



HAL
open science

Construction d'un Service Web d'Aide à la Décision Geo-Marketing à partir d'Outils OpenSource

Sylvain Willart, M. Calciu

► **To cite this version:**

Sylvain Willart, M. Calciu. Construction d'un Service Web d'Aide à la Décision Geo-Marketing à partir d'Outils OpenSource. Colloque Etienne Thil 2012, Nov 2012, Lille, France. halshs-00784016

HAL Id: halshs-00784016

<https://shs.hal.science/halshs-00784016>

Submitted on 2 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Construction d'un Service Web
d'Aide à la Décision Géo-Marketing
à partir d'Outils OpenSource**

Calciu, M.
michel.calciu@univ-lille1.fr
Willart, S.
sylvain.willart@univ-lille1.fr

Maîtres de Conférences
LEM
IAE Lille
Université Lille 1 (USTL)

Construction d'un Service Web d'Aide à la Décision Géo-Marketing à partir d'Outils OpenSource

Résumé

Cette recherche explique la construction d'un service web destiné à encourager les managers à utiliser des méthodes avancées de géomarketing. Le service est construit à partir d'outils open-source. Trois études démontrent son efficacité tant au plan technique et méthodologique qu'au plan de sa facilité d'utilisation.

Mots-clés

Géomarketing, Modélisation, Service Web

Elaboration of a Web Service for Geo-Marketing Decision Support using OpenSource Software

Abstract

This piece of research explains the elaboration of a web service aimed at fostering managers into using advanced géomarketing methodologies. The service is build using open-source software. Three studies demonstrate how well it performs on the technical and methodological side, as well as on its ease of use.

Keywords

Geomarketing, Models, Web Services

Construction d'un Service Web d'Aide à la Décision Géo-Marketing à partir d'Outils OpenSource

Résumé managérial

« *Things that are closer are more alike than things farther apart* ». La prise en compte, tardive en marketing, de ce principe du sens commun a permis l'émergence du géo-marketing. Ce dernier se définit simplement comme la prise en compte de la réalité spatiale du marché à des fins marketing.

Or, la plupart des données du marketing sont géo-localisées, ou tout au moins géo-localisables : les clients ont un lieu d'habitation et un lieu d'achat, ils se déplacent selon des itinéraires, les ventes se réalisent dans des magasins dont on connaît les adresses, et qui sont implantés dans des zones de chalandise, les livraisons de produits vont d'un point A à un point B tous deux connus, les affiches sont collées sur des panneaux publicitaires dont on connaît la localisation, et l'INSEE livre de nombreuses analyses au niveau des départements, communes, arrondissements, et IRIS... Et cette tendance va croissante avec la possibilité de suivre les déplacements des consommateurs via les smartphones, permettant l'apparition de technique telles que le géo-fencing.

D'un point de vue opérationnel, les logiciels permettant l'analyse géo-marketing ont également connu une petite révolution : leur orientation desktop a évolué vers une orientation web, avec le succès notamment de l'application Google Maps, qui permet, sans connaissance aucune et de façon extrêmement intuitive, d'afficher une carte interactive via une page web sans installer de programme sur son propre ordinateur. Cela étant, l'application de Google n'est pas orientée sur l'analyse, simplement sur la visualisation, et les services développés autour de cette application à des fins de géomarketing, s'ils existent sont encore très chers.

L'objectif de cet article est donc de construire, à partir d'outils open-source gratuits, un service web complet d'aide à la décision géomarketing. Ces outils intègrent une gestion des bases de données géographiques (PostgreSQL et son extension PostGIS), un puissant logiciel de calcul statistique (R, et son extension PL/R), et un service de création de cartes géographiques (MapServer).

Ainsi, n'importe quelle donnée de l'entreprise intégrant un caractère géographique peut être incluse sur cette plateforme, et consultée et analysée via une simple interface web depuis n'importe quel terminal connecté.

En s'appuyant sur trois exemples traités via cette plateforme, le présent article fait la démonstration des capacités de calcul et de rendu de l'outil développé : calcul de zones de chalandises et d'attractivité, modèles d'économétrie spatiale pour expliquer les ventes, visualisation de la clientèle... Cette recherche développe également un nouveau construit marketing : la segmentation spatiale d'une base de données clients. Cette métrique propose simplement de visualiser sur une carte, de façon statique ou dynamique, les différents segments de clientèle issus d'une analyse RFM.

Construction d'un Service Web d'Aide à la Décision Géo-Marketing à partir d'Outils OpenSource

Introduction

« *Things that are closer are more alike than things farther apart* ». Ainsi se résume le principe de base de la géographie. Aussi simple qu'il puisse paraître, il n'a été que tardivement pris en compte dans l'analyse statistique et marketing. Une des raisons étant certainement la complexité relative à sa mise en œuvre. Et la multitude des champs disciplinaires qui utilisent la géographie n'a pas contribué à sa simplification et à son unification.

Pourtant, la géographie est au cœur de la prise de décision des managers. Tonkin (1994) relève ainsi que 75 pourcent des données marketing dont les décideurs disposent contiennent au moins une composante spatiale : adresse des consommateurs, distribution de la population ou du pouvoir d'achat, zone de chalandise, localisation des concurrents, données du recensement...

Bien entendu, de nombreux modèles existent en marketing qui intègrent des informations spatiales, mais ils sont le plus souvent limités à l'étude et la résolution d'un problème précis : prix et promotion (Bronnenberg et Mahajan, 2001 ; Jank et Kannan, 2005), campagne publicitaire (Carton, 2003 ; Parker, 2004), marketing direct (Steenburgh, Ainslie et Engberston, 2003), choix de localisation d'un magasin (Hernandez et Bennisson, 2000 ; Ter Hofstede, Wedel et Steenkamp, 2002), organisation de la force de vente (Zoltners et Sinha, 2005).

Le principal objectif de ce papier n'est pas de traiter, ou re-traiter, une de ces questions de marketing spatial auxquelles les managers peuvent être confrontés, mais bien de proposer un cadre holistique permettant de toutes les appréhender.

Ce type d'objectif repose fortement sur la compréhension et l'explication des Systèmes d'Information Géographique et les approches « boîte à outils » vs « service » (I). Ayant expliqué nos choix en la matière, nous présenterons comment l'on peut mettre en place un service web intégrant l'ensemble des décisions relatives au géomarketing (II). Trois études, volontairement très différentes, seront ensuite présentées afin d'illustrer les vastes possibilités offertes par un tel service (III). En conclusion, nous verrons comment nous pouvons encore intégrer et étendre ce type d'analyses.

