

**Optimisation de la localisation commerciale :**  
**une application du traitement du signal et du modèle p-médian**

Jérôme Baray

L'auteur voudrait remercier en particulier les lecteurs anonymes pour leurs commentaires éclairés, le Professeur Gérard Cliquet pour ses précieux conseils, ainsi que Mme Dominique Maraine pour le suivi administratif des modifications.

Optimisation de la localisation commerciale : une application du traitement du signal et du modèle p-médian

**Résumé :** Cet article décrit un algorithme capable de détecter les meilleures localisations pour des points de vente. Améliorant les méthodes d'agrégation spatiale en prenant en compte une délimitation de la zone de chalandise, il se décompose en quatre étapes qui sont:

- (1) le géocodage des adresses de consommateurs à partir d'une base de données,
- (2) la délimitation de la zone de chalandise correspondante grâce à un filtre mathématique,
- (3) l'identification de centres de gravité d'aires formant cette zone de chalandise,
- (4) la résolution d'un modèle p-médian dans lequel les nœuds du réseau correspondent aux centres de gravité prédéterminés.

Intégrant le fait que les clients possèdent une certaine mobilité au sein de la zone de chalandise, cette méthode est rapide, précise et peut fournir les localisations optimales en particulier dans le cas d'un nombre très important de consommateurs potentiels.

Retail Store Location optimisation: an application of signal processing with the p-median model

**Abstract:** This paper presents an algorithm based on an improved spatial clustering procedure to find the best retail store locations based on 4 major steps:

- (1) a geocoding of customer addresses from a database,
- (2) a delimitation of the corresponding trade area thanks to a mathematical filter,
- (3) an identification of gravity centers of zones belonging to this trade area,
- (4) a solving of a p-median model in which the nodes of the network are the gravity centers.

Taking into account the concept of a consumer mobility in the trade area, the method is fast, accurate and can give optimal locations even for a huge number of potential consumers.

**Mots-clé :** p-médian, problème de localisation-allocation, zone de chalandise, optimisation des localisations commerciales

**Keywords:** p-median, location-allocation problem, trade area, retail store location optimisation

## OPTIMISATION DE LA LOCALISATION COMMERCIALE : UNE APPLICATION DU TRAITEMENT DU SIGNAL ET DU MODELE P-MEDIAN

### INTRODUCTION

Les regroupements massifs des réseaux de distribution ces dernières années et la conjoncture fluctuante et imprévisible ont rappelé que leur organisation n'était jamais figée et restait soumise à la dure loi de la concurrence. Des dizaines de grandes entreprises et leurs filiales se voient contraintes de restructurer selon leur activité, leur mode de production ou leur réseau de distribution. Pour des réseaux commerciaux comportant plusieurs centaines de points de vente ou même plusieurs milliers comme dans le cas des agences de compagnies d'assurance ou bancaires, cette réorganisation s'avère être une tâche longue et difficile. Chaque point de vente doit être passé à la loupe, être comparé en terme de rentabilité, de chiffre d'affaires et de zone de chalandise par rapport à ses plus proches voisins dans certains cas concurrents par le passé et désormais, partenaires d'un même réseau. De la configuration spatiale du nouveau réseau commercial va dépendre son pouvoir de capter la clientèle et une grande précision s'avère donc nécessaire dans la localisation des points de vente. A cette finesse, doit s'allier la rapidité de l'étude et de l'application sur le terrain des choix d'implantation pour bénéficier au plus tôt d'un réseau de vente optimisé possédant en principe une rentabilité commerciale accrue.

Or, les méthodes actuelles de détermination de localisations optimales ne répondent pas encore tout à fait à ces exigences. Dans un premier temps, seront présentées les principales méthodes théoriques et pratiques de délimitation des zones de chalandise et de détermination de localisations optimales, démarches souvent séparées l'une de l'autre. On montrera qu'en effet, ces procédures sont souvent peu précises ou compliquées pour être mises en pratique par les professionnels. La résolution d'un modèle de localisation-allocation tel que le p-médian (Kariv et Hakimi 1979, Reville et Swain 1970) est en effet susceptible d'engendrer un

nombre si important d'opérations mathématiques que les ordinateurs les plus puissants ne peuvent le résoudre en un temps raisonnable. Des techniques d'agrégation existent pour simplifier le modèle p-médian, mais elles ignorent la notion de zone de chalandise (Yip et Zhao 1996). Ainsi, sera présenté à ce niveau un nouvel instrument de décision précis et rapide sous la forme d'un algorithme intégrant harmonieusement la délimitation d'une zone de chalandise et la détermination des meilleurs emplacements commerciaux dans l'optique de la restructuration ou de la création d'un réseau de points de vente. Des méthodes originales de filtrage et de convolution couramment utilisées en traitement de l'image et en reconnaissance de forme seront mises à contribution dans cette démarche géomarketing pour simplifier l'expression du modèle p-médian à travers une amélioration des techniques d'agrégation. On vérifiera les qualités de précision et de rapidité de cette procédure, ainsi que sa fiabilité, en la comparant à l'utilisation habituelle du modèle p-médian, pour la délimitation d'une zone de chalandise de clients potentiels et la recherche de localisations optimales de magasins de produits biologiques dans l'Ouest parisien.

## ENJEUX ET PRATIQUES DE LA LOCALISATION COMMERCIALE

### *La zone de chalandise du point de vente et son appréciation*

Le choix d'une bonne localisation est sans doute l'une des décisions les plus importantes qu'un manager doit prendre car l'emplacement du point de vente est en effet un investissement fixé sur le long terme et son choix bon ou mauvais se ressentira sur la performance commerciale. Toute étude de localisation s'accompagne au préalable de l'identification et le repérage dans l'espace d'une clientèle potentielle qui constituera le fond de commerce du point de vente. En pratique, il s'agit de délimiter une aire géographique rassemblant l'essentiel de cette clientèle, dénommée zone de chalandise. Plusieurs méthodes ont été proposées par le passé pour apprécier au mieux la zone de chalandise dont les

méthodes normatives théoriques et les méthodes subjectives qui se fondent sur l'expérience et la connaissance des consommateurs. La principale méthode subjective est représentée par la *méthode du temps de conduite* qui considère que les clients ne sont prêts à parcourir qu'une distance ou qu'un temps limite pour rejoindre le point de vente : la zone de chalandise comprend alors les aires pas trop éloignées du point de vente selon ce critère de distance limite souvent mesuré en temps de conduite. Or, la proximité n'est dans les faits pas le seul critère de fréquentation d'un magasin. Encore faut-il en particulier, que la nature de l'offre proposée par le point de vente corresponde à une demande.

