

## OASIS - un Outil d'Analyse de la Ségrégation et des Inégalités Spatiales

Mihaï Tivadar<sup>1,2</sup>, Yves Schaeffer<sup>1</sup>, André Torre<sup>1</sup>, Frédéric Bray<sup>1</sup>

<sup>1</sup>Irstea, UR DTGR, 2 rue de la Papeterie-BP 76, F-38402 Saint-Martin-d'Hères, France

<sup>2</sup> auteur correspondant : mihai.tivadar@irstea.fr

### Résumé :

OASIS (Outil d'Analyse de la Ségrégation et des Inégalités Spatiales) est une interface internet automatisée qui offre un large éventail d'outils statistiques et cartographiques pour l'analyse des variables spatialisées. Par rapport aux logiciels existants, OASIS présente plusieurs avantages. Premièrement, c'est une plateforme complète, qui propose des outils nouveaux et regroupe une palette très large d'outils standards, actuellement dispersés sur plusieurs logiciels ou interfaces. Deuxièmement, OASIS a un caractère automatisé et donne accès à une série de variables utilisées fréquemment dans la littérature sur la ségrégation et les inégalités spatiales, qui sont disponibles au niveau communal sur l'ensemble de la France Métropolitaine (fonds de cartes inclus). L'originalité méthodologique est apportée par des simulations de Monte Carlo, qui permettent de tester la significativité statistique de l'ensemble des indices de ségrégation et d'autocorrélation spatiale.

### 1. INTRODUCTION

OASIS (Outil d'Analyse de la Ségrégation et des Inégalités Spatiales) est une interface internet automatisée disponible sur <http://oasis.irstea.fr/> qui offre un large éventail d'outils statistiques et cartographiques pour l'analyse des variables spatialisées. Il est destiné à un public très large, d'un niveau basique en statistique jusqu'à un niveau élevé : étudiants, autorités locales, chercheurs, etc.

La plateforme web OASIS a été configurée sous WordPress, tandis que les modules de calcul ont été programmés en php, en utilisant une base de données PostgreSQL9/postGIS1.5, le langage procédural plr, le logiciel R pour les calculs statistiques et son package spdep pour certains indices d'autocorrélation spatiale. Les modules de cartographie dynamique ont été programmés avec le langage php-mapscript pour mapserveur 6.

Par rapport aux logiciels existants, OASIS présente plusieurs avantages. Premièrement, c'est une plateforme complète, qui propose des outils nouveaux et regroupe une palette très large d'outils standards, actuellement dispersés sur plusieurs logiciels ou interfaces : applications pour le calcul des indices de ségrégation (Wong et Chong, 1998 ; Apparicio, 2000 ; Wong, 2003 ; Apparicio, 2008), logiciels ou modules de statistiques spatiales (PySAL, GeoDa, packages R, modules GIS, etc.), GIS ou interfaces cartographiques (MapInfo, ArcGIS, outils en ligne de la Datar et de l'Insee, etc.). Au-delà de l'intégration et l'automatisation de ces outils standards, OASIS apporte une originalité méthodologique à travers un outil de simulation Monte Carlo, qui permet de tester la significativité statistique de l'ensemble des indices de ségrégation et d'autocorrélation spatiale. Enfin, OASIS donne accès à une série de variables utilisées fréquemment dans la littérature sur la ségrégation et les inégalités spatiales (population, ménages, logements, famille, etc.), qui sont disponibles au niveau communal sur l'ensemble de la France Métropolitaine (fonds de cartes inclus).

Les outils proposés sont regroupés par thèmes. Après les analyses statistiques descriptives, OASIS dispose d'un puissant module cartographique qui réalise des cartes thématiques des variables analysées. Ces cartes sont entièrement personnalisables : aspect, palette de couleurs, nombre de classes et méthodologie de calcul des seuils de discrétisation, couches supplémentaires, etc.

Les *indices de ségrégation* (intragroupes, intergroupes et multigroupes) sont analysés par dimension de la ségrégation : l'égalité, l'exposition, la concentration, le regroupement et la centralité. Les résultats sont présentés sous forme de tables et de représentations graphiques inédites qui permettent

d'avoir une image rapide et claire des interactions entre variables. Le module « cartographie de la ségrégation » calcule et représente par cartes thématique les indices de ségrégation locale.

Les *analyses d'autocorrélation* spatiale incluent des indices statistiques et leurs tests standards de significativité. Les nuages de points et les droites de régression sur lesquelles reposent ces tests sont représentés graphiquement. Le module « cartographie de l'autocorrélation » établit les cartes des indices d'autocorrélation locale significatifs en fonction de leur type : autocorrélation positive ou négative.

Les derniers outils sont les *tests de Monte Carlo* des indices d'autocorrélation spatiale et des indices de ségrégation. Deux types de simulations sont faites. Le test des permutations permet de déterminer si la composante spatiale des indices est significative. Pour les indices non-spatialisés, les valeurs simulées sont identiques à l'indice calculé. Pour les autres, la composante spatiale est significative et doit être prise en compte lorsque l'indice calculé se trouve dans la zone de significativité de la distribution simulée.