I- Le choix d'un Système d'Information Géographique (SIG) orienté « services »

Les SIG en distribution

Un SIG est très simplement défini comme l'addition d'une carte et d'une information spatiale. Dès le moment où l'on représente sur un fonds de carte numérique des données ayant une dimension géographique, on fabrique un SIG. Les SIG les plus simples sont très utilisés en

marketing (Goodchild, 1991), et particulièrement en distribution et dans l'industrie bancaire. La plupart des enseignes de distribution utilisent des SIG comme support de décision (Hernandez et Biasiotto, 2001 ; Hernandez, 2007). La rencontre de l'offre et de la demande en distribution dépend en effet fortement de la distance qui sépare les consommateurs du point de vente. D'autre part, les distributeurs disposent en quantité importantes de données d'ordre géographique : variables démographiques de recensement, ventes et adresses des magasins de l'enseigne, adresse des consommateurs et profil de fidélité, mais également la localisation de leurs concurrents (Byrom et al., 2001 ; Hernandez, 1995)

Les SIG traditionnellement utilisés en distribution proposent une utilisation de ces données dans deux types d'analyses. Des analyses descriptives : localisation des consommateurs ou d'une zone de chalandise, segments de consommateurs, implantation des concurrents, et dans une certaine mesure, prévision des ventes et modélisation du comportement spatial des consommateurs (Hernandez, 2007). Et des analyses d'aide à la décision telles que l'ouverture d'un magasin, le remodeling, l'acquisition d'une enseigne concurrente, où l'expansion à l'international (Hernandez et Biasiotto, 2002).

En revanche, l'utilisation réelle des SIG est relativement restreinte. Comme c'est le cas pour la plupart des modèles marketing (Little, 1970), les SIG ne sont pas exploités comme ils le pourraient par les managers. Ces derniers se contentent souvent de quelques cartes générées automatiquement, sans aller vers la modélisation plus complexe et sophistiquée.

L'intérêt de l'orientation « service »

Si l'on peut blâmer les managers quant à leur faible utilisation des modèles de géomarketing disponibles dans certains SIG, le monde académique doit aussi assumer sa part de responsabilité quant aux choix de modélisation qui ne suivent pas nécessairement les règles avancées par Little (1970). Pour être adopté par les managers, les modèles doivent, entre autre, être faciles à contrôler et à adapter, et pouvoir évoluer en complexité au fur et à mesure de leur adoption par les managers.

Nous soutenons l'idée que l'utilisation d'un service web peut en ce sens être beaucoup plus adaptée.

Dans le cas des SIG, la tendance lourde est au passage d'une architecture centrée sur des « boîtes à outils » (*toolbox-centred* comme les logiciels classiques du champ : MapInfo, Arc/Info, ArcGis) à une architecture centrée sur des services en ligne (*service-centred* tels que Google Earth, Google Maps) (Bivand et al., 2008). Dans les architectures « boîtes à outils », les programmes et les données spatiales sont localisés sur l'ordinateur personnel de l'utilisateur. Alors que dans les architectures en ligne, les données et programmes sont situées sur des ordinateurs ou un réseau distant, auxquels l'utilisateur accède via sa connexion internet.

L'utilisation de ces applications en ligne se répandant très rapidement, les utilisateurs ont donc une grande habitude de ce type d'interfaces. D'autre part, elles ne nécessitent pas d'installation de programmes nombreux et complexes¹ puisqu'elles passent directement par le navigateur internet, que tout un chacun sait généralement utiliser de façon fluide. D'autre part, dans le cas d'analyses sur des bases de données de taille importante, il peut être beaucoup plus aisé de ne pas stocker ces données sur un ordinateur local : les données n'encombrent alors pas l'utilisateur, et les droits d'accès à ces données peuvent être gérés beaucoup plus

¹ Avant de pouvoir commencer la moindre analyse géographique sur un ordinateur local, un utilisateur doit en effet installer non seulement plusieurs programmes mais également un nombre important de « bibliothèques » pour ces programmes permettant de lire les données géographiques.

facilement par l'entreprise. Enfin, les données pouvant nécessiter des mises à jour très fréquentes (si chaque achat est enregistré par exemple), mieux vaut maintenir une base stockée sur un serveur, chaque utilisateur pouvant alors accéder aux données les plus actualisées lorsqu'il utilise le service.

L'orientation service offre à notre sens un outil plus simple et plus actualisé aux managers, donc peut-être plus utilisable et utilisé par ces derniers.

II- Mise en place d'un Service Web d'Aide à la Décision Géo-Marketing (SWAD-GM)

Cette partie, nécessairement technique, explique de façon succincte quels outils sont nécessaires à la mise en place d'un SWAD-GM, et comment ils doivent être combinés entre eux.

Un tel service doit répondre à plusieurs contraintes et repose de ce fait sur plusieurs éléments fondamentaux. Il doit tout d'abord être capable de gérer efficacement des bases de données relationnelles, lesquelles peuvent avoir une dimension spatiale. Il doit ensuite disposer d'un langage statistique assez développé pour permettre l'utilisation de modèles spatiaux complexes. Enfin, il doit être capable de produire des cartes de haute qualité, lisibles sans pixellisation à différentes échelles de précision.

Une dernière contrainte, posée par les auteurs, est qu'il doit reposer le plus possible sur des outils open-source.

Gestion des Bases de données

Le SGBDR (Système de Gestion des Bases de Données Relationnelles) est la pierre angulaire de tout service web. Il est utilisé en amont lorsque les données sont enregistrées dans le système, et en aval lorsque l'ordinateur personnel interroge le serveur (via le langage PHP) et génère la page qui sera affichée dans le navigateur. Entre les deux, le SGBDR sert à stocker les données dans plusieurs tables liées entre elles par un système de relations via des clés de classement. Plusieurs SGBDR existent, et les plus importants sont MySQL et Oracle pour les applications commerciales. Dans le monde open-source, c'est PostgreSQL qui apparaît comme la meilleure alternative. Développé à l'Université de Berkeley (projet POSTGRES), PostgreSQL a été le pionnier dans les applications les plus abouties du langage SQL ; il permet aujourd'hui l'utilisation de la majeure partie du langage SQL ainsi que les requêtes complexes, les clés étrangères, diverses vues, l'intégrité transactionnelle, et la compatibilité entre ses versions. D'autre part, il est extensible puisque chaque utilisateur peut y ajouter les fonctions qu'il désire, mais également de nouveaux types de données, ce qui se révèle particulièrement intéressant dans notre cas. Ce SGBDR peut être entièrement géré en ligne via l'interface phpPgAdmin.

Bases de données spatiales

Si les SIG doivent être des outils de collecte, de stockage, et de représentation de données spatiales, ils ne peuvent être réellement utilisés s'ils ne permettent pas la manipulation et

l'analyse de ces données.

Ainsi, dans l'optique d'une orientation service, les données spatiales doivent pouvoir être lues et intégrées au SGBDR afin de pouvoir être extraites, manipulées, et analysées dans le sens voulu par l'utilisateur.

Une telle fonctionnalité requiert que le SGBDR utilisé puisse avoir connaissance de ce type de données. Pour le cas de PostgreSQL, cette contrainte est résolue grâce à l'extension PostGIS. Cette extension créée, dans PostgreSQL un nouveau type de données : les données spatiales.

Ces données spatiales peuvent être diverses : points, lignes, polygones (simples et multiples), exprimées dans différents systèmes de coordonnées géodésiques (latitude / longitude, coordonnées, degrés, ...) et stockées sous différents formats (fichiers « .shp » notamment).

D'autre part, l'extension PostGIS ajoute au système PostgreSQL diverses fonctions et calculs répandus dans l'analyse spatiale : aires, distances, intersections... le tout dans un espace sphérique et non plan.

