

HyperSmooth : calcul et visualisation de cartes de potentiel interactives

Christine Plumejeaud*, Jean-Marc Vincent*, Claude Grasland**, Jérôme Gensel*,
Hélène Mathian**, Serge Guelton*, Joël Boulrier**

* Laboratoire d'Informatique de Grenoble, BP 72, 38402 Saint-Martin d'Hères
{prénom.nom}@imag.fr

** UMR Laboratoire Géographie-Cités, 13 rue du four, 75006 Paris
{prénom.nom}@parisgeo.cnrs.fr

Résumé. Le groupe de recherche Hypercarte propose HyperSmooth, un nouvel outil cartographique pour l'analyse spatiale de phénomènes sociaux économiques mettant en œuvre une méthode de calcul de potentiel. L'objectif est de pouvoir représenter de façon continue et en changeant d'échelle d'analyse une information statistique échantillonnée sur toutes sortes de maillages, réguliers ou non. Le défi technologique est de fournir un outil accessible sur le Web, interactif et rapide, ceci malgré le coût élevé du calcul, et qui assure la confidentialité des données. Nous présentons notre solution basée sur une architecture client serveur : le serveur calcule les cartes de potentiel en utilisant des techniques d'optimisation particulières, alors que le client est en charge de la visualisation et du paramétrage de l'analyse, et les deux parties communiquent via un protocole Web.

Introduction

Les avancées dans le domaine du Web ont ouvert de nouvelles perspectives dans le domaine de la cartographie interactive et dynamique (Koben, 2001, Josselin et Fabrikant, 2003). Aujourd'hui le web propose de multiples cartographies qui s'adaptent aux besoins d'un utilisateur, qui peut être, tour à tour, décideur, citoyen, voyageur... Les services d'itinéraires routiers en sont un exemple typique. Dans ce contexte, le groupe de recherche pluridisciplinaire HyperCarte¹ s'est donné pour objectif de concevoir et implémenter une collection cohérente de plates-formes interactives d'analyse spatiale et de représentations cartographiques de phénomènes sociaux, économiques, environnementaux, etc. (Grasland et al., 2005-b). Les applications visées se situent principalement dans le domaine socio-économique ou

¹ <http://www-lsr.imag.fr/HyperCarte/>

environnemental et l'aide à la décision en matière de prospective territoriale pour un public large : chercheurs en géographie, en sciences sociales et humaines, mais aussi décideurs politiques et grand public. L'enjeu est de proposer une cartographie s'adressant à des utilisateurs aux compétences diverses, néophytes ou spécialistes pour une cartographie exploratoire (Antoni et al, 2004). 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'exposé de cet article concerne plus précisément la réalisation d'un environnement, basé sur l'infrastructure du Web, capable de générer dynamiquement des cartes continues, c'est-à-dire affranchies d'éventuels maillages d'observation (administratifs, grilles de collecte, etc.). L'objectif est de visualiser la distribution spatiale des phénomènes analysés à un niveau macroscopique, c'est-à-dire largement supérieur aux maillages d'observation. La méthode que nous proposons, appelée *méthode de transformation par potentiel*, conserve la masse totale des données et en donne une représentation non biaisée. Définie empiriquement dans les années 1990 par les travaux de l'UMR Géographie-cités sur les conséquences de la réunification allemande (Grasland, 1991) et de la chute du mur de Berlin en Europe (Boursier-Mougenot et al. 1993), l'étude de cette méthode en vue de la création d'un outil d'analyse spatiale multiscalaire a abouti à une publication méthodologique de référence (Grasland et al., 2000). Le lissage est un cas particulier d'application de cette méthode globale, lorsque la portée du potentiel est petite.

Une des difficultés de diffusion de cette méthode d'analyse spatiale et de représentation cartographique vient, d'une part du coût élevé du calcul² qui empêche de répondre à l'exigence d'interactivité et, d'autre part de la contrainte sur la sécurité et la confidentialité des données analysées.

Cet article présente la solution conçue pour répondre à ces besoins. Elle s'est développée suivant trois axes : (i) distribution des calculs sur un serveur multi-processeur détenant les données, (ii) parallélisation des tâches de calcul, (iii) visualisation des cartes sur un client web interactif avec connexion sécurisée.

Cette présentation s'organise autour du plan suivant : dans un premier temps, les fondements théoriques et les applications possibles de la méthode du potentiel sont rappelés, et cette méthode est située en regard de ce qui existe. Suivent ensuite la justification de l'architecture retenue et la description de la stratégie d'optimisation des calculs. Enfin, les résultats de cette réalisation sont présentés ainsi que les perspectives d'amélioration qu'elle offre.

² Il faut plusieurs heures pour calculer une carte de résolution moyenne visualisant la densité de population en France, à partir du recensement sur les 36000 communes.

2 La méthode de transformation par potentiel

2.1 Principe

Les représentations cartographiques continues de phénomènes spatiaux discrets sont nécessaires lorsque l'on souhaite s'abstraire d'un maillage spatial, parce que le maillage est hétérogène ou qu'il n'est pas signifiant pour le phénomène étudié, afin de ne garder que l'organisation spatiale du phénomène, sans référence au découpage sous-jacent du territoire.

La méthode considère un certain espace géographique, sur lequel est plaqué un maillage constitué d'unités territoriales auxquelles sont associés des stocks, c'est-à-dire des variables statistiques résultant de dénombrement, comme la population, le nombre d'actifs, etc. Le maillage territorial peut être, par exemple, l'ensemble des communes françaises et leur équivalent européen dans la nomenclature NUTS de niveau 5, adoptée par l'*European Spatial Planning Observation Network*. Ce type de maillage s'emboîte généralement dans un maillage de plus haut niveau, par exemple le maillage départemental (NUTS-3) : un département est composé d'un sous-ensemble de communes. Les indicateurs considérés pour cette méthode possèdent une propriété additive, c'est-à-dire que la valeur d'un stock d'une maille de niveau supérieur est égale à la somme des valeurs des mailles qui la composent au niveau inférieur. L'objectif est de calculer en tout point de l'espace discrétisé la valeur du potentiel de chaque stock. La discrétisation est une division de l'espace en parcelles régulières, formant une grille, par exemple. En tout point de cet espace géométrique, la valeur du potentiel doit être comprise comme la valeur probable de l'indicateur considéré, qui dépend de la contribution de chaque maille de l'espace géographique, pondérée par sa distance au point calculé.

Si on note A l'ensemble des unités territoriales, a un élément de cet ensemble, S_a la valeur du stock sur cette unité, alors sachant que les effets des stocks s'additionnent et sont liés à la distance δ entre a et le point M , le potentiel $\Phi(M)$ est défini en tout point M de l'espace géométrique par :

$$\Phi(M) = \sum_{a \in A} S_a f(\delta(a, M)). \quad [1]$$

Par exemple, considérons A , l'ensemble des communes européennes, et S le nombre d'habitants centenaires par commune, et supposons que l'on veuille calculer le potentiel de centenaires en tout point de l'espace. Pour chaque commune a , S_a est le nombre de personnes centenaires, et g_a le centre de la commune.

On spécifie alors la distance $\delta(a, M)$ en mesurant la distance d entre M et g_a , un point représentatif de a (qui peut être son centre de géométrie, son centre administratif ou industriel, etc.). La contribution au potentiel de chaque élément a est pondérée par une fonction f de la distance d , car l'effet d'un stock diminue usuellement avec la distance : il est maximal à une distance nulle et nul à une distance infinie. Pour que le potentiel ait du sens, en particulier lorsqu'il dépendra d'un paramètre, on normalise celui-ci par l'équation [2], en tout point O de l'espace :

$$\int_{R^2} f(d(O, M)) . dM = 1, \text{ soit encore } \int_A \Phi(M) . dM = \sum_{a \in A} S_a . \quad [2]$$

La somme totale des stocks est égale à l'intégrale du potentiel, on obtient ainsi une redistribution de la masse sur l'espace considéré.

Dans une métaphore du modèle de gravité, $\Phi(M)$ s'interprète comme l'attraction exercée par l'environnement sur un mobile placé en M , dont le vecteur de déplacement serait alors $-\text{grad } \Phi$. Une interprétation duale serait aussi que $\Phi(g_a)$ mesure l'influence d'une masse placée en g_a sur l'ensemble des points M de son voisinage. Par exemple, le nombre de centenaires vivant à Nuoro, une ville de la Sardaigne, contribuera beaucoup plus à l'estimation d'un point M situé dans son voisinage que les centenaires habitant Rome. Du point de vue méthodologique, la méthode s'apparente aux méthodes de traitement du signal par déconvolution du signal échantillonné (Grasland, Vincent, 2006).

