

CIGI 2011

Analyses spatiales d'un réseau de distribution de points de vente : application à une entreprise canadienne de meubles distribués aux États-Unis

CECILE LOCART, BRUNO AGARD, NICOLAS SAUNIER

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

C.P. 6079, SUCC. CENTRE-VILLE, MONTREAL (QUEBEC), H3C 3A7, CANADA
cecile.locart@polymtl.ca; bruno.agard@polymtl.ca; nicolas.saunier@polymtl.ca

Résumé

Dans le contexte actuel de globalisation, la mise en place et le maintien d'un réseau de distribution performant sont essentiels pour les entreprises. Plusieurs critères permettent de juger de la qualité d'un réseau. Parmi ceux-ci, l'adéquation entre la localisation des points de vente et la position de la clientèle, à savoir, la couverture du marché, est prépondérante. Des techniques de data mining permettent de segmenter un marché en se basant sur les caractéristiques socioéconomiques des clients dans différentes zones géographiques. Dans cet article, nous proposons de mettre en place des mesures d'analyse spatiale et d'utiliser un système d'information géographique (SIG) pour qualifier l'intensité de recouvrement d'un réseau de distribution d'une entreprise. Dans le cadre de notre étude, nous étudions le réseau de distribution d'une entreprise de meubles haut de gamme avec des points de vente aux États-Unis. Il s'agit, *in fine*, de comparer, zone géographique par zone géographique, les résultats de la segmentation de la clientèle reflétant le potentiel commercial avec les indicateurs de recouvrement. Ce modèle intégrant les deux séries de résultats devient un outil d'aide à la décision pour l'entreprise dans le déploiement de son réseau.

Abstract

In the context of globalization, firms have to develop and maintain an efficient distribution channel in order to be competitive. Several criteria can be used to assess the quality of current networks. Among them is the market coverage of the network: customer locations should match store locations. Data mining tools allow us to segment a market by using socio-economic attributes of customers in each geographical area. In this paper, we develop spatial analysis measures and we use a geographical information system (GIS) to quantify the covering intensity of a firm's network. Herein, the channel distribution of an upscale furniture firm with stores in the United States of America is analyzed. Finally, segmentation and covering indicators results are compared geographical area by geographical area in order to make a decision support tool and allow us to help the firm to refine its network.

Mots clés – segmentation du marché, localisation de points de vente, géomarketing, data mining, analyse spatiale.

Keywords – segmentation, store localization, geomarketing, data mining, spatial analysis

1 INTRODUCTION

Selon Tellier (1985), «l'espace constitue une variable essentielle de l'univers économique, car, tant à l'échelle locale, régionale, nationale que mondiale, la réalité économique ne saurait être comprise profondément sans qu'une logique soit fournie qui explique pourquoi les activités de production et de consommation se trouvent à tel endroit plutôt qu'à tel autre».

Le géomarketing est issue d'une prise de conscience de la dimension spatiale des activités économiques qui repose sur des données marketing, sociologiques, démographiques, économiques, topographiques et comportementales distribuées dans l'espace (Latour & LeFloc'h, 2001). Selon (Douard & Heitz, 2004), les problématiques spatiales dans les décisions marketing peuvent être séparées en trois grands domaines :

- l'étude du comportement spatial du consommateur : connaître et modéliser les flux du consommateur pour l'achat de différents biens et services en fonction de son profil et de son environnement commercial,
- la mise en œuvre du management du marketing : adapter l'offre, la publicité et les prix en fonction de l'espace.

- et la localisation commerciale : établir un maillage des points de vente, d'agences et d'entrepôts en fonction de l'offre en présence, de la position des concurrents et des potentialités de marché.

De plus en plus de secteurs d'activité ont recours à des outils de géomarketing pour traiter les grandes quantités d'informations localisées qu'elles possèdent. Les applications opérationnelles sont multiples.

(Lendrevie et al., 2006) affirment que la mise en place d'un circuit de distribution performant est une opération longue et coûteuse. Aussi, l'anticipation est d'autant plus essentielle qu'une localisation défavorable est souvent une cause irrémédiable d'échec et qu'il n'est généralement pas possible de corriger, par d'autres mesures, l'erreur commise en choisissant une mauvaise localisation (Beaujeu-Garnier & Delobez, 1977). D'après (Lendrevie et al., 2006), il est possible d'évaluer un circuit de distribution selon plusieurs critères dont la couverture.

La présente étude concerne la localisation commerciale et plus particulièrement l'audit d'un réseau de distribution existant.

Il s'agit ici d'analyser la performance d'un réseau de distribution en étudiant sa couverture. Le circuit analysé permet-il de toucher une part importante de la clientèle visée ? Les acheteurs réguliers et importants sont-ils dans la clientèle du circuit ? Pour étudier la couverture du circuit de distribution, nous allons conjuguer des mesures d'analyse spatiale avec une segmentation de marché.

(Berry & Linoff, 1997) ont souligné que le résultat de la détection automatique de groupes doit être intégré à d'autres méthodes pour en dégager la signification. La plupart des applications de la segmentation du marché s'arrêtent à la constitution des groupes homogènes. Cependant, il est difficile de prendre une quelconque décision à partir de ce résultat partiel.

Dans cet article, l'objectif est de développer un modèle intégrant le résultat d'une segmentation de marché avec des indicateurs de couverture des points de vente afin d'aider une entreprise de meubles haut de gamme à identifier les zones géographiques où elle devrait revoir l'intensité de sa présence commerciale.

Dans une première partie, les principales approches de localisation commerciale sont abordées. Les différents indicateurs de couverture développés sont présentés dans la section méthodologie. Enfin, la dernière partie présente le contexte du marché des meubles aux États-Unis et les résultats de l'intégration de ces indicateurs avec la segmentation de la clientèle de l'entreprise.

2 REVUE DE LITTÉRATURE

2.1 *Revue des méthodes de géomarketing pour l'analyse d'un réseau de distribution*

La plupart des méthodes actuelles jugent de l'intérêt d'une implantation commerciale à des places prédéfinies en utilisant les modèles de localisation commerciale. Selon Cliquet (2001), les modèles de localisation commerciale sont pour beaucoup basés sur le concept d'attraction commerciale. Une analogie est faite avec l'attraction universelle : l'offre et non la masse est à l'origine de l'attraction d'un point de vente. On mesure l'attraction commerciale à travers l'étendue de sa zone de chalandise. Plus on s'éloigne du point de vente, plus l'attraction est faible. Les modèles classiques postulent que le comportement spatial du consommateur, centré sur son lieu de résidence, est fonction de la distance à parcourir. Pour se procurer les biens et les services qu'il souhaite, il minimise la distance de déplacement. D'après Cliquet (2001), la topographie de l'environnement et la compétition possèdent une influence sur l'attraction d'un point de vente.

