

ÉTUDE DE MÉTHODES D'ANALYSE SPATIALE ET ILLUSTRATION À L'AIDE DE MICRODONNÉES URBAINES DE LA GRANDE RÉGION DE MONTRÉAL

CATHERINE MORENCY

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

1. INTRODUCTION

Un point. Deux points. Trois points. Une multitude de points. Un portrait entier construit de points, petits et gros, semblables ou non, éparpillés ça et là autour de quelques agrégats.

Des points en modélisation des transports ? Il en existe plusieurs. Il existe toute une panoplie d'entités urbaines dont l'expression spatiale la plus simple est le point : les ménages échantillonnés lors des enquêtes Origine-Destination, les secteurs de dénombrement, plus petite entité statistique pour laquelle des données de recensement sont collectées, les noeuds du réseau de transport en commun, les extrémités de déplacements, les lieux d'emplois, les espaces de stationnement, les localisations successives exprimant la trace d'un véhicule muni d'un GPS.

D'abord perçus comme des ensembles sans cohérence, ces nuages de points deviennent des révélateurs de comportements et tendances lorsque examinés par le biais des outils disponibles dans les systèmes d'information géographiques et disciplinés selon des méthodes appropriées. L'identification et l'application de ces outils et méthodes ainsi que le choix et la structuration des données requises pour les utiliser sont des enjeux très actuels qui méritent d'être discutés.

Cet article contribue à ce travail. Il présente d'abord différents enjeux relatifs à l'exploitation des données spatiales. Différentes perspectives permettent de cerner les défis relevant de l'existence des données spatiales, de leur abondance, de leur structure ainsi que de leur état. Pour faire face aux différents défis énoncés, quelques méthodes sont présentées. L'exercice tente d'une part d'illustrer les potentialités de synthèse de différentes méthodes spatiales et statistiques et d'autre part de démontrer leur pertinence dans le contexte particulier de modélisation des interactions entre dynamiques spatiales et mobilité urbaine. En outre, cet article n'a pas la prétention de dénombrer l'ensemble des méthodes spatiales applicables aux données urbaines mais vise plutôt à contribuer à ce propos en illustrant l'usage de certaines de ces méthodes à l'aide de grandes bases de données de transport. Dans le contexte montréalais, les données des enquêtes Origine-Destination et des recensements canadiens, révélateurs à la fois des comportements de mobilité et de la socio-démographie, constituent un laboratoire privilégié d'expérimentation.

Sur cette base, l'article illustre différentes méthodes notamment :

- les mesures de centralité (centre moyen) et dispersion (écart type de la distance, ellipse de dispersion) qui synthétisent le plus simplement possible des nuages de points ;
- les mesures d'autocorrélation spatiale (indices de MORAN et GEARY) qui évaluent le degré de similarité entre observations en fonction de leur éloignement ;
- les mesures de classification automatique qui permettent d'identifier des zones de forte concentration à savoir des agrégats de points répondant à certains critères de proximité, révélant par conséquent certaines structures locales.

2. MISE AU POINT : DÉFINITION ET ENJEUX

2.1. EXPLORATION ET ANALYSE SPATIALE : QUELQUES CLARIFICATIONS

« *Exploratory data analysis is a detective work – numerical detective work – or counting detective work – or graphical detective work* ». En 1977, John TUKEY publiait un livre qui a vraisemblablement modifié la conception usuelle de l'analyse et de la visualisation des données. Le concept d'analyse exploratoire des données (EDA) a engendré plusieurs réflexions sur les

méthodes permettant d'utiliser les données et de révéler les patrons et tendances qu'elles contiennent. La métaphore semble réelle : plusieurs méthodes et instruments technologiques ont été développés afin d'assister la tâche d'exploitation des données, la tâche d'organisation des données pour des fins particulières, la tâche d'adaptation d'ensembles de données à des fins spécifiques d'analyse, de planification et de modélisation.

Le développement des systèmes d'information géographique et des multiples fonctionnalités de visualisation, d'acquisition et de traitement des données inhérentes, a rendu essentielle l'émergence de nouvelles techniques d'analyse exploratoire adaptées au caractère spatial des données (ANSELIN, 1995). Bien que les méthodes d'EDA permettent d'assister l'identification de patrons contenus dans un ensemble de données, elles ne permettent pas de traiter explicitement le caractère spatial des données. C'est ce que l'ESDA ou analyse spatiale exploratoire des données propose par le biais de différentes méthodes visant à révéler des tendances, relations et patrons observables dans l'espace. Le lecteur peut consulter BANOS (2001b) pour une introduction à la philosophie de l'approche spatiale exploratoire ainsi que BANOS (2001a) pour une démonstration de certaines fonctionnalités de l'ESDA appliquées à l'analyse de la demande de transport en commun. L'ESDA permet notamment de décrire les distributions spatiales, d'identifier les observations atypiques et de révéler des structures spatiales à l'aide de fonctionnalités de visualisation. Les nouveaux outils permettent dorénavant d'effectuer cette visualisation de façon interactive, sur différents sous-ensembles sélectionnés successivement (concept de *brushing*) (ANSELIN et alii, 2006).

La littérature fait état de différentes compréhensions du concept d'analyse spatiale. Tel que diffusé sur Hypergéô, encyclopédie électronique consacrée à l'épistémologie de la géographie, l'analyse spatiale permet de mettre « *en évidence des structures et des formes d'organisation spatiale récurrentes* ». Selon BAILEY et GATRELL (1995), il est question d'analyse spatiale de données lorsque les données sont d'une part localisées dans l'espace et d'autre part que cette organisation spatiale est considérée importante, explicitement, dans l'analyse ou l'interprétation des résultats. En outre, le concept d'analyse spatiale s'articule autour de trois composantes :

- la visualisation qui permet de représenter les données spatiales ;
- l'ESDA qui permet d'explorer et de synthétiser les données afin de révéler des configurations spatiales particulières (patrons, relations spatiales) ;
- la modélisation spatiale qui tente d'expliquer les configurations par la spécification d'un modèle statistique et l'estimation des paramètres.

La frontière entre ces composantes est parfois floue et relève souvent du niveau de complexité propre à la préparation des données nécessaires à l'application des méthodes. En outre, de multiples méthodes, issues de domaines variés, apparaissent comme de potentiels outils de visualisation et d'exploration. C'est dans cet esprit que cet article aborde le problème de

l'analyse des microdonnées spatiales urbaines disponibles dans la Grande Région de Montréal. Différentes méthodes sont illustrées afin de contribuer à l'identification des configurations spatiales singulières de différents phénomènes urbains.

2.2. DONNÉES EXPÉRIMENTALES

Tel que documenté dans la plupart des ouvrages de référence, il existe au moins trois types fondamentaux de données spatiales (BAILEY, GATRELL, 1995 ; CRESSIE, 1993):

- les nuages de points (*point patterns*) : les nuages de points représentent un ensemble d'événements se produisant en différentes localisations. C'est l'organisation spatiale de ces événements qui est d'intérêt et l'étude de ces données porte sur l'identification de configurations spatiales particulières.
- les données géostatistiques (*geostatistical data*) : les données géostatistiques réfèrent généralement à un échantillon de points (x,y), provenant d'une surface continue, pour lesquels un attribut (z) a été mesuré. La finalité de cet échantillonnage est souvent de modéliser les variations spatiales afin de prédire la valeur de l'attribut aux endroits non échantillonnés à l'aide d'un variogramme et de méthodes d'interpolation spatiale.
- les données agrégées (*regional data*) : ces données résultent de l'agrégation de données discrètes et sont représentées par des points ou des zones. Généralement, plusieurs attributs ont été mesurés pour ce découpage territorial et l'enjeu consiste à comprendre l'organisation spatiale de ces attributs afin d'identifier des tendances ainsi que les relations entre ceux-ci.

Les illustrations présentées dans cet article s'appuient sur deux principaux ensembles de données :

- Les données discrètes provenant des grandes enquêtes ménages Origine-Destination tenues environ tous les cinq ans dans la Grande Région de Montréal. Le taux d'échantillonnage étant d'environ 5 % des ménages résidant sur le territoire d'enquête (échantillon de 65 000 ménages en 1998), les observations sont pondérées afin de représenter la population de référence. Les illustrations proposées exploitent deux types d'objets : les ménages et les personnes. Chaque objet (ménage ou personne) est d'abord représenté par un point dans l'espace qui correspond à la localisation spatiale précise de son lieu de résidence (paire de coordonnées x-y). Il est par ailleurs caractérisé par un vecteur d'attributs stockant différentes propriétés par exemple, pour les personnes, l'âge, le sexe, la possession d'un permis de conduire et le facteur de pondération. Ces ensembles constituent des nuages de points.

