
Le projet Hypercarte : analyse spatiale et cartographie interactive

**Claude Grasland¹, Hervé Martin², Jean-Marc Vincent³
Jérôme Gensel², Hélène Mathian¹, Said Oulahal³
Olivier Cuenot², Euloge Edi³, Lliliane Lizzi¹**

(1) UMR Géographie-Cités

13 rue du four

75006 Paris

claudе.grasland@parisgeo.cnrs.fr

(2) LSR IMAG

BP 72 38402 Saint-Martin d'Hères

herve.martin@imag.fr

(3) ID IMAG

ZIRST 51, avenue Jean Kuntzmann

38330 Montbonnot

jean-marc.vincent@imag.fr

RÉSUMÉ. Les avancées dans le domaine du Web ont ouvert de nouvelles perspectives dans le domaine de la cartographie interactive. Néanmoins, les environnements existants ne supportent pas l'utilisation conjointe de grands volumes de données et l'exécution « au vol » de méthodes telles que celles développées en analyse spatiale nécessitant des calculs complexes. Ceci a pour conséquence de restreindre les possibilités de requêtes offertes aux utilisateurs et de ne pas proposer de méthodes trop consommatrices de ressources. Dans cet article, nous présentons les premiers résultats du projet Hypercarte qui vise à proposer certaines solutions novatrices à ces problèmes en travaillant simultanément sur les problèmes de modélisation spatio-temporelle, de gestion de caches (données et calculs), d'algorithmique parallèle et de visualisation cartographique, sur la base de questions liées aux organisations spatiales de phénomènes sociaux. Les applications visées sont dans le domaine socio-économique. Il s'agit, par exemple, de faire apparaître les gradients ou les discontinuités d'un territoire donné en spécifiant un maillage particulier ou la portée de l'analyse (rayon de lissage).

MOTS-CLÉS : cartographie interactive, analyse spatiale, technologies web

KEYWORDS: interactive cartography, spatial analysis, web technologies

1. Introduction

Les avancées dans le domaine du Web ont ouvert de nouvelles perspectives dans le domaine de la cartographie interactive et dynamique. Il existe de nombreux travaux dans de recherche dans ce domaine (Antoniou *et al.*, 2004) (Vincent *et al.*, 2004) (Josselin *et al.*, 2003) mais surtout une multitude de développements qui utilisent les possibilités offertes par le standard SVG (SVG, 1998) et les environnements de développement tels que MapServer (MapServer, 2004). En quelques années, nous sommes passés d'une cartographie papier à la production de véritables atlas électroniques.

Néanmoins, les projets existants ne sont souvent qu'une publication sur le Web de documents cartographiques simples décrits en SVG. Ce type d'utilisation avec ses possibilités de navigation voire d'animations convient dans de nombreux contextes. Mais, les limites de ces architectures apparaissent dès que les applications possèdent un sous-ensemble des caractéristiques suivantes :

- Contraintes de sécurité et de confidentialité nécessitant un cryptage des données. Dans ce cas, on utilise généralement des formats propriétaires ou binaires tels que MPEG-4.
- Grands volumes de données ne pouvant être transférées et traitées localement.
- Production d'une collection ou atlas de cartes répondant de manière complémentaire et cohérente à une requête posée. Il s'agit de générer à la demande une collection de cartes représentant un phénomène en fonction de différents critères (territoire à analyser, méthode d'analyse à utiliser, forme de représentation souhaitée).
- Requêtes complexes nécessitant des calculs en temps réel (aspect prédictif limité) avec application de méthodes nécessitant des temps de calcul importants.

La plupart des environnements ne supportent donc pas l'utilisation conjointe de grands volumes de données et l'exécution « au vol » de méthodes d'analyse spatiale nécessitant des calculs complexes. Ceci a pour conséquence de restreindre les possibilités de requêtes offertes aux utilisateurs et de ne pas proposer de méthodes trop consommatrices de ressources. De plus, les logiciels du marché qui offrent certaines garanties de performance comme Oracle 9i sont relativement fermés et supportent avec difficulté l'intégration de nouvelles méthodes d'analyse spatiale.