Les études préalables aux implantations commerciales concernent, premièrement la délimitation de la zone de chalandise puis deuxièmement l'évaluation du potentiel économique de la population pour le point de vente à l'intérieur de cette zone (Douard & Heitz, 2004).

Classiquement, il est considéré que la zone de chalandise est composée de plusieurs sous zones. Pour chaque zone, des taux d'emprise de la clientèle sont évalués. Parallèlement pour chaque zone, il est possible d'évaluer le nombre de ménages, de connaître leurs dépenses annuelles pour chaque type de produit vendu par le point de vente. Puis en tenant compte du profil sociodémographique des zones, il est possible de déterminer un marché théorique. C'est la pondération des marchés théoriques par les taux d'emprise de chaque zone qui permet d'évaluer le marché potentiel total du point de vente (Cliquet & Josselin, 2002).

De nombreuses critiques ont été émises sur ces méthodes. Les consommateurs sont de plus en plus mobiles, de moins en moins fidèles à un point de vente, et ont donc un comportement plus complexe. Le consommateur ne se déplace pas uniquement pour consommer, mais aussi pour travailler, pour ses loisirs... Il lui arrive souvent de faire des boucles autour de son domicile. La notion de zone de chalandise autour du lieu de résidence est donc remise en question. D'autre part, les décisions d'implantation sont des décisions multicritères et n'ont pas la maximisation de la rentabilité pour seul objectif. La couverture d'un territoire se fait aussi en fonction de la position de la concurrence. Une entreprise peut, par exemple, choisir de ne pas laisser un concurrent s'enraciner dans une zone.

Dans notre étude, on souhaite identifier les zones où il serait avantageux pour une entreprise de se développer et/ou de réduire son offre, étant donné un réseau de distribution déjà existant et très étendu. (Douard & Heitz, 2004) présentent une approche de géomarketing pour la recherche de sites d'implantation pour une chaîne de magasins de jardinage et de bricolage qui possède déjà plusieurs points de vente dans l'est de la France. Cette approche part du principe qu'une zone d'implantation sera intéressante à exploiter si sa morphologie (structure démographique, économique, sociale et urbaine) ressemble aux zones d'implantation actuelles qui réalisent les meilleures performances. Cette étude est basée sur une série de critères de 26 variables. Selon les auteurs, elle a permis d'identifier 9 zones potentiellement avantageuses pour l'entreprise sur les 60 zones de l'étude. Les zones sélectionnées sont ici des zones qui ne possèdent pas de points de vente. Davantage de sites potentiels sont envisagés par rapport aux modèles de localisation précédents mais aucune couverture du réseau n'est calculée dans cette méthode.

Dans le cadre de cet article, nous allons utiliser des techniques d'analyse spatiale pour qualifier la couverture du réseau de distribution donné. L'analyse spatiale est définie, par Brunet et al. (1993), comme l'ensemble des méthodes mathématiques et statistiques visant à préciser la nature, la qualité attachée aux lieux et aux relations qu'ils entretiennent - l'ensemble constituant l'espace -, en étudiant simultanément attributs et localisations. Selon Latour et Le Floch (2001) « la démarche de l'analyse spatiale consiste à aider à la compréhension et à la maîtrise de l'étude des phénomènes qui s'exercent sur les dimensions de l'espace ».

3 METHODOLOGIE

3.1 *Contexte et problématique*

À partir d'un cas d'étude développé en collaboration avec une entreprise qui conçoit et fabrique des meubles haut de gamme, commercialisés en majorité sur le territoire des États-Unis, nous montrons comment l'analyse spatiale et la segmentation peuvent être utilisées pour réaliser l'audit d'un réseau de distribution. Avec plus de 300 millions d'habitants, un territoire vaste et des cultures diverses, le marché des États-Unis est un défi pour l'analyse commerciale. Dans le cas du marché des meubles aux États-Unis, la plupart des entreprises de fabrication de meubles n'ont pas leurs propres magasins. Elles doivent, par conséquent, recourir à un canal de distribution passant par diverses chaînes de magasins.

3.2 *Données d'entrée*

Dans le cadre de notre étude, nous possédons des données d'entrée internes et externes à l'entreprise. Il s'agit :

- de l'étendue du marché de l'entreprise qui est un territoire fixé pour l'étude (États-Unis d'Amérique) et de la position des points de vente actuels de l'entreprise.
- de données sociodémographiques détaillées sur l'ensemble du territoire concerné grâce aux données du recensement américain (Census Bureau, 2010)

3.3 Différentes étapes de l'étude

3.3.1 Étape 1 : Segmentation de la population :

Les caractéristiques économiques et sociodémographiques sont des contraintes importantes sur un marché (Filion & Colbert, 1990). Plus précisément, Lihra (2009) a montré que les ventes de meubles sont liées aux aspects sociodémographiques. D'autre part, le marché des meubles en général est fortement corrélé avec les critères de logement et de l'économie (Volpe & Peluso, 2007).

Nous effectuons donc une segmentation de la clientèle cible, à partir des données sociodémographiques issues des bases de recensements (revenus, âge, ménages, caractéristiques des habitations, éducation, emploi) afin d'établir les potentialités de marché sur chaque zone. Nous appelons « critère » les données sociodémographiques utilisées pour la segmentation. A chaque zone est attribuée une valeur de « segment ». Une hypothèse importante de l'étude est de considérer que l'entreprise a une bonne connaissance du profil de son client et peut lister avec des experts du domaine, les critères sociodémographiques permettant de décrire son consommateur cible. Ceci a été validé dans une étude précédente (Le et al., 2009). Ici, nous effectuons une segmentation plus fine au niveau des comtés.

3.3.2 Étape 2 : Calcul d'indicateurs de recouvrement

Le but est de rendre compte de l'intensité de la couverture du réseau de distribution pour chaque zone du territoire en mesurant une quantité de points de vente et non une dispersion de ceux-ci à l'intérieur d'une zone. On appelle « indicateur » l'intensité de recouvrement des points de vente alors calculé pour chaque zone. Comme plusieurs méthodes sont mises en place, plusieurs séries d'indicateurs sont mesurées.