De l'équation [1], il découle que la complexité du calcul dépend à la fois de la taille de l'espace administratif (le nombre n d'éléments a), et de la résolution de l'image à produire (le nombre m de points M que l'on estime).

Le calcul dépend principalement de la fonction f , **fonction d'interaction spatiale**. Elle intègre les hypothèses concernant les lois de diffusion dans l'espace associées au phénomène étudié. Trois modèles de fonction paramétrée sont proposés : un modèle à support limité (disque et disque amorti), un modèle exponentiel (Gaussienne) pour des interactions proches - la décroissance de f se fait alors selon une exponentielle négative - ; et enfin un modèle à interaction de longue portée (Pareto) - la décroissance suit une puissance inverse. Cette méthode permet par exemple de modéliser et d'étudier la propagation d'épidémies : leur diffusion pourra se faire soit sur de longues distances, soit sur de courtes distances, suivant le rayon d'action de l'élément contaminant (Grasland *et al.*, 2005-a). L'utilisateur peut tester différents modèles en choisissant la fonction d'interaction à appliquer.

L'analyse du phénomène dépend aussi de la **portée** p de la fonction d'interaction. La portée est définie comme la distance moyenne d'action d'une masse sur son voisinage. Elle est reliée à la forme de la fonction d'interaction par l'équation suivante en tout point O de l'espace (après passage en coordonnées paramétriques, r représente le rayon) :

$$p = \int_{R^2} d(O, M) f(d(O, M)) . dM = \int_0^{+\infty} f(r) 2\pi r^2 . dr . \quad [3]$$

La portée peut être interprétée comme l'échelle spatiale de représentation choisie. Le tandem (fonction, portée) traduit concrètement les hypothèses économiques et sociologiques associées aux interactions entre les acteurs sur le territoire. À portées identiques, le calcul du ratio de deux potentiels de stocks différents s'interprète comme une densité. Par exemple, le potentiel Φ_p de population peut se rapporter au potentiel Φ_s de superficie, ce qui nous donne une densité de population sur une portée donnée et uniforme. (Voir figure 1).

D'autre part, il faut choisir le type de **distance** utilisée. Celle-ci dépend de la taille de l'espace couvert par la carte. Par exemple, à l'échelle d'un continent, on ne peut pas utiliser la distance euclidienne sans introduire des déformations. La

distance orthodromique convient alors mieux car elle tient compte de la sphéricité de la Terre. En outre, le type de phénomène analysé peut nécessiter parfois d'utiliser des distances tenant compte de l'anisotropie de l'espace (distance temps voiture, par exemple).

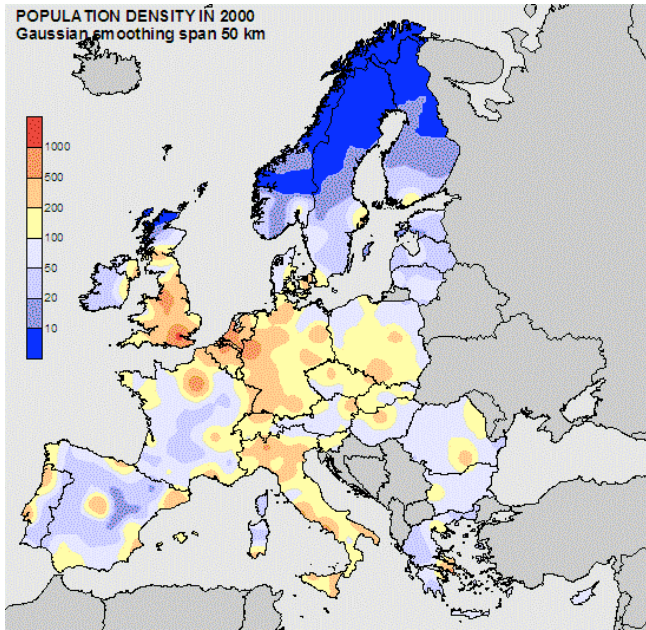


FIG. 1 - Cartographie de la densité de population en Europe en 2000, calculée par une fonction gaussienne avec une portée de 50 km.

2.2 Deux exemples d'application empirique

Nous avons déjà dressé un récapitulatif concernant des usages relativement classiques de la méthode, (Plumejeaud et al., 2007). Nous présentons ici des usages plus complexes, mettant en œuvre une modélisation de flux associés aux différentiels locaux, ou des distances non-isotropiques avec des effets de barrière.

Le premier exemple concerne la recherche d'une mesure des polarisations locales en Europe à partir d'une information très pauvre (population et superficie des communes en 1999) mais très détaillée spatialement, correspondant au niveau communal, (Dubois, Gloersen & al. 2007). Les communes au niveau européen sont particulièrement hétérogènes tant en superficie qu'en population. Une analyse des polarisations sur la base des densités de population communales risque d'une part d'être biaisée par les effets de maillage (*Modifiable Area Unit Problem* ou MAUP, cf. §2.3) et d'autre part de souffrir de l'hétérogénéité de la distribution de population sur l'ensemble de l'Europe. Par exemple, une densité de population de 100 habitants par kilomètre carré correspondra à un très fort niveau de concentration dans les

HyperSmooth : calcul et visualisation de cartes de potentiel interactives

périphéries scandinaves et au contraire à un niveau faible dans l'espace de forte concentration démographique de l'axe rhénan.

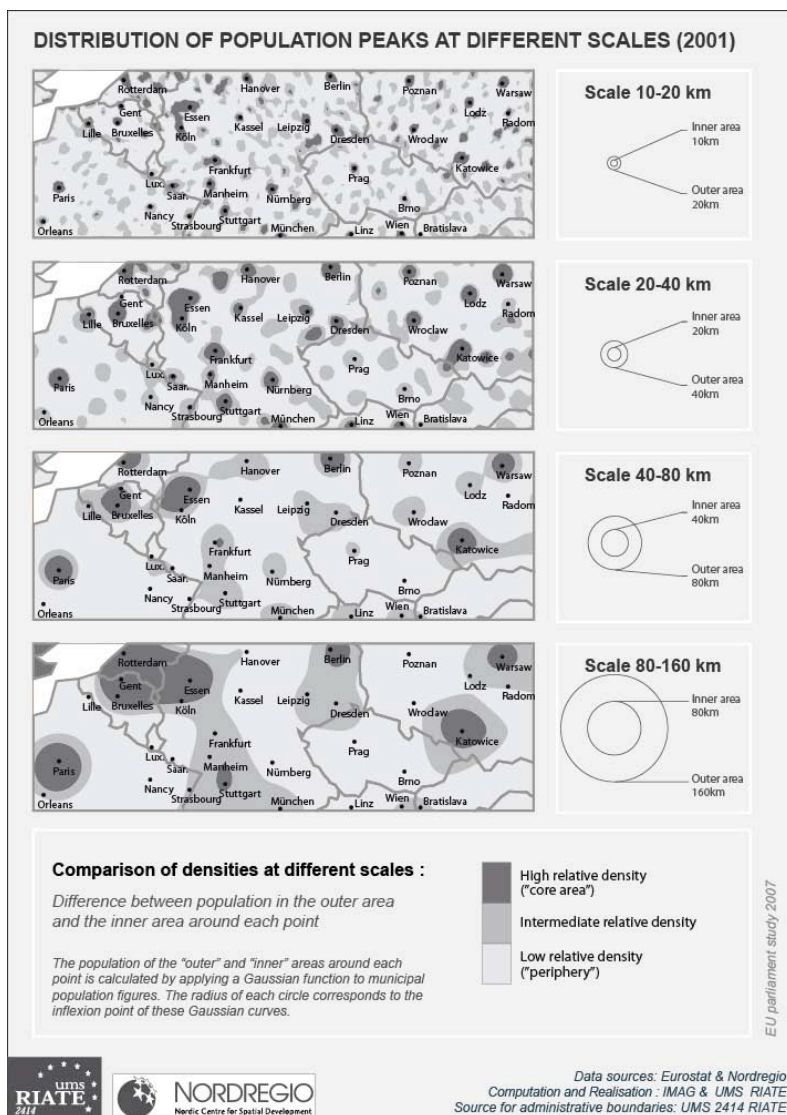


FIG. 2 – Analyse multiscale des pics de densités de population en Europe.
Source (Dubois, 2007).