Le cadre du travail que nous présentons est le projet Hypercarte, dont l'objectif est de mettre au point de nouveaux outils interactifs de production, de représentation et d'interrogation cartographique des phénomènes sociaux qui prennent en compte l'infini variété et l'extrême complexité des demandes sociales et politiques qui peuvent être adressées à la carte en sciences sociales. Pour la mise en œuvre de ces méthodologies d'analyse spatiales et souvent spatio-temporelles nous proposons certaines solutions novatrices à ces problèmes en travaillant simultanément sur les

problèmes de gestion de caches (données et calculs), d'algorithmique parallèle et de visualisation cartographique. En effet, il nous semble qu'actuellement seule une combinaison de ces différentes approches peut permettre de répondre de manière satisfaisante aux problèmes posés tout en gardant une flexibilité et une ouverture permettant de tester différentes méthodes d'analyse spatiale et d'interfaces de navigation. Nous avons donc travaillé sur une architecture parallèle de serveur cartographique sur le Web qui utilise de manière intensive la connaissance sur les données traitées (caches sémantiques) tout en essayant d'optimiser les calculs en prenant en compte l'aspect récurrent de certaines demandes.

Les applications visées sont principalement dans le domaine socio-économique et l'aide à la décision en matière de prospective territoriale. Le postulat d'"Hypercarte" est qu'il n'existe pas une représentation cartographique d'un phénomène social, mais un très grand nombre, selon la nature intrinsèque des phénomènes sociaux, en fonction des hypothèses du concepteur de la carte, en fonction des objectifs, des demandes, des pratiques ou des croyances des utilisateurs finaux de l'information cartographique (MacEachren, 1994) L'évolution des technologies associées au web offre d'énormes possibilités à ce type d'approche en particulier grâce à la souplesse qu'apporte une interactivité de haut niveau. L'enjeu est aussi de cibler simultanément un objectif de cartographie de communication "grand public" et une cartographie plus "exploratoire" stimulant la réflexion (Kraak, Brown, 2000). Le premier module développé dans le cadre de ce projet, propose un outil de représentation cartographique basé sur une méthode d'analyse spatiale des différenciations sociales. Nous présentons ici ce module, sur la base des résultats obtenus dans le cadre de l'Observatoire en Réseau de l'Aménagement du Territoire Européen (Orate¹). L'option qui est choisie dans ce module, est d'analyser ces différenciations dans le cadre d'une hiérarchie de maillages territoriaux. Il s'agit de faire prendre conscience à l'utilisateur/observateur que rares sont les cas où le niveau absolu d'un indicateur présente à lui seul un intérêt, alors qu'implicitement c'est son niveau relatif par rapport à différents contextes territoriaux que l'on cherche à évaluer. (Grasland et al., 2003). Par exemple savoir qu'une région peut être simultanément à -20% de la richesse européenne, mais à +30% de la richesse du pays auquel elle appartient, ou encore à +10% de la richesse de son voisinage. Ainsi la confrontation de différentes positions, par rapport à différents contextes spatiaux, permet de mettre en évidence une certaine combinaison de politiques jouant à des échelons différents, et mettant en jeu des acteurs différents: politiques au niveaux les plus hauts, économiques et sociaux aux niveaux les plus bas.

Dans le paragraphe suivant, nous présentons l'architecture du logiciel Hypercarte. La section 3 est dédiée à la présentation de notre proposition d'un point de vue fonctionnel, sur la base de l'exemple des régions européennes. L'article se termine par une rapide conclusion et quelques perspectives.

¹ Sur l'origine et les finalités du programme ORATE, cf. le site du programme : <http://www.espon.lu> et celui du point focal français <http://www.orate.prd.fr>

2. Travaux connexes

La cartographie interactive s'est beaucoup développée autour de standards tels que SVG (SVG, 1998). et GML (OGC, 2003) Ces standards ont permis une simplification de la mise en œuvre de véritables documents cartographiques échangeables. Il faut sans doute ajouter à ces standards la mise en ligne de nombreuses bibliothèques Java utilisables pour développer et distribuer des logiciels cartographiques. Ainsi les bibliothèques open source telles que JTS Topology Suite (JTS, 2004) qui propose des implémentations du langage SQL de l'OpenGIS ou les propositions OpenMap et GeoTools qui sont orientées vers les composants graphiques. Certains outils réalisations tels que GeoClip sont basés sur l'utilisation de *plug-ins* développés en format Flash de Macromedia.

Il existe un certain nombre d'environnements de développement cartographique qui permettent d'optimiser le temps de développement en fournissant une palette d'outils et de fonctionnalités facilement réutilisables. Ainsi MapServer (MapServer 2005) développé, entre autres, par l'Université du Minnesota et la Nasa est une solution intéressante qui supporte l'intégration des formats vectoriels les plus courants et propose une application CGI qui permet d'accéder à MapServer à partir de nombreux langages de programmation (C, Java, Perl, PHP, Python, etc.).