Pour la formulation mathématique des indicateurs, nous appelons Z l'ensemble des zones (comtés) et M l'ensemble des magasins. Différents indicateurs ont été envisagés et il a été choisi de les répartir en 2 types :

- Les indicateurs de « densité » qui sont le rapport direct entre une mesure absolue de la couverture du réseau et un attribut propre à la zone étudiée. Une zone qui possède un indicateur de « densité » élevé sera considérée comme une zone avec une couverture importante.
 - L'indicateur de « distance au plus proche magasin ». Contrairement aux indicateurs de densité, une zone qui possède un indicateur de distance élevé sera considérée comme ayant une couverture faible.
- D'autre part, aussi bien pour la mesure absolue de la couverture du réseau par les indicateurs de densité que par l'indicateur de distance, deux variantes peuvent être envisagées :

- Méthode directe : calculer directement la valeur de mesure pour la zone de découpage du territoire donné,
- Méthode par grille : établir une grille de découpage fine afin d'avoir la résolution désirée.

La mise au point d'une méthode d'analyse spatiale rend essentielle la considération de l'espace autrement qu'avec le découpage administratif. La division de l'espace en mailles carrées est alors très fréquente (Latour & LeFloc'h, 2001). Nous appelons alors G l'ensemble des mailles de la grille.

L'analyse des résultats à partir d'une grille nécessite alors le calcul pour chaque maille d'une valeur de segment. Cependant, une maille peut recouvrir plusieurs zones. Dans ce cas, pour déterminer une valeur de segment pour chaque maille, nous évaluons dans un premier temps les valeurs des critères à la base de la segmentation pour chaque maille en utilisant une moyenne pondérée par les surfaces. A titre d'exemple, on montre le calcul du critère « salaire moyen par famille » pour une maille recouvrant 3 zones. (Tableau 1).

Tableau 1. Exemple de calcul d'un critère pour une maille

| Zone | Segment de la zone | Salaire moyen par famille | % de la surface du comté au sein de la maille | Salaire moyen par famille Pondéré |
|---|--------------------|---------------------------|---|---|
| j | | $critère_{jl}$ | $\frac{Surface_{inter_{jk}}}{Surface_k}$ | $\frac{Surface_{inter_{jk}}}{Surface_k} * critère_{jl}$ |
| 1 | 2 | \$ 66 378 | 52% | \$ 34 830 |
| 2 | 1 | \$ 108 268 | 11% | \$ 11 508 |
| 3 | 2 | \$ 76 839 | 37% | \$ 28 353 |
| Salaire moyen par famille : $critère_{maille_{kl}}$ | | | | \$ 74 690 |

Si on appelle :

- C l'ensemble des critères pour la segmentation,
- $\forall l \in C, \forall j \in Z, critère_{jl}$ la valeur du critère l pour la zone j
- $\forall l \in C, \forall k \in G, critère_{maille_{kl}}$ la valeur du critère l pour la maille k
- $\forall j \in Z, \forall k \in G, Surface_{inter_{jk}}$ la surface d'intersection entre la zone j et la maille k de la grille
- $\forall k \in G, Surface_k$ la surface la maille k

Alors, $\forall l \in C, \forall k \in G,$

$$critère_{maille_{kl}} = \sum_{j \in Z} \frac{Surface_{inter_{jk}}}{Surface_k} * critère_{jl}$$

Dans un second temps, nous déterminons à partir des valeurs des critères calculées pour la maille le segment dont la maille se rapproche le plus.

On parlera alors d'« unités géographiques » pour désigner soit les zones issues des données d'entrée soit les mailles de la grille. U désignera alors l'ensemble des unités géographiques. La méthode par grille peut néanmoins servir d'intermédiaire. Il est possible de tirer profit de la résolution plus fine des indicateurs tout en effectuant une analyse des résultats par zone et ainsi ne pas avoir à calculer de valeur de segment pour chaque maille. En effet, on peut à partir des valeurs d'indicateurs obtenues pour chaque maille de la grille, déterminer la valeur de mesure pour la zone de référence par la moyenne des valeurs sur les mailles qui la composent (pondérée par les surfaces).

➤ Les indicateurs de densité

L'intensité de la couverture du réseau peut être évaluée à partir de plusieurs mesures. Pour une unité géographique donnée, il est possible de :

- Compter le nombre de magasins imputés à l'unité géographique en suivant deux règles d'imputation différentes, à savoir imputer un magasin à l'unité géographique donnée :

1. Si le magasin appartient physiquement à cette zone

$\forall i \in M, \forall j \in U X_{ij}$ binaire

$$X_{ij} = \begin{cases} 1, & \text{si le magasin } i \text{ appartient} \\ & \text{à l'unité géographique } j \\ 0, & \text{sinon} \end{cases}$$

2. Si le centroïde de l'unité géographique est le plus proche centroïde du magasin. Pour chaque point de vente, il faut déterminer le centroïde le plus proche du point de vente.

Dans le cas de mailles carrées, les résultats sont identiques à ceux avec l'imputation par appartenance.

$\forall i \in M, \forall j \in U, X_{ij}$ binaire,

$X_{ij} = 1$ si $j = \operatorname{argmin}_{k \in U} [\text{distance}(i, \text{centroïde}(k))]$,
0 sinon

Alors pour chaque unité géographique $j \in U$, pour les deux règles d'imputation, le nombre de magasin est :

$$\text{Imp}_j = \sum_{i \in M} X_{ij}$$

• Établir une mesure de la densité spatiale des magasins à l'aide des méthodes d'estimation par noyau. Cette méthode part du principe qu'un point de vente possède une certaine zone d'influence indépendamment des frontières de l'unité géographique à laquelle il appartient. Une zone donnée subit donc l'influence d'un ou plusieurs magasins et ces influences sont sommées pour obtenir une influence totale pour l'unité géographique. Un choix doit être fait pour la forme du profil d'influence choisi, appelé méthode d'interpolation et pour la largeur de la fenêtre d'influence. Il est possible d'envisager une méthode d'interpolation triangulaire, uniforme, normale, exponentielle négative... Les différences entre les méthodes d'interpolation peuvent être faibles (Levine, 2004). Choisir une largeur de fenêtre importante reviendra à considérer que le point de vente possède une attraction forte en distance et que le client est prêt à faire un déplacement important pour acquérir les produits vendus par l'enseigne. Le choix de ces paramètres dépend directement du type de produits étudié. Dans notre étude nous considérerons que seule la « présence » du magasin sera répartie sur une certaine surface.

