableau croisé dans lequel les entêtes de colonne et de rangée sont les catégories des deux variables et le contenu des cellules les valeurs du résultat. L'ajout d'une troisième variable rend déjà la représentation en 3 dimensions plus difficile à imaginer sans avoir recours à une représentation par couche pour chaque classe de la 3^{ème} variable. Au-delà de 3, la complexité devient trop grande pour même essayer d'y penser à moins que certaines variables n'interviennent que d'une simple façon (valeurs binaires par exemple) permettant de décomposer l'espace en parties plus « compréhensibles ».

OUTILS

MapInfo permet bien des choses mais il faut souvent bien des petites opérations pour mener à bien une manipulation simple. Il faut souvent commencer par créer une colonne et la mettre à jour; cependant la mise à jour est une application d'une expression (ensemble d'opérations arithmétiques fondamentales sur une ou plusieurs colonnes) à toute une table. Si la transformation est conditionnelle, il faut procéder par des « select » pour chaque état de la sélection et faire la mise à jour sur ces tables partielles, ce qui n'est pas très productif.

Une possibilité est d'utiliser un autre logiciel que MI pour faire les transformations de données qui n'impliqueraient pas des informations géographiques. Exporter la table en DBF pour la travailler sur Excel est une possibilité, mais les contraintes de la réimportation laissent préférer une telle solution que lorsque toutes les manipulations à faire ont été bien définies.

SPSS (www.spss.com) est un logiciel qui permet aussi de faire des traitements de codification et de transgénérations en plus des analyses statistiques des plus simples aux plus complexes. Il offre aussi une certaine interface avec MapInfo ce qui pourrait ajouter une autre ressource pour l'analyse et aussi la production de rapports.

Pour améliorer cet aspect, on pourrait aussi imaginer de couper les tables cellulaires en deux, une partie purement géographique plus identificateur et une tabulaire (en format DBF) avec le même identificateur et de faire une jointure au moment voulu. Ce format permet en particulier de gérer la partie tabulaire comme une banque de données à partir

de MI, ce que d'autres formats externes ne permettent pas. Il y a cependant une contre-indication pour cette solution si l'on prévoit que la définition de carte cellulaire pourrait subir des modifications ayant des conséquences sur la banque de données.

SPSS et IDRISI n'ont été donnés qu'à titre d'exemple d'outils d'appoint. Il ne faudrait pas limiter la recherche à ces deux seuls logiciels mais explorer ceux qui pourraient offrir des capacités de traitement préalable ou parallèle. En plus des formats d'échange de données, il faudrait regarder à leur capacité de manipuler efficacement les règles de synthèse ou de transformation, la variété des règles offertes l'efficacité à les appliquer.

D'un autre côté, on pourrait élaborer la spécification d'un outil universel de traitement de données tabulaires pour étendre les capacités de MapInfo dans le domaine de la transformation des données.

Il faudrait aussi expérimenter pour trouver la technique la plus efficace (simplicité, fiabilité, rapidité) pour échanger les données entre MI et le logiciel externe, puis entre celui-ci et MI. Il faut alors prendre en considération si la structure de la table a été modifiée (par ex. ajout de variables); il se peut que dans ces cas-là, l'utilisation de OBJTRANSFER (permettant d'assigner des objets graphiques à une table sans objet) soit aussi efficace que de faire une jointure entre l'ancienne table et la modifiée.

4 – Production des résultats

Les résultats d'analyses faites avec des cartes cellulaires sont comme dans bien des cas de deux natures : représentation spatiale et rapport écrit, la cartographie pouvant ou non être incluse dans un rapport. Les rapports ne se distinguent en rien de ce que les on imagine être en général. La cartographie est, elle, un peu plus spécifique car elle est essentiellement basée sur des cartes thématiques de cellules, avec souvent une superposition d'éléments géographiques descriptifs, ou des représentations à 3 dimensions.

4 – 1 Cartes thématiques

Les cartes thématiques sont construites sur des données numériques ou codifiées.

Dans le premier cas, il faut procéder à la définition des plages ou catégories de représentation. Si MI offre la possibilité d'enregistrer et de réutiliser des gabarits de thématiques, ces gabarits ne font que définir le nombre de catégories et leur style respectif, mais ne conservent pas la définition des catégories mêmes. Comme il est fort probable que les cartes produites feront un tout et qu'il est recommandé alors d'utiliser des définitions et une symbolique qui permettent des comparaisons faciles par leur cohérence, il faudrait prévoir la possibilité d'une procédure simple pour obtenir les résultats voulus à partir de paramètres préenregistrés.

Une particularité de la présentation graphique de cartes cellulaires est que les cellules objets ne devraient pas avoir de bordures visibles pour laisser toute la place à la couleur. Cela peut se faire en définissant dans la partie style graphique des gabarits le style de ligne comme étant « aucune ligne » ou ayant la même couleur que la cellule même. C'est probablement le « aucune ligne » qui est plus efficace et le plus simple à gérer.

Il faudrait imaginer un gestionnaire autonome de thématiques qualifiées (nombre de catégories, plages de valeurs, symbolique, étiquette de catégorie, reconnaissance et traitement des valeurs manquantes) pour disposer d'un moyen plus efficace que les gabarits thématiques offerts par MI. Une disposition similaire mais partielle a déjà été incluse dans un projet de gestion de la végétation le long de voies ferrées (N.S.W. Railroads, Sydney, Australie)

Une alternative à la production de cartes thématiques qui doivent être « reconstruites » chaque fois qu'elles sont affichées est la création de cartes de thématique fixée, c'est-à-dire de cartes dont les objets sont « coloriés » de façon permanente. Les avantages sont une plus grande rapidité d'affichage et une plus grande sécurité car aucun « document » n'est nécessaire (WOR par exemple) pour produire la carte. L'inconvénient majeur est la multiplication des cartes et donc de l'espace requis. Il est probable qu'un usage modéré de telles cartes maximiserait les avantages d'une telle solution. L'application FIXSHADE.MBX est déjà disponible pour cette transformation.