Ces différents travaux sont particulièrement adaptés au développement d'applications SIG traditionnelles. Néanmoins, comme nous l'avons souligné dans l'introduction, ces solutions ne permettent pas de résoudre ni les problèmes de traitement de gros volumes de données ni les problèmes de génération de véritables atlas de cartes et ni les problèmes de modélisation et de navigation dans les données temporelles. Le projet Hypercarte est centré sur les deux premiers aspects et propose une solution basée à la fois sur une architecture parallèle et sur la génération de plusieurs cartes en réponse à une requête complexe.

3. Architecture logicielle

3.1. Description de l'approche

L'un des objectifs principaux de l'environnement logiciel Hypercarte est sa capacité d'interaction avec le géographe, c'est-à-dire sa capacité à fournir en temps réel utilisateur les cartes demandées avec les informations statistiques estimées. Si le contexte actuel du logiciel Hypercarte correspond à l'analyse d'un ensemble d'environ 1000 unités de bases (l'union européenne étendue avec un niveau d'observation nuts 3), l'objectif final est de pouvoir traiter des bases de données de plus grande taille comme par exemple des données l'espace européen avec un niveau d'observation nuts 5 (100 000 unités) ou encore les données issues du Corine Landcover (environ 106 unités). La masse des données à traiter génère ainsi des

volumes de calcul très élevés et il est nécessaire de construire des stratégies permettant de réduire la latence des calculs. L'approche suivie dans le projet consiste à exporter le calcul sur des serveurs spécialisés distants (grappe de PC) permettant d'exécuter des calculs en parallèle. La figure 1 illustre la décomposition de la latence vue par l'utilisateur :

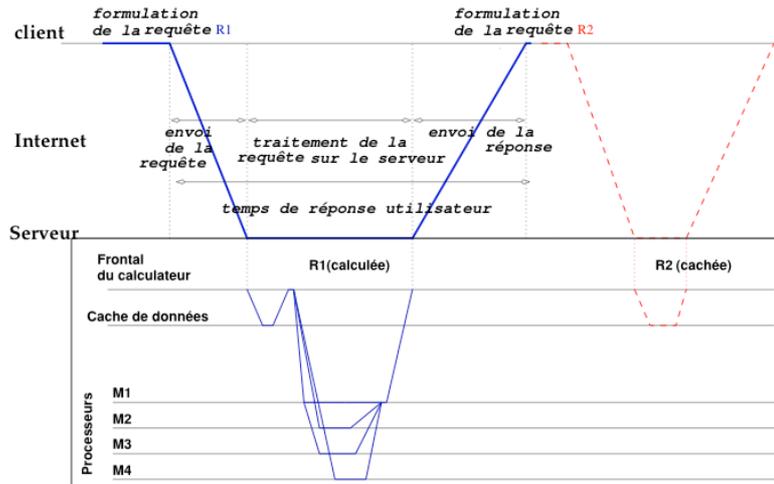


Figure 1 : décomposition de la latence.

L'objectif est donc de réduire le temps de réponse des requêtes utilisateur en jouant à la fois sur la parallélisation des calculs et sur l'utilisation de caches aux différents niveaux de la requête.

3.2. Stratégies d'accélération des calculs

L'amélioration du temps de calcul d'un ensemble de cartes se fait en deux étapes. Une première étape consiste à écrire un code parallèle de calcul. On se base sur le formalisme de graphe de tâches et chaque requête exprimée par l'interface est décomposée en tâches élémentaires pouvant être dépendantes (dépendance de données) mais dont certaines peuvent s'exécuter en parallèle (Edi et al., 2004).

La parallélisation des calculs repose essentiellement sur la décomposition spatiale des données. Dans le cas du module MTA, le calcul des ratios se fait par rapport à des valeurs de référence et donc la granularité du calcul se traduira par un niveau de l'arbre des données. Par exemple pour calculer un ratio de deux stocks, on

peut faire tous les calculs sur chaque unité territoriale en parallèle. Cependant, le surcoût de parallélisation (coût de gestion des tâches) est prohibitif. Il est préférable de faire le calcul en séquentiel pour toutes les unités d'une zone et de faire le calcul des zones en parallèle. Cette approche est d'autant plus intéressante que l'on peut affecter des priorités aux tâches en fonction des données à visualiser.

Lors de la décomposition de requêtes sous la forme de graphe de tâche (figure 2), nous avons remarqué que, lors d'une session, l'utilisateur envoyait des séquences de requêtes présentant des similarités fortes. En particulier, la traduction sous forme de graphe de tâches des requêtes montre une forte redondance des tâches. Celle-ci peut donc être exploitée au niveau du serveur de calculs multiprocesseurs.