Les méthodes normatives comprennent la *méthode analogique* (Applebaum 1968), le *modèle de régression*, la *méthode par les surfaces enveloppantes* (Peterson 1974) et la *méthode des nuées dynamiques* (Roger 1983). La méthode analogique sous-tend que deux magasins placés à deux emplacements identiques auront la même performance commerciale : la zone de chalandise est considérée comme étant le disque dont le rayon a une dimension telle qu'il concentre au minimum X % de la clientèle (X=80 % par exemple). Les méthodes précédemment citées supposent que la zone de chalandise est une zone géographique compacte, mais dans les faits, les barrières naturelles (infrastructure routière, parcs) ont tendance à fragmenter cette zone en de multiples aires géographiques. Le *modèle de régression* cherche à mesurer l'influence de variables quantitatives socio-économiques, environnementales et marketing sur un paramètre de performance commerciale. La *méthode par les surfaces enveloppantes* (Peterson 1974) consiste à représenter les taux de pénétration sur une carte quadrillée en zones de manière à obtenir un relief dont la surface est évaluée par des courbes ou des plans et en particulier par des équations dont les coefficients sont déterminés grâce à un modèle de régression. La *méthode des nuées dynamiques* (Roger 1983) construit itérativement une classification d'un nuage de points. Les deux dernières méthodes engendrent un nombre important de calculs, tout cela pour ne réussir qu'à approcher, à l'aide

de fonctions mathématiques idéales, la forme de la zone de chalandise. D'après des entretiens menés auprès de responsables d'entreprises, ces deux dernières méthodes sont rarement appliquées dans les cas concrets de réorganisation ou de création de réseaux commerciaux.

De fait, toutes ces méthodes de description de la zone de chalandise, une phase primordiale dans la recherche d'une bonne localisation, sont compliquées, approchées et réduisent le plus souvent la zone de chalandise à une aire compacte et centrée autour du magasin alors que cette dernière est généralement morcelée du fait des irrégularités géographiques (barrières naturelles, infrastructures routières, séparation de la population en quartiers).

### *Théorie et pratique de la localisation commerciale*

La phase de recherche d'une localisation commerciale s'opère après que la société se soit attachée à déterminer sur quel marché elle souhaite s'implanter et en ayant à l'idée le nombre exact de points de vente à créer en fonction de la saturation du marché pour le produit ou le service proposé et des capacités financières disponibles. Une première catégorie de modèle de recherche théorique de localisation se fonde sur l'analogie gravitaire : l'un des critères principaux de ces modèles est la distance et l'on cherchera ainsi à se placer au plus proche de la masse des clients de manière à en attirer le plus grand nombre. Citons parmi ces modèles dits modèles d'interaction spatiale, la *théorie des places centrales* (Christaller 1935) qui considère l'espace géographique comme comprenant une répartition uniforme des clients ayant tous un comportement identique : Selon cette théorie, l'implantation optimale se situe au centre d'un hexagone dont les sommets sont occupés par six magasins élémentaires. La *méthode des secteurs proximaux* est une variation de la théorie précédente qui utilise le principe des polygones de Thiessen (1911) ou de Dirichlet (1850). Les modèles d'interaction spatiale comprennent également la *loi de Reilly* (1931) qui prend l'hypothèse que la population intermédiaire I localisée entre deux pôles urbains A et B sera attirée par chacun de

ces pôles proportionnellement à leur taille et en proportion inverse de la distance entre la zone I et les pôles urbains A et B. Le *modèle de Huff* à l'approche probabiliste introduit dans la loi précédente la notion de surface de vente du magasin qui selon lui, joue un rôle tout aussi important dans son attractivité vis-à-vis des clients que sa proximité. Le *modèle MCI* ou Modèle Interactif de Concurrence est une généralisation du modèle de Huff qui évalue la probabilité qu'un consommateur fréquente un point de vente donnée. Il tient compte encore d'autres paramètres d'attraction que la distance ou que la surface de vente comme le prix des articles, le nombre de caisses ou même la perception des consommateurs dans le cas du MCI subjectif (Cliquet 1995).

La recherche de localisations utilise une deuxième catégorie de modèles nommés modèles de localisation-allocation qui tentent de trouver les emplacements les plus proches des clients grâce à des algorithmes mathématiques. Tout modèle de localisation-allocation comprend (1) une *fonction objectif* quantifiant l'éloignement du magasin aux clients ou plus généralement mesurant un critère de performance du point de vente par rapport à sa localisation ; (2) des *points de demande* ou nœuds de demande représentant l'importance de la clientèle et son désir d'acheter un produit ou un service particulier ; (3) les *emplacements potentiels* ou nœuds d'emplacement potentiel correspondant aux emplacements possibles du ou des points de vente à localiser ; (4) la *matrice d'éloignement ou de temps* mesurant les distances kilométriques ou temporelles entre les points de demande et les emplacements potentiels ; (5) la *règle d'allocation* qui spécifie à quelles conditions les points de demande représentant les clients fréquenteront les points de vente placés au niveau d'emplacements potentiels.

Le modèle de localisation-allocation le plus utilisé pour la recherche de localisations commerciales optimisées est le modèle p-médian dont les champs d'applications vont du transport à la grande distribution, en passant par les services bancaires et l'assurance : sa problématique est de trouver les localisations pour un nombre p d'activités devant fournir une

panoplie de services ou de produits à  $n$  clients de telle manière que la somme de l'ensemble des distances séparant les clients de l'activité voisine soit minimale. La formulation mathématique du modèle  $p$ -médian est la suivante :

$$\text{Minimiser } \sum_i \sum_j a_i d_{ij} x_{ij} \quad \text{représente la fonction objectif,} \quad (1)$$

$$\text{avec } \sum_i x_{ij} = 1, \quad \forall i, \quad \text{assure que tous les clients sont assignés à une activité et une seule,} \quad (2)$$

$$x_{ij} \leq y_j, \quad \forall i, j \quad \text{empêche d'assigner un client à une activité si elle n'est pas ouverte,} \quad (3)$$

$$\sum_j y_j = p, \quad \text{le nombre total d'activités est } p, \quad (4)$$

$$x_{ij}, y_j \in \{0,1\}, \quad \forall i, j \quad \text{nature binaire des variables } x_{ij}, y_j \quad (5)$$

où

$a_i$  : la demande au nœud  $i$ ,

$d_{i,j}$  : la distance du nœud  $i$  au nœud  $j$ ,

$p$  : le nombre d'activités à localiser,

$x_{i,j} = 1$ , si le nœud  $i$  est assigné à l'activité  $j$  et 0 autrement,

$y_j = 1$ , si l'activité  $j$  est ouverte et 0 autrement.