Le calcul des potentiels de population et de superficie pour des voisinages gaussiens de portée croissante (10, 20, 40, 80, 160 km) a permis de résoudre le premier problème (élimination de l'hétérogénéité liée au découpage initial) tandis que la comparaison des densités de population dans des voisinages de portées successives a permis de résoudre le second problème en repérant les zones localement plus denses que les espaces environnants. La série de cartes ainsi obtenues (voir figure 2) permet de mettre en valeur les pics relatifs de concentration de population à différents niveaux de généralisation et d'observer comment des pics locaux (différentiels de densité entre les lissages à 10 et 20 km) fusionnent progressivement dans des pics régionaux ou globaux. Ce résultat est fondamental en terme d'aménagement du territoire européen car il permet de donner une formalisation objective d'un concept politique central qui est celui du polycentrisme. Même si celui-ci n'est envisagé ici que sous une forme morphologique (l'approche devant également être menée en termes de flux et de réseau) la méthodologie proposée ouvre de très importantes perspectives pour l'identification de réseaux de villes et de coopérations transfrontalières.

Le second exemple concerne l'étude de la vulnérabilité des régions européennes face à la mondialisation³. La méthode des potentiels a été utilisée en tenant compte de l'anisotropie de l'espace puisque les distances ont été mesurées en temps routier et que d'éventuels effets de barrières liés au franchissement des frontières internationales ont été introduits. A partir d'un indicateur de vulnérabilité connu pour l'ensemble des régions européennes (part des actifs dans les secteurs industriels menacés tels que le textile, l'électronique et la mécanique), la vulnérabilité de chaque région a été comparée à celle des régions voisines en tenant compte d'une part, de leur accessibilité routière, (matrice de distance temps), et d'autre part, du fait que ces régions voisines étaient situées dans le même pays ou dans un pays voisin. Nous avons ainsi défini 7 niveaux de vulnérabilité correspondant à une discrétisation de l'espace autour de la région elle-même selon différents temps de trajets (moins de 2h, entre 2 et 4h, entre 4 et 8h), y compris le franchissement d'une frontière. Comme l'illustre la figure 3, on identifie ainsi la région elle-même (Reg), une auréole de régions situées à moins de 2h du même pays (N1nat) ou d'un pays voisin (N1int), puis une deuxième zone de régions situées entre 2 et 4h (N2nat, N2int) et enfin une troisième zone de régions situées entre 4 et 8 h (N3nat, N3int). Les seuils de 2h, 4h et 8h ont été choisis pour tenir compte des différents points de vue des voyageurs considérés :

- pour les travailleurs : 2h représente le temps maximal acceptable pour se rendre à son travail
- pour les entreprises : 4h représente le temps maximal acceptable pour effectuer un aller-retour dans une journée
- pour les fournisseurs : 8h représente le seuil maximum de transport franchissable en une journée.

³ Etude en cours de réalisation pour la DG Région coordonnée par l'IGEAT-ULB, http://www.ulb.ac.be/igeat/igeat/ulb_igeat/hp/hp_fr.htm

La distinction national/international est quant à elle fondamentale en raison des politiques plus ou moins protectionnistes de chaque pays en matière de compétition industrielle et d'ouverture du marché de l'emploi.

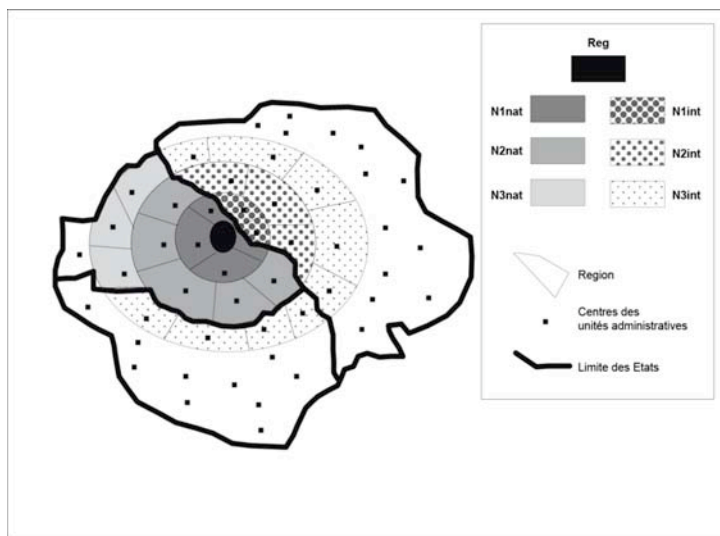


FIG. 3 – Définition de voisinages tenant compte des temps de trajets et des franchissements de frontières.

Les figures 4 et 5 présentent les résultats pour deux régions typiques permettent de saisir l'intérêt de l'approche ainsi proposée par rapport à la thématique de la vulnérabilité régionale face à la mondialisation. La région française des Ardennes et la région italienne de Modène présentent grosso modo des niveaux voisins de vulnérabilité, chacune ayant environ 15-16% de ses emplois dans des secteurs industriels menacés. Toutefois, elles s'inscrivent dans des contextes radicalement différents si on examine la situation dans leurs différents voisinages fonctionnels comme définis précédemment.

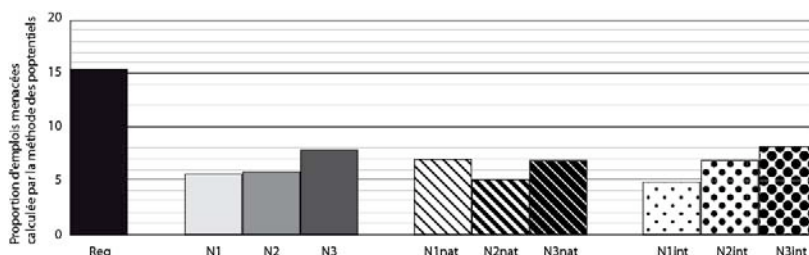


FIG. 4 - Situation de la région des Ardennes (France).

La région des Ardennes (voir figure 4) est un pic isolé de vulnérabilité, entouré de régions qui ont très peu d'emplois dans les secteurs vulnérables, aussi bien à l'intérieur du territoire français que dans les espaces voisins de Belgique, Luxembourg, Belgique ou Pays-Bas. Une crise des secteurs industriels vulnérables pourra donc être compensé par la recherche d'emplois dans les régions voisines. Mais les entreprises touchées par la crise ne pourront pas compter sur une solidarité des territoires voisins.

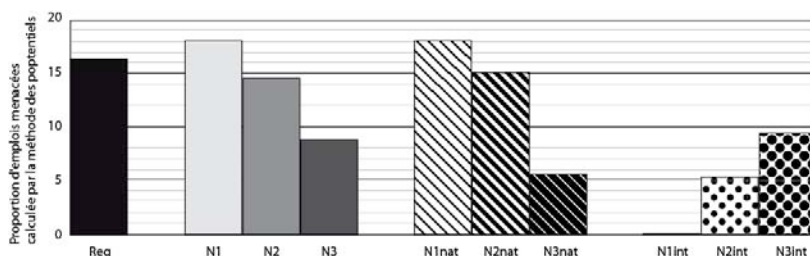


FIG. 5 - Situation de la région de Modène (Italie).

La région de Modène (voir figure 5) se situe au contraire à l'intérieur d'un véritable bastion de régions vulnérables dans un cadre qui est essentiellement national (aucune région étrangère à moins de 2h de route). Les régions les plus proches (moins de 2h de route) sont d'ailleurs encore plus vulnérables que celle de Modène. Plus on s'éloigne, plus la vulnérabilité diminue. À l'inverse de la région des Ardennes, les employés des secteurs vulnérables risquent d'avoir des difficultés à trouver un emploi dans les régions voisines si la crise frappe l'ensemble du secteur. D'un autre côté, les entreprises pourront plus facilement faire bloc à l'échelle de plusieurs régions voisines...

2.3 Situation par rapport à d'autres méthodes

Les vertus de la méthode ne s'expriment pas sur un maillage régulier (taille des unités quasi-égales, et distance relative des centres toujours identique), mais au contraire lorsque l'on traite des données issues de recensements sur des maillages irréguliers. En effet, la forme et l'échelle du support des données jouent un rôle très important sur l'estimation produite : ces facteurs sont la source de problèmes identifiés comme le *Modifiable Areal Unit Problem* (MAUP) ou plus généralement le *Change Of Support Problem* (COSP), qui ont donné lieu à divers développements méthodologiques et discussions sur leurs apports et contraintes respectives [Gotway et al., 2002], et qui sont liés à la nature des données traitées.