Enfin, une fois calculée la mesure absolue rendant compte de l'intensité de la couverture du réseau sur une zone, il convient de relativiser cette mesure en la rapportant à un attribut propre à la zone (reflétant le potentiel de consommation de cette zone). En effet, deux zones peuvent posséder un même score absolu, par exemple un même nombre de magasins imputés mais une clientèle potentielle très différente. Plusieurs attributs propres à la zone peuvent être envisagés :

- La population résidentielle
- La population ambiante : la population parcourant la zone dans la journée pour différents motifs (loisir, travail, résidence) et susceptible d'y faire un achat. Cette « population à risque » est utilisé dans les études de criminologie (Andresen, 2006).
- Une certaine population ciblée par l'entreprise

➤ L'indicateur de distance au plus proche magasin

Pour chaque unité géographique, on calcule les coordonnées du centroïde. Il s'agit ensuite de déterminer quel est le magasin le plus proche de ce centroïde et de calculer la distance entre ces deux points.

Pour chaque unité géographique $j \in U$, l'indicateur s'écrit :

$$d_j = \min_{i \in M} [\text{distance}(i, \text{centroïde}(j))]$$

On calcule ici la distance euclidienne à partir des latitudes et longitudes des deux points. Elle ne tient pas compte du réseau de transport et d'éventuelles barrières naturelles.

Zones non-connexes

Enfin, il arrive que la forme des comtés soit très irrégulière voire que les comtés ne soient pas connexes. La notion de centroïde utilisée dans le cadre de la méthode directe pour l'indicateur densité avec imputation par centroïde et l'indicateur distance est alors hasardeuse. Dans le cas de comtés non connexes, on a choisi de prendre en compte le centroïde de chaque sous élément. Pour l'indicateur densité, on somme, pour tout le comté, les nombres de magasins imputés à chaque centroïde des sous éléments et on obtient un nombre total qui devient valable pour tout le comté. Pour l'indicateur

distance, on mesure la distance au magasin le plus proche de chaque centroïde des sous éléments et on effectue une moyenne pondérée de ces distances en prenant en compte la surface des sous-éléments du comté.

3.3.3 Étape 3 : Analyses et comparaisons

C'est la comparaison, unité géographique par unité géographique, de l'indicateur des points de vente et des résultats de la segmentation qui va permettre ensuite d'énoncer des recommandations sur la pertinence des localisations.

4 EXPERIMENTATION / CAS D'ETUDE

4.1 Contexte du marché du meuble canadien aux US

Le cas d'étude porte sur une entreprise canadienne de meubles haut de gamme distribués aux États-Unis. Les États-Unis sont le plus grand marché des meubles au monde en plus d'être le plus concurrentiel avec plus de 120 pays exportateurs (Florence, 1990). Les États-Unis sont le premier pays de destination des meubles canadiens avec 93% de la valeur des exportations canadiennes mondiales en 2010. Cependant, les exportations canadiennes de meubles chutent. Elles ont diminué de 14,5 % en 2008, passant de 5,2 milliards de dollars en 2007 à 4,4 milliards de dollars (Industrie Canada, 2011). Une des grandes préoccupations de l'industrie du meuble est l'augmentation des importations, aussi bien au Canada qu'aux États-Unis, de meubles en provenance de pays où les coûts de production sont plus faibles. En 2006 notamment, la Chine a surpassé le Canada en tant que premier exportateur de meubles vers les États-Unis (50%); le Canada prend alors la deuxième place avec une part de 16% (Murillo, 2007). Il est donc primordial pour les entreprises de meuble du Canada de mettre en place des stratégies performantes pour regagner des parts de marché aux États-Unis.

4.2 Données

4.2.1 Données géographiques

Une précédente étude portant sur le croisement de l'historique de ventes réelles de l'entreprise avec une segmentation de marché basée sur les caractéristiques socioéconomiques des clients, à l'échelle des états américains, avait prouvé que l'application des méthodes de data mining s'applique bien au marché des meubles (Le et al., 2009). La présente étude avec un découpage plus fin du territoire permettra d'établir des recommandations plus localisées et précises. Ainsi, nous avons choisi un découpage du territoire américain (hors Alaska et Puerto Rico) selon les comtés (ou équivalents : les paroisses en Louisiane et les 42 villes indépendantes). Le marché est donc analysé selon un fractionnement de l'espace en 3109 zones.

4.2.2 Données pour la segmentation de marché

Dans le cas de notre étude, avec l'aide de l'entreprise concernée, 34 critères ont été choisis et calculés, pour chaque comté, à partir des données fournies par le Census Bureau américain. L'entreprise sait que le client type de ses produits est globalement un client aisé, plutôt âgé et propriétaire d'une maison relativement grande. Aussi, on a utilisé des données comme la taille de la population, son évolution démographique, le pourcentage de personnes dans certaines tranches d'âges, des caractéristiques de logement, de situation familiale, de revenus, et d'emplois.

4.2.3 Position des points de vente et géocodage

L'entreprise a fourni les adresses de ses 13051 différents points de vente : adresse, code postal, état. Chaque adresse a ensuite été convertie en vecteur (latitude, longitude) avec les services de géocodage de Google Maps. A partir de ces

données, il est alors possible, grâce à un système d'information géographique, de visualiser sur une carte le positionnement du réseau de distribution de l'entreprise (Figure 1).



Figure 1. Position des points de vente

4.3 Segmentation de la population étudiée

La détermination de la segmentation a été réalisée par étape avec une segmentation grossière à l'aide de la méthode des k-moyennes. Le résultat de cette dernière a été utilisé pour construire une classification hiérarchique qui montre qu'un partage de la population en 4 groupes est approprié. Grâce à la méthode des k-moyennes avec un nombre de segments égal à 4 permet de partitionner l'ensemble des comtés (Figure 2).