S'il s'agit de valeurs codifiées, les cartes sont plutôt basées sur le principe des « valeurs individuelles », ou une couleur par valeur. Les thématiques de MI sont alors insuffisantes car elles n'enregistrent que la séquence des codifications graphiques sans aucune valeur attachée; si une carte ne comporte pas une certaine valeur, cette symbolique particulière sera utilisée pour la valeur présente suivante. IVT.MBX (Individual Value Thematics) a été conçu spécifiquement pour éviter un tel problème et permettre la définition et la réutilisation facile de thématiques complètes (paramètres graphiques, valeurs, étiquettes, ...).

4 – 2 Fusion de cellules

Alors que pour l'analyse l'utilisation des cellules fragmentent l'espace, lors de la présentation des résultats on pourrait penser à simplifier la représentation de cet espace en fusionnant les cellules de même « couleur ». Cela allège la définition de la carte réduisant ainsi les besoins de stockage et accélère l'affichage.

On pourrait ajouter cette fonction dans le cadre de l'application FIXSHADE.

5 – VARIA

5-1 Fragmentation/fusion d'une carte cellulaire avec attributs

Quand on veut fragmenter ou fusionner une carte cellulaire à laquelle des attributs sont déjà assignés, il faut travailler sur les deux fronts : réaliser l'opération géographique et ré-assigner les attributs aux cellules affectées. Chaque type d'opération est semblable à celle faite pour la définition d'une carte ou pour l'assignation des attributs, cette dernière pouvant être simplifiée étant donné la géographie particulière de ces cartes (surfaces nouvelles multiples ou sous-multiples de l'originale)

Une fonction de désagrégation des attributs pourrait être prévue dans la ligne de l'application CELLMAP.

De même, si tout outil ou tout procédure de fusion devrait prendre en considération l'agrégation de données existantes.

5-2 Combinaison de deux cartes cellulaires avec attributs

Quand il s'agit de combiner les attributs de 2 cartes cellulaires basées sur la même trame sur une seule, nous nous trouvons dans une situation semblable à l'assignation d'attributs à partir d'une carte ordinaire, au point qu'il serait pratiquement inutile d'en faire un sujet spécial. Cependant, si les deux cartes sont construites sur la même définition de trame (même origine de la trame, même taille minimale sous-jacente), on peut tirer profit des caractéristiques de la géographie des cartes et utiliser les centroïdes des cellules (de l'une comme de l'autre carte) pour faire les assignations d'attributs.

On ne peut pas négliger la possibilité d'utiliser des identificateurs uniques qui, s'ils sont fiables, pourraient être utilisés au lieu de faire une jointure géographique.

Si la combinaison exige aussi le redéfinition de la base géographique par intersection des deux cartes, il faut d'abord construire la carte et ensuite lui assigner les attributs extraits des deux cartes originales, comme indiqué au paragraphe précédent. L'intersection des deux cartes en ayant recours aux commandes standard de MapInfo peut donner lieu à des problèmes topologiques (en particulier celui de points intermédiaires pas nécessairement sur la ligne); il faudrait prévoir une approche tenant compte de la géographie particulière de ces cartes pour en tirer profit.

5-3 Exclusion de zones d'une carte

Si la région à cartographier contient de zone à exclure (l'extérieur de la région, ou à l'intérieur comme des lacs), il est important que la carte suive au mieux les contours à conserver de la région. Un outil de génération de trame devrait comprendre cette fonction d'exclusion. Mais cela va nécessiter de préciser la procédure et les critères à suivre dans de tels cas.

Par exemple le principe général pourrait être de retenir toutes les cellules qui se trouvent pas « complètement » dans la région, d'exclure celles qui se trouvent « complètement » en dehors de la région et de fragmenter celles à cheval que l'on retient-exclue-fragmente progressivement.

« Complètement » peut être qualifié par au moins X% ou au moins N (en unités de surface), cette qualification pouvant varier avec le niveau de fragmentation. Les tailles maximale et minimale des cellules doivent être connues. La fragmentation peut procéder par niveaux binaires (une cellule est coupée en 4) ce qui peut entraîner une contrainte entre les tailles maximale et minimale (1000, 500, 250, 125 ou 800, 400, 200, 100) ou par niveaux ad hoc (fragmentation inégale) qui permet d'adapter les niveaux aux tailles maxi-mini au prix de « sauts » parfois assez grands (1000, 200, 100 ou 1000, 500, 100).

Pour guider ces choix, il faut se rappeler que le but des cartes cellulaires est de conserver le plus grand nombre possible de cellules des plus grandes tailles. Avec le dernier exemple, on peut imaginer qu'un saut de 1000 à 200 conservera/exclura un plus grand nombre de cellules de 200 qu'un de 1000 à 500; il devrait donc y avoir plus de cellules de niveau intermédiaire que dans le deuxième cas.

On devrait faire un test sur l'île de Montréal avec ces 2 séquences et en décalant l'origine de la trame de 350 m deux fois en x et y. On pourrait jouer aussi avec les seuils d'inclusion/exclusion. Voici quelques mesures de performance à prévoir : surface totale de la région selon les limites originales et selon les cellules retenues, surfaces en plus et en moins des cellules partielles par rapport aux limites originales de la région, distribution des cellules par taille.