Le premier facteur d'erreur est un effet de l'agrégation : les résultats de l'estimation de la répartition spatiale des données varient en fonction de la taille des unités. Par exemple, le traitement d'un jeu de données communales au niveau de son découpage initial ne donne pas des résultats identiques à celui qui serait fait en regroupant les données au niveau départemental. Ce qu'on appelle communément

l'« effet d'échelle » (« *scale effect* ») s'explique par le fait que la variance de l'échantillon initial (communal dans l'exemple) est mathématiquement plus élevée que la variance de la moyenne de ces échantillons. Ce qui signifie que sur des maillages fins, on perçoit une hétérogénéité des valeurs plus forte que sur des maillages plus grossiers, effets du pour partie à un simple effet mécanique, et pour partie à l'organisation du phénomène à cet échelon. Le second facteur, dénommé « effet du découpage » (« *zoning effect* »), dérive du mode de regroupement des données à une échelle fixée : on prouve que la corrélation entre des données spatiales varie en fonction du découpage territorial (comparaison faite entre entités de surfaces comparables) (Openshaw et al., 1979).

Les méthodes classiques d'interpolation telles que triangulation, moyenne locale, interpolation polynomiale (splines, Bezier) ou méthode de Shepard subissent les effets du MAUP : elles sont bien adaptées à l'estimation de variables continues dans l'espace pour lesquelles il existe des points de mesures (comme la température, ou l'altitude), mais ne le sont pas pour des variables résultant d'un comptage sur une zone délimitée à l'intérieur de laquelle la répartition effective de la population est inconnue. Notre démonstration (Grasland, et al, 2006) s'appuie sur la comparaison de cartes obtenues à partir d'un maillage de niveau NUTS 2 ou NUTS 3 en employant la méthode de Shepard. Elles ne font pas disparaître le maillage sous-jacent, et au contraire, elles pourraient faire croire que les disparités observées sont le fait du phénomène étudié, alors qu'en réalité, elles sont la résultante de l'hétérogénéité du maillage. Ces méthodes sont les plus usuellement implémentées dans les SIG.

Pour solutionner le problème du COSP, des auteurs ont discuté les avantages des méthodes géostatistiques (krigeage) qui modélisent la variabilité spatiale des observations, et utilisent ces résultats pour inférer une surface continue, (Journel et al., 1978). Ces méthodes ont l'avantage de fournir une mesure sur l'incertitude des résultats. Elles s'implémentent via un filtre de Kalman, et sont implantées dans de nombreux SIG. A contrario, la méthode du potentiel propose un modèle de diffusion a priori défini par un couple (fonction, portée) proposé par l'utilisateur, et la structure spatiale du phénomène n'entre pas en compte. Cependant si les méthodes géostatistiques ont un caractère prédictif efficace, elles restent des méthodes locales qui contrairement à la méthode du potentiel ne permettent pas de prendre en compte l'impact de toutes les mesures sur un point donné de l'espace géographique. D'autre part, les modèles géostatistiques doivent être simplifiés lorsque la quantité de données à traiter est importante.

Les méthodes basées sur les modèles bayésiens hiérarchiques sont aussi des solutions envisageables pour l'estimation de variables dans le cadre du MAUP. Les hypothèses sous-jacentes sont que la variable à estimer dépend (et ceci s'exprime sous forme de probabilité conditionnelle) d'autres variables dont on fournit une loi de dispersion spatiale (loi de Poisson, ou Gaussienne par exemple), et dont les données de recensement sont fournies. Il est ainsi possible de modéliser des interactions complexes entre variables pour affiner l'estimation (Wickel, 2002). Ce type de méthodes s'inscrit donc dans des démarches plus explicatives que celle que nous proposons. Le cadre proposé pour la méthode du potentiel ne nécessite pas de statistiques auxiliaires pour l'estimation des variables d'intérêt.

La méthode la plus similaire est la méthode pycnophylactique (Tobler, 1979) qui, elle aussi, conserve la masse des données. Elle réalloue les stocks mesurés à l'intérieur de chaque maille de façon à :

- (i) garantir la continuité avec une maille voisine,
- (ii) conserver sur chaque maille la masse mesurée.

La courbe continue de la figure 6 montre comment peut se passer le réajustement de la répartition des stocks à l'intérieur de chaque unité. C'est un cas simple, on peut employer des méthodes d'ajustement plus complexes où les surfaces obtenues sont au moins dérivables, voire de classe supérieure. Rase donne des développements intéressants de cette méthode (Rase, 2001). Mais en raison du prédicat (ii), la visualisation dépend alors du maillage utilisé pour l'analyse : les résultats sont différents selon le niveau d'agrégation des données, alors que ce n'est pas le cas avec le potentiel.

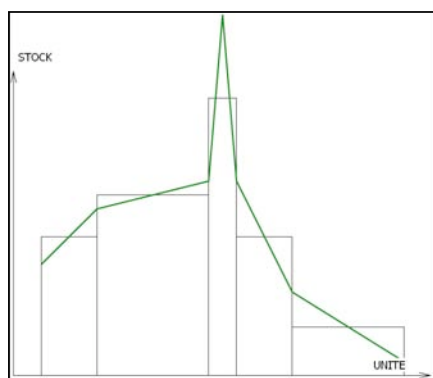


FIG 6. Réajustement de la répartition des stocks sur des unités d'aires voisines via la méthode pycnophylactique.

La méthode des potentiels s'apparente aux méthodes de traitement du signal par transformation de Fourier, visant à découvrir la structure du signal échantillonné et ceci indépendamment de la fréquence d'échantillonnage. Dans le cadre de l'analyse spatiale, cette fréquence doit être comprise comme la mesure de la régularité des recensements dans l'espace, évaluée via la distance entre chacun des centres administratifs des unités spatiales du maillage support de l'étude. Pour cette raison, la méthode des potentiels s'inscrit dans la lignée des méthodes proposées pour solutionner les problèmes d'instabilité des résultats, notamment les corrélations, en fonction de la résolution spatiale choisie.

Cependant, en dessous d'une certaine portée, la méthode devient imprécise ; la portée minimale peut être calculée par l'analogue du théorème de Nyquist : elle vaut deux fois la taille maximale des mailles (Nyquist, 1928). Enfin, un maillage trop hétérogène conduit à faire un compromis entre les portées minimum associées à chaque classe de taille d'unité.

Par ailleurs, la méthode de transformation par potentiel n'intègre pas encore les effets de barrières (cet aspect fait partie de nos futurs travaux de recherche). Par exemple, les montagnes sont des barrières géographiques qui jouent un rôle dans la dispersion spatiale des phénomènes étudiés. De même, la mer est un espace géographique non constructible, et en ce sens, les côtes maritimes pourraient être modélisées comme des barrières infranchissables lorsque on étudie la densité d'habitat.

3 Réalisation d'un prototype

Un prototype a été réalisé pour des utilisateurs disposant de données socio-économiques géo-référencées, des géographes par exemple, et désireux de tester leurs hypothèses de diffusion spatiale des phénomènes étudiés. Pour l'instant, ces analyses utilisent la distance orthodromique, qui permet de travailler sur de vastes espaces d'étude, et requiert un format de données suffisamment simple. Le prototype se base sur une architecture distribuée client serveur.

En effet, le volume de données à traiter et les ressources exigées pour les calculs sont, de fait, importants. Par conséquent, la partie graphique (qui inclue la visualisation et la configuration de l'analyse) est déportée sur un client Web Java, tandis que les calculs et le traitement lourd de données sont effectués sur un serveur accessible à distance, via le protocole SOAP (*Simple Object Access Protocol*) avec sécurisation des échanges via SSL (*Secure Socket Layer*). Le traitement des données coté serveur est optimisé de façon à produire des résultats intermédiaires dans des temps n'excédant pas quelques secondes. La figure 7 donne une vue globale de cette architecture.

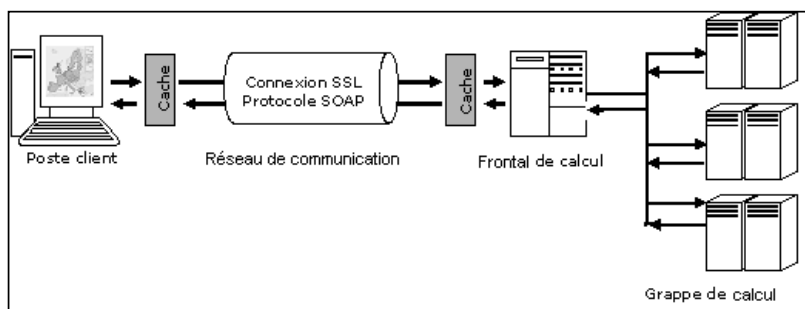


FIG. 7 - Vue générale de l'infrastructure distribuée.