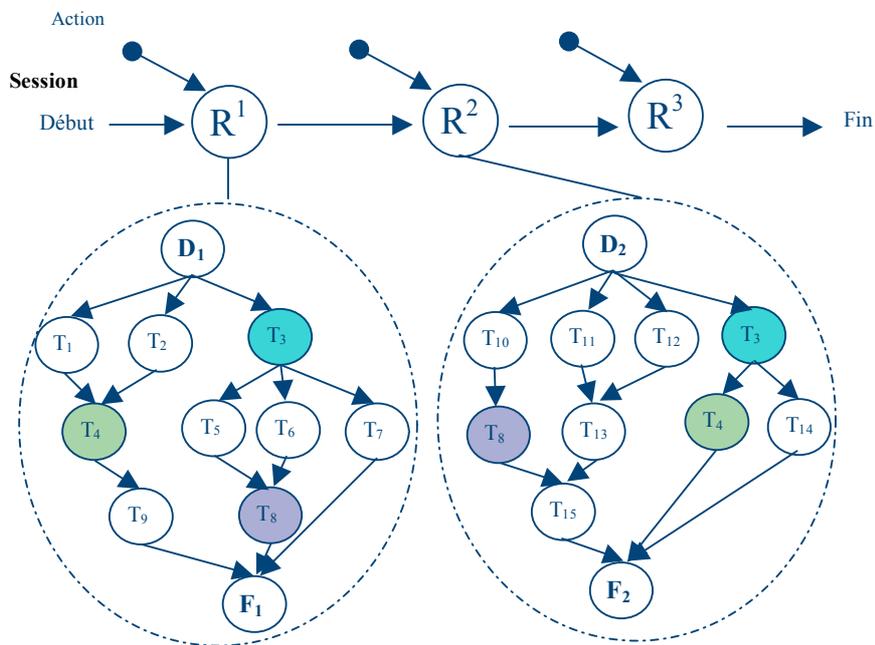


Figure 2 : *graphe des tâches*

Les caches peuvent être insérés à différents niveaux dans le traitement de la requête :

- au niveau de l'interface utilisateur pour les transformations simples liées à la visualisation,
- au niveau du réseau internet, proxy web à usage général,

- au niveau du frontal du serveur, traitement de requêtes presque identiques,
- au niveau du calculateur parallèle, analyse de redondance fine des tâches .

Nous nous donc sommes concentrés sur ce dernier niveau, sachant que pour les autres il existe des stratégies connues donnant de bons résultats. Il a été montré que le problème d'optimisation des temps de calculs d'un flot de requêtes avec usage de caches est un problème difficile (prouvé NP-dur) aussi, nous avons développé des schémas heuristiques permettant d'assurer une bonne utilisation à la fois des ressources en calcul et des ressources en volume de stockage (espace de cache).

4. Analyse Territoriale Multiscaleire

Dans cette partie, nous présentons les résultats de la partie du projet qui concerne l'Analyse Territoriale Multiscaleire (ATM) sur la base de l'analyse de la répartition du PIB par habitant dans les régions européennes. Ce module à spécifiquement pour objectif de proposer plusieurs cartographies d'un même phénomène: pour un découpage spatial ou maillage élémentaire donné d'un espace d'étude, il s'agit de décliner un indicateur mesurant le rapport entre 2 quantités (ici la richesse d'une région et la population résidant dans la région) relativement à trois niveaux de contextes territoriaux, que l'utilisateur choisit: un contexte global, un contexte intermédiaire et un contexte local.

Ainsi, les paramètres d'entrées originaux sont la définition des trois contextes spatiaux, qui vont servir tour à tour de référentiel aux valeurs absolues. On met alors en évidence des organisations différentes, pour un même phénomène observé à la lumière de ces trois filtres territoriaux. Ces représentations cartographiques, en se complétant, viennent donner une vue plus globale du phénomène. Une réflexion particulière a été menée sur la conception de l'interface et le choix des outils de cartographie interactive, de sorte que l'utilisateur soit guidé au mieux, mais contraint au moins, tout en restant dans les règles de la cartographie.

Actuellement, les logiciels existants ne permettent pas d'obtenir en quelques manipulations différentes cartes permettant une telle analyse. Les possibilités de l'architecture présentée précédemment pour développer un module spécifique (module ATM) permettant de générer une série de sept cartes associées de manière dynamique à un ensemble de paramètres choisis par l'utilisateur.