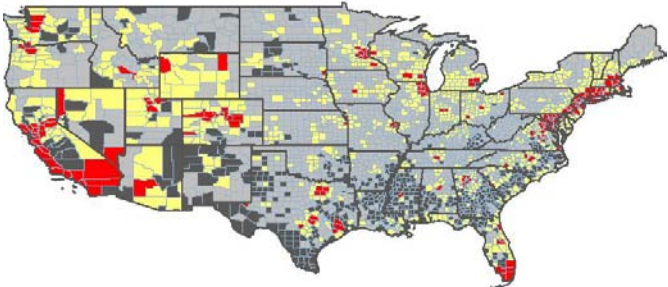


Figure 2. Segmentation des comtés en 4 segments (1 : Rouge, 2 : Jaune, 3 : gris clair, 4 gris foncé)

À partir des données sociodémographiques du Census Bureau, on peut caractériser le profil des segments (Tableau 3). La fonctionnalité du SIG de superposition de couches devrait nous permettre de visualiser en même temps le réseau de distribution et les résultats de la segmentation. Cependant, le nombre de points de vente est trop élevé et l'échelle des comtés trop petite pour faire des recommandations directement. Latour et Le Floc'h (2001) mettent en avant l'insuffisance de ces représentations. Il est essentiel de calculer des indicateurs de couverture du réseau pour chaque comté pour avoir une information claire, fiable et précise.

4.4 Calcul des indicateurs

Dans cet article nous présentons les résultats de 5 méthodes (Tableau 4). Nous avons choisi de ne présenter que les méthodes par grille intermédiaire et directe par comté.

4.4.1 Représentation des indicateurs

Lorsqu'il existe une forte amplitude dans la répartition des données de la série que l'on souhaite représenter, il est difficile d'identifier les nuances de la série de données avec des écarts fixes. N'apparaissent alors que les zones avec des valeurs extrêmes. Nous avons donc préféré utiliser une échelle à 3 classes pour la représentation par cartes des indicateurs. Pour cela, les proportions des groupes issues de la segmentation (en regroupant à l'intérieur d'une même classe les groupes 3 et 4) ont été suivies. Pour chaque indicateur, les valeurs ont été classées et réparties en 3 classes (Tableau 2):

Tableau 2. Classes pour la représentation des indicateurs

| Classe | Légende | % des valeurs | Densité | Distance |
|--------|---------|---------------|---------|----------|
| 1 | | 7% | ++ | -- |
| 2 | | 29% | + | - |
| 3 | | 64% | | |

4.4.2 Indicateurs de densité

Pour tous les indicateurs de densité, les valeurs absolues mesurées sont rapportées à la population du comté.

Tableau 3. Profil et répartition des groupes

| Segment | Caractéristiques du groupe |
|--|--|
| 1 Majorité de Très riches : 214 comtés (7%) | <ul style="list-style-type: none"> • Les salaires et la valeur de la maison sont beaucoup plus élevés que la moyenne américaine. • Les couples mariés de 35 à 60 ans et la population de 35 à 65 ans sont plus représentés que dans le reste des Etats-Unis. • La taille moyenne des familles et des foyers est élevée. |
| 2 Moyenne supérieure : 894 comtés (29%) | <ul style="list-style-type: none"> • La part de population sous le seuil de pauvreté (familles et personne) est inférieure à la moyenne. • Le salaire médian des comtés de ce groupe est 14% plus élevé que le salaire médian américain. |
| 3 Moyenne inférieure : 1445 comtés (46%) | <ul style="list-style-type: none"> • Les revenus de ce groupe sont moins élevés que la moyenne : la part de foyers dont le salaire pour le foyer annuel est supérieur à 75000 dollars est de 9% contre 13% pour la part nationale. • De même la part de familles dont le salaire pour la famille annuel est supérieur à 75000 dollars est de 12% contre 17% pour la part nationale. |
| 4 Majorité de revenus faibles : 556 comtés (18%) | <ul style="list-style-type: none"> • Les revenus y sont beaucoup plus faibles que dans les autres groupes. • La taille des ménages et des familles est élevée. La part de population sous le seuil de pauvreté (familles et personne) est supérieure à la moyenne. • Les maisons sont plus petites et le nombre de foyers avec moins de une personne par pièce est plus faible que la moyenne américaine. • Les couples mariés de 35 à 60 ans et la population de 35 à 65 ans sont moins représentés que dans le reste des Etats-Unis. |

Tableau 4. Méthodes utilisées pour le cas d'étude

| METHODES | Directe par comté | Avec une grille intermédiaire (carrés de 0.1 degrés en longitude et latitude) |
|--|--|---|
| Nombre d'unités géographiques | 3109 | 83636 |
| Densité (rapport à la population du comté) | Compte du nombre de magasins avec imputation selon l'appartenance au comté | X |
| | Compte du nombre de magasins avec imputation selon le centroïde du comté | X |
| | Estimation par noyau | X |
| Distance | Distance au magasin le plus près | X |

Imputer les magasins aux différentes unités géographiques selon l'appartenance (Figure 3 a.) possède un désavantage certain : pour chaque unité géographique, les points de vente des unités géographiques voisines ne sont pas pris en compte. En réalité, les frontières des unités géographiques sont ouvertes et un consommateur d'une unité géographique donnée a la possibilité de se rendre dans une autre unité géographique pour y effectuer ses achats.

Le type de bien acheté a une influence sur le comportement et donc les déplacements du consommateur et les distances parcourues. Nelson (1970) classe les biens en deux grandes catégories : d'une part les produits de commodité achetés à fréquence élevée, ne nécessitant pas de recherche d'informations et pour lesquels les distances parcourues sont faibles et les produits de comparaison pour lesquels le consommateur fera un travail de recherche avant d'effectuer son achat et des déplacements plus nombreux et importants.

La probabilité que le consommateur franchisse les frontières du comté est d'autant plus importante que l'achat du bien nécessite le parcours de distances élevées. Il serait donc, pour cette raison, plus réaliste d'imputer les magasins selon le centroïde (Figure 3 b.). On voit que sur les cartes (Figure 3 a., Figure 3 b.) certains comtés n'appartiennent pas à la même classe d'indicateurs (rectangles rouges).

Les méthodes d'estimation par noyau permettent aussi de considérer que le consommateur n'est pas borné par les frontières de son comté. Nous avons choisi de présenter la méthode d'interpolation triangulaire avec une largeur de fenêtre de 15 miles. En effet, une étude a montré qu'en moyenne un consommateur (belge, britannique ou français) parcourt 28 km pour se rendre jusqu'à un magasin de meubles spécialisé (Rizet, 2008). Ce choix pourrait être affiné par une étude des comportements d'achat des clients spécifiques de l'entreprise. Nous présentons le résultat intermédiaire des calculs de densité à l'aide de l'estimation par noyau par grille pour montrer comment, premièrement les résultats par grille (Figure 3 c.) sont moins agrégés qu'avec la méthode par comté (Figure 3 d.) et deuxièmement comment le rapport à la population relativise la valeur de l'indicateur.