Les tests des localisations aléatoires permettent de déterminer la significativité de l'indice par rapport à une distribution aléatoire. Les localisations sont simulées sans contrainte (probabilité identique pour chaque unité spatiale) ou avec contraintes spatiales (probabilités déterminées par différents critères : la superficie totale des unités spatiales, ou leur superficie artificialisée). En comparant la distribution des valeurs simulées avec la valeur calculée de l'indice, nous pouvons conclure qu'il est le résultat d'un processus aléatoire ou d'un schéma de localisation déterminé. Si l'indice calculé se trouve à l'intérieur de la distribution (au-dessous du seuil de significativité), il n'est pas significatif. Ces tests résolvent certaines incohérences entre les dimensions de la ségrégation spatiale. La comparaison entre les valeurs des indices associés à ces dimensions soulève généralement des contradictions (par exemple concernant le groupe le plus ségrégué dans un territoire, distinct selon que l'on considère l'égalité ou l'exposition). Ces contradictions disparaissent très souvent lorsqu'on ne tient compte que des indices significatifs selon les tests de Monte Carlo.

## 2. INTRODUCTION DES DONNEES ET ANALYSES DESCRIPTIVES

Lors de l'activation du menu Outils, vous êtes invités à introduire dans un premier temps la zone d'étude. Dans la version actuelle de la plateforme, l'échelle des unités spatiales est communale et couvre la France métropolitaine. Vous avez la possibilité de choisir entre deux modalités : par import de fichier ou par sélection manuelle. Lors de la sélection manuelle, à partir d'une liste déroulante, vous avez le choix entre départements entiers, parties des départements, grands pôles d'aires urbaines et grands pôles et couronnes d'aires urbaines. La plateforme permet de faire des sélections multiples en utilisant un ou plusieurs critères de sélection.

Une fois la sélection de la zone d'étude validée, vous passez à la deuxième étape : l'introduction des données. Tout d'abord, on vous demande d'indiquer une commune-centre (nécessaire pour le calcul de certains indices de ségrégation) dans une liste déroulante, triée en ordre décroissante par rapport à la population. Nous vous proposons ensuite une liste de variables d'étude, regroupées par thème (population, ménages, logements, famille, etc.). Les variables sont issues des enquêtes de recensement de la population, à partir de 1990. Pour certaines variables, nous vous proposons également une agrégation, afin d'homogénéiser les données (en termes de structure) et permettre des comparaisons temporelles. Une fois le groupe de variables sélectionné, une liste des variables vous est proposée à droite de la fenêtre, avec différentes années disponibles. Pour chaque année, vous avez le choix de garder l'ensemble des données ou seulement celles qui vous intéressent. La version actuelle ne permet pas l'importation des données propres. Cette composante sera introduite ultérieurement dans la plateforme. L'outil dispose également d'une option « Restitution des données », qui permet de visualiser et sauvegarder les données de la zone d'étude sélectionnée.

Les analyses descriptives sont regroupées en deux types d'outils. Premièrement, OASIS permet de calculer et exporter diverses statistiques descriptives, pour les données en valeurs absolues ou relatives. La fenêtre d'affichage présente une série de résultats principaux, tandis que l'enregistrement

des résultats sur le disque dur permet d'accéder à la liste complète des indicateurs disponibles. Cet outil fournit des analyses habituelles sous forme de tables et de graphiques (indicateurs de position : min, max, moyenne, quartiles, BoxPlot ; indicateurs de dispersion : variance, écart-type, coefficient de variation, coefficients de corrélation).

Ensuite, le module cartographique réalise des cartes thématiques des variables analysées. Ces cartes sont personnalisables : aspect, palette de couleurs, nombre de classes et méthodologie de calcul des seuils de discrétisation, couches supplémentaires, etc.

### 3. ANALYSE DE LA SEGREGATION SPATIALE

Les indices de ségrégation permettent de mesurer la ségrégation spatiale d'une population formée de plusieurs classes ou groupes. Ils s'appliquent dans des domaines très divers, tels que l'économie, la sociologie, la biologie, l'écologie, etc...

Les indices de ségrégation sont aujourd'hui de plus en plus utilisés, notamment en France. Dans un article de langue française souvent cité, Apparicio (2000) propose une présentation synthétique de leurs origines et de leurs principes. Les recherches sur la ségrégation trouvent leur commencement aux Etats-Unis, où le thème de la concentration des groupes ethniques est abordé dès les années 1920. Mais c'est à partir des années 1940 qu'une série de travaux quantitatifs débouche sur un ensemble d'indicateurs de la ségrégation sociale. Parmi ces indicateurs, on trouve les classiques indices de dissimilitude et delta de Duncan (1955) et les indices d'interaction de Bell (1954). Ultérieurement, dans les années 1980-1990, d'autres chercheurs américains (Morgan, White, Morill et Wong) développent de nouveaux indices intégrant la dimension spatiale dans la mesure de la ségrégation. Ces indices a-spatiaux et spatiaux, initialement conçus pour étudier la ségrégation d'un seul ou de deux groupes de population, ont été généralisés pour permettre l'étude la ségrégation multi-groupes (Reardon et Firebaugh, 2002 ; Reardon et O'Sullivan, 2004).

Dans un article canonique de cette littérature, Massey et Denton (1988) regroupent les manifestations spatiales de la ségrégation en cinq dimensions bien distinctes : l'égalité, l'exposition, la concentration, le regroupement (ou agrégation spatiale) et la centralisation.