- les trois premières illustrent la répartition de l'indicateur choisi ainsi que des deux stocks dont il est issu,
- les trois suivantes sont les représentations du phénomène relativement aux trois contextes choisis,
- enfin, la dernière, synthétise pour chaque région, ses positions relatives par rapport aux trois contextes précédents.

4.1. Les entrées et les sorties

Dans le cadre du module ATM, les données nécessaires sont :

- la géométrie du maillage élémentaire,
- les statistiques associées aux unités élémentaires,
- le dictionnaire décrivant les emboîtements entre les unités,
- la définition du voisinage de chaque unité élémentaire,
- la définition d'espaces types (ici Union des 25, Pays de l'ex-Europe de l'est...).

A la fin de chaque session, l'utilisateur peut exporter sous forme d'un rapport (en format XHTML) l'intégralité des cartes créées, ainsi que les données calculées.

4.2. Les paramètres et représentations cartographiques associées

Le module ATM est composé d'une seule fenêtre, avec dans sa partie supérieure, trois groupes de paramètres (figure 3) : espace et maillage, indicateurs et contextes. A chaque paramètre sont associées une ou plusieurs représentations cartographiques. L'ensemble des paramètres est dynamiquement lié à la série des sept cartes qui sont proposées dans la partie basse de la fenêtre sous forme d'onglets. L'utilisateur peut jouer simultanément sur le niveau géographique d'observation (maillage élémentaire / elementary zoning), sur l'indicateur quantifiant le phénomène observé (indexés), et sur les différents contextes par rapport auxquels les quantités absolues vont être relativisées pour être interprétées en terme d'écart (contexts for the deviation).

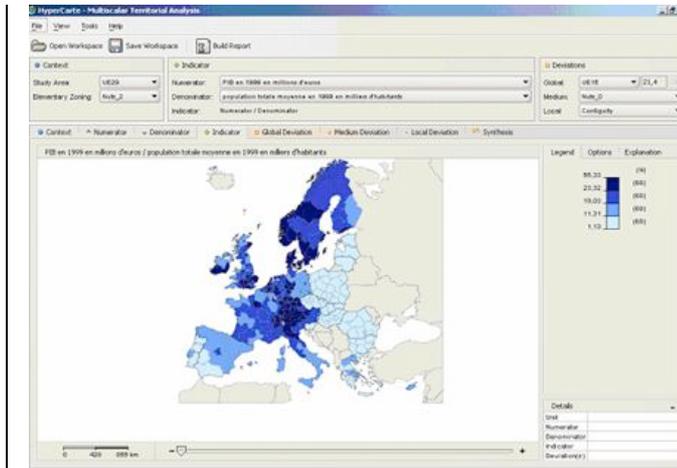


Figure 3: La fenêtre du module ATM

4.2.1. Définition de l'espace d'étude et du maillage élémentaire

La première boîte de dialogue propose de choisir le niveau des entités élémentaires sur lesquelles on veut travailler ainsi que l'extension spatiale du phénomène à représenter. La figure 4 présente différentes combinaisons de ces 2 paramètres. Chacun des paramètres est à choisir dans une liste prédéfinie lors de l'implémentation du module sur la base de données. Toutes les combinaisons ne sont pas proposées, certaines étant incompatibles. Au fur et à mesure des choix de l'utilisateur l'extension spatiale (grisé) et le maillage sont représentés sur le 1er onglet de la partie basse de la fenêtre.