Le problème est que le  $p$ -médian est réputé appartenir à la classe des problèmes connus comme étant NP-complets (Kariv et Hakimi 1979): ses solutions issues d'algorithmes linéaires deviennent insolubles au fur et à mesure que le nombre des variables (activités à localiser et clients) augmente puisqu'on assiste à une progression exponentielle de la taille du problème. Il existe néanmoins un certain nombre d'heuristiques pour trouver un optimum acceptable au problème  $p$ -médian malgré le fait que toutes ces solutions ne convergent que vers des optima locaux et non vers une solution globale, et qu'il ne soit pas possible a priori de connaître le niveau d'optimalité de cette solution. Dans les algorithmes de résolution fondamentaux, on trouve l'algorithme flou, l'algorithme de recherche de voisinage, l'algorithme par les multiplicateurs de Lagrange, l'algorithme génétique, une heuristique de

substitution (Teitz et Bart 1968) et ses variantes (Goodschild et Noronha 1983). D'une façon générale, les heuristiques se rangent en deux classes (Golden et alii 1980): les *algorithmes de construction* qui permettent de rechercher des localisations avec un risque élevé d'obtenir une solution sous-optimale et les *algorithmes d'amélioration* destinés comme leur nom l'indique, à améliorer les résultats fournis par les algorithmes de construction. Ainsi l'algorithme flou est du type construction de même que l'algorithme génétique et par les multiplicateurs de Lagrange. L'heuristique de substitution et l'algorithme de recherche de voisinage ont une approche d'amélioration.

#### *La recherche de localisations optimales en pratique*

Dans la pratique, pour éviter la mise en œuvre complexe des modèles mathématiques précédents, les professionnels utilisent des méthodes statistiques en général assez simples et rapides, mais approximatives puisqu'elles relèvent du calcul statistique élémentaire et ne tiennent pas compte de la localisation précise des clients (ces méthodes n'exploitent en particulier pas les bases de données de clients au niveau de l'adresse). Citons parmi elles la *méthode par les parts de marché et les surfaces de vente* qui consiste à évaluer la taille de la zone de chalandise à l'emplacement pressenti puis à calculer le chiffre d'affaires escompté en pratiquant une règle de trois après avoir comparé cette zone de chalandise à celles de magasins existants. La *méthode analogique* consiste à calculer le chiffre d'affaires par client pour un magasin et à l'extrapoler à celui de l'emplacement prévu du nouveau point de vente après avoir évalué le nombre de clients potentiels sur lesquels il pourra compter. Le *modèle de régression multiple* cherche à corrélérer une mesure de la performance d'un magasin (chiffre d'affaires, résultat) avec divers critères comme le revenu des ménages dans la zone, le niveau de concurrence, la part de marché estimée,... Il s'agit ensuite de transposer l'équation de régression obtenue aux sites potentiellement intéressants pour retenir celui ayant la performance la plus élevée. La *méthode du discriminant* est une variation du modèle de

régression qui sépare grâce à une analyse multidiscriminante un ensemble de magasins en plusieurs groupes selon leurs résultats commerciaux, en l'occurrence les magasins ayant une performance acceptable et les autres plutôt inacceptables. L'analyse discriminante s'apparente à l'analyse de régression multiple, mais s'applique dans le cas où la variable dépendante est de nature qualitative. On recherchera ainsi ultérieurement à créer un nouvel établissement dans un endroit ayant les caractéristiques d'une zone où est implanté un magasin performant avec éventuellement les caractéristiques de ce type de magasins. La *méthode du marché potentiel* découle directement du principe de gravité du commerce de détail de Reilly : souvent utilisée par la grande distribution, ce mode de recherche par tâtonnements tente de minimiser la distance du futur emplacement aux consommateurs potentiels et la maximise par rapport à celle des concurrents. Enfin, le *modèle p-médian* lui-même, sert à localiser des points de vente en repérant les adresses ou les origines géographiques de clients potentiels le plus souvent par des enquêtes de terrain. Dans un premier temps, lors de la construction du modèle, on cherche à assigner chaque client de la base de données d'adresses à un quartier ou à un secteur urbain rassemblant un ensemble de rues ou d'îlots, de manière à réduire le nombre de cellules d'analyse. Ainsi, c'est le centre de gravité de chacun de ces secteurs qui servira de nœud au réseau du modèle. Le nombre de nœuds est alors considérablement réduit par rapport au cas où l'on aurait créé pour chaque client un nœud particulier. Le problème est que ce découpage en secteurs géographiques correspond plus à une logique administrative qu'à une logique commerciale et que d'une manière générale, cette procédure de simplification à l'extrême de la réalité commerciale nuit considérablement à la précision et même à la fiabilité des résultats obtenus suite à la résolution du modèle. Il existe d'autre part des procédures automatiques d'agrégation spatiale de clients comme la méthode des k plus proches voisins (Yip et Zhao 1996), mais celles-ci ont tendance à entacher d'erreurs les résultats de la localisation des points de vente (Hodgson et Shmulevitz 1996, Murray et

Gottsegen 1997, Erkuz et Bozkaya 1999) car la logique de regroupement ne prend en général pas en compte la notion de zone de chalandise. Outre que le nombre d'agrégats doit parfois être spécifié au préalable, les points de demande sont considérés comme étant fixes. Or, si l'on considère les bases de données de domicile, un client ne reste pas en général complètement immobile au niveau de son adresse, mais possède une certaine probabilité d'être présent au voisinage de cette adresse. C'est donc en terme de zone géographique qu'il faut considérer la demande représentant la clientèle et non selon une série de points précis de demande comme le font les méthodes actuelles d'agrégation.

Mais, malgré les travaux de recherche, il est vrai que l'ensemble des méthodes utilisées en pratique fait de plus largement appel à l'intuition, ce fait ayant été démontré par un certain nombre d'études (Clarke, Horita et Mackaness 2002, Hernandez 1998). Cela a conduit à proposer une nouvelle méthode plus fiable, rapide et facile d'utilisation, méthode faisant appel au modèle p-médian, aux méthodes d'agrégation et aux notions de traitement du signal.

#### UNE NOUVELLE APPROCHE POUR LA RECHERCHE DE LOCALISATIONS OPTIMALES

Les constatations précédentes ont posé le problème de trouver une approche souple et fiable pour en premier lieu délimiter précisément les zones de chalandise ou même toute autre zone géographique possédant des caractéristiques commerciales, économiques, sociologiques propres puis rechercher des emplacements optimisés pour la création ou la redistribution de points de vente. Cette nouvelle méthode devait à la fois être rapide, précise, de préférence convertissable en un algorithme et un logiciel pouvant fonctionner sur les calculateurs courants du type micro-ordinateur.