**L'égalité** renvoie à la distribution d'un ou de plusieurs groupes de la population à travers les unités spatiales de la zone d'étude : plus un groupe est inégalement réparti entre ces unités spatiales (surreprésenté dans certaines et sous-représenté dans d'autres), plus il est ségrégué. **L'exposition** mesure la possibilité d'interactions (plus exactement la probabilité de partager une même unité spatiale) entre les membres d'un même groupe de la population ou entre les membres de deux groupes distincts : plus elle est forte, plus le groupe est ségrégué. **La concentration** fait référence à l'espace (la superficie) occupé par un groupe de la population : plus il occupe une partie faible de la zone étudiée, plus il est concentré, donc ségrégué. **Le regroupement** évalue le degré selon lequel les membres d'un groupe de la population occupent des unités spatiales contiguës (voisines) : plus ils sont regroupés, et forment ainsi une enclave au sein la zone d'étude, plus ce groupe est ségrégué. **La centralisation** correspond au niveau de proximité d'un groupe de la population au centre de la zone étudiée : plus ce groupe est localisé près du centre-ville, plus il est ségrégué.

Pour chaque dimension, on distingue (i) les indices **intragroupes**, qui mesurent la répartition d'un groupe par rapport à l'ensemble de la population, (ii) les indices **intergroupes**, qui comparent la répartition d'un groupe à celle d'un autre groupe et (iii) les indices **multigroupes**, qui mesurent le niveau de ségrégation « global » de l'ensemble des groupes qui composent la population.

A côté de ces indices, qui mesurent le niveau de ségrégation spatiale sur l'ensemble de la zone d'étude, les indices de **ségrégation locale** permettent d'avoir une image spatialisée de la ségrégation à l'aide des cartes thématiques.

#### 3.1. L'égalité

Parmi les indices de ségrégation, les indices d'égalité sont les plus souvent utilisés dans la littérature. Ils permettent d'analyser la distribution spatiale d'un ou plusieurs groupes de population. Dans le cas où un seul groupe est étudié, l'indice d'égalité est une mesure de la surreprésentation ou la sous-représentation de ce groupe dans l'espace par rapport à la population totale. Lorsque plusieurs groupes sont considérés, il s'agit d'une mesure de la séparation spatiale entre ces groupes.

Généralement, les indices ont une valeur comprise entre 0 et 1 et expriment ainsi la part du groupe qui devrait déménager pour obtenir une distribution parfaitement proportionnelle à la distribution spatiale de la population totale (ou à celle des autres groupes étudiés). Le calcul de ces indices est reporté dans les Annexes.

#### *a) Les mesures d'égalité intragroupes*

Les indices d'égalité intragroupes mesurent le niveau de ségrégation d'une catégorie à la fois, par rapport au reste de la population, sans s'intéresser aux éventuelles interactions entre classes.

L'indice d'égalité le plus célèbre est *l'indice de ségrégation de Duncan* (1955). Il varie entre 0 (distribution parfaitement proportionnelle d'un groupe par rapport à la population totale) et 1 (distribution ségrégative maximale). La valeur de l'indice exprime la part de la catégorie qui devrait déménager afin d'obtenir une distribution spatiale parfaitement identique à celle de la population totale. L'indice de Duncan n'a pas de dimension spatiale : toute permutation de la population entre les unités spatiales donne le même niveau de ségrégation.

Pour résoudre cette limite, plusieurs indices « spatiaux » ont été proposés dans la littérature. *L'indice de Morrill* (1991) prend en compte les interactions spatiales en intégrant le voisinage entre les unités spatiales : si la configuration spatiale fournit des opportunités aux membres d'une catégorie d'interagir avec les membres de leur groupe localisés dans les unités spatiales adjacentes (par exemple lorsque les unités spatiales où ce groupe est sous-représenté sont voisines de celles où il est surreprésenté), la valeur de l'indice de ségrégation doit être diminuée. Il soustrait ainsi aux indices de ségrégation de Duncan cette possibilité d'interaction en utilisant une matrice de contiguïté. Les éléments de la matrice de contiguïté ont la valeur 1 si deux unités sont contiguës et 0 si elles ne sont pas adjacentes. L'indice de Morrill varie également entre 0 et 1.

*L'indice de Wong* (1993) est assez semblable à l'indice de Morrill, la différence résidant dans le calcul de l'opportunité d'interactions interzonales. Wong considère que le voisinage entre unités spatiales est insuffisant et que les éventuelles interactions spatiales sont mieux décrites en tenant compte de la frontière commune entre les unités adjacentes (une frontière commune plus importante implique des interactions plus fortes), ainsi que de la taille et la forme de chacune d'elles (plus une unité spatiale est petite et compacte, plus la possibilité d'interaction est forte).

Un indice alternatif à ceux calculés sur la base de l'indice de ségrégation de Duncan a été proposé par *Gorard* (2000), dans le but de résoudre le problème de la sensibilité de ces derniers à la taille de la population : si on modifie tous les effectifs de la population de manière proportionnelle, le niveau de ségrégation devrait rester le même. Cet indice s'interprète toujours comme la part d'individus appartenant à un groupe qu'il faut déplacer pour atteindre une distribution spatiale égalitaire.

La ségrégation est considérée comme « une forme d'inégalité », or les premières mesures de l'inégalité datent du début du vingtième siècle, avec *l'indice de Gini* (1921). Cet indice est déduit de la courbe de Lorenz et représente la surface comprise entre la courbe de concentration et la diagonale d'équirépartition. Cet indice a donc été utilisé pour la mesure de la ségrégation. L'indice varie entre 0 (ségrégation nulle) et 1 (ségrégation totale).