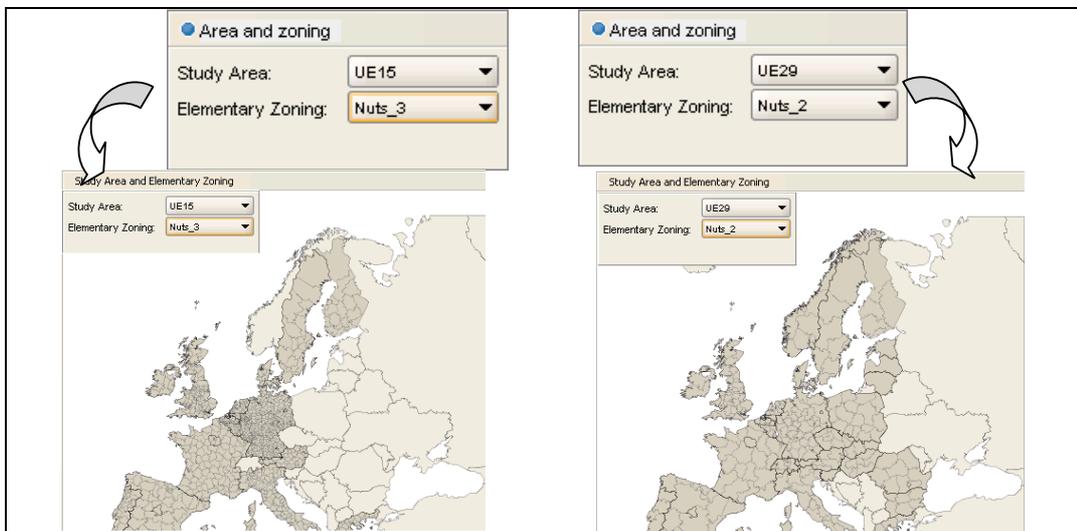


Figure 4 : *Différentes combinaisons d'espaces et maillages*

4.2.2. Les indicateurs

L'application ne travaille qu'avec des variables de stock (effectifs ou comptage). Le phénomène analysé est défini par le rapport entre deux stocks, illustrés par les 2 premières cartes en ponctuels de la figure 4. Elles donnent une idée de la répartition des masses La troisième est la représentation "classique" du rapport entre les deux.



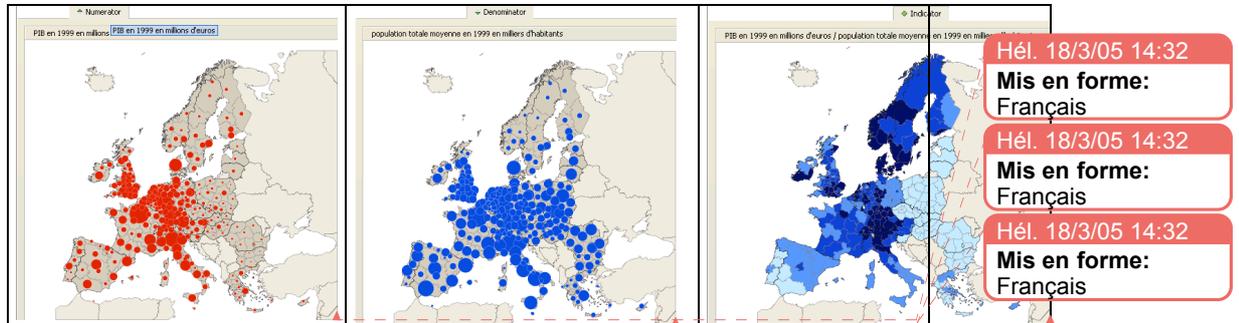
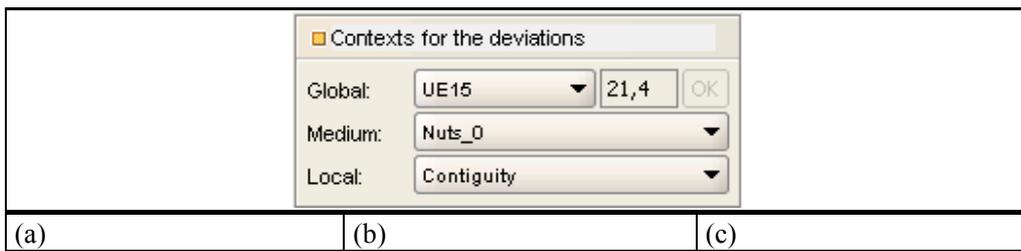


Figure 4 : Choix des indicateurs et cartes élémentaires associées aux valeurs absolues

4.2.3. *Les cartographies des valeurs relatives selon les différents contextes*

La troisième boîte de paramètres permet à l'utilisateur de construire une information relative (figure 5). La valeur mesurée en chaque unité élémentaire va être mise en relation avec trois niveaux géographiques.

- un niveau global: les valeurs associées aux unités élémentaires sont relativisées par rapport à une même valeur de référence (ici l'union des 15, fig. 5a).
- un niveau intermédiaire: C'est un maillage supérieur au maillage élémentaire choisi, ici les pays (NUTS0). La valeur d'une unité est mise en relation avec la valeur de l'entité de niveau supérieur à laquelle elle appartient. Dans le cadre de notre exemple, la figure 5b se lit donc comme l'écart des régions à la valeur nationale.
- un niveau local: Chaque unité est mise en relation avec son voisinage local. Ce point de vue permet de s'affranchir de la hiérarchie administrative des maillages, et d'identifier localement les minimum ou maximum au regard de l'indicateur choisi. La carte (5c) représente donc les écarts entre la richesse par habitant des unités et la richesse de leur voisinage. Le voisinage choisi ici est basé sur la contiguïté. On peut aussi envisager d'autres types de voisinage.