- Les données agrégées issues des recensements canadiens quinquennaux. Ces données sont diffusées selon différents niveaux d'agrégation spatiale, notamment au niveau du secteur de dénombrement¹. Les résultats sont disponibles sous forme de tables, plus de cinquante en 1996, exprimant l'occurrence simple ou croisée de différentes variables. Chaque secteur est délimité spatialement par le biais de frontières, ce qui en permet la représentation spatiale sous forme de zones (polygones). Ces frontières délimitent le territoire impliqué dans la compilation des tables agrégées. Chaque secteur peut aussi être géolocalisé par le biais d'un point de type centroïde. Les données concernant les unités de logement sont utilisées pour illustrer les mesures d'auto-corrélation spatiale.

Selon la définition de MILLER (2003), ces données expérimentales ont trois types d'existence :

- Une existence **spatiale** : toute donnée à caractère spatial est géoréférencée selon un système de projection cartographique. Le système UTM (projection cartographique de Mercator transverse universelle) agit comme référentiel spatial dans la présente étude. Chaque objet est spatialement localisé par le biais d'une paire de coordonnées (x,y). Ce point représente la localisation de l'objet en soi (un lieu de résidence, une extrémité de déplacement) ou la synthèse d'une entité spatiale (centre de gravité, point représentatif, centroïde de zone).
- Une existence **temporelle** : les données utilisées proviennent de séries mixtes² puisqu'elles permettent d'étudier à la fois les tendances spatiales et temporelles. Chaque donnée est liée à la période à laquelle elle réfère : l'année de réalisation de l'enquête, le jour et l'heure de déplacement.
- Une existence **sémantique** : toute donnée révèle un état. L'existence sémantique est la valeur, l'intensité, l'attribut révélé par la donnée : le genre d'un individu, le nombre d'unités de logement d'un secteur de dénombrement.

2.3. ENJEUX RELATIFS À L'EXPLOITATION DES DONNÉES SPATIALES

Bien que la localisation spatiale absolue d'un événement ou d'une observation puisse être pertinente, c'est souvent la localisation relative des événe-

¹ Secteur de dénombrement (SD): jusqu'au recensement de 1996, plus petite entité statistique pour laquelle des données de recensement sont diffusées. En 1996, environ 4 600 SD, comptant chacun en moyenne 300 ménages, couvrent le territoire de la Grande Région de Montréal. En 2001, la plus petite entité statistique est l'aire de diffusion, composée d'un ou de plusieurs pâtés de maisons avoisinants et regroupant de 400 à 700 habitants.

² Séries temporelles : séquence de mesures au même point de l'espace prises à différentes périodes temporelles ; Séries spatiales : ensemble de mesures faites simultanément ou non en différents points de l'espace ; Séries mixtes : ensemble de séries temporelles obtenues de différents points dans l'espace (SHAROV, 1996).

ments les uns par rapport aux autres, ou en regard d'un autre événement (la position du centre-ville par exemple), qui devient critique dans l'analyse spatiale des phénomènes urbains. L'appréciation du positionnement relatif entre observations fait appel aux concepts de distance (mesurable selon différentes métriques) ou d'adjacence (observations qui partagent une frontière), concepts qui permettent de définir des voisinages différents selon le type d'observation. Le voisinage d'une observation particulière définira souvent l'ensemble des observations qui influencent ses attributs, cadrant l'effet d'interaction spatiale entre celles-ci.

Le phénomène d'interaction spatiale est un des deux principaux problèmes qui ont engendré la nécessité de s'intéresser davantage à l'incidence de l'espace sur la nature et l'intensité des mesures faites en différents points (JAYET, 2001 ; ANSELIN, 1999). Ces problèmes sont :

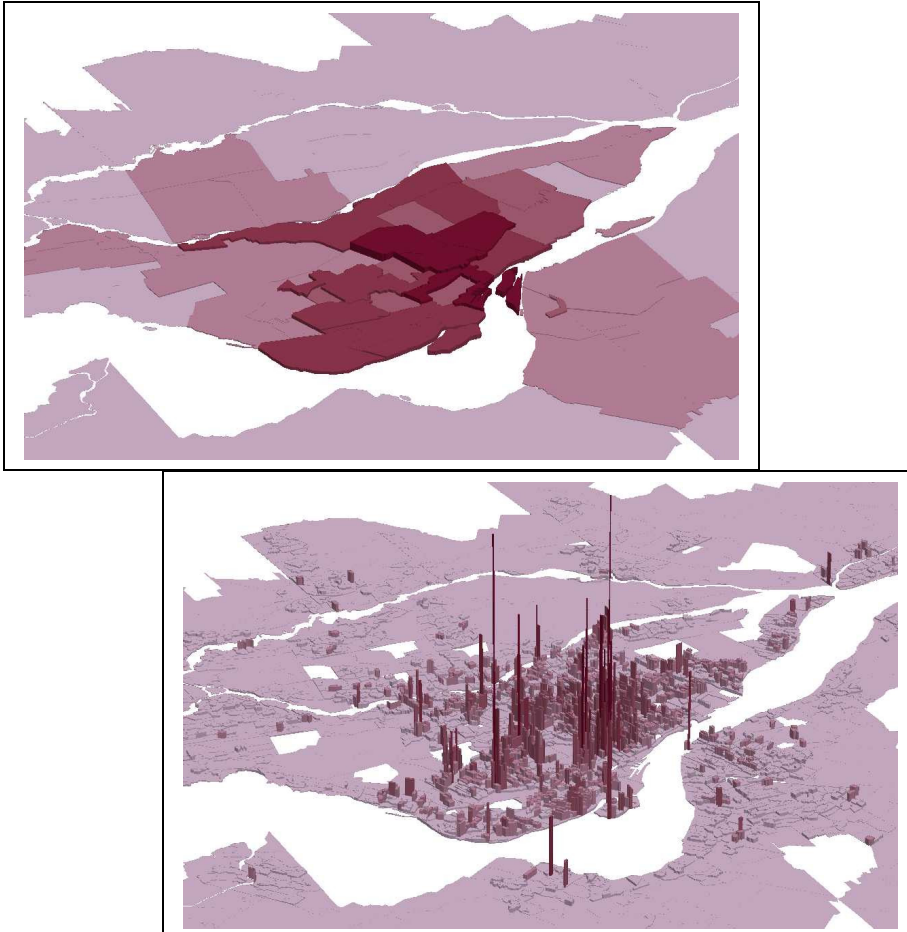
- **l'hétérogénéité** des observations : tel qu'exprimé par JAYET (2001), « toute analyse statistique d'une population suppose que les éléments de cette population ont des points communs, sur lesquels on peut fonder des comparaisons et asseoir des régularités ». Cependant, les données spatiales présentent généralement une forte hétérogénéité qui implique que la valeur des observations varie dans l'espace. JAYET (2001) distingue différents types d'hétérogénéité : de taille (ordre de grandeur des agrégats), de forme/position (morphologie spatiale, urbain/périurbain) et de structure (socio-démographie de la population, niveau de scolarité, revenus).
- les **interactions** liées à la proximité : la première loi de géographie de TOBLER (1970), « *Everything is related to everything else, but near things are more related than distant things* », exprime simplement le concept d'interaction spatiale selon lequel des observations faites dans une localisation particulière dépendent des observations faites dans d'autres localisations et que cette dépendance est fonction de la distance séparant ces localisations.

La considération désagrégée des données spatiales demeure un enjeu critique pour le raffinement des modèles de transport et d'utilisation du sol (MILLER et alii, 2004 ; TIMMERMANS, 2003). De nombreuses discussions se tiennent autour des effets de l'agrégation (spatiale, temporelle et sémantique) dans la mesure des phénomènes, notamment l'estimation des modèles (une littérature abondante porte sur le sophisme écologique³ (*ecological fallacy*) et le problème de la zone modifiable (*MAUP : Modifiable Areal Unit Problem*)). En fait, les résultats des analyses dépendent du niveau d'agrégation choisi. Les cartes de densité de population présentées à la Figure 1 illustrent la perte

³ L'action d'inférer que les relations à l'échelle individuelle sont les mêmes que celles observées à un niveau agrégé est appelée « sophisme écologique » (erreur d'inférence due à l'absence de distinction entre différents niveaux d'organisation). (Définition tirée du glossaire du Guide Canadien des incidences sur la santé - SANTÉ CANADA, 2004).

de variabilité avec l'augmentation du niveau d'agrégation spatiale (65 secteurs municipaux versus plus de 4 000 secteurs de dénombrement). Face à l'inconsistance des mesures agrégées, de nouvelles pistes d'analyse sont examinées afin de révéler l'information que contiennent les données désagrégées ou microdonnées sans toutefois faire abstraction de leur grande variabilité.