L'infrastructure de gestion de données, d'authentification et de gestion d'erreurs restant encore à développer, l'installation des fichiers de données dans un répertoire dédié à l'utilisateur se fait manuellement du côté du serveur. Le format du fichier de données est le suivant : il contient une liste des points de mesure connus par leur latitude et leur longitude, et la valeur du stock associé, et il est nommé d'après le nom du stock. Le client Web est fourni à l'utilisateur avec le certificat de sécurité adéquat, avec le fond de carte approprié qu'aura fourni l'utilisateur au format

MIF/MID. Ainsi, il peut utiliser librement et à tout moment l'infrastructure, sur l'espace d'étude qui l'intéresse. Le serveur de calcul est installé sur une machine de type SMP, dont les ressources sont partagées par d'autres projets de recherche, et qui est accessible à travers cette connexion Web sécurisée. Le serveur est prévu pour le partage de charge dynamique, il n'y a donc pas en théorie de limitations sur le nombre d'utilisateurs, même si dans les faits, il est encore restreint à la communauté des chercheurs en géographie, analyse spatiale et aménagement du territoire qui doivent s'adresser au projet HyperCarte pour son utilisation.

3.1 Communication entre le serveur et le client

Le format des données renvoyées par le serveur est contraint par les fonctionnalités attendues sur le client. Outre l'interactivité sur le choix des palettes, du nombre de classes et du type de progression dans la distribution des couleurs, on souhaite que le client puisse générer à la demande un rapport numérique (sous forme de fichier texte ou bien HTML) contenant les coordonnées géographiques de chaque point M , avec la valeur de son potentiel. Le client a donc tout avantage à récupérer et sauver la grille matricielle des valeurs de potentiel calculées plutôt qu'une image au format PNG, JPEG, ou GIF. Il peut ainsi redessiner une image rapidement en cas de changement de préférences graphiques.

La grille de potentiels transmise en valeurs flottantes non tronquées que calcule le serveur dépend des paramètres de calcul (résolution, cadrage, fonction d'interaction et portée) transmis par le client. La nature et le domaine de validité de ces paramètres sont établis par un contrat commun partagé par le serveur et le client.

3.2 Stratégie d'optimisation

Remarquons, en premier lieu, que les stratégies classiques de gestion de cache ne sont pas adaptées à notre cas car chaque requête doit générer un résultat global qui ne peut être pré-calculé puisqu'il dépend des paramètres de l'analyse. Cependant un examen approfondi des tâches de calcul sur le serveur montre qu'il existe des redondances que nous pourrions exploiter pour optimiser certaines parties du calcul. (Grasland, 2005). En effet, deux problèmes se posent lorsque nous calculons le potentiel $\Phi(M)$ (confère l'équation [1]) : d'une part, la somme se fait sur un nombre important d'éléments, et, d'autre part, le calcul des distances orthodromiques⁴ $d(M, g_a)$ est coûteux puisqu'il fait intervenir des calculs d'angles en arccosinus, cosinus et sinus.

Nous réduisons la somme en pratiquant une politique d'élagage (*cut-off*) algébrique. Contrairement au *cut-off* géométrique des géographes qui limite le calcul sur un certain rayon de portée, nous tenons compte de points éloignés dont le poids

⁴ Rappel : la distance orthodromique entre 2 points A et B de coordonnées respectives $A(a1, a2)$ et $B(b1, b2)$, avec r rayon terrestre et les latitudes, longitudes en radians vaut : $d(A,B) = arccos(sin(latA) sin(latB) + cos(latA) cos(latB) cos(longB-longA)) * r$.

S_a (valeur statistique) influence le résultat du calcul. Notre algorithme de *cut-off* utilise une méthode numérique par recherche dans un arbre de type *quadtree*.

A chaque feuille de l'arbre de profondeur n est associé un point M valorisé avec ses coordonnées et son stock (cf. figures 8 et 9). Chaque point est ensuite sommé par groupe de 4 (des voisins sur la grille). Le calcul du potentiel $\Phi(M)$ s'effectue récursivement en cumulant le produit des stocks associés aux feuilles, par leur distance au point M .

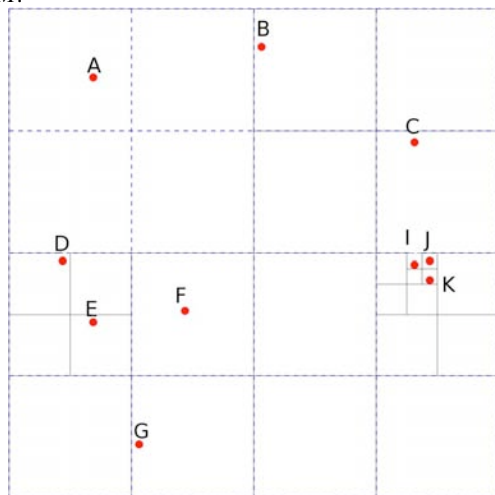


FIG. 8 - Espace géographique discrétisé avec les mesures localisées.

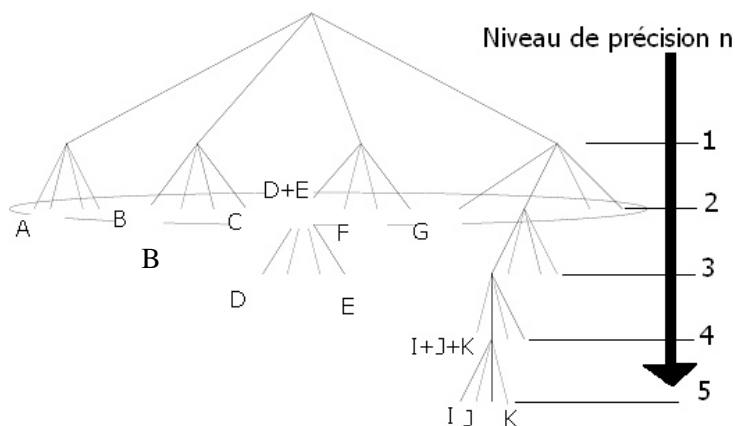


FIG. 9 - Exemple de construction d'arbre à partir d'un espace d'étude discrétisé.

La visite de chaque branche de niveau $n-1$ dépend de la réussite du test suivant, qui vérifie si le poids des enfants du nœud $n-1$ est négligeable ou non :

$$\left(\sum_{\text{nœud}_{n-1}} S_a \right) * d_{\min} \leq \varepsilon * \Phi(M)_{\text{cumulé}}. \quad [4]$$

Le point délicat est d'ajuster la valeur de l'*epsilon* de manière à ne pas négliger trop de points. Par défaut, la valeur est fixée à un 1/1000 de la somme totale des stocks. Des études plus approfondies restent à mener sur l'influence d'*epsilon* sur, d'une part la précision des calculs obtenus, et, d'autre part, la durée du calcul. Intuitivement, on suppose qu'une petite valeur d'*epsilon* allongerait le temps de calcul, mais aussi augmenterait la précision des résultats. Inversement, un grand *epsilon* provoquerait une forte simplification du calcul, qui serait alors ramené à une moyenne mobile sur moments gaussiens, et diminuerait l'intérêt de la méthode.

Le second facteur de lenteur de calcul (l'évaluation d'expressions contenant des termes en arccosinus, cosinus et sinus d'angles) peut être contourné en tabulant de façon fine ces fonctions. C'est-à-dire que les valeurs des fonctions sur des angles correspondant à une division régulière et fine d'un intervalle I donné sont pré-calculées, et pour tout angle la valeur de la fonction est approximée par la borne inférieure de la division à laquelle il appartient. Le grain de cette tabulation est fixé à l'avance par un paramètre du programme.