Langage statistique

Si les fonctions de PostGIS sont nombreuses, elles restent génériques. Et l'on peut avoir besoin, pour la modélisation, de fonctions plus spécifiques. Il est donc nécessaire d'ajouter à ce couple PostgreSQL/PostGIS un langage statistique performant capable de pourvoir aux besoins nombreux et complexes de la modélisation spatiale. Dans la communauté open-source, la meilleure alternative est sans doute de langage R.

Le langage R peut opérer aussi bien localement que dans une architecture web. De plus, il peut être intégré à PostgreSQL via l'extension PL/R. Cette extension permet d'exécuter des instructions R directement sur les bases de données gérées par PostgreSQL.

Avec R, on intègre ainsi à l'architecture web tout un ensemble de bibliothèques (« sous-programmes » de R) dédiées à la modélisation spatiale (sp, spatstat, spacetime, maps...). Ces bibliothèques, qui fondent le système R, lui permettent d'être toujours à jour des derniers développements en terme de modélisation. D'autre part, il permet d'utiliser très simplement des méthodologies issues d'autres champs disciplinaires. Par exemple, on peut avoir accès aux bibliothèques développées par des géographes, des épidémiologues, des physiciens, le tout dans une « grammaire » unifiée.

Production de cartes

Le dernier élément nécessaire à la création du SWAD-GM est un programme capable de restituer les données stockées et les analyses réalisées sous forme de cartes. La référence open-source en la matière est le programme MapServer. Ecrit en C (langage de très bas niveau), il est capable d'interagir avec de nombreux autres langages de programmation de plus haut niveau (PHP, Python, Perl, Ruby, Java, .Net) et est paramétrable à l'envie. Il permet la production de cartes « à la volée », de haute qualité, et éventuellement dynamiques (sous forme de films plutôt que d'image). MapServer est notamment capable de produire une carte suite à une requête SQL en y incluant plusieurs « couches » (informations issues de différentes tables de données), et des analyses statistiques (réalisées par PL/R). L'exemple ci-dessous illustre cette fonctionnalité en présentant, pour une ville de Bretagne, des informations relatives aux routes principales, aux eaux et forêts, au découpage administratif,

aux « points d'intérêts », à la localisation de divers magasins, et de divers consommateurs, et des diagrammes de Voronoï calculés sous R (Figure 1)

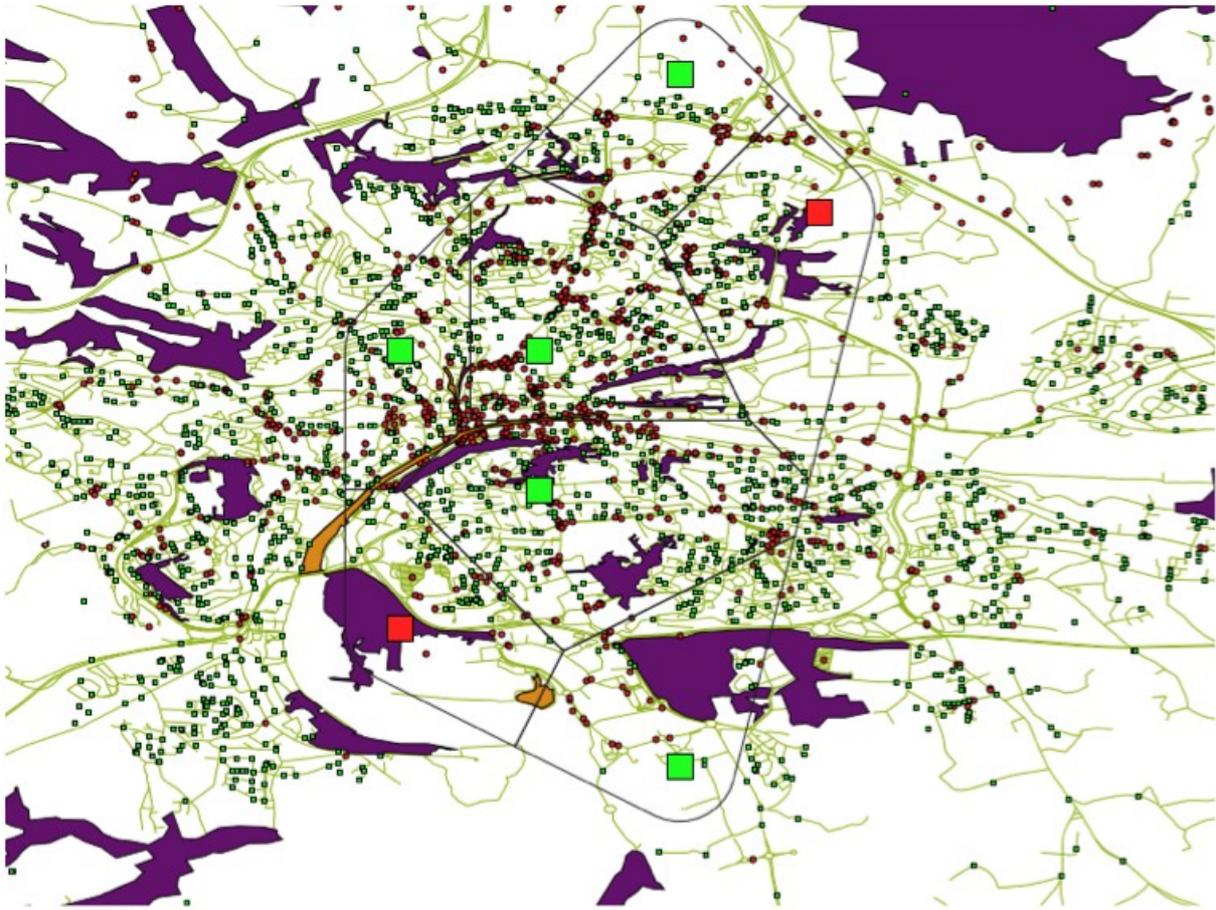


Figure1 : Exemple d'une carte complexe générée par MapServer

Données géographiques

Un dernier élément est nécessaire à l'analyse géo-marketing en sus du SWAD-GM précédemment explicité : les données géographiques.

Les données géographiques sont de deux ordres : les informations de fonds de carte, et les informations managériales. Le fonds de carte est constitué d'informations relatives aux routes, lignes ferroviaires, fleuves, plans d'eau forêts, découpages administratifs, points d'intérêts (touristiques ou économiques). Ces données sont accessibles à tous via Google Maps (web service bien connu...), mais elles ne sont pas exploitables librement et ne peuvent ainsi être gratuitement utilisées dans le cadre d'application commerciale. Une alternative open-source est OpenStreetMap (OSM). Sous forme de wiki, ce site agrège dans le monde entier des informations spatiales uploadées par des utilisateurs et libres de droits. Bien qu'encore incomplète, cette source de données connaît une expansion sérieuse et est, sur certains points, beaucoup plus précise que Google Maps. D'autre part, les données d'OSM sont téléchargeables sous de nombreux formats et facilement intégrables dans un service de type SWAD-GM.

Pour les informations managériales, elles sont récoltées auprès des entreprises, et bien

évidemment limitées dans leur utilisation par le secret professionnel. Ces données sont rarement géolocalisées et se présentent plus généralement sous forme de fichiers d'adresse qu'il faut convertir en données de longitude et de latitude. Plusieurs services existent sur le web permettant une telle transformation, les plus connus étant les API de Google, Yahoo, et Bing.