Figure 1 : Densité de population : comparaison entre une représentation basée sur les secteurs municipaux (65 zones) et une représentation basée sur les secteurs de dénombrement ($\approx 4\ 000$ zones) (1998)



3. ÉTUDE ET ILLUSTRATION

Dans la perspective de contribuer à l'évolution des modèles usuels utilisés dans l'étude des dynamiques spatiales et des comportements individuels de mobilité, différentes méthodes simples, permettant de synthétiser des ensembles de données spatiales, sont expérimentées. Ces méthodes sont souvent

implantées au cœur de systèmes d'information géographique ou sont accessibles via des applications *ad hoc* respectant les standards de stockage de données (format de fichier). Outre les fonctionnalités offertes par les systèmes d'information géographiques commerciaux (ESRI, MapInfo), différents logiciels gratuits soutiennent l'analyse de données spatiales notamment GeoDa, développé par le Center for Spatially Integrated Social Science (ANSELIN et alii, 2006), CrimeStat, développé sous la supervision du Dr. LEVINE (2004). Les illustrations présentées dans cet article ont été produites à l'aide des fonctionnalités du logiciel CrimeStat.

Les objectifs poursuivis par cette démarche sont d'une part de valider l'applicabilité de ces méthodes à de grands ensembles de données, pour assister l'exploration spatiale, et d'autre part d'apprécier leur pertinence analytique pour fins de modélisation des dynamiques urbaines.

3.1. LES MESURES DE CENTRALITÉ ET DE DISPERSION

Les mesures statistiques descriptives les plus usuelles sont la moyenne et la variance. L'introduction d'une composante spatiale permet de caractériser la centralité et la dispersion de nuages de points référencés selon un système de projection cartographique donné.

Les mesures de centralité

Le centre moyen est le descripteur le plus simple d'une distribution ; c'est l'équivalent de la moyenne, en deux dimensions. Les coordonnées du centre moyen d'un nuage de points représentent la moyenne des deux dimensions en présence :

$$x_{cm} = \frac{\sum_{i=1}^n x_i}{n}, y_{cm} = \frac{\sum_{i=1}^n y_i}{n}$$

où n est le nombre de points et x, y les coordonnées géographiques du point i .

La présence d'une troisième dimension, à savoir une intensité observée en un point particulier de l'espace (existence sémantique), ou un poids lorsque le point est une synthèse d'un sous-ensemble territorial (identificateur spatial d'un agrégat), permet d'estimer un centre moyen pondéré (centre de gravité) :

$$x_{cg} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, y_{cg} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

où w est l'intensité (poids) observée aux différents points.

Les mesures de dispersion

La première mesure de dispersion consiste à mesurer l'écart type (*Standard Deviation*) dans chacun des axes (x,y), de façon distincte :

$$s_x = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}}, \quad s_y = \sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}}.$$

Une seconde méthode permettant de synthétiser la dispersion en une seule mesure statistique est l'écart type de la distance (*Standard Deviation Distance*). Cette statistique mesure l'écart type de la distance de chaque point par rapport au centre moyen. C'est donc l'équivalent, en deux dimensions, de l'écart type classique d'une distribution.

$$s_{xy} = \sqrt{\sum_{i=1}^n \frac{(d_{icm})^2}{n-2}}$$

où d_{icm} est la distance entre chaque point et le centre moyen.

La dispersion d'un nuage de points peut toutefois présenter certaines tendances spatiales. L'ellipse de dispersion (EDS : *Standard Deviation Ellipse*) permet de tenir compte de cette anisotropie⁴. Selon cette mesure, les écarts-types dans les deux directions (x,y) sont orthogonaux et définissent une ellipse. Ces axes sont pivotés de façon à minimiser la somme des distances quadratiques entre les points et les axes. La longueur de chaque axe correspond à deux fois l'écart type des données dans cet axe. L'orientation de l'ellipse permet d'identifier les axes de dispersion maximale et minimale. Environ 68 % des observations se retrouvent à l'intérieur de l'EDS (1 écart-type).

Il existe différentes méthodes de détermination de ces ellipses ; la méthode appliquée est la suivante :

- l'axe Y est pivoté d'un angle (dans le sens des aiguilles d'une montre)

où

$$\theta = \frac{\text{ARCTAN} \left\{ \frac{\left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right) + \left[\left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right)^2 + 4 \left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)^2 \right]^{1/2}}{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})} \right\}}{\left(2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)}$$

- les deux écarts-types sont calculés, un dans l'axe x transposé, l'autre dans l'axe y transposé :

⁴ Anisotropie : variance d'état selon la direction considérée.

$$s_x = \sqrt{\frac{\sum_{i=1}^n ((x_i - \bar{x})\cos\theta - (y_i - \bar{y})\sin\theta)^2}{n-2}}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n ((x_i - \bar{x})\sin\theta - (y_i - \bar{y})\cos\theta)^2}{n-2}}$$

- la longueur de chaque axe de l'ellipse correspond à deux fois l'écart-type dans la direction correspondante.

Illustration

Deux mesures classiques, le centre moyen et l'ellipse de dispersion, sont utilisées pour caractériser la distribution spatiale de différents segments de population. Des tendances spatiales lourdes, bien connues, sont confirmées à l'aide de ces mesures de synthèse estimées à partir de nuages de points de grande envergure : 65 000 points représentant les ménages et 165 000 points représentant les personnes.

Le Tableau 1 présente une synthèse de ces deux mesures pour différents segments de ménages et personnes (enquêtes ménages de 1987 et 1998). Les propriétés de centralité et de dispersion de chaque segment peuvent être comparées avec celles de la population de référence (ménages ou personnes) afin d'apprécier l'effet discriminant de la variable de segmentation en regard de l'espace. La position du centre moyen pondéré est résumée par son éloignement du centre-ville alors que la dispersion s'apprécie d'une part par la superficie de l'ellipse de dispersion et d'autre part par le ratio entre les deux axes de l'ellipse (mesure d'anisotropie spatiale).

Des exemples d'EDS sont présentés aux Figures qui suivent (Figure 2 : dispersion de référence des ménages et personnes, Figure 3 : taille des ménages, Figure 4 : usagers du transport en commun). Ces données permettent notamment d'observer que :

- La dispersion des ménages augmente avec leur taille, les célibataires (personnes vivant seules) étant plus concentrés autour du centre-ville que les ménages de grande taille. En outre, cette dispersion s'est grandement accrue avec le temps pour toutes les tailles de ménage (Cf. Figure 3).
- La distribution spatiale des segments de ménages selon leurs revenus annuels est plutôt caractéristique : les ménages les plus fortunés sont surtout localisés dans la partie ouest de l'Île de Montréal, qui présente aussi une forte dominance d'anglophones. Les ménages les plus dispersés sont ceux gagnant entre 50 000\$ et 75 000\$ par année. Finalement, l'évolution temporelle de la dispersion des ménages moins

fortunés est remarquable, passant d'une forte concentration dans la partie centrale de l'Île de Montréal en 1987, à une dispersion dans les banlieues proches en 1998.