(a)

(b)

(c)

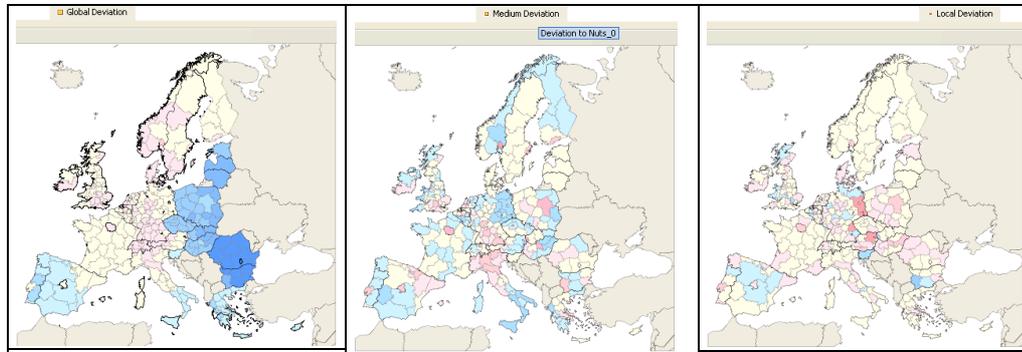


Figure 5 : Un indicateur, trois organisations spatiales relatives à trois contextes

4.2.4. La carte de synthèse

Une dernière carte propose une synthèse des trois précédentes cartes. Elle permet de classer les unités selon leur trois différentes positions relatives. La cartographie de cette synthèse nécessite un choix préalable de point de vue (figure 6). Cette carte s'actualise selon les choix précédents.

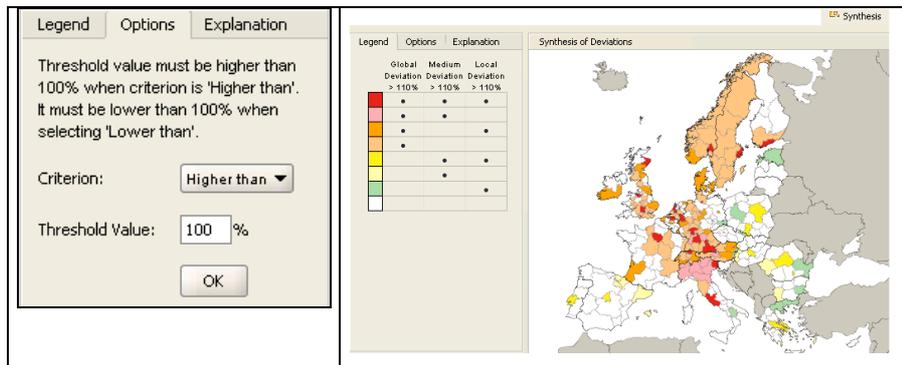


Figure 6 : La carte de synthèse

4.3. Les outils associés

Un certain nombre d'outils viennent compléter ces cartes. D'une part l'utilisateur dispose à tout moment des outils classiques associés au curseur tels que l'identification des unités (fenêtre information), le zoom, le déplacement... Mais associés à chaque carte, on trouve une série de trois onglets (figure 7) contenant la

légende, des outils permettant à l'utilisateur de faire ses propres choix cartographiques ("options"), et une aide à l'interprétation ("explanation").

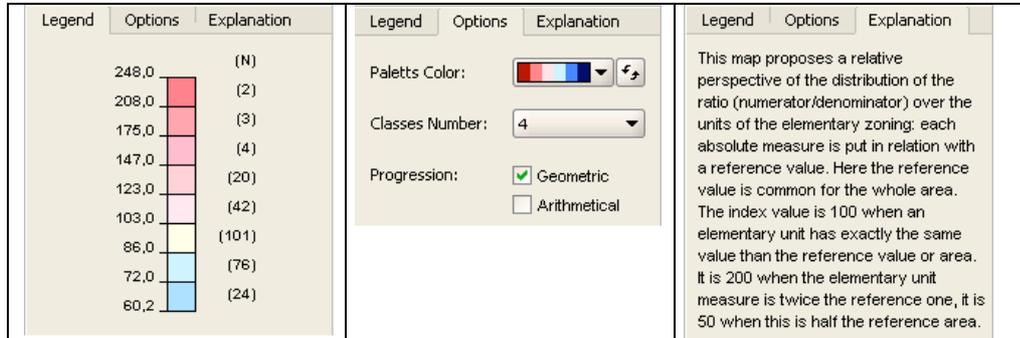


Figure 7 : Représentation cartographique et interactivité.