#### *b) Les mesures d'égalité intergroupes*

*L'indice de dissimilarité* (Duncan, 1955) compare les distributions de deux groupes entre les unités spatiales et mesure leur séparation spatiale. Sa valeur exprime la part d'une des deux catégories qui

devrait déménager pour obtenir des distributions identiques. L'indice varie entre 0 (similitude spatiale parfaite) et 1 (dissemblance spatiale maximale).

### 3.2. L'exposition

L'exposition mesure la possibilité d'interactions entre les membres d'une même catégorie (indices intragroupes) ou entre les membres de deux catégories distinctes (indices intergroupes). Les premiers indices d'exposition ont été proposés par Bell (1954).

#### a) *Les mesures d'exposition intragroupes*

*L'indice d'isolement* (Bell, 1954) varie de 0 à 1 et mesure la probabilité qu'un individu partage la même unité spatiale avec un membre de son propre groupe. Une valeur nulle signifie qu'un individu ne partage jamais la même unité spatiale avec les individus de son propre groupe. Au contraire, une valeur unitaire signifie que les membres d'un groupe sont localisés dans des unités spatiales où l'on ne trouve que des membres de leur propre groupe. Ainsi, une valeur plus élevée de l'indice d'exposition implique un niveau de ségrégation plus élevé : les membres d'une catégorie ont la tendance à s'isoler par rapport au reste de la population. Par exemple, une valeur de 0,25 signifie qu'en moyenne, un individu est localisé dans des unités spatiales où un quart de la population fait partie de son propre groupe. L'indice d'isolement est influencé par la structure de la population : un individu membre d'un groupe majoritaire aura une probabilité plus importante de partager les mêmes unités spatiales que les membres de son propre groupe qu'un individu membre d'un groupe minoritaire.

*L'indice d'isolement ajusté* modifie l'indice d'isolement en tenant compte de la proportion du groupe dans la zone d'étude, de manière à éviter les effets liés à la composition de la population (Massey et Denton, 1988; Streans et Logan, 1986).

#### b) *Les mesures d'exposition intergroupes*

*L'indice d'interaction* (Bell, 1954) est semblable à l'indice d'isolement, la différence résidant dans le fait qu'il mesure la probabilité qu'un individu appartenant à un groupe partage la même unité spatiale avec un membre d'un second groupe. Comme l'indice d'isolement, il varie entre 0 et 1 mais son interprétation est opposée : une valeur plus élevée implique une intégration plus importante entre les deux catégories (une probabilité plus grande de partager la même unité spatiale). L'indice d'interaction est également influencé par la structure de la population : un individu d'un groupe sera confronté à une probabilité plus élevée de partager la même unité spatiale avec les membres du groupe majoritaire.

### 3.3. La concentration

Une autre forme de ségrégation est décrite par la concentration spatiale, c'est-à-dire la superficie occupée par une catégorie de population. Plus une catégorie occupe une faible partie de la superficie totale de la zone d'étude, plus elle est concentrée dans l'espace. Les indices de concentration ont des valeurs comprises entre 0 (concentration spatiale minimale) et 1 (concentration spatiale maximale).

#### a) *Les mesures de concentration intragroupes*

*L'indice Delta* de (Duncan et al, 1961) est une mesure de concentration intragroupe qui est calculée comme la moitié de la somme des différences (en valeurs absolues) entre les parts relatives de la population d'une catégorie dans chaque unité spatiale (par rapport à la population totale de la catégorie respective) et la part de la superficie de l'unité spatiale par rapport à la zone d'étude. Cette somme d'écart entre les proportions des populations et de superficie entre unité spatiale est divisée par deux, parce qu'un écart négatif dans une unité spatiale implique un écart positif dans une ou plusieurs autres unités spatiales (comptage double). Si la distribution spatiale de la catégorie correspond parfaitement à la répartition des superficies des unités spatiales, la valeur de l'indice est égale à zéro (pas de

concentration). Plus il y a des différences entre les deux distributions, plus le niveau de concentration spatiale augmente. L'indice est interprété comme la proportion de la catégorie qui devrait délocaliser afin d'obtenir une densité uniforme.

Lors du calcul de l'indice Delta, les superficies totales des unités spatiales sont prises en compte. Or, une partie importante du territoire des unités spatiales est souvent inhabitable et ainsi les densités calculées ne sont représentatives. Afin d'éviter ce problème, nous proposons dans OASIS une version adaptée de l'indice, où ne sont prises en compte que les superficies artificialisées (d'après les données Corine Land Cover).

Massey et Denton (1988) proposent *l'indice de concentration absolue*, plus complexe que le précédent. Il compare la superficie totale habitée par une certaine catégorie avec le minimum et le maximum d'unités spatiales où pourraient résider les membres de la catégorie dans les cas de concentrations maximales et minimales. L'indice a des valeurs comprises entre 0 et 1, qui correspondent à une concentration spatiale minimale et une concentration maximale. Comme pour l'indice Delta, nous avons développé dans OASIS une version de l'indice de concentration absolue plus pertinente, où les superficies habitées sont déterminées par les superficies artificialisées (données Corine Land Cover), et non par les superficies des unités spatiales.