Le traitement du signal dont les algorithmes sont capables de traiter les informations en dynamique a fourni de nombreuses applications dont en particulier la plus élaborée, très gourmande en information et en temps de calcul, la vision artificielle, une technique complémentaire de la robotique. La nouvelle approche proposée utilise les principes du

traitement du signal associés au modèle p-médian pour déterminer des emplacements optimisés pour la création de points de vente. Cette approche duale analysant géographiquement la demande pour ensuite optimiser l'offre se compose de quatre phases, (1) le géocodage et la représentation cartographique des localisations de clients potentiels, (2) la délimitation des zones de chalandise en utilisant les principes de traitement du signal, (3) le calcul des caractéristiques de chaque aire formant dans leur ensemble la zone de chalandise, ces caractéristiques formant un réseau p-médian avec ses nœuds (centres de gravité de chaque aire) et ses niveaux de demande (étendue des aires) et enfin, (4) la résolution de ce réseau p-médian simplifié par les heuristiques traditionnelles. Eventuellement, on pourra ensuite se focaliser au niveau de chaque aire identifiée comme pouvant constituer une localisation optimale, en effectuant la même procédure mais à une échelle plus fine pour encore mieux préciser la localisation commerciale (Figure 1).

Ainsi, la partie centrale et novatrice de cette méthode est constituée par la délimitation des zones de chalandise (zones denses de clientèle) au niveau desquelles il sera possible de définir un centre de gravité caractérisant la localisation spatiale de ladite zone. Ces centres de gravité constitueront les nœuds d'un modèle p-médian. Des études ont montré que le centre de gravité des agrégats, en l'occurrence ici les aires de chalandise, est leur meilleur représentant pour constituer les nœuds d'un modèle p-médian simplifié représentatif de la clientèle initiale (Plastria 2001). Il est donc fondamental de définir avec précision les contours de ces zones de chalandise: une évaluation grossière des frontières induirait nécessairement un mauvais repérage des centres de gravité ce qui ne manquerait pas de se répercuter sur le résultat final de localisation des p points de vente à placer. La deuxième amélioration découle du fait que l'on simplifie le problème en l'appréhendant de manière globale grâce à l'identification de "paquets" de clients, et qu'ensuite seulement, on pousse l'analyse dans le détail en recherchant dans les zones détectées comme potentiellement intéressantes, des emplacements plus précis,

toujours en utilisant la même technique de délimitation de zones de chalandise et de construction d'un modèle p-médian. On utilise dans cette méthode, à la fois des principes de la localisation discrète (à travers le modèle p-médian) et du modèle planaire (à travers le calcul de centres de gravité).

### *La délimitation des zones de chalandise par traitement du signal*

Le géocodage et la représentation des données géomarketing (phase 1) issues le plus souvent d'enquêtes ne posent en général pas de problème. La deuxième phase de délimitation de la zone de chalandise mérite une description plus approfondie. Elle se compose d'un *prétraitement des données par un filtrage* directement sur la carte représentant la localisation des clients (figure 2) pour accentuer les contours de la zone de chalandise. De nombreux types de filtrage existent comme les filtres sigma, Nagao ou médian déjà utilisé par McKay (1973) sur des cartographies de clients. Le *filtrage en moyenne* est le plus simple des filtres : il consiste à remplacer chaque point de la carte par la valeur moyenne d'une petite aire centrée sur ce point et à reconstruire ainsi toute la carte. Le *filtre médian* prend lui la valeur médiane de chaque petite aire au lieu de la valeur moyenne. Le *filtre sigma* étudie la statistique de dispersion des valeurs en chaque point et prend le plus souvent comme nouvelle valeur en chaque point, la moyenne de ses voisins, en ne considérant que ceux dont la différence de niveau avec le point considéré ne dépasse pas en valeur absolue deux fois la variance de la région considérée. Le *filtre de Nagao* (1979) est intéressant car il conserve beaucoup mieux la forme des contours des objets d'une image que les autres filtres : il revient à analyser un certain nombre de configurations spéciales (tangentes, horizontales ou verticales au voisinage de chaque point) et à retenir celle qui est la plus adaptée comme nouvelle valeur du point balayé. Enfin, mentionnons la transformation de *dilatation* qui consiste à augmenter la surface de chaque point en le grossissant (chaque point représentant un client sur l'image)

jusqu'à ce que ceux-ci se touchent ce qui permet de remplir les interstices : ce prétraitement est fondamental pour obtenir une représentation générale de la zone de chalandise lorsque les clients ou points sont très séparés les uns des autres dans l'espace géographique et qu'il n'est alors pas possible de repérer visuellement les formes pleines de la zone de chalandise.

Après avoir amélioré l'aspect global de l'image représentant la dispersion des clients en accentuant les contours de la zone de chalandise, on procède dans la suite de la méthode, à une délimitation de la zone de chalandise par un masquage (un type spécial de filtrage) réalisé grâce à une opération matricielle sur l'image, nommée convolution. Rappelons que l'opération de convolution entre deux matrices, l'image  $f_{i,j}$  et le filtre ou masque  $h_{i,j}$ , est définie par :

$$g_{i,j} = f_{i,j} * h_{i,j} = \sum_{\alpha=i-k}^{i+k} \sum_{\beta=j-k}^{j+k} f_{\alpha,\beta} h_{i-\alpha, j-\beta}$$

Les masques de Sobel sont parmi les plus courants pour délimiter les frontières des objets d'une image. Ils sont représentés par les deux matrices suivantes appliquées successivement par convolution sur l'image:

$$S_x = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad ; \quad S_y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

Matrice de détection horizontale ;                      et matrice de détection verticale

Les masques du gradient ou du laplacien constituent des alternatives aux masques Sobel moins précises dans la délimitation des frontières. Ainsi, la simple convolution de l'image avec de tels masques permet d'obtenir immédiatement les contours de la zone de chalandise (figure 3).

La troisième phase de la méthode consiste à repérer les coordonnées des centres de gravité des différentes aires constituant la zone de chalandise préalablement délimitée et de calculer la superficie de ces aires de manière à construire un réseau p-médian. Pour cela, le plus simple est de balayer l'image représentant les contours de la zone de chalandise avec un algorithme

de parcours qui va suivre les frontières de chaque aire de la zone en notant les coordonnées de chacun de ses points (figure 4).