La juxtaposition des cartes issues des différentes méthodes de calcul des indicateurs de recouvrement (Figure 3) montre bien que les résultats diffèrent. Globalement, pour toutes les méthodes, les points de vente sont davantage présents à l'est du territoire, le long de la côte. Cependant, le choix de la méthode a une incidence sur les valeurs des indicateurs :

- Des comtés, au Texas, sont identifiés comme appartenant à la deuxième classe de l'indicateur par compte du nombre de magasins selon le centroïde alors que ces mêmes comtés appartiennent aux classes 3 ou 4 de l'indicateur par estimation par noyau (rectangles verts)

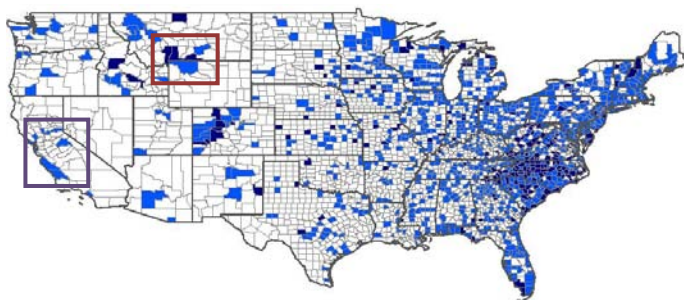
- Des comtés, en Californie, sont identifiés comme appartenant à la classe 2 de l'indicateur par compte du nombre de magasins selon l'appartenance alors que ces mêmes comtés appartiennent à la classe 3 ou 4 de l'indicateur par estimation par noyau (rectangles violets)

4.4.1 Indicateurs de distance

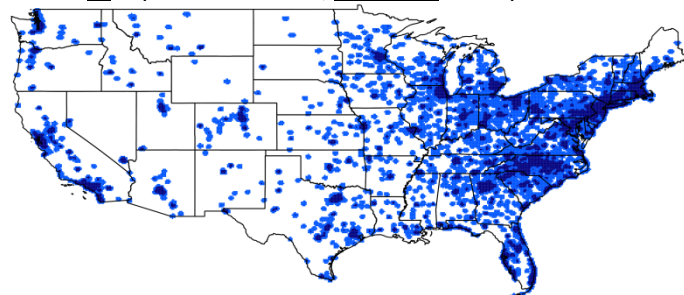
L'utilisation de l'indicateur de distance n'est pas adaptée pour deux raisons :

- quand la taille des zones est très variable (uniquement pour la méthode directe). En effet, la distance n'est pas évaluée avec la même résolution pour les zones petites et grandes. C'est une limite qui n'apparaît que dans le cas de la méthode directe par zone. Dans le cas de notre étude, la superficie des comtés est très variable. Les comtés sont plus étendus à l'ouest qu'à l'est. Au recensement de 2000, la superficie médiane est de 1 611 km², le plus grand comté est celui de San Bernardino en Californie (52 073 km²), et le plus petit est celui de New York dans l'État de New York (59,47 km²) (en dehors de l'Alaska et de Hawaï).

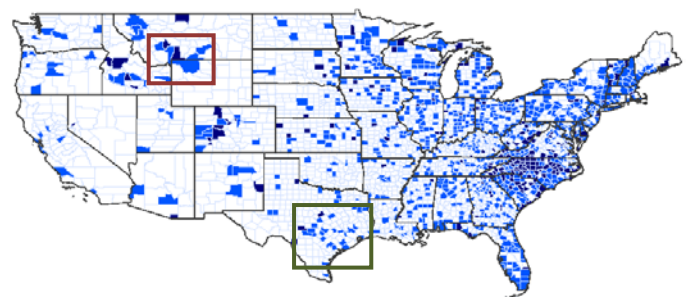
- puisqu'il peut arriver que l'ensemble des points de vente soient concentrés localement à un endroit de l'unité géographique (pour la méthode directe et la méthode par grille) en outre, l'utilisation de l'indicateur distance. Si cette concentration est située en périphérie de l'unité géographique, la valeur de l'indicateur distance est élevée par rapport au recouvrement réel des magasins. Pour contourner ce problème, il pourrait être envisagé de déterminer la position du centroïde en tenant compte de la répartition de la population.



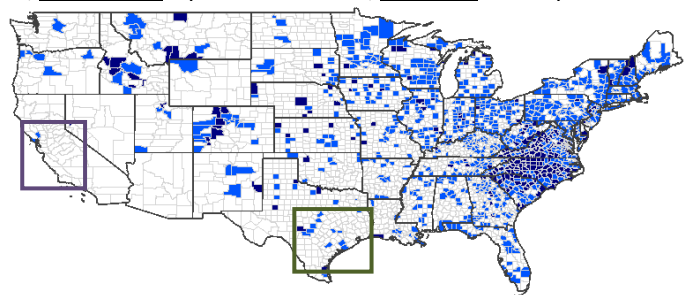
a. Imputation : nombre de magasins appartenant au comté ; **Rapporté à :** Population du comté ; **Méthode :** directe par comté



c. Estimation par noyau : triangulaire, 15 miles ; **Rapporté à :** Pas de rapport ; **Méthode :** grille



b. Imputation : plus proche centroïde de comté du magasin ; **Rapporté à :** Population du comté ; **Méthode :** directe par comté



d. Estimation par noyau : triangulaire, 15 miles ; **Rapporté à :** Population du comté ; **Méthode :** grille intermédiaire

Figure 3. Visualisation des résultats pour les méthodes par indicateurs de densité

Tableau 5. Moyennes réelles des indicateurs par segment.

| MOYENNES REELLES | Densités (rapport : pour 1000 habitants) | | Distance au plus proche magasin (en degrés) | | |
|------------------|--|-----------|---|-----------------|----------------------|
| | Compte du nombre de magasins, imputation | | Estimation par noyau | | |
| Segment | appartenance | centroïde | Triangulaire, 15 miles | Méthode directe | Grille intermédiaire |
| 1 | 0.05454 | 0.07034 | 0.04778 | 0.14248 | 0.15928 |
| 2 | 0.04274 | 0.05075 | 0.04586 | 0.22301 | 0.25301 |
| 3 | 0.02384* | 0.02730* | 0.02905 | 0.48456* | 0.48775* |
| 4 | 0.02470* | 0.02907* | 0.02728 | 0.38814* | 0.39961* |

4.5 Analyse des données

4.5.1 Comparaison des résultats par comté

Deux cas de figure, opposés et nécessitant une restructuration du réseau de distribution, peuvent se présenter pour chaque comté (Tableau 6):

Tableau 6. Restructurations du réseau de distribution

| Cas pour chaque comté | « segment recouvrement » | « segment recouvrement » |
|-----------------------|---|--|
| Analyse spatiale | faible représentation de l'enseigne densité -, distance + | forte représentation de l'enseigne densité +, distance - |
| Segmentation | Clientèle cible | Pas la clientèle cible |
| Recommandation | Réseau de distribution pas assez développé, ajout d'implantations | Redondance dans les points de vente, supprimer des implantations |

Il a été choisi de calculer des indicateurs par comté : méthode directe et par grille intermédiaire pour limiter le nombre de comparaisons. Il s'agit de 3109 comparaisons avec des indicateurs par comté contre 83636 avec la méthode par grille.