5. Conclusions et perspectives

Cet article traite du problème de la génération au vol de cartes sur le Web. Le problème est rendu difficile par le volume de données traitées et la nature des méthodes utilisées. Les méthodes utilisées dans HyperCarte, reposent essentiellement sur la formalisation de deux types de relations entre les unités :

- des relations horizontales: ce sont les relations entre unités de même niveau. Elles sont ici associées à des relations de voisinage, basées sur des mesures de proximités ou de contiguïté. Ces relations formalisent les interactions potentielles entre les unités, liées à la proximité entre elles.
- des relations verticales entre les unités: relations d'appartenance à des entités de niveaux supérieurs illustrant des relations d'emboîtement entre les différents maillages. Elles formalisent par exemple les relations d'administration des entités de niveau supérieur sur celles de niveau inférieur, et réciproquement, la "participation" des entités de niveau inférieur à la composition d'une entité de niveau supérieur.

Dans les outils spécialisés dans la gestion et la représentation des données spatiales, il est bien entendu possible de formaliser de telles relations. Mais elles doivent être formalisées dans le cadre d'un modèle plus général, sans être prises en compte en tant que telles. Par exemple une appartenance à un niveau d'agrégation supérieur est représentée comme un champ qualitatif. En revanche les relations de voisinage se représentent plus difficilement. Dans les outils appelés "d'analyse spatiale" (jointure spatiale, buffers) dans les SIG, il est possible d'exprimer des relations entre les entités, mais il n'est pas simple de les calculer de manière exhaustive. Ainsi, quel que soit le type de relation, leur utilisation demande un apprentissage minimal de l'ensemble du fonctionnement de l'outil pour in fine une sous-utilisation de la majeure partie des fonctionnalités. De plus, l'esprit de la

méthodologie proposée, s'inscrit dans une démarche exploratoire, qui demande une grande facilité pour répéter les cartographies et une grande réactivité des représentations cartographiques aux différents choix. Ainsi on se situe plus dans le cadre d'une démarche de GéoVisualisation que de gestion de données spatiales.

Afin de répondre de manière satisfaisante aux problèmes posés, nous avons proposé une architecture parallèle qui est couplée avec une interface cartographique développée en Java. Cette approche est également en cours d'évaluation sur d'autres territoires (Tunisie, Cameroun) et sur d'autres méthodes d'analyse spatiale telle que les méthodes de lissage par la méthode de potentiel (Grasland et *al.*, 2000).

7. Bibliographie

- Antoniou B. , Tsoulos L., « Converting raster images to XML and SVG », *3rd Annual Conference on Scalable Vector Graphics*, Tokyo, Japan, 2004.
- Edi E., Trystram D., and Vincent J-M. "A new cache approach based on graph for web server" In *Parallel Computing System*, pages 393–398, Collima, Mexico, September 2004. IEEE, Computer Society.
- Grasland C., Lizzi, L., Martin, H., Mathian, H., Vincent, J.M., "Hypercarte : un outil d'analyse spatiale multiscalair des inégalités régionales en Europe". XXXIXème colloque de l'Association de Science Régionale de Langue Française, Lyon, France, 1, 2 et 3 Septembre 2003.
- Grasland C., Mathian, H., Vincent, J.M., " Multiscalar analysis and map generalization of discrete social phenomena: statistical problems and political consequences". *Statistical journal of the European Community* 2000.
- Josselin D., Fabrikant S., *Cartographie animée et interactive*, Revue Internationale de Géomatique, Editions Hermès, 2003.
- JTS JTS Topology Suite, 2004, <http://www.vividsolutions.com/JTS/JTSHome.htm>
- Kraak M-J., Brown A. (eds), *Web Cartography*, Taylor&Francis, London, 2000
- MAcEachren A.M, Taylor D.R.F, (eds) *Visualization in modern cartography*, Oxford, 1994
- MapServer 2004, site web de MapServer. <http://mapserver.gis.umn.edu/>.
- OGC Open GIS Consortium, 2003. OpenGIS® Geography Markup Language (GML) Implementation Specification, Version 3.0. Spécification de GML, 2003.
- SVG 1998, spécifications du langage Scalable Vector Graphics, <http://www.w3.org/Graphics/SVG/>
- Vincent M., Adcock T., « Implementing an integrated SVG application for real time, dynamically generated Internet mapping», *3rd Annual Conference on Scalable Vector Graphics*, Tokyo, Japan · Sept 7-10, 2004.