Ensuite, il suffira de prendre la valeur moyenne des coordonnées des points de frontières pour obtenir les coordonnées du centre de gravité de l'aire. Si la frontière a été décrite parallèle en terme d'orientation Nord, Sud, Est, Ouest, la superficie de l'aire est donnée par l'algorithme suivant :

- 1) Au départ;  $u = 0$  et  $t = 0$
- 2) **De**  $i = 1$  à  $n$  **Faire**

A) <b>Si</b> $a_i = \text{Nord}$	<b>Alors</b> $t = t + 1$	<b>Sinon Aller en B)</b>
B) <b>Si</b> $a_i = \text{Sud}$	<b>Alors</b> $u = u + t$	<b>Sinon Aller en C)</b>
C) <b>Si</b> $a_i = \text{Ouest}$	<b>Alors</b> $t = t - 1$	<b>Sinon Aller en D)</b>
D) <b>Si</b> $a_i = \text{Est}$	<b>Alors</b> $u = u - t$	
- 3)  $S_g = u \times S$

où la *valeur du paramètre u* à la fin de la procédure est le nombre de points contenus dans la zone de chalandise considérée,  $t$  un paramètre de comptage,  $a_i$  les orientations géographiques successives décrivant le contour et  $S$  la superficie géographique unitaire d'un point.

Le repérage de coordonnées des centres de gravité correspondant aux futurs nœuds du modèle p-médian dans le présent algorithme, donnera aussi par un simple calcul de longueur, les distances de chaque segment du réseau ce qui permettra d'obtenir tous les paramètres du réseau p-médian (figure 5).

Il suffira donc en dernière étape de résoudre ce modèle p-médian composé d'un nombre limité de nœuds par les heuristiques classiques (algorithmes génétiques, de substitution,...) pour avoir les localisations optimales. Une étape facultative est de réitérer le même processus au niveau des aires mises en évidence par cette résolution pour affiner la position des localisations optimales. La méthode a été utilisée avec succès pour trouver les emplacements optimisés de magasins de produits biologiques dans l'ouest de la région parisienne.

UNE APPLICATION DU MODELE P-MEDIAN ASSOCIE AU TRAITEMENT DU SIGNAL DANS LE DOMAINE DE LA DISTRIBUTION DES PRODUITS BIOLOGIQUES

La vente de produits biologiques progresse en France et aux Etats-Unis au rythme de 20 % par an et les experts tablent qu'à moyen terme, 5 % des dépenses des français seront consacrés au bio (Sylvander 1999), soit encore une marge de progression de 400 %, ce qui laisse encore des opportunités pour créer des commerces vendant ce type de produits. Une base de données d'adresses de 10 211 clients potentiels intéressés par l'achat de produits biologiques dans les communes de Boulogne-Billancourt, Issy-Les Moulineaux, Neuilly-Sur-Seine, Paris 7<sup>ème</sup>, Paris 15<sup>ème</sup>, Paris 16<sup>ème</sup> et Paris 17<sup>ème</sup> a été utilisée pour mettre en pratique la méthode de localisation.

Dans la pratique, les spécialistes des problèmes de localisation réalisent beaucoup d'approximations pour simplifier l'approche et réussir à construire un modèle p-médian à partir d'une telle base. Selon la procédure classique, l'Ouest parisien est découpé en communes de périphérie et en arrondissements tels que Boulogne-Billancourt, Issy-Les-Moulineaux, Neuilly-Sur-Seine, Paris 7<sup>ème</sup>, Paris 15<sup>ème</sup>, Paris 16<sup>ème</sup>, Paris 17<sup>ème</sup>. Chaque élément de ce découpage comporte un nœud placé au centre de gravité du secteur géographique ou encore plus grossièrement un repère géographique connu du lieu considéré comme la mairie ou l'église. Les distances inter-nœuds sont ensuite calculées et à chaque nœud est affectée une pondération liée au nombre de clients potentiels habitant dans le secteur s'y rattachant. Un découpage généralement subjectif et distinct des divisions administratives classiques, ou un zonage plus fin du type quartier ou îlot Insee rend la tâche de construction du modèle p-médian plus compliquée. Il est nécessaire dans ce cas d'affecter par une procédure fastidieuse, chacun des 10 211 clients aux différentes cellules géographiques. Une alternative est d'affecter automatiquement les adresses grâce à un programme spécifique (écrit par exemple en MapBasic<sup>®</sup>). Mais, là encore, on n'échappe pas à la spécification précise lors de l'utilisation d'un tel logiciel, des limites de chaque zone ce qui peut prendre un temps considérable. Dans le cas étudié, un découpage géographique en arrondissements et

communes de périphérie obtenait ainsi un réseau de 7 nœuds soit, *un nœud pour Boulogne-Billancourt (poids de 1732 clients) ; un nœud pour Issy-Les-Moulineaux (poids de 952 clients) ; un nœud pour Neuilly-Sur-Seine (poids de 398 clients) ; un nœud pour Paris 7ème (poids de 517 clients) ; un nœud pour Paris 15ème (poids de 3456 clients) ; un nœud pour Paris 16<sup>ème</sup> (poids de 1087 clients) ; un nœud pour Paris 17ème (poids de 2069 clients).*

La résolution du modèle p-médian plus simple compte-tenu du très faible nombre de nœuds dans le réseau, préconise en examinant toutes les combinaisons possibles, les nœuds de Boulogne-Billancourt et des 15<sup>ème</sup>, 16<sup>ème</sup> et 17<sup>ème</sup> arrondissement comme localisations optimales de points de vente. Le problème est qu'une telle démarche a surestimé l'intérêt du 16<sup>ème</sup> arrondissement en considérant le 17<sup>ème</sup> arrondissement dans sa globalité, ne lui attribuant alors qu'un seul point de vente. Or, en réalité, le 17<sup>ème</sup> arrondissement de forme longiligne comporte deux pôles distincts à fort potentiel avec à l'Ouest, les beaux quartiers (avenue de la Grande Armée / Ternes) et bien plus à l'Est, des quartiers plus modestes (Guy Môquet / Batignolles). Chacun de ces pôles concentre autant de clients que le 16<sup>ème</sup> arrondissement où le potentiel est plus dispersé. Une procédure fiable devrait être capable de distinguer ces pôles pour éventuellement leur affecter un point de vente grâce au p-médian si leur potentiel le justifie.

L'algorithme fondé sur le traitement du signal détecte justement de façon automatique, les aires géographiques significatives sans introduire le biais d'un découpage administratif ou issu d'un jugement intuitif. Selon la description faite précédemment, les données ont d'abord été géocodées grâce au logiciel MapPoint<sup>®</sup>, et représentées sur une carte (figure 6). Le prétraitement a constitué en une dilatation des points représentant les adresses des clients de manière à combler les vides dans les zones denses de clientèle, en utilisant le programme de traitement du signal SCI<sup>®</sup> (figure 7). L'image ainsi traitée a alors subi une convolution par un filtre Sobel de manière à faire apparaître les frontières de la zone de chalandise. Les aires

constitutives ont ensuite été analysées en termes de coordonnées de leurs centres de gravité et de leurs superficies pour construire le réseau p-médian correspondant : chaque centre de gravité constitue alors un nœud du réseau, la demande en chaque nœud ayant la valeur de la superficie de l'aire considérée.