4.5.1 Moyennes

L'établissement des moyennes des indicateurs pour chaque segment permet de vérifier la cohérence de l'ordre des tendances (Tableau 5). Idéalement par exemple, si le réseau de distribution était cohérent, le segment 1 devrait correspondre aux plus grandes densités et aux distances les plus faibles. L'ordre est respecté à l'exception des densités par compte du nombre de magasins (appartenance et centroïde) et l'indicateur distance. En effet pour ces indicateurs, la densité est plus élevée pour le segment 4 que pour le segment 3 et la distance est plus faible pour le segment 4 que pour le segment 3 (moyennes indiquées avec un astérisque).

D'autre part, pour chaque segment j , on calcule la moyenne théorique idéale de l'indicateur (notée μ_j^{th}). C'est la moyenne qu'on aurait dû obtenir si tous les points de vente avaient été positionnés en respectant la valeur des segments et avec des valeurs des indicateurs identiques à celles obtenues. La valeur du segment des comtés n'est alors pas prise en compte. On trie toutes les valeurs d'indicateurs par ordre croissant ou décroissant (densités ou distances) et on les assigne à chaque segment selon les proportions issues de la segmentation (par exemple, les premiers 7% au segment 1 etc.). La moyenne théorique d'une classe pour un indicateur est alors la moyenne des valeurs à l'intérieur de la classe. Cette moyenne théorique est bien évidemment différente de la moyenne obtenue pour chaque segment (Figure 4).

On s'aperçoit que, par exemple pour l'indicateur distance avec méthode par grille intermédiaire, les moyennes théoriques des classes 1, 2 et 3 sont plus faibles que les moyennes des classes réelles (issues de la segmentation) et que la moyenne théorique de classe 4 est plus élevée que la moyenne réelle.

4.5.1 Histogrammes

Pour chaque indicateur, on trace l'histogramme de distribution des valeurs pour chaque segment.

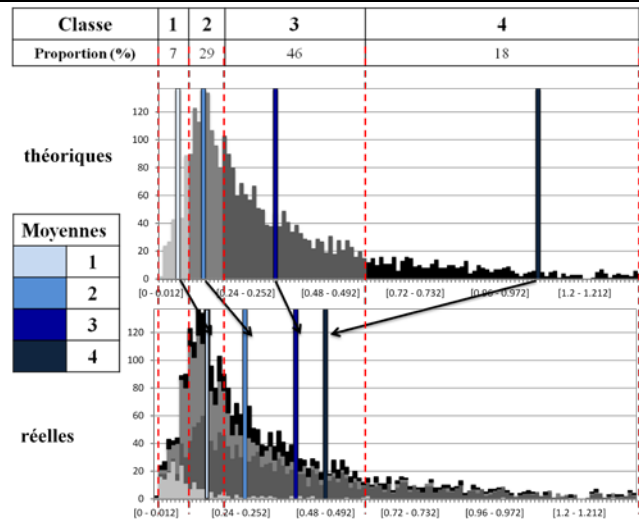


Figure 4. Moyennes réelles et théoriques pour chaque segment pour l'indicateur distance (grille intermédiaire)

| Classe indicateurs | Segment | | | |
|--------------------|---------|--------|--------|--------|
| | 1 | 2 | 3 | 4 |
| 1 | Green | Yellow | Orange | Red |
| 2 | Yellow | Green | Yellow | Yellow |
| 3 | Yellow | Yellow | Green | Yellow |
| 4 | Red | Orange | Yellow | Green |

Figure 5. Code de lecture des histogrammes

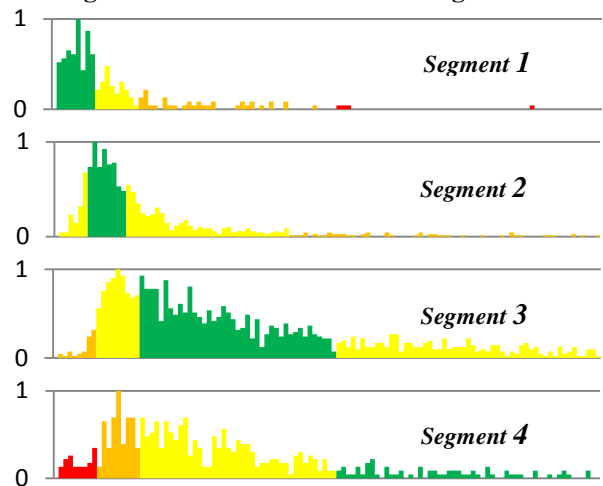


Figure 6. Distribution par classes de l'indicateur distance pour chaque segment (méthode par grille intermédiaire)

Parallèlement, on détermine les limites des 4 classes des valeurs des indicateurs en respectant les proportions des segments (classes utilisées pour la représentation des indicateurs sur les cartes). Si toutes les localisations des points de vente étaient bien adaptées au potentiel de consommation du comté, il devrait y avoir coïncidence entre les classes des indicateurs et les segments. Le code couleur de la Figure 5 a été utilisé pour identifier les comtés dont l'écart entre la valeur de segment et la valeur de l'indicateur considéré est élevé, soit en orange et rouge. La couverture sur les comtés en question est trop faible pour les segments 1 et 2 et trop élevée pour les segments 3 et 4. Si l'on considère l'indicateur de distance avec

la méthode par grille intermédiaire (Figure 6), on voit que l'histogramme du segment 4 présente de nombreux comtés avec des distances faibles et dont la couverture devrait être revue à la baisse.