Tableau 1 : Mesures de centralité et de dispersion de différents segments de population (ménages et personnes) population de 1987 et 1998 (territoires respectifs)

	% Observations		Centre moyen (CM) Dist. CM-CV (km)		Ellipse de dispersion Aire (km ²) Axes Long / Court			
	1987	1998	1987	1998	1987	1998	1987	1998
Population de référence (ménages et personnes)								
Ménages	53 384	65 227	5.18	6.16	572.5	1 022.9	1.29	1.21
Personnes	137 956	164 075	5.51	6.57	652.9	1 111.1	1.29	1.21
Taille du ménage								
1 personne	21.2%	24.4%	4.13	5.07	331.89	753.69	1.28	1.20
2 personnes	34.6%	33.6%	5.07	6.10	526.11	1032.44	1.30	1.22
3 personnes	18.7%	17.6%	5.64	6.71	665.19	1150.51	1.27	1.20
4 personnes +	25.4%	24.4%	6.01	7.22	789.69	1246.28	1.30	1.22
Revenus annuels moyens des ménages								
< 25 000\$	7.7%	8.2%	3.18	3.69	103.04	535.48	1.03	1.35
25 000\$ - 49 999\$	62.7%	56.4%	5.08	6.44	496.68	992.07	1.12	1.12
50 000\$ - 74 999\$	24.7%	28.6%	6.42	6.67	879.37	1280.25	1.59	1.36
75 000\$ +	4.9%	6.8%	6.51	6.37	436.78	706.35	2.00	1.86
Age des personnes								
05-15 ans	14.8%	15.7%	6.00	7.34	826.29	1298.68	1.29	1.21
16-24 ans	15.2%	13.2%	5.39	6.09	573.25	1016.88	1.28	1.21
25-34 ans	20.0%	15.7%	5.50	6.53	638.11	1017.92	1.26	1.17
35-54 ans	29.3%	35.2%	5.55	6.63	700.53	1173.55	1.30	1.22
55-64 ans	10.2%	9.5%	5.36	6.41	521.16	1070.85	1.32	1.25
65 ans +	10.4%	10.7%	4.85	5.91	441.29	887.11	1.30	1.21
Accessibilité à l'automobile par les personnes dans le ménage								
0 auto	16.2%	12.4%	3.91	4.33	240.33	424.94	1.17	1.06
< 1 auto / 16 ans+	58.9%	48.7%	5.52	6.35	646.98	1015.82	1.27	1.20
>=1 auto / 16 ans+	25.0%	38.9%	6.63	7.75	931.60	1489.31	1.37	1.24
Usager du transport en commun								
Usager TC	19.9%	13.1%	4.22	4.29	279.72	325.14	1.22	1.14
Non Usager TC	80.1%	86.9%	5.84	6.97	746.68	1242.16	1.30	1.21

- Lorsque l'âge agit comme variable de segmentation, les personnes âgées de 65 ans et plus sont les moins dispersées bien que leur dispersion ait doublé en 10 ans (superficie de l'EDS). Les jeunes, suivis des 35-54 ans, segment de population agissant plausiblement comme parents, sont les plus dispersés.
- Le niveau d'accès à l'automobile a un impact évident sur le choix d'un lieu de résidence (ou l'inverse...). Les individus sans accès à l'automobile sont concentrés dans la partie centrale de l'Île de Montréal où les services de transport en commun sont les plus présents (particulièrement le métro). Néanmoins, la dispersion de ce segment de population

s'est aussi légèrement accrue avec le temps. Par opposition, les personnes ayant un accès complet à l'automobile se dispersent davantage.

Figure 2 : Ellipses de dispersion et centres moyens des ménages et personnes, pour 1987 et 1998 (dispersion moyenne)

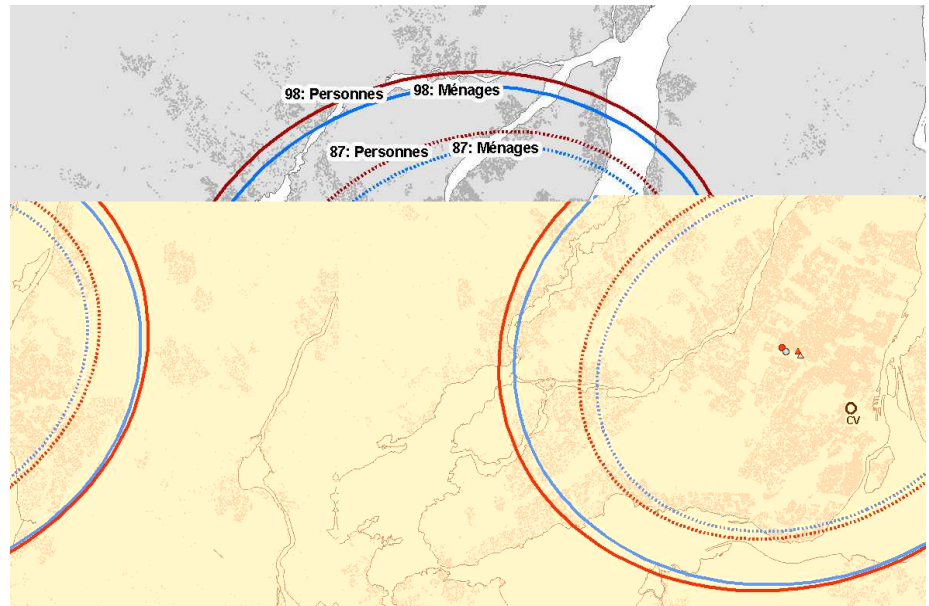
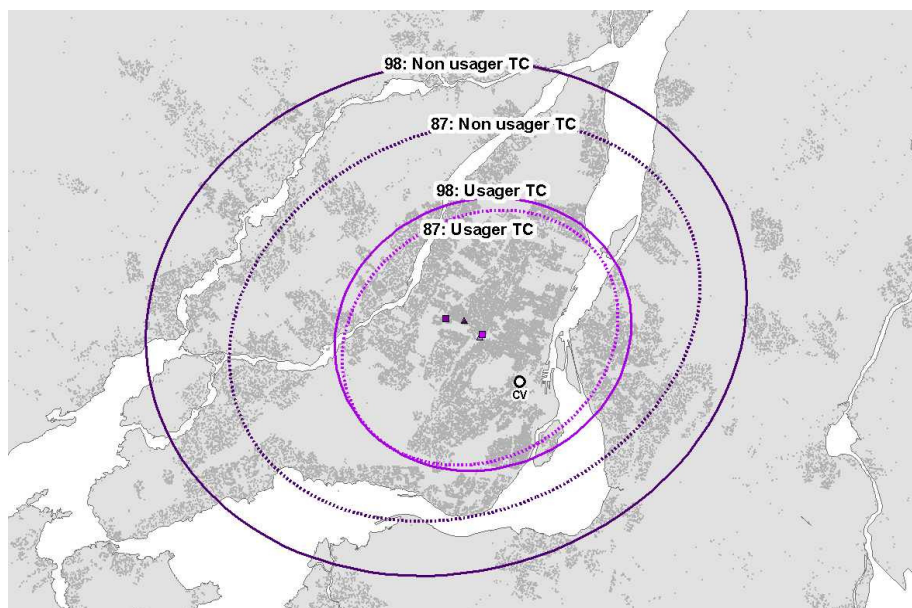


Figure 3 : Ellipses de dispersion et centres moyens des ménages appartenant à deux classes de taille : 1 personne et 4 personnes et plus (1987 et 1998)

Figure 4 : Ellipses de dispersion et centres moyens des usagers et non-usagers du réseau régional de transport en commun (1987 et 1998)



- Finalement, l'utilisation du transport en commun est liée à une concentration plus forte dans la portion centrale de l'Île de Montréal, là où les services de transport en commun sont les plus présents. Fait troublant, le segment des usagers du transport en commun est, de tous les segments de population, celui dont la dispersion est demeurée la plus stable dans le temps. Le marché du transport en commun ne semble donc pas suivre la vague de l'étalement urbain (Figure 4).

3.2. LES MESURES D'AUTOCORRÉLATION SPATIALE

L'autocorrélation spatiale est un concept fondamental dans l'étude des données spatiales et fait l'objet d'une littérature abondante (voir notamment CLIFF, ORD, 1973 ; CLIFF, ORD, 1981). La notion d'autocorrélation spatiale mesure essentiellement la ressemblance entre observations voisines et est, conséquemment, analogue à celle d'autocorrélation appliquée aux séries temporelles. Pour expliquer ce concept, ANSELIN (1999) parle de « *coïncidence de valeur avec la coïncidence de localisation* ». De multiples statistiques permettent de mesurer l'autocorrélation spatiale, les deux plus simples étant l'indice de MORAN (I) et l'indice de GEARY (C). Le premier est généralement considéré plus robuste et moins sensible aux variations locales (LEVINE, 2004). Ces indices permettent de valider l'existence d'autocorrélation spatiale pour une variable donnée, et ce pour l'ensemble des points.

Ces mesures sont expérimentées à l'aide de données agrégées portant sur les secteurs de dénombrement (SD), notamment sur les attributs des unités de

logement disponibles dans les recensements canadiens. La prise en compte de l'hétérogénéité de taille de ces secteurs (nombre d'unités de logement) s'effectue par la transformation des variables exprimant l'intensité (nombre d'unités de chaque type) en données de type intervalle (proportion d'unités de chaque type).

3.2.1. Autocorrélation globale

Indice de MORAN (I)

L'indice de MORAN est le ratio de la covariance⁵ entre deux observations (produit croisé des écarts à la moyenne), pondérée par une mesure d'éloignement entre ces observations, sur la variance totale de l'ensemble de données à l'étude.