#### *b) Les mesures de concentration intergroupes*

*L'indice de concentration relative* (Massey et Denton, 1988) compare la concentration spatiale (en termes d'espace occupé) d'une catégorie sociale par rapport à une autre catégorie et varie de  $-1$  à  $1$ . Il présente des valeurs négatives lorsque la concentration de la première catégorie est plus forte que celle de la deuxième catégorie, et des valeurs positives dans le cas inverse. Il est égal à  $0$  dans la situation particulière où les deux catégories sont également concentrées dans l'espace. L'indice de concentration relative compare seulement les distributions entre deux catégories, sans tenir compte d'aucune interaction entre elles.

### **3.4. Le regroupement**

Les indices de regroupement (ou d'agrégation spatiale) prennent en compte, non seulement la concentration spatiale des membres d'un groupe de population, mais aussi les proximités géographiques de ces unités spatiales. On dit que plus une catégorie occupe des unités spatiales contiguës (voisines), formant ainsi une enclave dans la zone d'étude, plus elle est regroupée et donc ségréguée. A l'inverse, plus une catégorie occupe des unités spatiales éloignées les unes des autres, moins elle est ségréguée. Les indices de regroupement et la concentration spatiale offrent des informations différentes : une population peut apparaître très concentrée (la majorité de ses membres sont localisés dans quelques unités spatiales) tout en étant très peu regroupée si ces unités sont éloignées les unes des autres.

#### *a) Les mesures de regroupement intragroupes*

*L'indice de regroupement absolu*, développé par Massey et Denton (1988), exprime le nombre moyen de membres d'une catégorie de population dans une unité spatiale comme une proportion de la population totale des unités spatiales voisines. Deux unités spatiales sont considérées voisines si elles partagent une frontière commune, quelle que soit sa longueur. Ainsi, l'indice de regroupement absolu fait appel à la matrice de contiguïté spatiale. L'indicateur varie de  $0$  (pas de regroupement) à  $1$  (regroupement maximal).

#### *b) Les mesures de regroupement intergroupes*

*L'indice de regroupement relatif* (Massey et Denton, 1988) compare la proximité moyenne des membres d'une catégorie à celle des membres d'une autre catégorie. La proximité moyenne est mesurée en utilisant une matrice des distances entre les centroïdes des unités spatiales : elle correspond

pour un groupe à la somme de ces distances pondérées par la population de ce groupe dans les unités spatiales. L'indice de regroupement relatif est positif quand le niveau d'agrégation du premier groupe est supérieur à celui du second groupe et négatif dans le cas contraire. Si le niveau d'agrégation des deux catégories est identique, la valeur de l'indice est nulle.

### 3.5. La centralisation

La centralisation est une autre forme de ségrégation spatiale. Les indices de centralisation mesurent la proximité de la catégorie de population au centre de zone d'étude (le centre peut être défini sur des raisons géographiques, économiques, sociales, etc.). Plus un groupe est localisé vers le centre, plus il est ségrégué selon cette dimension.

#### a) *Les mesures de centralisation intragroupes*

La proportion du groupe dans l'unité spatiale centrale donne une première information (très simple mais basique et incomplète) sur la tendance de localisation centrale d'un certain groupe. L'information offerte par cet indicateur est insuffisante, parce qu'on ne connaît pas la distribution de la catégorie autour du centre.

C'est pour cette raison que Duncan et Duncan (1955) ont proposé *l'indice de centralisation absolue*, qui mesure la part d'un groupe qui devrait changer de localisation afin d'obtenir une densité uniforme du groupe autour du centre de la zone d'étude. Sa valeur est négative lorsque les membres d'une catégorie ont tendance à résider en périphérie de la zone, et positive dans le cas inverse. Une valeur nulle signifie que le groupe est parfaitement distribué dans la zone d'étude par rapport au centre.

#### b) *Les mesures de centralisation intergroupes*

*L'indice de centralisation relative* (Duncan et Duncan, 1955) mesure la part d'un groupe qui devrait changer de localisation afin d'obtenir le même degré de centralisation que celui d'un autre groupe. Il varie de  $-1$ , lorsque les membres de la première catégorie sont localisés plus loin du centre que les membres d'une autre catégorie, à  $1$  dans la situation inverse.

### 3.6. Les indices de ségrégation multi-groupes

Nous préférons traiter séparément les indices multigroupes pour plusieurs raisons. Tout d'abord, ces indices permettent de calculer un niveau de ségrégation « global » dans le territoire étudié, pour toutes les catégories. Deuxièmement, il n'existe pas d'indices multigroupes pour toutes les dimensions de la ségrégation spatiale.

#### a) *Les indices multigroupes d'égalité*

Une série d'indices multigroupes provenant des analyses de la biodiversité peuvent être utilisés pour mesurer la diversité (sociale ou biologique) d'une zone. Ils peuvent être utiles pour comparer l'hétérogénéité de la population de différentes unités spatiales ou zones d'études. Dans cette catégorie, on trouve notamment *l'indice de Shannon-Wiener* (Shannon, 1948), basé sur la notion d'entropie, dont les valeurs s'établissent entre  $0$  et  $\ln(K)$ , où  $K$  est le nombre des catégories présentes. Une valeur plus élevée implique une hétérogénéité de la population plus importante. Pour pouvoir faire des comparaisons entre zones avec des nombres de catégories différentes, on utilise *l'indice de Pielou* (avec des valeurs comprises entre  $0$  et  $1$ ), qui normalise l'indice de Shannon-Wiener. Par ailleurs, *l'indice d'interaction de Simpson* (Simpson, 1949) mesure la probabilité (entre  $0$  et  $1$ ) que des individus sélectionnés de manière aléatoire n'appartiennent pas au même groupe.