4.5.2 Priorisation des modifications

Pour chaque comté et pour chaque méthode de calcul d'indicateur de recouvrement, on compare la valeur de segment et la valeur de classe de l'indicateur. L'implantation du réseau est satisfaisante pour les comtés qui ont des valeurs de segment et de classes d'indicateur identiques. Pour les autres comtés, on mesure la divergence entre la valeur de segment et la valeur de l'indicateur en calculant en nombre d'écart type de l'indicateur σ , l'écart entre la valeur de l'indicateur obtenue et la moyenne théorique de la classe auquel il devrait appartenir. Le signe dépend du type d'indicateur : + pour les indicateurs de densité et - pour les indicateurs de distance.

On définit $\forall j \in Z, S_j$ la valeur de segment du comté j et $divergence_j$

$$= \begin{cases} 0, & \text{si } S_j \text{ est égale à la classe de l'indicateur de } j \\ \pm [\text{indicateur}_j - \mu_{S_j}^{th}] / \sigma, & \text{sinon} \end{cases}$$

Les comtés pour lesquels la divergence est la plus élevée (en valeur absolue) sont ceux dont le réseau de distribution est le moins bien adapté. On normalise ensuite l'ensemble des indicateurs. On peut donc calculer la moyenne de la divergence des 5 indicateurs pour chaque comté. Des recommandations ainsi qu'une carte des comtés pour lesquels les modifications sont prioritaires peuvent alors être établies (Figure 7).

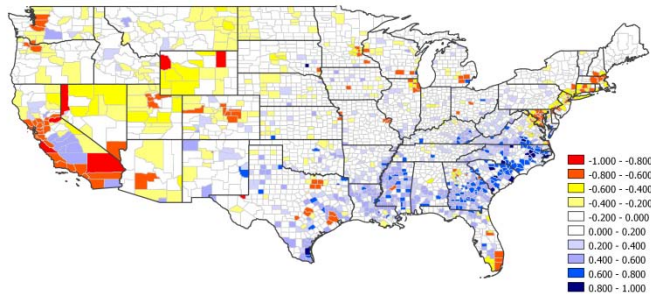


Figure 7. Priorisation des modifications. Rouge : ajout de points de vente, Bleu : retrait de points de vente

5 CONCLUSION

Les résultats confirment que l'intégration d'une segmentation de marché avec des méthodes d'analyse spatiale permet de formuler des avis sur la localisation des points de vente d'un réseau de distribution pour une entreprise.

Les zones qui se distinguent par une couverture trop ou pas assez importante sont directement mises en valeur sur la carte. A partir du résultat final, l'entreprise pourra décider s'il y a lieu d'adapter le réseau de distribution dans ces zones. En effet, si la clientèle joue un rôle prépondérant dans le choix d'un emplacement, il existe d'autres principes à respecter et d'autres facteurs à considérer. L'entreprise doit tenir compte, par exemple, du principe d'attraction cumulative (Lewison & Delozier, 1986) selon lequel un regroupement de commerces similaires crée une synergie.

Différentes perspectives de poursuite peuvent être envisagées :

- Pondérer les points de vente selon leurs surfaces ou leurs chiffres d'affaires pour représenter leurs performances
- Prendre en compte la sectorisation de la force de vente et la localisation des centres de distribution

- Intégrer des mesures d'analyse spatiales par rapport à la concurrence directement dans le modèle.

6 REFERENCES

- Andresen, M. A. (2006). Crime measures and the spatial analysis of criminal activity. *Brit. J. Criminol.* 46(2), pp. 258-285
- Beaujeu-Garnier, J., & Delobez, A. (1977). *Géographie du commerce*. Paris: Masson.
- Berry, M. J., & Linoff, G. (1997). *Data Mining: For Marketing, Sales, and Customer Support*: John Wiley & Sons.
- Brunet, R., Ferras, R., & Théry, H. (1993). *Les mots de la géographie, dictionnaire critique*, 2nde Ed, La Documentation française - Collection dynamiques du territoire.
- Census Bureau. (2010). Decennial Census and American Community Survey
- Cliquet, G. (2001). *Le géomarketing : Méthodes et stratégies du marketing spatial*: Hermes Science Publications - Aménagement et gestion du territoire.
- Cliquet, G., & Josselin, J.-M. (2002). *Stratégies de localisation des entreprises commerciales et industrielles : de nouvelles perspectives*, 1ère Ed, De Boeck.
- Douard, J.-P., & Heitz, M. (2004). *Le Géomarketing : au service de la démarche marketing*: Dunod.
- Filion, M., & Colbert, F. (1990). *Gestion du marketing*, 2nde Ed, Boucherville: Gaëtan Morin.
- Florence, M. (1990). International Trade and the Furniture Industry. *Alabama International Trade Center*, 18-32.
- Industrie Canada (2011). Exportations totales canadiennes - SCIAN 337 - Fabrication de meubles et de produits connexes, from <http://www.ic.gc.ca/>
- Latour, P., & LeFloc'h, J. (2001). *Géomarketing : Principes, méthodes et applications*: Editions d'Organisation.
- Le, T. T. H., Agard, B., & Deveault, S.). Application du data mining à la segmentation du marché des meubles aux États-Unis. 8ème Congrès International de Génie Industriel – GI 2009, Bagnères de Bigorre, France, 10-12 juin, 2009.
- Lendrevie, J., Lévy, J., & Lidon, D. (2006). *Mercator*, 8ième Ed, Théorie et pratique du Marketing - Dunod.
- Levine, N. (2004). *CrimeStat : a spatial statistics program for the analysis of crime incidents location*. 3rd Ed, Washington (DC) : The National Institute of Justice.
- Lewison, D. M., & Delozier, M. W. (1986). *Retailing*: Merrill Pub. Co.
- Lihra, T. (2009). La personnalisation de masse pour l'industrie du meuble en bois. Doctorat en sciences du bois, Université Laval.
- Murillo, L. M. (2007). Manufacturers-Retailers: The New Actor in the U.S.Furniture Industry. Characteristics and Implications for the Chinese Industry. *World Academy of Science, Engineering and Technology*.
- Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 78, 311-329.
- Rizet, C. (2008). Chaines logistiques et consommation d'énergie : Cas des meubles et des fruits & légumes. *Institut National de recherche sur les transport et la sécurité, Département Economie et Sociologie des Transports (DEST)*.
- Tellier, L.-N. (1985). *Économie spatiale: rationalité économique de l'espace habité*: Gaëtan Morin.
- Volpe, A., & Peluso, M. (2007). The kitchen furniture market in the united states: CSIL reports S34, Centre for Industrial Studies.