L'indice de MORAN varie entre -1 et 1 : plus les points à proximité ont des valeurs similaires, plus la valeur de l'indice est élevée. En outre, l'appréciation plus juste de l'ampleur de l'autocorrélation spatiale s'appuie sur la moyenne théorique de l'indice. Lorsque l'indice calculé est supérieur à la moyenne théorique, les données présentent de l'autocorrélation spatiale positive alors qu'il y a autocorrélation négative (plus grande dissimilitude entre observations situées à proximité) lorsque l'indice calculé est inférieur à la moyenne théorique. Plus formellement, l'indice de MORAN s'estime à l'aide de l'équation suivante :

$$I = \frac{N \sum_i \sum_j W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{ij}) \sum_i (X_i - \bar{X})^2}$$

pour $j \neq i$; où N est le nombre d'observations (nombre de points), X_i la valeur d'une variable donnée au point i, X_j la valeur de la même variable au point j, W_{ij} un poids appliqué à la comparaison entre les deux localisations. Les poids entre paires d'observations peuvent prendre la forme d'une matrice d'adjacence où seules les localisations contiguës ont une valeur unitaire (les autres paires recevant une valeur nulle) ou d'une matrice de poids exprimant

les distances entre observations : $W_{ij} = \frac{1}{d_{ij}}$.

La moyenne théorique de l'indice correspond à $E(I) = -\frac{1}{N-1}$.

Indice de GEARY (C)

L'indice de GEARY est le ratio de la variance entre observations (déviations entre observations) sur la variance totale de l'ensemble de données. La valeur de l'indice de GEARY varie généralement entre 0 et 2 mais cette borne supé-

⁵ La covariance mesure la façon dont deux variables, X et Y, varient simultanément.

riure n'est pas stricte. La valeur théorique de cet indice est de 1, valeur indiquant l'absence d'autocorrélation spatiale. Les valeurs inférieures à 1 indiquent une autocorrélation spatiale positive alors que celles supérieures à 1 indiquent une autocorrélation négative entre observations. L'indice C de GEARY s'estime à l'aide de l'équation suivante :

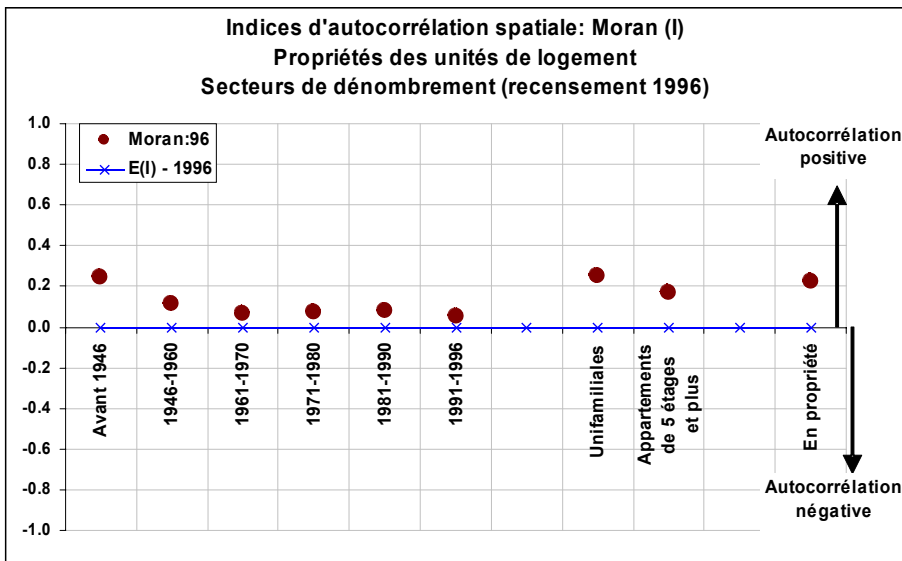
$$C = \frac{(N-1) \left[\sum_i \sum_j W_{ij} (X_i - X_j)^2 \right]}{2 \left(\sum_i \sum_j W_{ij} \right) \sum_i (X_i - \bar{X})^2} \text{ pour } j \neq i.$$

Illustration

Ces indices ont été estimés pour différents attributs des unités de logement, obtenus par secteur de dénombrement dans le recensement de 1996. La significativité statistique des indices estimés ci-dessous est appréciée par le biais d'un test de normalité qui compare la distribution empirique à une distribution normale (Z).

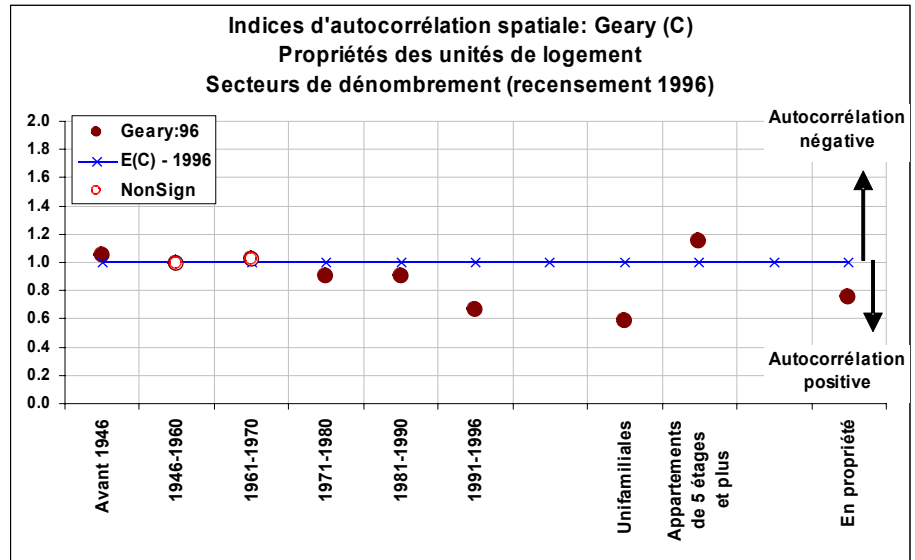
La Figure 5 présente les indices de MORAN estimés. Ces indices révèlent tous la présence d'autocorrélation spatiale positive (valeurs supérieures à la moyenne théorique E(I)) à des degrés toutefois faibles, les valeurs les plus élevées étant observables pour les unités de logement construites avant 1946, les unifamiliales (bungalow) ainsi que les unités en propriété. Ces indices sont tous hautement significatifs ($Z(I) > |10|$).

Figure 5 : Indices d'autocorrélation de MORAN (I) estimés pour différents attributs des unités de logement (secteurs de dénombrement : recensement de 1996)



Les indices de GEARY ont été estimés pour les mêmes propriétés (Figure 6). Deux de ces indices ne sont pas significatifs ($Z(C) < |2|$) à savoir ceux portant sur les unités de logement construites aux périodes 1946-1960 et 1961-1970. Par ailleurs, le portrait révélé par ces indices est quelque peu différent alors que deux des propriétés présenteraient de l'autocorrélation spatiale négative.

Figure 6 : Indices d'autocorrélation de GEARY (C) estimés pour différents attributs des unités de logement (secteurs de dénombrement : recensement de 1996)



Tel que précisé précédemment, l'indice de GEARY est plus sensible aux variations locales. Les constats différents produits par ces deux indices suggèrent l'existence de petites zones présentant de l'autocorrélation spatiale négative, pour les propriétés concernées (unités de logement construites avant 1946 et appartements dans des immeubles de 5 étages et plus), ce à quoi le second indice aurait été sensible. La carte choroplèthe de la proportion d'unités de logement construites avant 1946, présentée à la Figure 7, permet de mieux cerner cette différence et d'observer de petites zones, à forte ou faible incidence, enclavées dans des voisinages dissimilaires.

Les constats faits à l'aide d'indices globaux sont une invitation à explorer davantage les données afin de mieux cerner les différentes configurations spatiales locales. Il est en effet évident que les phénomènes urbains sont sujets à de multiples dynamiques locales ainsi qu'à des tendances spatiales de nature radio-concentrique. Les indicateurs locaux permettent d'approfondir ces questions.

Figure 7 : Carte choroplèthe des proportions d'unités de logement construites avant 1946 – données provenant du recensement 1996, agrégées en secteurs de dénombrement



Proportion d'unités construites avant 1946

3.2.2. Autocorrélation spatiale locale (*L-MORAN*)

Les indicateurs locaux d'association spatiale (LISA) sont discutés de façon notable par ANSELIN (1995). Celui-ci propose une implantation locale de l'indice d'autocorrélation spatiale de MORAN (*L-MORAN*) afin d'apprécier le degré de similarité par petits voisinages. Cette méthode requiert aussi que les données soient agrégées par zone; son application est donc appropriée aux données de recensement. La procédure calcule l'indice d'autocorrélation de MORAN pour chaque observation, dans ce cas-ci chaque secteur de dénombrement, en fonction des observations localisées à proximité (voisinage); ceci permet d'obtenir une mesure des similarités et différences locales.