3.3 Visualisation interactive

La méthode prend tout son intérêt dans un contexte d'utilisation interactive, qui permet une exploration des structures spatiales révélées selon un certain nombre de paramètres :

- la fonction d'interaction : une liste déroulante présente la liste des fonctions implantées sur le serveur : *disk*, *amortized_disk*, *gaussian*, *pareto*, *exponential*.
- la portée moyenne en kilomètres de l'analyse, sélectionnée via une barre de glissement, ou par saisie numérique. Ce choix n'est pas borné.
- la résolution de visualisation définie en nombre de points en largeur et hauteur désirés pour la grille, et qui se rapporte au pas de discrétisation de l'espace étudié.
- le cadrage (le territoire ciblé) pour l'analyse, correspondant actuellement à l'espace visualisé. Un zoom ou un déplacement modifie donc le cadrage.
- les stocks proposés dans deux listes déroulantes séparées, une pour le numérateur, l'autre pour le dénominateur, afin de pouvoir ensuite calculer un ratio de potentiel. Le client interroge le serveur lors de l'authentification de l'utilisateur pour connaître la liste des stocks disponibles.

Une collection d'onglets est créée, chacun contenant une carte spécifique : le premier présente simplement l'aire d'étude et le maillage administratif. Les trois suivants présentent pour une portée V_1 donnée le potentiel du numérateur, du dénominateur et leur ratio Z_1 . Les trois autres suivants de même mais sur une portée différente V_2 . Enfin le dernier onglet compare les densités entre les deux voisinages en visualisant leur différence : $Z_2 - Z_1$.

L'utilisateur dispose en plus des fonctions de base de l'interactivité : zoom, déplacement dans une carte, navigation dans l'atlas de cartes ainsi produit. La visualisation d'une carte de potentiel se fait en superposition avec un fond vectoriel présentant les limites administratives de l'espace analysé. Ces cartes présentent une gradation de couleurs, suivant l'intensité du phénomène, paramétrable indépendamment pour chaque onglet, via une proposition à gauche de la zone de visualisation pour le choix de palette, le type de progression, et le nombre de classes dans la distribution. La figure 10 donne un aperçu de l'interface du client.

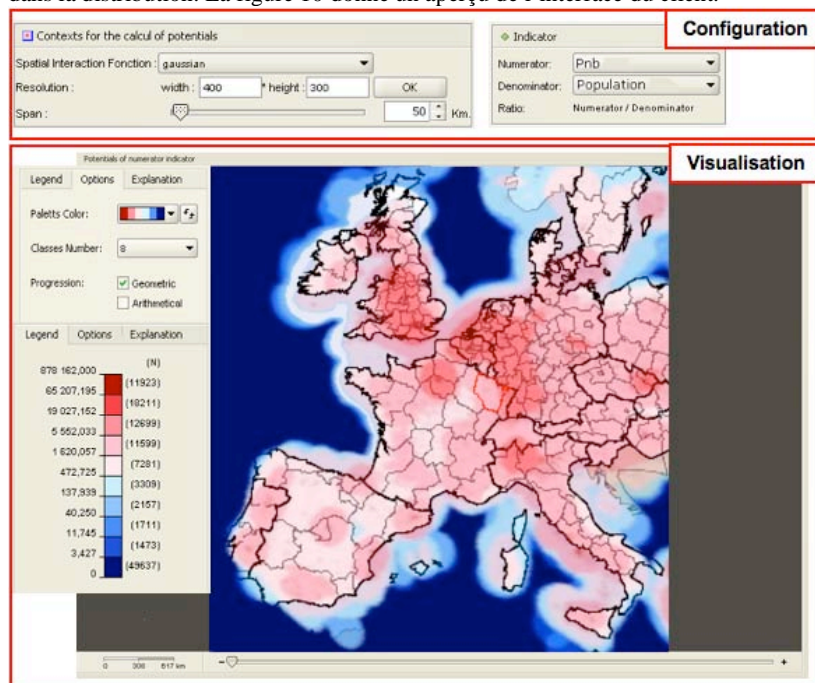


FIG. 10 - Aperçu de l'interface, avec un ratio du PNB par habitant sur un calcul gaussien de portée 50 km.

4 Bilan de la réalisation

4.1 Performances et expérimentations pour un usage local

Le serveur peut s'utiliser de façon autonome par rapport au client. Les expérimentations suivantes ont été réalisées sur un serveur bi-processeurs sous Linux (Pentium 4 à 2,6 Ghz, avec 1 Go de mémoire), une machine de travail pour un utilisateur ordinaire, avec comme jeu de test la population en Europe recensée au niveau communal, ce qui représente 116203 entités géographiques. Par exemple,

une carte de population est calculée pour une fonction gaussienne de portée 80 km avec la résolution forte de (800 x 600) en 1 minute 38 secondes.

La complexité de l'algorithme est vérifiée lors des mesures sur les temps de calcul, comme le montre les mesures du tableau 1. Mathématiquement, elle est linéairement proportionnelle à la résolution (nombre de points n à calculer sur l'image). Elle est aussi linéairement proportionnelle à la portée p . La connaissance de complexité en $O(np)$ permet d'estimer le temps de calcul d'une carte en fonction de la résolution demandée n et de la portée p requise. Cette durée reste raisonnable pour un usage sur des machines classiques : 2 minutes environ pour une longue portée (100 km) et une résolution fine de (800 x 600).

Résolution	Portée (km)	Temps de calcul (s)
200x100	100	5
400x300	100	32
800x600	100	130
800 x 600	25	33
800 x 600	50	57
800 x 600	100	130

TAB. 1 - *Temps de calcul des matrices sur le serveur.*

4.2 Performances et expérimentations pour un usage serveur

Nous avons implanté des options pour l'exécution parallèle du calcul sur un ensemble de processeurs, dans le cas où le serveur de calcul est partagé par plusieurs usagers, et où il se déploie sur une machine multi-processeurs. L'algorithme de calculs des potentiels est une itération sur l'ensemble des points de la grille qui sont arrangés dans un tableau, dont la taille n égale la valeur de la résolution. Chaque point peut-être calculé indépendamment des autres : pour cette raison la parallélisation du programme est très aisée. Elle est fondée sur la distribution de tâches (le calcul d'une entrée du tableau) à un nombre k de processeurs. Le nombre de points dévolus à un processeur est donc de n/k , et le calcul se termine lorsque chaque processeur a terminé le calcul de sa portion de tableau. Cette distribution naïve des tâches montre cependant deux limitations majeures. En premier lieu, un tel algorithme n'est pas résistant aux perturbations : si les capacités d'un processeur tombent à 50% brusquement, le temps de calcul est alors rallongé du temps équivalent à la moitié de sa tâche, puisqu'il n'y a pas de procédé de rééquilibrage des charges de travail automatique entre processeurs. D'autre part, notre grille est très hétérogène : les points situés dans des zones à faible densité de mesure sont calculés bien plus rapidement grâce au procédé d'élagage, et certains processeurs terminent donc leur travail plus vite.

Une solution pour la répartition de charge dynamique a été imaginée, par un mécanisme de transmission de charge adaptatif (Roch, 2006) : chaque processeur « libre » vole la moitié de la tâche de travail restante p au processeur occupé. Cette redistribution n'a lieu que lorsque le processeur occupé achève la portion de travail

indivisible qui lui est attribuée $\alpha \log(p)$. α est un paramètre configurable, qui est ajusté de manière à ce que le temps de contention (*lock* pour accéder au tableau de données) soit largement inférieur au temps de calcul de la portion du tableau. Cette notion de quantité minimale de travail est introduite pour éviter les problèmes de contention lors de l'accès au tableau de points : lorsque qu'un processeur prend une tâche, il actionne un sémaphore sur le tableau et bloque l'accès des autres processeurs au tableau. Ce temps d'inter-blocage est court, et la définition d'une charge de travail minimale optimisée évite la répétition fréquente d'inter-blocages. Cet algorithme a été mis en œuvre avec PTHREAD dans notre implémentation et, testé sur une machine multi-processeurs de type SMP, huit cœurs avec mémoire partagée sur la population européenne, à une résolution fixée de 800*600 et une portée gaussienne de 100 km. Il donne les résultats confirmant l'efficacité de la méthode, (voir la figure 11). En outre, nous signalons que cet algorithme a été implémenté avec deux bibliothèques différentes pour l'ordonnancement de tâches, TBB⁶ (*Intel Threading Building Blocks*, <http://threadingbuildingblocks.org/>) et Kaapi (Danjean, 2007). Quelque soit la bibliothèque, les performances sont améliorées de façon similaires : le gain en temps de calcul est linéaire avec le nombre de processeurs mobilisés, tant que la redistribution n'est pas saturée (observée à partir de 5 processeurs avec PTHREAD, elle est repoussée avec Kaapi car l'algorithme de répartition de tâches est mieux optimisé).