Pour mesurer la ségrégation entre les unités spatiales d'une zone d'étude, les indices d'égalité les plus utilisés sont dérivés de l'indice de Gini et de l'indice de dissimilarité de Duncan. *L'indice de Gini*

*multigroupe* a été développé par Reardon (1998) et l'**indice de dissimilarité multigroupe** par Morgan (1975) et ultérieurement Sakoda (1981).

#### *b) Les indices d'exposition multigroupes*

Il existe également des versions multigroupes des indices d'exposition. Il s'agit d'une part de l'**indice d'isolement normalisé** (Massey et Denton, 1988), et d'autre part de l'**indice de diversité relative** (Carlson 1992, Reardon 1998), obtenu à partir de l'indice d'interaction de Simpson.

### **3.7. Les indices locaux et cartographie de la ségrégation**

Les indices présentés précédemment donnent une image de la ségrégation spatiale sur l'ensemble du territoire étudié, mais ils ne fournissent aucune information sur la distribution spatiale de cette ségrégation, qui pourrait être plus intense dans certaines zones que dans d'autres. Ainsi, les indices de ségrégation locaux permettent d'avoir une image spatialisée de la ségrégation à l'aide des cartes thématiques.

**Les quotients de localisation** (Isard 1960) sont calculés pour chaque groupe séparément et répondent aux questions liées à la sur-représentation ou la sous-représentation d'un groupe dans le territoire. Le quotient de localisation est calculé comme le rapport entre la proportion du groupe dans la population de chaque unité spatiale et la proportion du groupe dans la zone d'étude. Si l'indice est inférieur à 1, le groupe est sous-représenté dans l'unité spatiale par rapport à la structure de la population totale et inversement si l'indice est supérieur à 1. En cas de valeur unitaire, la part du groupe dans l'unité spatiale est identique à la représentation du groupe dans la population totale.

**L'indice d'entropie ou de diversité** (Theil 1971, 1972) est une adaptation « locale » de l'indice de Pielou, un indice multigroupe qui calcule l'entropie sur l'ensemble du territoire. Cet indice mesure la diversité ethnique ou sociale dans chaque unité spatiale de la zone d'étude. L'indice varie entre 0 et 1. Une valeur égale de 0 représente une homogénéité parfaite de l'unité spatiale (un seul groupe y réside). Pour une valeur proche de l'unité, l'unité spatiale est caractérisée par une forte hétérogénéité de populations.

## **4. L'AUTOCORRELATION SPATIALE**

L'autocorrélation spatiale est l'absence d'indépendance entre observations géographiques. Ainsi, on constate très souvent que les variables spatialisées sont soumises à des dépendances spatiales (ou interactions spatiales), qui sont d'autant plus fortes que les localisations sont plus proches : "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

Les mesures d'autocorrélation spatiale permettent d'estimer la dépendance spatiale entre les valeurs d'une même variable en différents endroits de l'espace. Pour la mettre en évidence, les indices prennent en compte deux critères : la proximité spatiale et la ressemblance ou la dissemblance des valeurs de cette variable dans les unités spatiales de la zone d'étude. On fait la distinction entre la mesure de l'autocorrélation spatiale globale d'une variable dans un territoire donné et celle de l'autocorrélation locale dans chaque unité spatiale. Cette dernière correspond à l'intensité et la significativité de la dépendance locale entre la valeur d'une variable dans une unité spatiale et les valeurs de cette même variable dans les unités spatiales environnantes (plus ou moins proches).

Dans OASIS, nous proposons le calcul des indices d'autocorrélation spatiale pour les données en valeurs absolues, mais aussi en valeurs relatives. Les indices les plus utilisés sont l'**indice de Moran** et l'**indice de Geary**. Dans la littérature, le coefficient de Moran est souvent préféré à celui de Geary en raison d'une stabilité générale plus grande (Upton et Fingleton, 1985). Avant de calculer ces indices, il est nécessaire de définir les interactions spatiales prises en compte.

#### *a) La matrice d'interactions spatiales*



Comme le nom l'indique, la matrice d'interactions spatiales est une matrice carrée qui permet de mesurer les interactions entre les unités spatiales, indépendamment de la variable étudiée. Le nombre de lignes et de colonnes est égal au nombre d'unités spatiales dans la zone d'étude. Sur la diagonale la matrice a des valeurs nulles. Chaque élément de la matrice mesure l'interaction entre les unités spatiales correspondante en ligne et en colonne. Sa valeur est d'autant plus élevée que la « proximité » spatiale entre les unités est forte.

Pour utiliser une matrice d'interactions spatiales, il est nécessaire de donner *a priori* une forme fonctionnelle spécifique aux interactions spatiales. C'est pour cette raison que nous proposons dans OASIS les trois formes les plus utilisées dans la littérature :

- La **matrice de contiguïté** : généralement, les éléments de la matrice de contiguïté d'ordre  $k$  sont définis comme le nombre minimal de frontières qu'il faut franchir pour aller d'une unité spatiale à l'autre. La matrice de contiguïté la plus utilisée est d'ordre 1 : les éléments de la matrice sont égaux à 1 si deux unités sont voisines et sont nuls dans le cas contraire. C'est la forme la plus simple d'interactions spatiales, souvent suffisante pour déterminer les niveaux d'autocorrélation spatiales d'une variable.