L'indice de MORAN appliqué de façon locale s'estime comme suit :

$$I_i = \frac{(Z_i - \bar{Z})}{S_Z^2} * \sum_{j=1}^N [W_{ij} * (Z_j - \bar{Z})]$$

où \bar{Z} est la moyenne globale (pondérée), Z_i l'intensité observée au point i , Z_j l'intensité de chaque observation $j \neq i$, S_Z^2 la variance globale et W_{ij} un poids appliqué à la comparaison entre les deux localisations.

Un indice normalisé est ensuite approximé à partir des variances et indices locaux. De forts indices normalisés $Z(I_i)$ supérieurs et inférieurs à zéro

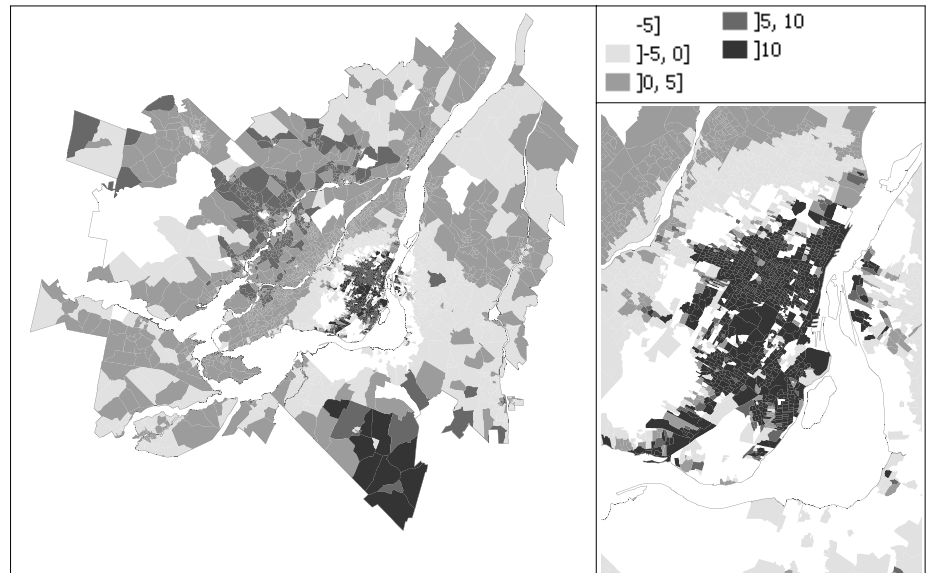
révèlent respectivement d'importantes similarités et différences.

Les indices locaux d'autocorrélation spatiale permettent de détecter des rassemblements de secteurs présentant une similarité particulière mais aussi de détecter les zones ségréguées, différentes de leur voisinage.

Illustration

La proportion d'unités de logement construites avant 1946, variable examinée précédemment, est de nouveau utilisée pour expérimenter cette mesure locale. Les cartes de la Figure 8 représentent les indices locaux (normalisés) de MORAN estimés pour les quelque 4 600 secteurs de dénombrement de la région de Montréal. Le degré d'autocorrélation spatiale est exprimé par le biais d'un dégradé de couleurs révélant, en pâle, des zones différentes (autocorrélation spatiale négative) et, en foncé, des zones similaires (autocorrélation spatiale positive). Plus la valeur est élevée, plus le degré d'autocorrélation spatiale est élevé.

*Figure 8 : Cartes choroplèthes des indices locaux d'autocorrélation spatiale des proportions d'unités de logement construites avant 1946 (recensement 1996)
Trame de couleur exprimant la ségrégation (très pâle) et la similarité (très foncé)*



Proportion d'unités construites avant 1946

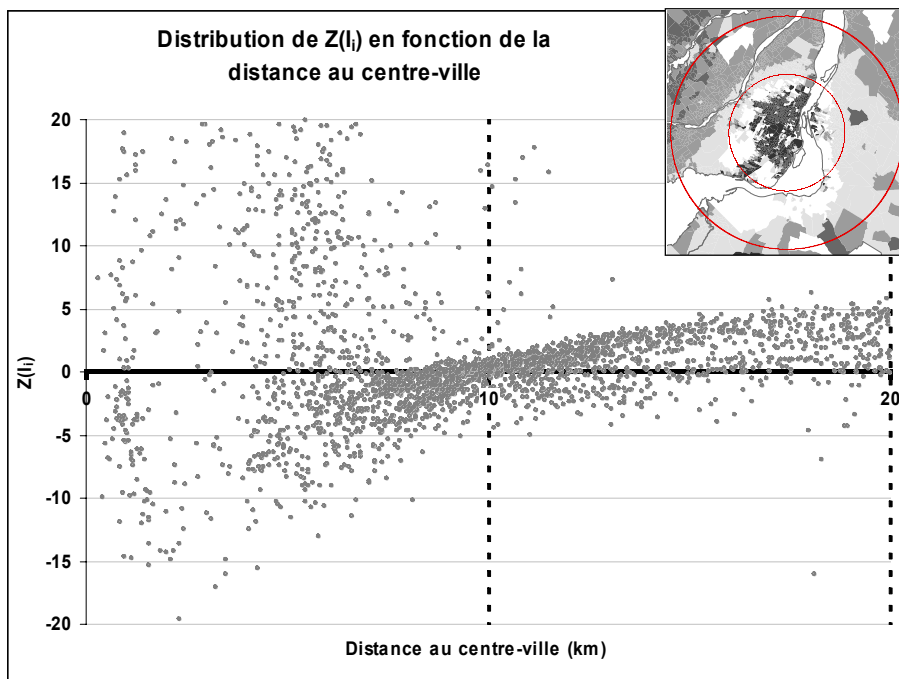
Combiné à la distribution spatiale de l'attribut à l'étude (voir la carte de la Figure 7), ces indices locaux permettent d'identifier les ensembles de secteurs similaires mais surtout les secteurs présentant une rupture en regard de leur voisinage. Alors que les secteurs similaires, sous la perspective de l'attribut à

l'étude, peuvent vraisemblablement être agrégés sans perte de variabilité, les secteurs dissimilaires sont riches en questionnement.

Dans le cas à l'étude (Figure 8), on observe une concentration des unités de logement construites avant 1946 dans les secteurs centraux ainsi qu'en certaines zones périphériques. Le noyau central est cerné par une bande de secteurs montrant de l'autocorrélation spatiale négative importante, ce qui suggère une mixité d'unités de logement construites à différentes périodes. En outre, au cœur même du noyau central sont détectées des secteurs très différents qui mériteraient, dans une perspective de modélisation, d'être examinés de façon distincte.

Un nuage de points mettant en relation l'indice local standardisé de MORAN et la distance au centre-ville, pour les secteurs de dénombrement localisés à moins de 20 km du centre-ville, permet d'identifier les secteurs ségrégués ainsi que l'anneau d'autocorrélation négative situé entre 5 et 10 kilomètres du centre-ville. Cette projection selon la distance au centre-ville est très fréquente dans l'étude des phénomènes urbains, plusieurs présentant une structure radioconcentrique.

*Figure 9 : Distribution de l'indice local normalisé de Moran ($Z(I_i)$) en fonction de la distance au centre-ville – indice mesuré pour la proportion d'unités de logement construites avant 1946
Données provenant du recensement canadien 1996*



3.3. LES MESURES DE CLASSIFICATION AUTOMATIQUE

Les mesures de classification automatique assistent aussi l'identification de structures locales. Cette démarche mène généralement à la construction d'agrégats (clusters) de données répondant à certains critères de proximité.

Ces techniques spatiales sont similaires à celles utilisées pour délimiter des groupes de comportement homogène. Dans ces cas, la similarité entre observations est évaluée à l'aide d'une distance composée des écarts entre différents attributs jugés significatifs (analyse factorielle, composante principale). La prise en compte de la spatialité des phénomènes implique des techniques similaires qui doivent alors intégrer la distance spatiale entre observations.