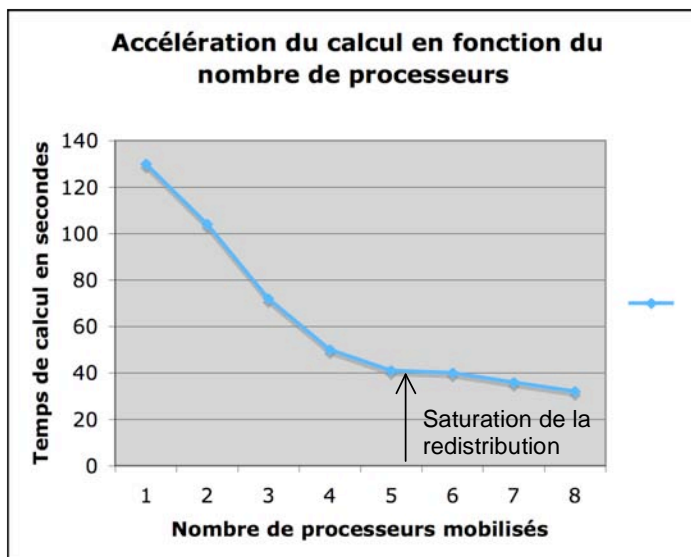


FIG 11 - Accélération du calcul des potentiels en fonction du nombre de processeurs

Le mode de répartition dynamique de charge décrit ne tient pas compte de la dimension spatiale des données : une répartition de charge basée sur un découpage

spatial statique paraît plus efficace. Afin de vérifier cette hypothèse, une seconde version de l'algorithme de répartition de charge a été implémentée et testée : l'espace d'étude est découpé en tuiles contiguës de tailles égales, et chaque tuile est traitée en parallèle par un des processeurs mobilisés pour le calcul. La figure 12 illustre les résultats comparés des deux modes de répartition de charge sur un banc d'essai de 16 processeurs. Les points des courbes situent le taux d'occupation des processeurs au cours du calcul, en fonction du temps. On constate que l'algorithme de répartition dynamique de charge termine plus tôt (une centaine de secondes d'écart), et occupe avec un meilleur rendement les processeurs que l'algorithme de répartition statique. Une première explication est suggérée par l'interprétation du palier observé au niveau du taux d'occupation des processeurs pour l'algorithme statique : certains blocs de calculs avec une densité de données moindre auront terminés plutôt, laissant des processeurs inactifs, dans l'attente de la fin du calcul.

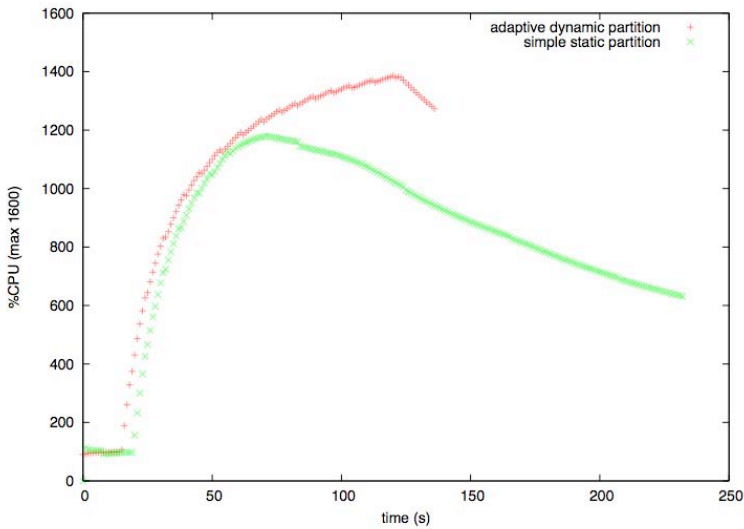


FIG. 12 – *Comparaison des performances de deux modes de répartition de charge pour le calcul d'une carte de potentiel sur un ensemble de 16 processeurs.*

La version statique pourrait être améliorée en découpant les tuiles proportionnellement à la densité de données présente, de manière à équilibrer la charge sur les différents processeurs. Cependant, un partitionnement statique, même prenant en compte la répartition spatiale des données, ne permet pas de s'adapter dynamiquement aux variations de charges CPU induites par les autres utilisateurs de l'environnement de calculs intensifs, contrairement à l'algorithme adaptatif. Par contre, la combinaison d'un partitionnement statique pour la première partie du

calcul et d'un équilibrage de charge dynamique pour la fin du calcul est une stratégie qui porte souvent ses fruits lorsqu'on travaille sur des volumes importants de données (Jafar, 2006).

Les temps de latence réseau sont bons, malgré le fait que nous n'avons pas encore mis en œuvre la compression des données échangées via *gzip*. En fait, on mesure un temps correspondant à l'emballage de la réponse, l'encryptage et le déballage de 4s localement avec une résolution de 300 x 400. Côté client, la reconstruction de l'image à partir d'un tableau de flottants prend très peu de temps (128ms). Cette matrice est re-calculée dès lors que l'utilisateur change un des paramètres d'analyse, agrandi, réduit ou se déplace dans la zone de visualisation. Elle est cachée avec les paramètres de la requête correspondante afin de pouvoir l'exploiter ultérieurement. La résolution demandée étant le plus souvent inférieure à celle de l'image vectorielle (1027 x 688), une interpolation bi-quadratique est appliquée sur l'image produite afin de la caler sur le fond vectoriel.

4.3 Perspectives d'amélioration

Plusieurs pistes s'offrent pour améliorer les performances du serveur de calcul. Par exemple, le calcul de la distance orthodromique peut bénéficier d'un pré-calcul supplémentaire. Sans entrer dans les détails, il faut encore exploiter les travaux de recherche concernant la quête du compromis idéal entre consommation d'espace mémoire et gain de temps obtenu par tabulation. De même, la troncature des flottants transmis dans la grille matricielle réduirait le volume des données échangées, et accorderait donc un gain de temps sur la transmission des données. Un algorithme qui automatise la détermination du niveau de troncature est en cours d'élaboration.

D'autres pistes, moins classiques, sont liées à l'usage de notre algorithme de *cut-off*. Il s'agit de préparer une stratégie de sous-échantillonnage des données : par exemple, adapter la valeur du seuil d'élagage dynamiquement au lieu de fixer arbitrairement sa valeur, ou bien limiter la visite de l'arbre à un certain niveau, sans aller jusqu'aux feuilles, lorsque que le niveau d'information requis est très grossier. Dans ce but, l'usage de cubes de données organisés en fonction de la hiérarchie spatiale des maillages territoriaux, tels ceux développés dans le domaine de l'analyse de données en ligne (OLAP) serait d'un très grand intérêt (Rigaux, 2005).

Enfin, la sélection de la distance pourrait se faire de manière adaptative, selon l'échelle d'étude retenue. Par exemple, employer la distance euclidienne sur la région Rhône-Alpes n'introduirait pas de biais notables.

De façon identique, quelques améliorations pour le client sur le plan de l'ergonomie sont prévues. Comme, par exemple, décorréler la zone de visualisation et la zone de traitement, en utilisant un cadre de sélection de la surface à traiter. Nous souhaitons donner une interprétation plus directe de la résolution : indiquer un pas en nombre de kilomètres, plutôt qu'un nombre de points sur une grille. Enfin, le calcul et la représentation de courbes de niveau à partir de l'image matricielle amélioreraient sensiblement le fondu avec le fond vectoriel représentant le maillage administratif.

Le client offre aussi un champ de réflexions en ce qui concerne l'usage de caches autorisant un raffinement progressif des images, grâce à l'émission de requêtes successives. En effet, tant que l'utilisateur ne modifie pas ses souhaits, il serait judicieux de construire une image dont la résolution augmente dans le temps, et d'anticiper un zoom sur la zone lissée.

5 Conclusion

Cet article présente HyperSmooth, une mise en œuvre du calcul et de la visualisation de potentiels dans un contexte de cartographie interactive. L'objectif étant de fournir un outil accessible à des utilisateurs via le Web, tout en assurant la sécurité et la confidentialité de ces données. Nous montrons que la méthode est adaptée pour étudier la propagation spatiale de phénomènes sociaux, environnementaux, ou économiques. Nous soulignons que le paramétrage d'une telle analyse peut s'avérer délicat, et nécessite donc de développer un outil adapté afin de rendre les choix plus compréhensibles pour un utilisateur néophyte.