Synthèse du fonctionnement du SWAD-GM

La Figure 2 reprend les principaux éléments du système décrit ci-avant, et les relations qui existent entre eux.

Les pointillés rouges indiquent la frontière du système en lui-même, qui est installé sur un serveur distant. Cette installation non-locale permet paradoxalement une meilleure utilisation des capacités du SWAD-GM sur n'importe quel poste de travail : les utilisateurs n'ont aucun logiciel particulier à installer localement, l'ensemble fonctionnant via une connexion internet. Les flèches bleues pointillées décrivent le chemin principal du « workflow ». Les données sont d'abord envoyées sur le serveur en provenance des entreprises, et de l'INSEE pour les données non-géographiques, et de diverses sources, dont OSM, pour les données géographiques. Les données géographiques sont pré-traitées par PostGIS. Les données d'adresse client, ou magasin, peuvent également être géocodées et renvoyées sur le serveur via PostGIS.

L'utilisateur peut alors, via une requête SQL, ou plus simplement via une interface graphique sous forme de page web, demander telle ou telle analyse au serveur. Notons ici que les analyses les plus simples sont pré-programmées afin de permettre une utilisation plus intuitive. Cette requête envoyée va être traitée en temps réel : le serveur va sélectionner les données pertinentes dans ses tables, réaliser une analyse statistique, et produire une carte à la volée. Cette carte sera alors incluse dans une page web que l'utilisateur pourra consulter depuis son navigateur.

L'intérêt d'un tel système est qu'il est totalement flexible et évolutif. Par exemple, si l'utilisateur demande une analyse à un moment t , puis la même analyse à un moment $t+1$, la réponse peut être différente si entre temps de nouvelles données, plus récentes, ont été téléversées sur le serveur. On peut ainsi voir l'impact en temps réel sur une zone de chalandise de l'arrivée d'un concurrent par exemple.

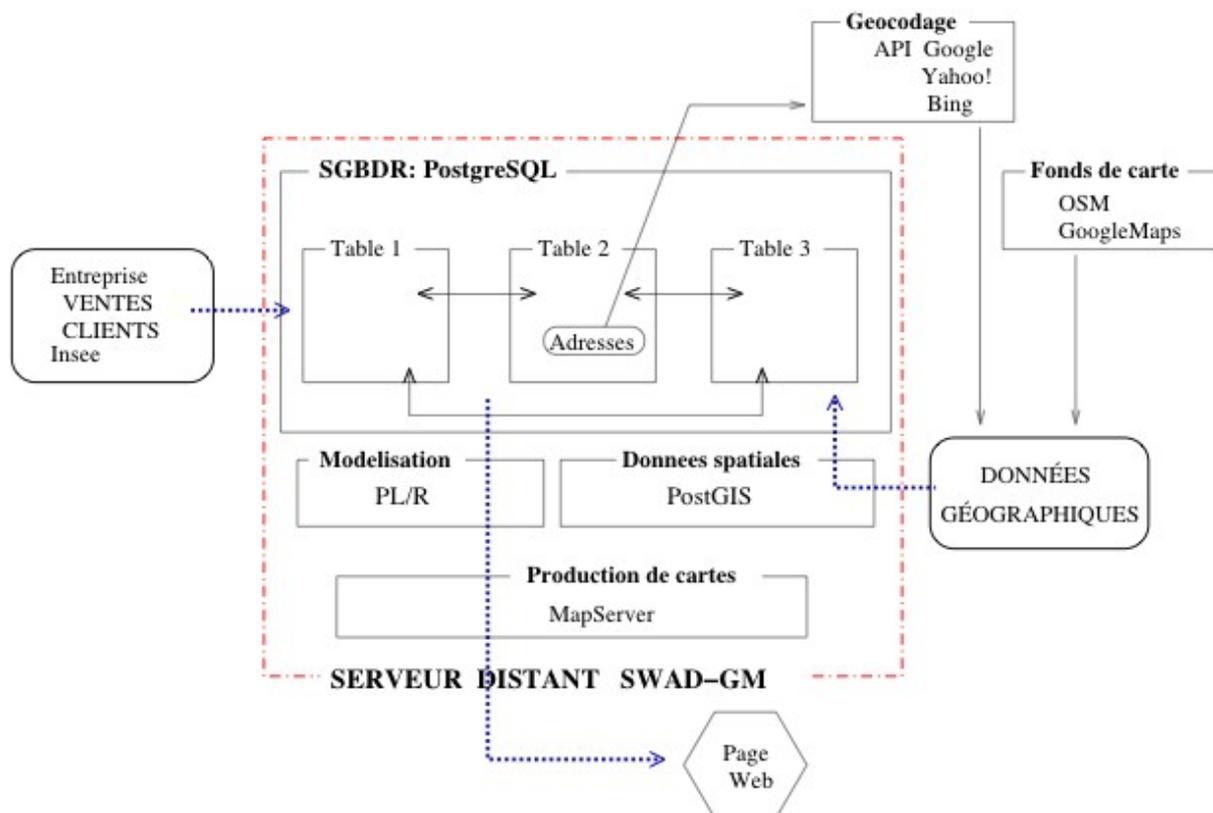


Figure 2 : Schéma récapitulatif du SWAD-GM

Les études présentées par la suite reposent sur ces différents types de données et illustrent l'étendue des analyses possibles grâce au SWAD-GM développé.

III- Trois études de géomarketing – une démonstration de l'étendue des possibilités d'un SWAD-GM

Etude 1 : Les polygones gravitaires : une opérationnalisation de la théorie des places centrales

L'objet de cette première étude est de proposer une méthode permettant de retracer sur une carte les zones de chalandise calculées à partir de l'étude l'attractivité des points de vente.

Une première méthode de découpage des zones de chalandise a été proposée par Christaller (1935), puis formalisée par Lösch (1954) ; elle est connue sous le nom de « théorie des places centrales » et livre des zones de chalandise d'aspect hexagonal. La théorie des places centrales permet également d'établir une hiérarchie des emplacements en fonction de leur potentiel d'attractivité.

La forme hexagonale, quoique simple, peut paraître quelque peu limitative quant à la réalité des zones de chalandise. Nous proposons ici d'étendre cette méthode en adoptant une plus grande flexibilité dans la forme de ces zones au travers de l'utilisation des polygones de Voronoï.

L'objet que nous proposons emprunte donc aux places centrales la prise en compte de l'attractivité, et aux polygones de Voronoï leur forme non-contrainte. Ce double emprunt permet de développer des polygones gravitaires, de forme non-contrainte, et d'aire proportionnelle à l'attractivité des points de vente. Ils sont à notre sens la meilleure alternative pour la modélisation automatique des zones de chalandise.

Cette étude est appuyée par une analyse des comportements de la clientèle d'une banque sur l'ensemble des agences d'une métropole française. Il s'agit plus précisément de l'intégralité des contrats souscrits dans les 70 agences de la banque par ses 254744 clients de la métropole. Pour des raisons évidentes d'anonymat, nous ne disposons pas des adresses des clients, mais seulement de leur localisation dans l'une des 5600 zones géographiques de la métropole. Au plan géographique, les consommateurs ne sont donc pas représentés par des « points », mais par des « polygones ».

Les données et l'analyse réalisée à l'aide du SWAD-GM apparaissent sur les Figures 3a et 3b.