- La **matrice des distances** est une mesure d'intensité spatiale qui peut être utilisée quand les données sont spatialisées sous forme de points. Lorsqu'on étudie des unités spatiales, la distance prise en compte est généralement la distance entre leurs centroïdes. Cette distance peut être intégrée sous différentes formes : linéaire, quadratique ou logarithmique. Dans OASIS, la matrice d'interactions spatiales basée sur la distance utilise les distances linéaires.

- Une autre manière plus détaillée de décrire les interactions spatiales consiste à prendre en compte non seulement le voisinage (des valeurs 0 ou 1), mais de tenir compte aussi de l'intensité de cette proximité spatiale, qui dépend de l'importance de la **frontière commune**. Les éléments de la matrice sont ainsi des coefficients proportionnels à longueur de la frontière commune.

#### **b) L'indice I de Moran**

L'indice de Moran (Moran 1950) permet de mesurer le niveau d'autocorrélation spatiale d'une variable et de tester sa significativité. Il est égal au ratio de la covariance entre observations contiguës (définies par la matrice d'interactions spatiales) à la variance totale de l'échantillon (Jayet, 2001).

L'indice a des valeurs comprises entre -1 (indiquant une dispersion parfaite) à 1 (corrélation parfaite). Une valeur nulle signifie que la distribution spatiale de la variable étudiée est aléatoire dans le territoire. Les valeurs négatives (positives) de l'indice indiquent une autocorrélation spatiale négative (positive).

Pour le test d'hypothèse statistique, l'indice I de Moran peut être transformé en Z-scores, pour lesquels les valeurs plus grandes que le seuil de significativité positive ou plus petites que le seuil de significativité négative indiquent une autocorrélation spatiale significative, avec un taux d'erreur correspondant au seuil (hypothèse de loi normale).

Une interprétation facile de l'indice de Moran peut être faite avec l'aide du diagramme de Moran, qui représente, sous la forme d'un nuage de points, les couples de valeurs correspondant à la valeur de la variable dans chaque unité spatiale (en abscisse) et la moyenne des valeurs des zones contiguës (en ordonnée) définies par la matrice d'interactions spatiale (cette moyenne est appelée *spatial lag* ou décalage spatial). Dans OASIS, les unités spatiales où l'autocorrélation spatiale n'est pas significative sont représentées sous la forme de cercles, tandis que les unités avec une autocorrélation locale statistiquement significative sont en forme de losange, avec en libellé le nom de l'unité spatiale. Le graphique est complété avec la droite de pente égale à l'indice de Moran. Une pente négative signifie une autocorrélation spatiale négative tandis qu'une pente positive signifie une autocorrélation spatiale positive.

### c) *L'indice de Geary*

Une autre alternative pour mesurer l'autocorrélation spatiale est l'indice de Geary qui est, à un facteur  $\frac{1}{2}$  près, égal au ratio de la variance des écarts entre observations contiguës à la variance totale (Jayet, 2001). Si l'indice de Moran est une mesure de l'autocorrélation spatiale globale, l'indice de Geary est plus sensible à l'autocorrélation spatiale locale.

L'indice de Geary varie de 0 à l'infini et vaut 1 s'il y a indépendance spatiale. Pour toute valeur inférieure à l'unité, il y a une autocorrélation spatiale positive, et inversement pour des valeurs supérieures à l'unité. L'indice de Geary varie en sens inverse de l'indice de Moran. Comme pour l'indice de Moran, l'indice de Geary peut être testé statistiquement, par une transformation en Z-scores et en déterminant son seuil de significativité.

### d) *L'autocorrélation spatiale locale (LISA)*

L'analyse de l'autocorrélation spatiale locale a été introduite par Anselin (1995). *L'indice Moran local* permet de mesurer le degré de corrélation spatiale au niveau local pour chaque unité spatiale. Comme dans le cas de l'indice de Moran, on peut calculer les Z-scores et tester la significativité du degré d'autocorrélation spatiale locale. Les enregistrements significatifs peuvent être représentés sous formes de cartes.

## 5. TESTS DE MONTE CARLO

Les simulations de Monte Carlo permettent de tester l'hypothèse que la valeur d'un indicateur est le résultat d'un processus aléatoire. Si cette hypothèse est rejetée, alors la valeur de cet indicateur est jugée significative.

Le principe de base est relativement simple : les valeurs d'une variable sont distribuées de manière « aléatoire » (suivant un critère de répartition aléatoire choisi) entre les unités spatiales, afin de pouvoir calculer une valeur simulée de l'indicateur. Les simulations sont répétées un nombre suffisamment élevé de fois pour que les valeurs simulées puissent être approximées par une loi de distribution. Cette distribution des valeurs simulées sera utilisée comme base de comparaison pour la valeur « observée » de l'indicateur. Si la valeur « observée » se trouve à l'intérieur des valeurs critiques données par les seuils de significativité, alors on peut conclure qu'elle est non significative. Au contraire, si la valeur « observée » de l'indicateur se trouve en dehors des valeurs critiques, elle n'apparaît pas comme étant le résultat d'un processus aléatoire et peut donc être considérée comme significative.