La résolution de ce modèle en utilisant le logiciel Sitation développé par l'université Northwestern aux Etats-Unis et vendu avec l'ouvrage du Professeur Daskin sur les algorithmes de localisation (1995), donne quelle que soit l'heuristique classique utilisée (algorithme flou, de voisinage, multiplicateurs de Lagrange et algorithme génétique) pour la recherche de 3 localisations optimales les aires 1 [Paris 17<sup>ème</sup> Guy Môquet / Batignolles], 10 [Paris 15<sup>ème</sup>], 18 [Boulogne-Billancourt Est] (voir figure 8). Cependant, la recherche de 4 localisations optimales fait apparaître en plus la partie Ouest du 17<sup>ème</sup> arrondissement recelant un fort potentiel de clients, au lieu du 16<sup>ème</sup> arrondissement donné par la résolution classique du p-médian comme on l'a vu au paragraphe précédent. La même procédure a été réitérée au sein des 3 aires parisiennes à une échelle plus fine: l'observation sur le terrain des commerces distribuant des produits biologiques a alors conduit à noter de façon étonnante que des points de vente de ce type étaient déjà implantés à moins de 150 mètres des préconisations calculées par le modèle dans les aires 10 et 1 des 15<sup>ème</sup> et 17<sup>ème</sup> arrondissements de Paris. Il est à mentionner que la base de données d'adresses utilisée a été constituée il y a 4 ans, c'est-à-dire antérieurement à la création de ces points de vente « bio » localisés d'après nos renseignements de manière empirique, souvent par conversion de supermarchés traditionnels déjà existants. Ainsi, les résultats de la démarche valident des décisions de localisation prises intuitivement par des gérants connaissant bien ces quartiers parisiens. Il reste que la qualité d'expertise de ces décideurs reste très locale alors que la présente méthode peut s'appliquer sans connaissance préalable du marché et à différentes échelles géographiques L'aire 18

correspondant au quartier Est de Boulogne-Billancourt laisse de plus entrevoir une possibilité d'implantation non encore exploitée à ce jour.

#### APPORTS, LIMITES ET PERSPECTIVES DE RECHERCHE

La méthode par le traitement du signal offre un bon compromis en termes de rapidité et de précision quant aux résultats obtenus par rapport au fait de considérer chaque client comme un nœud ou à l'autre extrême, de simplifier le modèle en associant à chaque nœud une division administrative du type arrondissement ou une commune. Le tableau 1 résume les avantages de la méthode dans le cas étudié et pour cinq localisations optimales recherchées, par rapport à d'autres manières d'appréhender le modèle p-médian et de le résoudre. Comme on l'a un peu évoqué, le traitement du signal permet d'agrèger les clients en aires de chalandise. La clientèle ayant une certaine mobilité, au moins pour rejoindre les points de vente, ces aires caractérisées par leurs centres de gravité sont mieux représentatives de la demande que le fait de considérer les clients comme fixes géographiquement.

En revanche, une des principales limites de la présente approche coïncide d'une manière générale avec celle du modèle p-médian, à savoir que les localisations optimales sont déterminées essentiellement sur le critère de proximité entre les clients et les points de vente à placer. Ainsi, une orientation future de cette recherche sera d'illustrer la possibilité d'associer le traitement du signal à des modèles plus sophistiqués comme le modèle interactif de concurrence (Nakanishi et Cooper 1974) susceptible de prendre en compte d'autres variables jouant sur la fréquentation comme la surface de vente, le service de paiement par carte bancaire, le nombre d'allées du magasin, le nombre de caisses, l'emplacement à une intersection ou même des facteurs subjectifs comme l'image du magasin dans le modèle interactif de concurrence subjectif (Cliquet 1995).

Les centres de gravité de chaque aire ont également été supposés appartenir à la zone de chalandise constituaient des localisations potentielles pour la création d'un point de vente. Dans les faits, il aurait fallu indiquer précisément au modèle p-médian les localisations potentielles intéressantes et disponibles en effectuant un repérage sur le terrain au sein de chacune de ces aires mises en évidence par le processus de filtrage.

Une autre limite relative à l'exemple pris pour illustrer l'algorithme est que les clients sont tous supposés avoir la même probabilité de fréquenter le magasin. Ces clients étaient donc représentés sur la carte sous forme de points noirs (représentation binaire). Or, la fidélité des clients et leur inclination à consommer des produits « bio » sont généralement très variables. On pourra, dans une recherche ultérieure, associer à chaque client un poids et une couleur évaluant leur niveau de participation au chiffre d'affaires des futurs magasins. Des filtres comme le filtre médian, pouvant traiter des images de points en niveaux de couleurs (en niveau de gris par exemple) seront alors appliqués pour bâtir un modèle de localisation intégrant le poids de chaque client.

## CONCLUSION

La méthode de localisation associant modèle p-médian et traitement du signal comportant un processus intégré de géocodage des clients, de délimitation des zones de chalandise et de résolution d'un modèle de localisation-allocation simplifié permet de se rendre très rapidement compte du double emploi ou non de certains points de vente ou bien de l'opportunité de créer un magasin dans certaines zones lacunaires délaissées par la concurrence. On l'a vu, alors que la méthode classique du p-médian a ici ignoré la répartition du 17<sup>ème</sup> arrondissement en deux pôles de clientèle en ne détectant qu'une opportunité d'implantation, l'algorithme a au contraire bien respecté les frontières naturelles des aires de chalandise en identifiant deux concentrations de clients potentiels et deux opportunités d'implantations dans cet arrondissement. Le modèle p-médian utilisé selon un schéma

classique a en effet le fort inconvénient de nécessiter, par souci de simplification, le jugement subjectif du manager qui doit lui-même découper les secteurs en fonction de sa logique personnelle pour créer un réseau de nœuds.

D'un autre côté, la mise en œuvre du modèle p-médian associé au traitement du signal anticipe la gestion quotidienne du point de vente : la détermination précise (au niveau de la rue) des frontières de la zone de chalandise est susceptible de faciliter considérablement la tâche du directeur commercial dans le ciblage géographique des campagnes de promotion, que ces dernières prennent la forme de prospectus distribués dans les boîtes aux lettres ou de panneaux publicitaires à placer aux endroits stratégiques. L'un des objectifs de la recherche future est de combiner le traitement du signal à d'autres modèles de localisation-allocation que le p-médian, tel que le modèle interactif de concurrence non limité au critère de proximité, ou encore aux modèles p-centré et de couverture maximale, de manière à étudier la validité d'une telle méthode pour des problématiques de logistique (optimisation de centres de livraison ou d'un réseau de livraison) ou de création de services de proximité (hôpitaux, services d'urgence, casernes de sapeur-pompiers). La présente méthode pourra également être adaptée dans les traitements statistiques pour effectuer des analyses typologiques. Le regroupement en classes ou amas aussi homogènes que possible d'individus associés à des caractéristiques s'apparente en effet à la délimitation de clients en aires disjointes d'une zone de chalandise.