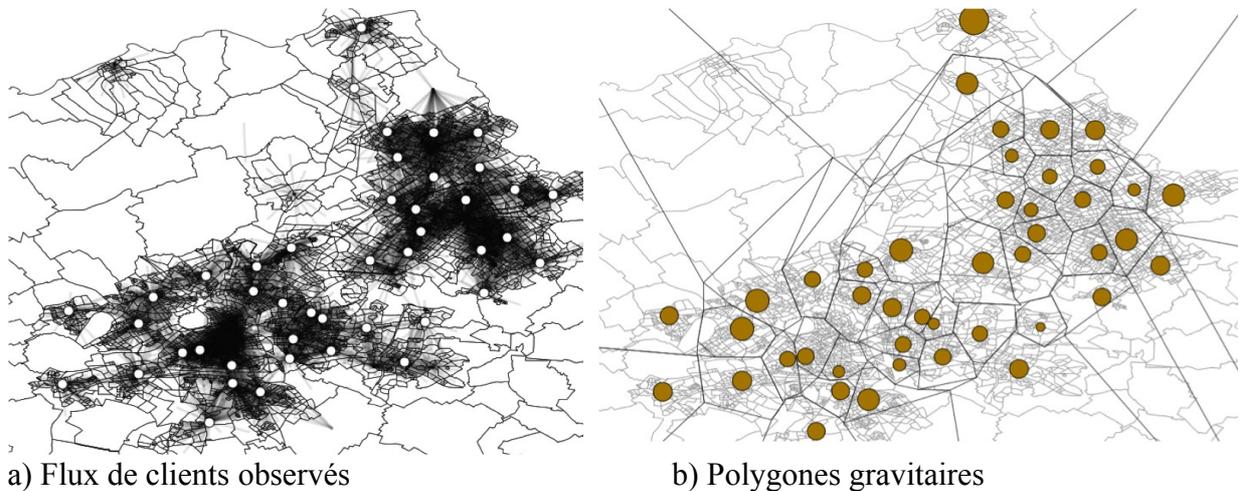
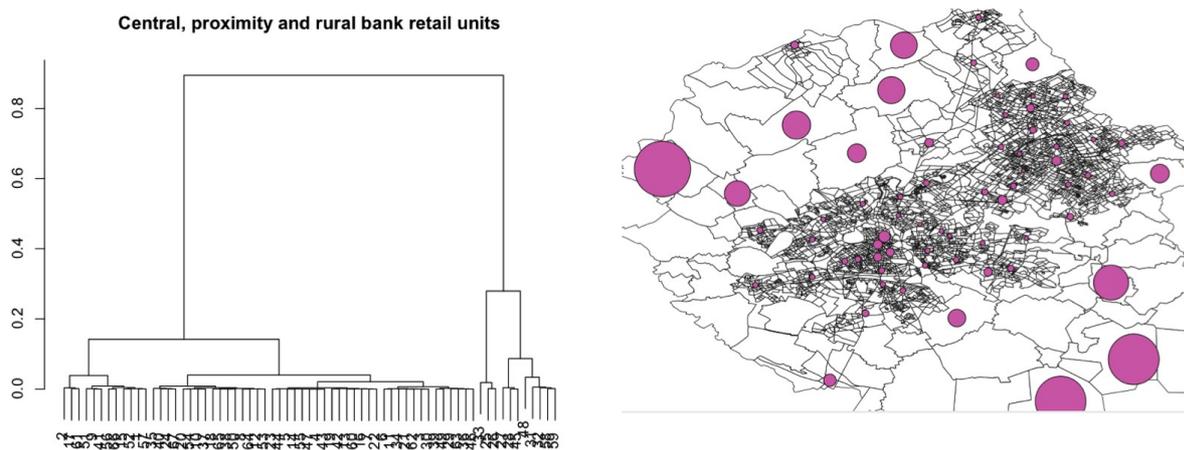


Figure 3 : Analyse des zones de clientèle des agences bancaires

A partir de ces informations, il est possible d'opérer une classification des agences bancaires pouvant conduire à des choix stratégiques spécifiques. L'un des points signalés comme important par les managers de la banque est la distance qui sépare le client de son agence bancaire. Une fois les données intégrées au SWAD-GM, il est aisé de calculer, pour chaque client la distance parcourue, et pour chaque agence la distance moyenne parcourue par ses clients. Sur cette base, on peut opérer une classification hiérarchique (Figure 4a), et produire une carte sur laquelle les banques sont représentées par un point de taille proportionnelle à la distance moyenne parcourue par leurs clients (Figure 4b).



a) Classification hiérarchique

b) résultat cartographié

Figure 4 Classification des agences selon la distance parcourue par les clients

Cette classification a notamment permis au manager de préciser la notion d'agence « rurale » et d'agence « urbaine ». Une analyse complémentaire a été réalisée indiquant quels facteurs ont le plus d'influence sur l'attractivité des agences des différents clusters. Mais il s'agit là d'une analyse non-spécifiquement géographique.

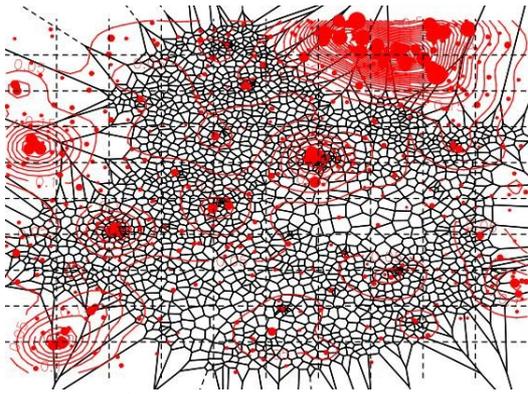
Cette méthode de calcul des zones de chalandise par les polygones gravitaires a été incluse au SWAD-GM *via* l'écriture d'une librairie R. Elle permet de répliquer l'analyse sur n'importe quel jeu de données similaire qui seraient versé sur le serveur du SWAD-GM comme l'illustre l'étude suivante.

Etude 2 : Assortiments et compétition locale dans la grande distribution

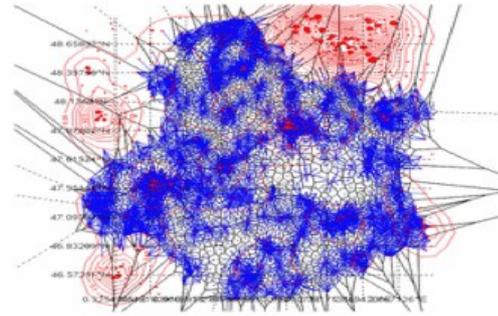
Notre deuxième étude se concentre sur l'analyse de l'assortiment comme levier de la concurrence locale dans la grande distribution. Elle s'appuie dans un premier temps sur le découpage des zones de chalandise selon la méthode présentée ci-avant, et dans un second temps sur une classe de modèle économétrique spécifiquement développée pour l'analyse des données géographiques. L'objectif est de construire un modèle explicatif des ventes prenant en compte la composition de l'assortiment et la concurrence locale à laquelle est confrontée chaque point de vente.