La réalisation d'un tel prototype nécessite une réflexion avancée et coordonnée sur l'architecture, les modalités d'optimisation des calculs, et de paramétrage de l'analyse pour offrir une visualisation interactive. En effet, le coût du calcul, et les contraintes de confidentialité sur les données nous placent face à un défi technique.

Le bilan de notre réalisation s'avère positif quant au choix de l'architecture distribuée : répartition des calculs sur un serveur parallèle d'un côté, visualisation et paramétrage de l'analyse sur un client Java intelligent de l'autre, les deux parties étant connectées via un protocole de plus en plus répandu et présentant une accessibilité et sécurité maximale : SOAP, couplé avec un cryptage SSL. De plus, les calculs sont accélérés grâce à l'utilisation d'une méthode de *cut-off* algébrique, et une tabulation des distances orthodromiques. L'ensemble du code et un manuel du serveur sont accessibles librement sur Internet à l'adresse suivante : <http://hyantes.gforge.inria.fr/>

Une question d'ordre théorique et algorithmique porte sur l'introduction d'autres distances, et sur l'extension de la méthode de façon à illustrer les propriétés anisotropiques de l'espace géographique réel. Comme l'a montré une étude des phénomènes de tempêtes, aux flux très orientés (Boulier, 2003), il est intéressant de développer sur la base de la méthode générale une vision anisotropique de l'espace. Pour quantifier le potentiel en un lieu, il s'agit de tenir compte à la fois de la distance de la source d'information au point à estimer, mais aussi de l'orientation du flux d'information qui relie les deux points. Dans le premier exemple, appliqué à la forêt, les corrélations entre les potentiels ainsi déterminés et les dégâts constatés sont très nettes. Cette modification de la méthode générale permet d'intégrer de nouvelles formalisations des mouvements dans l'espace (vents dominants, courant marin, etc.) complétant celles qui sont associées à la fonction d'interaction spatiale. Cette extension du modèle peut également alors rendre compte des barrières géographiques naturelles comme la mer ou bien les montagnes pour l'étude de la dispersion de phénomènes dans l'espace, ou de phénomènes de polarisation.

Remerciements

Les auteurs tiennent à remercier les relecteurs de l'article pour les commentaires constructifs et les références supplémentaires fournies. De même, merci à tous les membres du groupe de recherche Hypercarte qui ont contribué à l'avancement de ces travaux, et en particulier :

- Liyun Guelton
- Saïd Ouhalal
- Nicolas Lambert

Par ailleurs ces travaux ont bénéficié de financements partiels de ESPON (*European Spatial Planning Observation Network*).

Références

- Antoni, J-P., O. Klein, et S. Moisy, (2004) Cartographie interactive et multimédia : vers une aide à la réflexion géographique. *Cybergeo* 288 <http://www.cybergeo.eu/index2621.html>
- Boulrier, J. et C. Grasland, (2003) La forêt face au risque tempête : vers d'autres apports de l'analyse spatiale. *Colloque Théoquant*, Besançon
- Boursier-Mougenot, I., N. Cattan, C. Grasland C., et Rozenblatt C., (1993) Images de potentiel de population en Europe, *L'Espace Géographique*, 4:333-345.
- Danjan, V., R. Gillard, S. Guelton, J.-L. Roch et T. Roche, (2007) Adaptive Loops with Kaapi on Multicore and Grid: Applications in Symmetric Cryptography. *In Parallel Symbolic Computation'07 (PASCO'07)*, ACM publishing, London, Ontario, Canada
- Dubois, A., J. Gensel, T. Hanell, C. Schürmann, N. Lambert, C. Zanin, R. Ysebaert, C. Grasland, O. Damsgaard, K. Lähteenmäki-Smith, E. Gloersen, et R. Thomas, (2007) Observing the structure of European territory in relative terms *DG-IPOL, Regional Disparities and Cohesion : What Strategies for the future*, EU Parliament, report IP/B/REGI/IC/2006_201
- François, J.-C., (1996) Diffusion et dynamique des discontinuités: les élèves d'origine africaine dans les collèges de l'agglomération parisienne, *Mappemonde* 4
- Gotway, C. et L. Young, (2002) Combining Incompatible Spatial Data, *Journal of the American Statistical Association*, 97(458):632-648
- Grasland, C., (1991) Potentiel de population, interaction spatiale et frontières : des deux Allemagnes à l'unification, *L'Espace Géographique*, 3:243-254.

- Grasland, C., F. Guérin-Pace, et C. Terrier, (2005-a) La diffusion spatiale, sociale et temporelle des pièces euros étrangères : un problème complexe *Actes des journées de Méthodologie Statistique*
- Grasland, C., Martin H., Vincent J.-M., Gensel J., Mathian H., Oulahal S., Cuenot O., Edi E., et Lizzi L., (2005-b) Le projet Hypercarte : analyse spatiale et cartographie interactive, *SAGEO 2005*
- Grasland, C. et J.-M. Vincent, (2006) *ESPON 3.4.3 : Final report for the MAUP*. 151-168
- Jafar S., L. Pigeon, T. Gautier, et J.-L. Roch. (2006) Self-Adaptation of Parallel Applications in Heterogeneous and Dynamic Architectures. In *IEEE, editor, ICTTA'06 IEEE Conference on Information and Communication Technologies: from Theory to Applications, Damascus, Syria*, 3347-3352
- Josselin D. et Fabrikant S., (2003), La "cartactive " en mouvement : un nouveau domaine de recherché pluridisciplinaire ou un pan de la géomatique. In *Cartographie animée et interactive*, Josselin D., Fabrikant S.(eds), Revue internationale de Géomatique, 13(1).
- Journal, A.-G. et C.-J. Huijbregts, (1978) *Mining geostatistics*. Academic Press. London
- Kobben, B. (2001) *Publishing maps on the web. Chapter 6 in Web Cartography*, M-J. Kraak and A. Brown (eds.), New York: Taylor and Francis
- Lacaze, M. et F. Nirascou, (2000) Ces terres qui nous entourent..., *Les données de l'environnement 51*
- Nyquist, H. (1928) Certain Topics in Telegraph Transmission Theory, *Proceedings of the Institute of Electrical and Electronics Engineers*, 90(2):280-305, réédition en 2002
- Openshaw, S. et Taylor P.J., (1979) A million or so correlation coefficients *Statistical methods in the spatial sciences*, Pion, London, 127-144.
- Plumejeaud, C., J.-M. Vincent, C. Grasland, J. Gensel, H. Mathian, S. Guelton et J. Boulrier (2007) HyperSmooth : calcul et visualisation de cartes de potentiel interactives. *SAGEO 2007*.
- Poulain, M., G. Mario Pes, C. Grasland, C. Carru, C. Ferrucci, G. Baggio, C. Franceschi, et L. Deiana, (2004) Identification of a geographic area characterized by extreme longevity in the Sardinia island: the AKEA study, *Experimental Gerontology*, 39(9):1423-1429
- Rase, W.-D. (2001) Volume-preserving interpolation of a smooth surface from polygon-related data, *Journal of Geographical Systems*, 3(2):199-213
- Rigaux, P. et Scholl M. (1995) Multi-Scale Partitions : Application to Spatial and Statistical Databases. In : 4th International Symposium on Advances in Spatial Databases, Springer-Verlag 170-183

- Roch, J.-L., D. Traore, J. Bernard (2008) Processor-oblivious parallel stream computations *16th Euromicro International Conference on Parallel, Distributed and network-based Processing*, Toulouse, France
- Tobler, W, (1979) Smooth Pycnophylatic Interpolation for Geographical Regions, *Journal of the American Statistical Association*, 74(367):519-536
- Wickel, C., (2002) Hierarchical Models in Environmental Science, *International Statistical Review*. 71(2):181-199.
- Wong, DWS., (1996) Aggregation effects in geo-referenced data. Griffiths D (ed) *Advanced spatial statistics*. CRC press, Baton Rouge, 83-106

Summary

This paper presents a new cartographic tool for spatial analysis of social data, using the potential smoothing method. The purpose of this method is to view the spreading of a phenomenon (demographic, economical, social, etc.) in a continuous way, at a macroscopic scale, from data sampled on administrative areas. We aim to offer an interactive tool, accessible through the Web, but guarantying the confidentiality of data. The biggest difficulty is induced by the high complexity of the calculus, dealing with a great amount of data. A distributed architecture is proposed: map computation is made on server-side, using particular optimization techniques, whereas map visualization and parameterisation of the analysis are done on a web-based client, the two parts communicating through a Web protocol.