## REFERENCES

- APPLEBAUM W. (1968) The Analog Method for Estimating Potential Store Sales, *Guide to Store Location Research*, Addison-Wesley, Reading, Mass. by Kraus Reprint Ltd Nendeln Lichtenstein in 1967.
- CHRISTALLER W. (1935) *Die Zentralen Orte in Süddeutschland*, Ed. G.Fischer, Germany, Jena.
- CLARKE I., HORITA M. and MACKANESS W. (2002) Intuition et Evaluation des Sites Commerciaux : Appréhender la Connaissance des Commerçants, *Stratégies de localisation des Entreprises Commerciales et Industrielles : De Nouvelles Perspectives*, Gérard Cliquet et Jean-Michel Josselin éditeurs, De Boeck Université, Bruxelles,107.
- CLIQUET G. (1995), Implementing a subjective MCI model: An application to the furniture market, *European Journal of Operational Research*, 84, 279-91
- DASKIN M.S. (1995) *Network and Discrete Location: Models, Algorithms, and Applications*, Ed. John Wiley and Sons, New York.
- DIRICHLET G. L. (1850), Über die Reduktion der Positiven Quadratischen Formen mit Drei Unbestimmten Ganzen Zahlen, *Journal für die Reine und Angewandte Mathematik* 40, 216.
- ERKUT E. et BOZKAYA B. (1999) *Analysis of Aggregation Errors for the p-Median Problem*, Computers & Operations Research, 26, 1075-1096.
- GOLDEN B, BODIN L., DOYLE T. et STEWART Jr. (1980) Approximate Travelling Salesman Algorithms, *Operations Research*, 28, 694-711.
- GOODSCHILD M.F. et NORONHA V. (1983), Location-Allocation for Small Computers, University of Iowa, Monograph N°8.
- HERNANDEZ J.A. (1998) The Role of Geographical Information Systems Within Retail Location Decision Making, PhD Thesis, The Manchester Metropolitan University, Manchester.
- HODGSON et SHMULEVITZ F. (1996) Aggregation Error in Location-Allocation Modeling: The Case of Edmonton, papier présenté à la Conférence INFORMS Atlanta.
- KARIV O et HAKIMI S.L. (1979) An Algorithmic Approach to Network Location Problems, Part 2: The p-median", *SIAM Journal of Applied Mathematics* 37, 539-560.
- MCKAY D.B. (1973) Spatial Measurement of Retail Store Demand, *Journal of Marketing Research*, 10, 4, 447-453.
- MURRAY A.T. et GOTTSEGEN J.M. (1997) The influence of data aggregation on the stability of p-median location model solutions, *Geographical Analysis* 29, 200-213.
- NAKANISHI M. et COOPER L.G. (1974) Parameter Estimates for Multiplicative Competitive Interaction Models: Least Square Approach, *Journal of Marketing Research* 11, 303-311.
- NAGAO M., (1979) Edge Preserving Smoothing, *CGIP*, 9, 94-407.
- PETERSON R.A. (1974) Trade Area Analysis Using Trend Surface Mapping, *Journal of Marketing Research*, XI, 338-42.
- PLASTRIA F. (2001) On the Choice of Aggregation Points for Continuous p-Median Problems : A Case for the Gravity Centre, *TOP*, Sociedad de Estadística e Investigación Operativa Madrid, 9, 2, 217-242.
- REILLY W. J. (1931) *The Law of Retail Gravitation*, Ed. W. Reilly, New York.
- REVELLE C. et SWAIN R. (1970) Central Facilities Location, *Geographical Analysis*, 2, 30-40.
- ROGER P. (1983) Description du Comportement Spatial du Consommateur, Thèse de Doctorat, Université de Lille I, Lille.

- SYLVANDER B. (1999) Les Tendances de Consommation de Produits Biologiques en France et en Europe : Conséquences sur les Perspectives d'Evolution du Secteur, papier présenté au colloque l'Agriculture Biologique Face à Son Développement, Lyon.
- TEITZ M.B. et BART P. (1968) Heuristic Methods for Estimating Generalized Vertex Median of a Weighted Graph, *Operations Research*, 16, 955-961.
- THIESSEN A.H. et ALTER J.C. (1911), Precipitation Averages for Large Areas, *Monthly Weather Review*, 39, 1082-1084.
- YIP K. ZHAO F. (1996) Spatial Aggregation: Theory and Applications, *Journal of Artificial Intelligence Research*, 5, 1-26.