Les données utilisées proviennent de l'IRI et portent sur les ventes dans 317 magasins de la région Centre pour 42 catégories de biens de grande consommation. Nous disposons également des adresses de ces magasins, que nous avons soumis à une étape de géolocalisation *via* l'API de Google. Nous avons également recueilli auprès de l'INSEE les données socio-démographiques pour l'ensemble des IRIS de la région Centre.

A partir de ces données, on peut estimer, *via* un modèle de Huff (1964, 1966) l'attractivité de chaque point de vente pour chaque IRIS avec pour variables explicatives la surface du magasin et la distance IRIS-magasin. Ces mesures d'attractivité sont ensuite intégrées à un modèle de part de marché pour calculer la probabilité pour les consommateurs de chaque IRIS de visiter chaque magasin. On peut ainsi inférer le comportement spatial des consommateurs de la région (Figure 5b)



a) Magasins et IRIS



b) Comportement spatial inféré

Figure 5 Magasins et consommateurs de la région Centre

La matrice de probabilités ainsi calculée décrit comment la clientèle de chaque IRIS se répartit entre les différents magasins voisins, mais également dans quelle mesure chaque IRIS fait partie de la zone de chalandise d'un magasin (probabilité inverse). On peut donc connaître de manière assez précise quelle est la composition d'une zone de chalandise en termes d'IRIS, et donc en termes de profil socio-démographique (via l'information de l'INSEE par IRIS), et en termes de concurrence (via l'information sur la localisation des points de vente). Pour chaque magasin, on calcule alors des indicateurs de concurrence : le nombre d'IRIS sur lesquels s'étend sa zone de chalandise ($nbiris$), et la surface totale de ses concurrents ($Tsurf$). Pour aller plus loin, et prendre en compte les aspects stratégiques plus qu'économiques de la concurrence, nous calculons également un indice de présence de chaque enseigne de distribution dans la zone de chalandise de chaque magasin. Cet indice, d'abord construit comme une variable discrète au niveau de chaque IRIS (1 : telle enseigne est présente, 0 : elle est absente), est agrégée au niveau des magasins en prenant en compte le poids dudit IRIS dans la zone de chalandise du magasin considéré. Cela révèle, pour chaque magasin, quelle est l'enseigne qui lui fait localement le plus de concurrence.

Au niveau de l'assortiment, nous nous concentrons sur les variables les plus importantes dans le choix des consommateurs, à savoir les prix et la variété (Hoch et al., 1999). Le prix correspond ici à la médiane des prix constatés dans l'assortiment ($Mprix$), et la variété au ratio des variantes d'un produit offert par le magasin sur le nombre de variantes totales de ce produit disponibles dans l'ensemble des magasins (il s'agit d'un indicateur très similaire à la Distribution Numérique, $Pvar$). Ces variables sont calculées pour chaque magasin dans chaque catégorie, puis centrées-réduites et agrégées au niveau de chaque magasin. Cette méthode, quoique simple, permet, du fait du grand nombre de catégories prises en compte (42), d'avoir une image assez précise de l'assortiment offert par chaque magasin.

Les données ainsi calculées relatives à la concurrence et aux assortiments sont intégrées dans un modèle explicatif des ventes :

$$Ventes_i = \alpha + \beta_1' Mprix_i + \beta_2' Pvar_i + g_1' nbiris_i + g_2' Tsurf_i + t' X_{enseigne} + e_i$$

Les ventes correspondent au logarithme des ventes en euros par magasin (i), les coefficients β sont associés aux variables d'assortiment, g aux variables de concurrence. La matrice $X_{enseigne}$ indique la présence de chaque enseigne, et e_i est un résidu spatialement autocorrélé.

Le résidu est ici particulier et correspond à une méthode de modélisation développée spécifiquement dans le domaine de l'économétrie spatiale. Il s'agit de la classe de modèle

SAR (*Spatially Auto-Regressive*). C'est en fait une extension des modèles temporels classiques de type AR (Auto-Régressif). Mais, ici, au lieu de considérer qu'il existe un lien spécifique entre un individu au temps t et ce même individu au temps $t-1$, on considère qu'il existe un lien entre un individu et son plus proche voisin dans l'espace. Ce sont donc les individus et non les périodes qui sont corrélées. Si cette idée est relativement simple, sa mise en œuvre n'a été que tardive car, alors que les périodes peuvent être définies comme des vecteurs, l'espace est défini comme une matrice : lorsqu'une observation est située dans le temps par un chiffre (t), elle est située dans l'espace par deux chiffres (longitude et latitude). Enfin, de la même façon qu'il existe des modèles AR(λ) indiquant le nombre de périodes autocorrélées, il existe des modèles SAR(p) indiquant le nombre des plus proches voisins auxquelles une observation est corrélée. La méthode d'estimation des modèles SAR(p) calcule automatiquement la meilleure valeur de p , laquelle peut également être choisie par l'utilisateur suite à l'analyse du semi-variogramme (pendant de la fonction d'autocorrélation pour les modèles AR).

Le modèle ci-dessous a été estimé en utilisant la librairie `spdep` du logiciel R, intégré à la plateforme du SWAD-GM développé auparavant.

Les résultats livrent un effet positif de la variété sur les ventes ($t=16,54$), des prix sur les ventes ($t=2,63$, rappelons ici qu'il s'agit de ventes en euros et non en volume...). Au niveau de la concurrence, l'étendue de la zone de chalandise a également un effet positif (nbiris, $t=9,88$), de même que la surface totale des concurrents (Tsurf, $t=3,57$). Ce dernier effet valide l'hypothèse de l'attractivité plus grande des magasins située dans une zone commerciale dense. Au niveau des enseignes, il est intéressant de remarquer que certaines ont un effet positif sur les ventes de leurs concurrents (Auchan, Champion, Intermarché), alors que d'autres impactent négativement les ventes des magasins concurrents (Casino, Cora, Carrefour, Géant). Dernier point, la présence des Hard-Discount a tendance à augmenter les ventes en valeur des hyper- et supermarchés de leur zone ($t=4,02$). Ce dernier résultat, quelque peu contre-intuitif peut aider les managers à voir dans l'implantation d'un concurrent hard-discount une opportunité de relever la qualité de leurs assortiments et d'augmenter leurs ventes en valeur.

Etude 3 : Analyse spatiale des métriques du marketing direct

La troisième étude se concentre sur la représentation spatiale en temps réel des principales métriques utilisées dans le marketing direct. Elle insiste également sur les aspects conviviaux du SWAD-GM qui peuvent favoriser son adoption par les managers.

Un des objectifs est donc, à partir d'une base de données clients, les statistiques d'ancienneté, de récence, de fréquence et de montant (ARFM), et de projeter ces informations sur une carte.

Pour cette analyse, les données proviennent d'une base de données clients d'une chaîne de magasins spécialisés. Elles concernent les six points de vente de la région Bretagne. Ce sont des données de carte de fidélité relative à un peu plus de 30 000 clients sur une période de 57 mois (presque 5 ans) recensant plus de 300 000 actes d'achat. La Figure 6 reprend la décomposition temporelle de ces données de vente sur les 6 points de vente agrégés (analyse effectuée via PL/R). On peut voir que l'activité de l'enseigne se porte plutôt bien et que la composante saisonnière n'est pas négligeable.