Il existe de nombreuses techniques pour classifier les observations spatiales en agrégats cohérents et révélateurs de similarité. Ces techniques peuvent être classées en catégories générales dont les plus fréquentes sont (LEVINE, 2004, LEBART et alii, 1979) :

- Les techniques de classification hiérarchique ou algorithmes ascendants qui procèdent à la construction des agrégats par agglomération successive des observations et qui fournissent une hiérarchie de classification (plusieurs générations de groupement) ;
- Les techniques de partitionnement ou d'agrégation autour de centres mobiles qui permettent de segmenter l'ensemble de données en un nombre spécifique (k) d'agrégats. Ces techniques diffèrent par les hypothèses d'initialisation des centres d'agrégats (valeurs références pour l'estimation des similarités/proximités) ainsi que par les critères d'optimisation des groupements.

3.3.1. Classification hiérarchique de voisinage

La classification hiérarchique de voisinage s'applique sur des ensembles de points bruts; elle ne tient donc pas compte de l'existence sémantique des données (poids). Cette limitation restreint a priori sa pertinence analytique en modélisation urbaine étant donnée la nature des données disponibles (agrégats, observations pondérées). Dans le cas des données d'enquêtes Origine-Destination, cette technique s'applique néanmoins de façon intéressante à des segments de l'échantillon de ménages ou personnes (observations non pondérées). À titre d'exemple, cette méthode a été appliquée à un des segments de population exploités à la section 3.1. à savoir celui des usagers du transport en commun.

Cette méthode groupe successivement les données en agrégats sur la base de deux critères, un premier relatif à la distance entre voisins, un second fixant le nombre minimal d'observations par agrégat. L'agrégation se poursuit jusqu'à ce que tous les points soient groupés en un seul agrégat ou jusqu'à ce que les critères échouent. Plusieurs générations d'agrégats peuvent donc être formées, des agrégats de génération précédente pouvant être groupés en un

nouvel agrégat si leur association répond au critère de distance.

Deux critères de classification doivent être spécifiés :

- Le critère de distance (maximale) entre voisins : cette distance seuil agit comme filtre d'admissibilité à l'agrégation ; seuls les points séparés d'une distance inférieure au seuil fixé sont plausibles d'être groupés.
- Le critère de nombre minimal d'observations par agrégat : ce critère spécifie le nombre minimal d'observations qu'un agrégat doit contenir pour être retenu comme solution d'une génération donnée et passer à l'itération suivante. Ce critère traduit quantitativement la perception de proximité/similarité non aléatoire de points en fixant arbitrairement le seuil au-delà duquel une situation mérite l'intérêt. Logiquement, la diminution de ce critère aura comme conséquence d'augmenter le nombre d'agrégats retenus.

La spécification des critères est du ressort de l'analyste et dépend directement de la nature des données ainsi que de l'objectif poursuivi lors de l'utilisation de la méthode. En outre, les données de population présentent une structure spatiale qui exprime déjà, en soi, la densité du développement urbain. Il ne s'agit donc pas d'une distribution aléatoire. Conséquemment, l'utilisation d'un seuil de distance constant pour l'ensemble du territoire implique nécessairement que des agrégats seront davantage générés là où la densité de personnes est supérieure. Ainsi, cette méthode de classification permet d'identifier les rassemblements importants de personnes présentant une similitude d'attributs, que ce soit dû à la forte densité de population ou à l'occurrence plus importante d'une certaine classe d'individus.

Il est cependant possible de tenir compte d'une distribution spatiale de référence afin de ne faire ressortir que les zones où la concentration est supérieure à ce qui serait attendu de cette configuration spatiale. Une des approches possibles est de faire varier le seuil de distance, dans l'espace, en fonction de la densité de référence. Des détails concernant l'application de cette méthode peuvent être trouvés dans LEVINE (2004).

Illustration

La routine de classification hiérarchique de voisinage, avec et sans prise en compte de la densité de population, a été utilisée afin d'identifier les zones de forte incidence en termes d'usagers du transport en commun. Ce sous-échantillon compte 24 418 observations. Le Tableau 2 présente les résultats inhérents. Tel qu'attendu, la proportion des observations qui se retrouvent, ultimement, dans un agrégat (dernière génération) est nettement moindre lorsqu'une configuration spatiale de référence, à savoir la densité de population, est considérée (37,2 % vs 11,7 %). Par ailleurs, le nombre d'agrégats de première génération est aussi moins élevé dans le second cas puisque le critère de distance fluctue avec la densité de population, devenant plus sévère dans les zones plus denses et plus souple dans les zones moins peuplées. La

représentation spatiale des agrégats formés à l'aide des deux approches permet d'observer, de façon évidente, ces différences (voir Figure 10). La spatialisation simultanée des agrégats formés et du réseau de métro montréalais confirme par ailleurs l'effet structurant de cette infrastructure sur le niveau d'utilisation du système de transport en commun.

*Tableau 10 : Comparaison entre les résultats de la classification hiérarchique de voisinage appliquée au sous-ensemble des usagers du transport en commun, avec et sans prise en compte de la densité de population
Données provenant de l'enquête Origine-Destination de 1998*

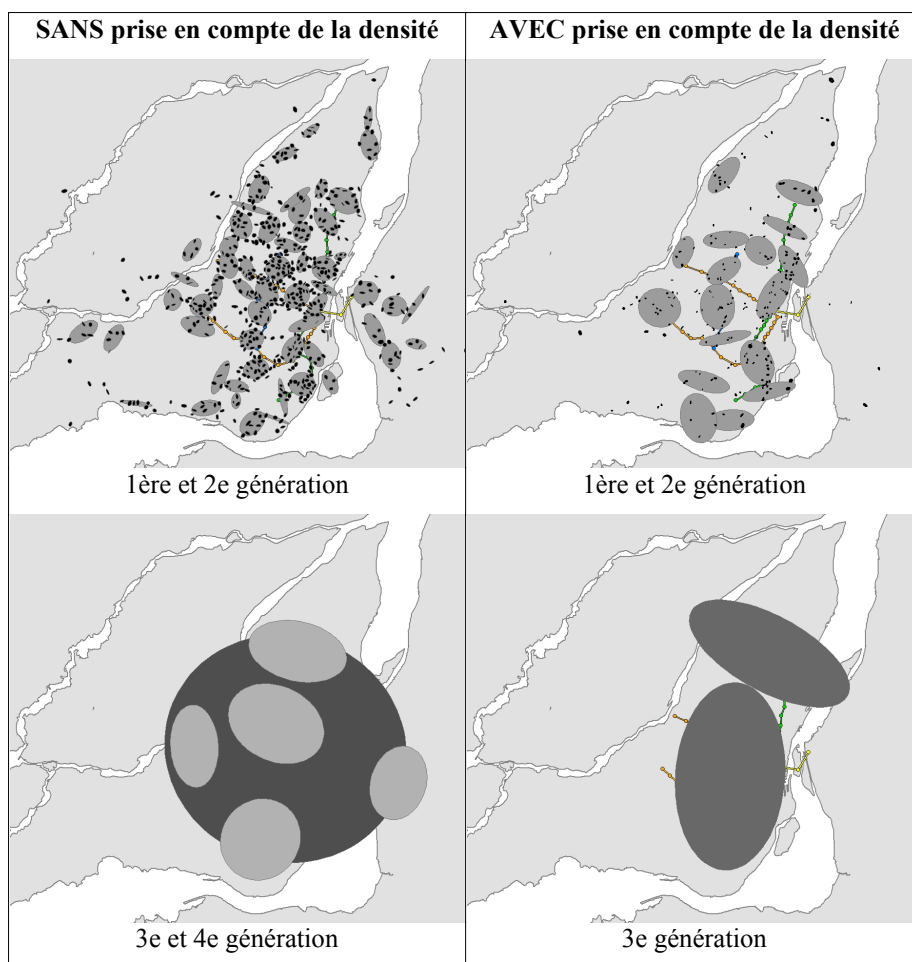
Usagers du transport en commun (OD 1998)				
24 419 observations				
	Génération d'agrégats			
	1	2	3	4
Nombre d'agrégats				
SANS densité	617	57	5	1
AVEC densité	222	17	2	
Proportion des observations appartenant aux agrégats				
SANS densité	52.6%	47.1%	37.2%	37.2%
AVEC densité	13.0%	11.7%	11.7%	

3.3.2. Classification par k-moyennes

Les méthodes de classification descendante permettent de partitionner un ensemble de points en un nombre prédéterminé d'agrégats. Similairement à la méthode précédente, un point ne peut appartenir qu'à un seul agrégat. Cependant, la classification autour de centres mobiles génère une partition concernant tous les points de l'ensemble.