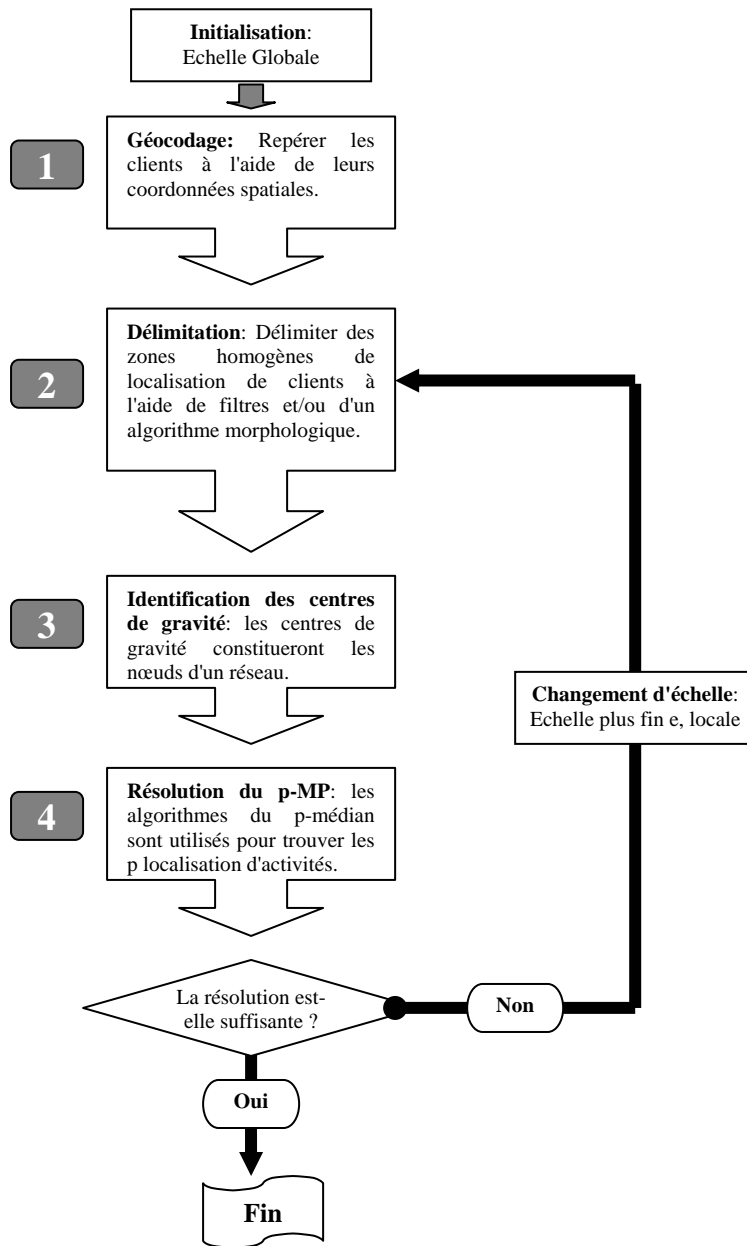


Figure 1 - Organigramme de la nouvelle méthode de recherche de localisations proposée

Figure 2 - Adresses clients associées aux fréquentations

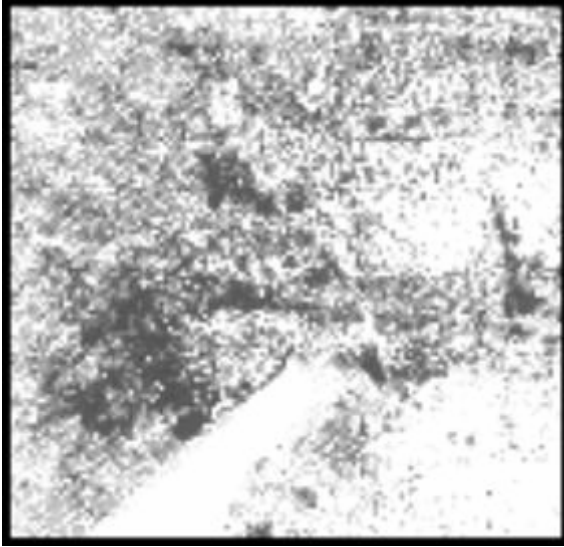


Figure 3 - La représentation traitée par filtre Sobel après 2 filtres médian (les contours sont superposés à l'image filtrée par 2 filtres Nagao)

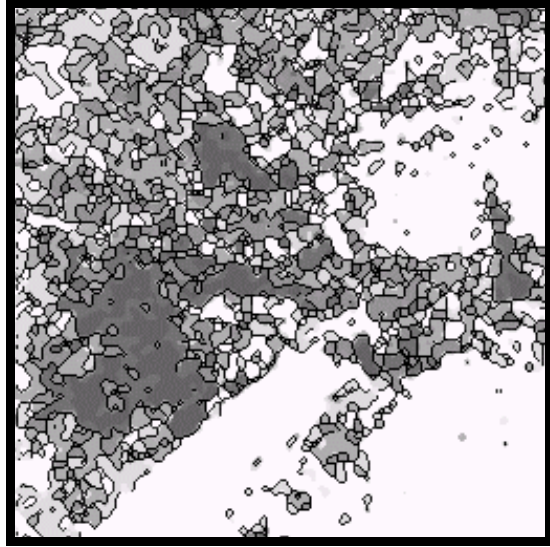


Figure 4 - Exemple de parcours de frontière de zone de chalandise à l'aide de l'algorithme

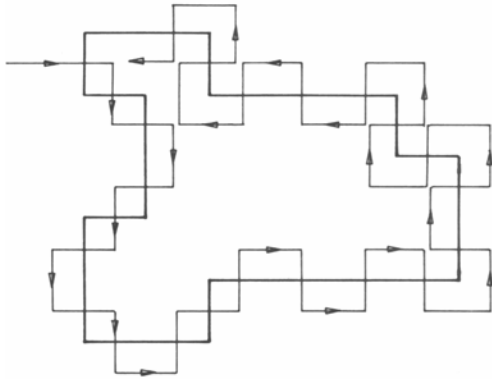


Figure 5 - Le p-médian modélisé après délimitation de la zone de chalandise et calcul des centres de gravité.

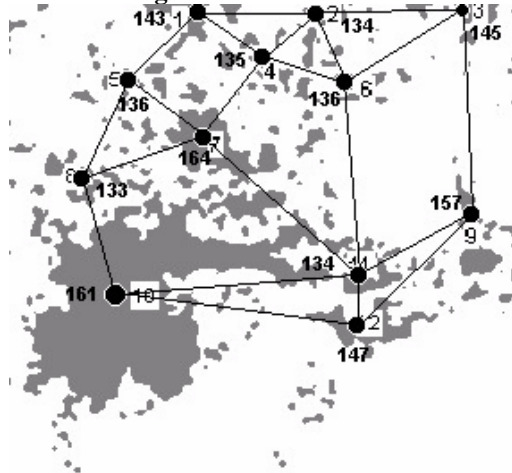


Figure 6 - Géocodage des 10 211 clients potentiels de l'ouest parisien



Figure 7 - Filtrage par dilatation de points-client



Figure 8 - Numérotation et repérage du centre de gravité des aires constitutives de la zone de chalandise



**Tableau 1 – Performances des méthodes de résolution des différentes approches du modèle p-médian**

<b>Méthode</b>	<b>Traitement du Signal et P-médian</b>	<b>P-médian direct</b> avec résolution par examen exhaustif des solutions	<b>P-médian</b> avec résolution par heuristique	<b>P-médian simplifié</b> (1 nœud par secteur géographique)
<b>Taille du réseau à résoudre</b>	25 nœuds dans réseau principal puis sous-réseaux	10211 nœuds (un par client)	10211 nœuds (un par client)	7 nœuds (un par arrondissement ou commune)
<b>Temps de résolution</b>	4 minutes	impossible avec les moyens actuels	plusieurs heures	Quelques minutes
<b>Nombre de combinaisons possibles à examiner en théorie (pour 5 localisations à rechercher)</b>	53130	$83 \times 10^{16}$	$100 \times 10^6$ (algorithme de voisinage)	21
<b>Exactitude de la recherche</b>	certaines sites détectés correspondent à l'emplacement effectif de magasins	résolution exacte ardue	exactitude inconnue	1 localisation incohérente sur 4 localisations obtenues
<b>Précision des localisations</b>	au numéro de rue	à l'adresse client	à l'adresse client	à l'arrondissement ou à la commune