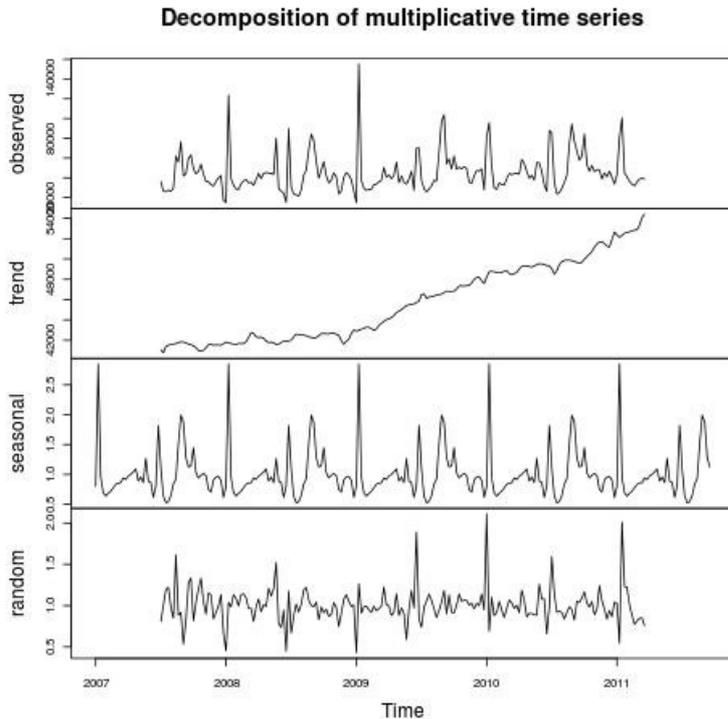


Figure 6 : Décomposition de la série temporelle des ventes des 6 points de vente

Si l'on se concentre sur une des villes d'implantation de l'enseigne en Bretagne, on peut également estimer des zones de chalandise à l'aide de la méthodologie des polygones gravitaires précédemment explicitée. Ces zones apparaissent à la Figure 1 où figurent, en rouge, les points de vente de l'enseigne, et en vert ses principaux concurrents.

Ces deux aspects (décomposition temporelle des ventes et zones de chalandise) soulignent le fait que ces données ont une dimension temporelle et une dimension spatiale. L'une des façons de mettre en scène ces deux dimensions est la création de cartes animées. Ce genre d'animation peut être créée sous PL/R. Le lien ci-après pointe vers une vidéo de ce type où l'on voit la succession des actes d'achats (figurés par des points) pour un des magasins de l'enseigne au cours du temps.

<http://www.dailymotion.com/GeoMarket#videoId=xmi48u>

Figure 7 : vidéo des actes d'achat pour un magasin

Ce genre de cartes animées permet de visualiser intuitivement comment la clientèle a réagit à tel ou tel événement survenu à un moment précis. Même si sa portée explicative est limitée, elle constitue un aspect important dans l'adoption des méthodologies spatio-temporelles par les managers.

Dans cette même idée (adoption du modèle par les managers), nous avons construit une interface utilisateur via une page web où le manager peut, de façon interactive construire la carte qui l'intéresse (Figure 8).

Sur la cette interface, on voit apparaître à gauche les éléments que l'utilisateur peut ajouter à sa carte (Magasins de l'enseigne, concurrents, zones d'attraction ou contours pour les zones de chalandises, éléments du fond de carte, niveau de zoom). En cliquant sur « Go », la carte est générée en temps réel, et l'utilisateur la voit apparaître sur le côté droit.

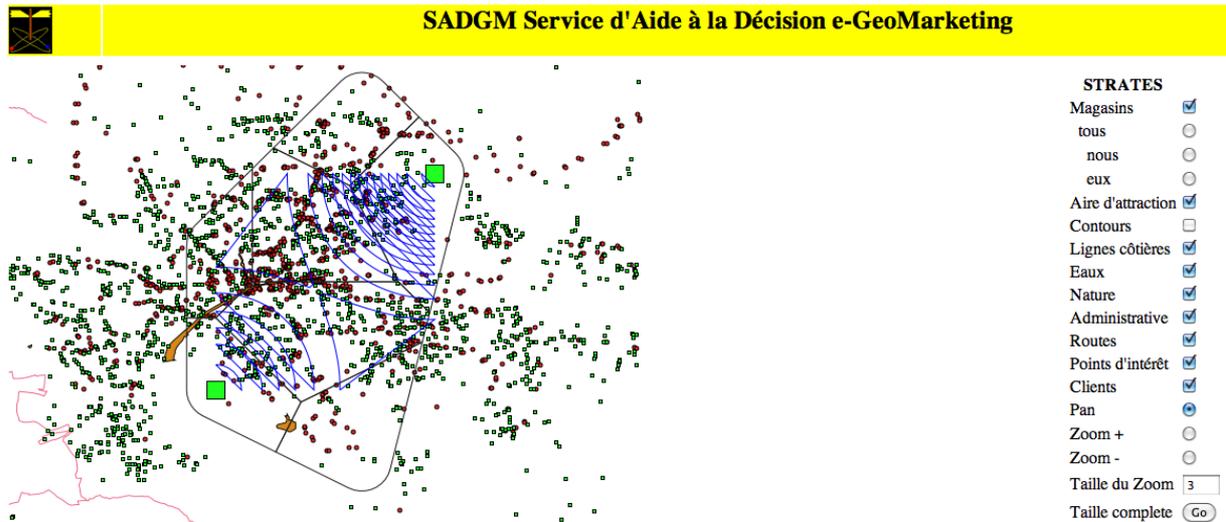


Figure 8 : interface utilisateur

Enfin, des éléments méthodologiquement plus complexes ont également été introduits. Il s'agit ici, comme nous l'avions évoqué, de présenter une visualisation spatiale de la principale métrique du marketing direct : la segmentation ARFM (Ancienneté, Récence, Fréquence, Montant). Ces quatre variables ont été calculées classiquement à partir de la base de données clients. Par la suite, les variables Ancienneté et Récence ont été soustraite l'une à l'autre pour donner une mesure à notre sens plus pertinente : la longueur de la relation client (Récence moins Ancienneté).

A partir de ces trois variables, nous avons réalisé un clustering par la méthode des k-means. Le Tableau 1 reprend les valeurs moyennes des trois clusters livrés par l'analyse. Ils se différencient notamment par la longueur de la relation client et la fréquence de visite des points de vente.