Cette méthode est particulièrement intéressante lorsque l'analyse requiert ou se prête à un certain contrôle du nombre de groupes à construire puisque le choix préalable du nombre d'agrégats est arbitraire : identifier la localisation optimale d'un nombre fixe d'équipements urbains par exemple. Toute routine de classification par k-moyennes exige l'initialisation (spatiale) des agrégats puis procède à l'optimisation du partitionnement par minimisation de la variance intragroupe. La nécessité de fixer a priori un nombre d'agrégats est aussi la plus grande limitation de cette approche puisqu'il est toujours possible que le nombre d'agrégats soit inapproprié pour le phénomène à l'étude et qu'il suggère des patrons artificiels. Cette approche peut néanmoins assister la synthèse d'ensemble de points à différents niveaux de résolution par l'application successive de la méthode avec différents paramètres d'initialisation.

Figure 10 : Agrégats formés lors de la classification hiérarchique de voisinage avec et sans prise en compte de la densité de population - Echantillon des usagers du transport en commun - Données de l'enquête Origine-Destination 1998



4. POINT FINAL...?

Bien qu'apparemment en marge d'un exercice classique de modélisation des transports, les méthodes d'analyse spatiale expriment pourtant les perspectives fondamentales vers lesquelles tend l'exploitation de microdonnées urbaines. En effet, l'état de la pratique et de la recherche confirme que plusieurs efforts s'activent à perpétuer l'idée qu'un modèle mathématique, fondé sur une réduction (parfois artificielle) de la réalité, permettra de prévoir les multiples facettes de la dynamique urbaine. En parallèle, les constats de complexité, incertitude, variabilité, détermination dynamique se multiplient.

Les moyens technologiques maintenant disponibles, tels que systèmes

d'information géographique et outils d'analyse spatiale, permettent pourtant d'approfondir réellement la compréhension des phénomènes observables sans devoir recourir immédiatement à des techniques de simplification masquant les données. Le rôle de la modélisation, dans un contexte informationnel où la réalité se traduit de plus en plus en données, est de révéler l'information contenue dans ces données afin d'assister le professionnel dans son travail d'analyse. « *The greatest value of a picture is when it forces us to notice what we never expected to see* » (TUKEY, 1977). Il ne s'agit donc pas de substituer un modèle à l'analyste mais de l'instrumenter correctement en lui donnant des moyens interactifs pour assister sa réflexion.

Techniquement, les expérimentations dont cet article fait état, illustrent le type de perspectives pouvant être construites à l'aide de microdonnées examinées à l'aide de techniques spatiales simples. L'applicabilité de ces méthodes est donc confirmée et déjà, il semble que celles-ci pourraient s'intégrer dans le cycle de production et d'analyse des données de transport.

Sommairement,

- Les mesures de centralité et dispersion permettent de comparer rapidement des nuages de points propres à des segments clés de la population. L'effet discriminant ou non de l'espace peut être observé.
- Les mesures d'autocorrélation spatiale, appliquées localement, permettent de cerner les secteurs similaires et dissimilaires. Le caractère distinctif ou non d'un secteur en regard de son voisinage peut faciliter la formulation de modèles, surtout dans le choix d'un découpage spatial. L'identification, a priori, des secteurs différents peut faciliter le travail d'analyse corrélative.
- Les mesures de classification automatique permettent d'abord d'identifier des zones de forte concentration d'observations puis de raffiner cette identification par la prise en compte d'une structure spatiale de référence afin de n'identifier que les zones où l'occurrence dépasse celle qui serait normalement attendue étant donnée cette structure spatiale.

Finalement, ces méthodes viennent souvent confirmer des relations très intuitives ou plus simplement observables à l'aide de modèles classiques basés sur des données agrégées. Face à un tel constat, quiconque est en droit de se questionner sur l'utilité et la pertinence de ces méthodes, voire même des microdonnées spatiales qui en permettent l'application. Cependant, la richesse du détail spatial semble de plus en plus faire consensus et ce, dans plusieurs champs d'application. En outre, les différents moyens déployés pour exploiter les ensembles de données spatiales confirment le besoin et la nécessité de structurer l'usage qui en est fait afin de favoriser l'émergence de nouvelles compréhensions des comportements urbains. Il est de notre avis que ces méthodes spatiales, intégrées dans une démarche exploratoire, ont la

capacité d'aller au-delà des grandes tendances spatiales bien connues en identifiant des structures locales, fournissant ainsi l'information nécessaire pour intervenir plus efficacement là où des dynamiques spécifiques l'exigent.

REMERCIEMENTS : L'auteur tient à remercier les institutions montréalaises responsables de la poursuite des exercices de collecte de données que sont les enquêtes Origine-Destination. Ces exercices de grande envergure permettent d'entretenir, dans la grande région de Montréal, une culture d'analyse et d'observation de la mobilité réelle : STM, RTL, STL, AMT et MTQ.

L'auteur remercie par ailleurs le Pr. Robert CHAPLEAU pour sa contribution générale à l'avancement des méthodes d'analyse des réseaux de transport ainsi que pour ses contributions spécifiques dans l'application des méthodes spatiales et des systèmes d'information géographique.

L'auteur souligne finalement la contribution des organismes subventionnaires FCAR (Fonds pour les Chercheurs et l'Aide à la recherche) et CRSNG (Conseil de recherches en sciences naturelles et en génie du Canada).

RÉFÉRENCES

ANSELIN L. (1995) Local Indicators of Spatial Association – LISA. **Geographical Analysis**, Vol. 27, pp. 93-115.

ANSELIN L. (1999) **Spatial econometrics**. 32 p.
(http://www.csiss.org/learning_r.pdf).

ANSELIN L., SYABRI I., KHO Y. (2006) GeoDa: An Introduction to Spatial Data Analysis. **Geographical Analysis**, Vol. 38, pp. 5-32
(<http://sal.uiuc.edu/pdf/geodaGA.pdf>).

BANOS A. (2001a) **Le lieu, le moment, le mouvement : pour une exploration spatio-temporelle désagrégée de la demande de transport en commun en milieu urbain**. Université de Franche-Comté, Besançon, 356 p. (Thèse de Géographie).

BANOS A. (2001b). A propos de l'analyse spatiale exploratoire des données. **Cybergeog**, n° 197 (<http://www.cybergeog.presse.fr/modelis/banos/article.htm>).

BAILEY T.C., GATRELL A.C. (1995) **Interactive Spatial Data Analysis**. Harlow, Longman Scientific & Technical, 436 p.

CLIFF A.D., ORD J.K. (1973) **Spatial Autocorrelation**. Londres, Pion.

CLIFF A.D., ORD J.K. (1981) **Spatial Processes: Models and Applications**. Londres, Pion.

CRESSIE N. (1993) **Statistics for Spatial Data**. NY, Wiley (Revised edition).

HYPERGEO **Encyclopédie électronique consacrée à l'épistémologie de la géographie** (<http://hypergeo.free.fr/>).

- JAYET H. (2001) Économétrie et données spatiales : une introduction à la pratique. **Cahiers d'Economie et Sociologie Rurales**, n° 58-59, pp. 105-129.
- LEBART L., MORINEAU A., FÉNELON J.-P. (1979) **Traitement des données statistiques. Méthodes et programmes**. Paris, Dunod, Bordas, 512 p.
- LEVINE N. (2004) **CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations**. Houston/Washington DC, Ned Levine & Associates/the National Institute of Justice.
- MILLER E.J., DOUGLAS-HUNT J., ABRAHAM J.E., SALVINI P.A. (2004) Micro-simulating urban systems. **Computers, Environment and Urban Systems**, Vol. 28, pp. 9-44.
- MILLER H. J. (2003) What about people in geographic information science? Editorial. **Computers, Environment and Urban Systems**, Vol. 27, pp. 447-453.
- MORENCY C. (2004) **Contributions à la modélisation totalement désagrégée des interactions entre dynamiques spatiales et mobilité urbaine**. École Polytechnique de Montréal, 405 p. (projet de thèse de doctorat).
- SANTÉ CANADA (2004) **Guide canadien d'évaluation des incidences sur la santé**. 160 p.
(http://www.hc-sc.gc.ca/ewh-semt/pubs/eval/handbook-guide/vol_1/index_f.html).
- SHAROV A. A. (1996) **Quantitative Population Ecology**. Virginia Tech, On-line lecture (<http://www.ento.vt.edu/~sharov/PopEcol/>).
- TIMMERMANS H. (2003) The Saga of Integrated Land Use-Transport Modeling: How Many More Dreams Before We Wake Up? Resource paper, presented at **the 10th International Conference on Travel Behaviour Research**, Lucerne, Suisse.
- TOBLER W. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography**, Vol. 46, n° 2, pp. 234-240.
- TUKEY J. (1977) **Exploratory Data Analysis**. Addison-Wesley, 691 p.