MOYENNES	fréquence (sur 5 ans)	montant (€)	longueur (jours)
Cluster 1	2,33	41,8	97,94
Cluster 2	8,84	38,3	753,45
Cluster 3	22,35	38,9	1473,52

Tableau 1 : valeurs moyennes des trois clusters

Les clients étant géolocalisés, il peut être intéressant de projeter ces clusters, non pas dans le plan factoriel comme il est classiquement fait, mais sur une carte, afin de voir où se situent principalement les clients de chaque clusters. Le résultat de cette projection est livré par la Figure 9

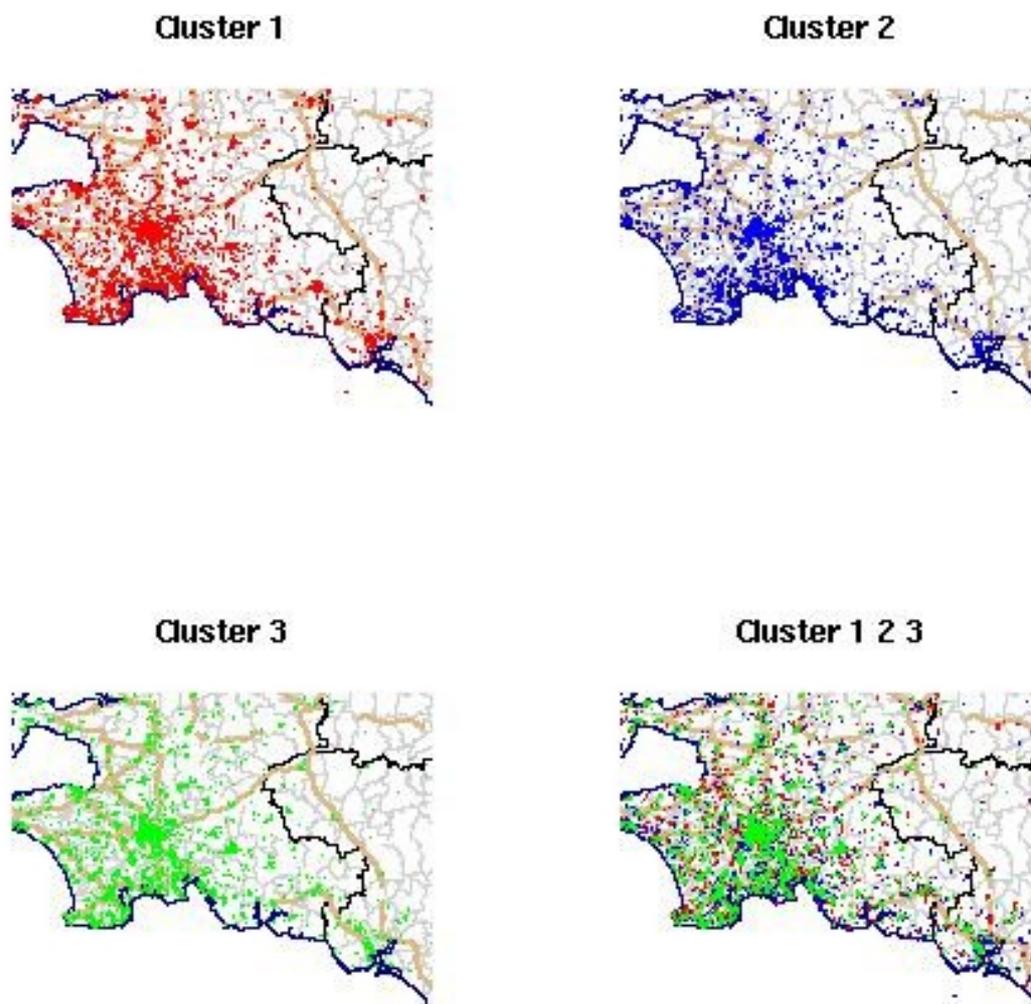


Figure 9 : Projection des clusters de clientèle sur une carte

Conclusion

L'objectif de cet article n'est pas de développer en détail une méthodologie, mais bien plutôt de présenter un système permettant l'utilisation de méthodes connues en marketing par les managers.

Il se fonde sur le double constat que les managers disposent de nombreuses données géographiques, et que les services web de géolocalisation ont connu un succès croissant suite à l'apparition de Google Maps. Sur cette base nous sommes intéressés à la façon dont les universitaires pourraient proposer des modèles complexes de géomarketing aux managers en les intégrant dans des services en ligne afin de faciliter leur adoption et leur utilisation.

Le service que nous avons mis au point (SWAD-GM) est donc à la fois complet (et extensible) au plan des méthodes, et d'utilisation relativement aisée par les managers (tout au moins pour les routines les plus simples, ou pré-programmées par nos soins).

Les études que nous avons présentées démontrent à notre sens le vaste champ des possibilités d'application de ce service.

Il existe bien entendu plusieurs limites à cette recherche encore en cours. Notamment, la focalisation est mise pour le moment sur la facilité d'utilisation. Nous n'avons pas encore intégré les modèles réellement plus complexes de l'économétrie spatiale. Par exemple, en se fondant sur des travaux d'autres domaines scientifiques, il est possible de développer une

modélisation spatiale pour la prévision des prospects (les biologistes utilisent par exemple ces modèles pour la prévision de la diffusion des espèces en fonction des caractéristiques des lieux d'habitat, Franklin 2009).

Cependant, nous pensons que la complexification des méthodes doit se faire pas à pas, en veillant à chaque étape à conserver la facilité d'utilisation.

Bibliographie

Bivand, R.S., Pebesma E.J. and Gómez-Rubio V. Applied Spatial Data Analysis with R, Springer,

Campo, K., Gijbrecchts, E., Goossens, T., & A.Verhetsel (2000) The impact of location-specific factors on the attractiveness and optimal space shares of product categories, International Journal of Research in Marketing, 17, 4, 255-279

Conway E.J. (2009) PL/R User's Guide - R Procedural Language, ..

Delaunay, B. (1932), Neue Darstellung der geometrischen Krystallographie, Zeitschrift für Krystallographie, 84, 109-149.

Delaunay, B. (1934), Sur la sphère vide, Izv.Akad.Nauk SSSR Otdelenie Matemat. Estestvennyka Nauk , 7, 793-800.

Dirichlet, G.L. (1850), Über die reduktion der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen, Journal für angewandte Mathematik, 40, 216.

Franklin, J. (2009), Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, UK

Hoch, S.J, Bradlow, E.T., & Wansink, B. (1999), The variety of an assortment, Marketing Science, 18, 4, 527-546

Huff, D.L. (1963), A Probabilistic Analysis of Consumer Spatial Behaviour, in William S. Decker (ed.), Emerging Concepts in Marketing, Chicago: American Marketing Association, pp. 443-461.

Huff, D.L. (1964): Defining and estimating a trade area, Journal of Marketing, 28, 34-38.

Huff, D.L. (1966), A programmed solution for approximating an optimum retail location, Land Economics, 42, 293-303.

Jain, A.K., & Mahajan V. (1979), Evaluating the competitive environment in retailing using multiplicative competitive interactive models, in Sheth (ed.) Research in Marketing, Greenwich, 217-235.

Luce, R. (1959), Individual choice behaviour, New York: John Wiley & Sons.

Montgomery, A. (1997), Creating micro-marketing pricing strategies using supermarket scanner data, *Marketing Science*, 16(4), 315-337

Nakanishi, M., & Cooper, L.G. (1974), Parameter estimate for multiplicative interactive choice model: least squares approach, *Journal of Marketing Research*, 11, 303-311.

R Development Core Team. (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Viena, Austria, <http://www.R-project.org>

Reilly, W.J. (1931), *The law of retail gravitation*, New York, Knickerbocker Press.

Thiessen, A.H., & Alter, J.C. (1911), Precipitation averages for large areas, in: *Monthly Weather Review*, 39, 1082-84.

Voronoi, G. (1907), Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire: Sur quelques propriétés des formes quadratiques parfaites", *J. Angew. Math.*, 133, 97-178.

Voronoi, G. (1908), Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire: Recherches sur les parallélogrammes primitifs, *J. Angew. Math.*, 134, 167-171.