Deux types de simulations sont faites, en fonction du critère de répartition aléatoire choisi. Le premier permet de tester plus spécifiquement si la composante spatiale de l'indicateur est significative ou non : il est basé sur le test des permutations des données, proposé par Anselin (1995) pour les indices d'autocorrélation spatiale et utilisé par Feitosa et al. (2007) pour les indices de ségrégation. Le second permet de tester la significativité globale de l'indicateur : il est basé sur des localisations aléatoires avec ou sans contrainte.

### 5.1. Test des permutations

Pour le test des permutations, on garde la même structure des données, mais pour chaque simulation les données sont permutées de manière aléatoire entre les unités spatiales. Cette méthode permet ainsi de tester si la composante spatiale de l'indicateur est significative ou non.

Dans le cas des indicateurs d'autocorrélation spatiale, la composante spatiale est donnée par la matrice de décalage spatial. Dans OASIS, l'utilisateur peut choisir la définition des interactions spatiales qui

lui paraît la plus appropriée parmi les modalités proposées : matrice de contiguïté, matrice de distances et matrice des frontières communes.

Dans le cas des indices de ségrégation, la composante spatiale des indicateurs est incluse dans leur définition. Certains prennent en compte la forme des unités spatiales (périmètres et superficies) ou les interactions spatiales entre unités (distances, frontières communes, contiguïté). Les indices qui ne tiennent pas compte de manière explicite des caractéristiques de l'espace sont insensibles aux tests des permutations et ont toujours la même valeur simulée, égale à l'indice observé.

Une fois les simulations avec données permutées réalisées, l'indice observé peut être comparé avec la distribution des valeurs simulées : son rang dans cette distribution permet de calculer une pseudo p-value et de conclure quant à la significativité de la composante spatiale de l'indice.

## 5.2. Test des localisations aléatoires

Les tests des localisations aléatoires permettent de déterminer la significativité globale de l'indice. Les localisations sont simulées sans contrainte (probabilité identique pour chaque unité spatiale) ou avec contraintes spatiales (probabilités déterminées par différents critères : la superficie totale des unités spatiales, ou leur superficie artificialisée). En comparant la distribution des valeurs simulées avec la valeur calculée de l'indice, nous pouvons conclure qu'il est le résultat d'un processus aléatoire ou d'un schéma de localisation déterminé. Si l'indice calculé se trouve à l'extérieur de la distribution (au-dessus du seuil de significativité), il n'apparaît pas comme étant le résultat d'un processus aléatoire et peut donc être considéré comme significatif.

## 6. CONCLUSIONS

La cohésion ou la mixité urbaine, définie comme la coexistence harmonieuse de catégories sociales différenciées dans un même espace, constitue un objectif politique éminent, mis en exergue en France par le Conseil d'Analyse Economique (Fitoussi et al., 2004) et à la source de lois et de politiques publiques conséquentes, telles que la loi relative à la solidarité et au renouvellement urbains (SRU), qui impose 20% de logement social à toutes les grandes communes, ou encore le programme national pour la rénovation urbaine (PNRU), dont le budget s'élève à plus de 40 milliards d'euros sur 10 ans. Dans ce contexte, rendre les outils de mesure de la ségrégation et des inégalités spatiales plus accessibles et plus fiables est une nécessité. Comme le remarque à juste titre Stiglitz et al. (2009) : « ce que nous mesurons et la manière dont nous le mesurons importe à la fois pour le choix des politiques à mener, pour la formation de nos jugements et pour déterminer ce qui fonctionne et ce qui ne fonctionne pas dans nos sociétés ».

Malgré l'existence d'outils web ou logiciels disponibles pour les analyses des variables spatialisées, ces outils sont dispersés et nécessitent un temps de formation et d'utilisation important de la part des usagers. Ainsi, il nous a semblé nécessaire de créer une plateforme unique, qui rassemble ces outils et permet de faire ces analyses de manière complètement automatisée. De plus, OASIS est un outil en ligne ne nécessitant aucune installation sur l'ordinateur des utilisateurs, tous les calculs étant réalisés sur nos serveurs. Au-delà de cette facilité d'utilisation et de son caractère intégrateur, OASIS propose des tests nouveaux, qui n'existent pas encore dans les outils précédents, ainsi qu'une batterie de données fréquemment utilisées dans les analyses de la ségrégation, disponibles pour la France entière (fonds de cartes inclus).

OASIS se trouve à sa première version et des améliorations et développements seront réalisés prochainement. Dans un premier temps, l'outil sera développé pour permettre d'inclure des fonds de carte et des données propres, de manière à ce que les analyses ne soient plus limitées à la France métropolitaine. Par la suite, de nouveaux indices et tests statistiques seront inclus régulièrement.

## BIBLIOGRAPHIE

- ANSELIN, L. (1995): "Local indicators of spatial association", *Geographical Analysis*, 27, 93–115
- ANSELIN, L. (1996): "The Moran scatterplot as an ESDA tool to assess local instability in spatial association". pp. 111–125 in M. M. Fischer, H. J. Scholten and D. Unwin (eds) *Spatial analytical perspectives on GIS*, London, Taylor and Francis;
- APPARICIO P. (2000), "Les indices de ségrégation résidentielle : un outil intégré dans un système d'information géographique", *Cybergeo : European Journal of Geography* [En ligne], Espace, Société, Territoire, article 134.
- APPARICIO P., PETKEVITCH V., CHARRON M. (2008), "Segregation Analyzer : a C#.NET application for calculating residential segregation indices", *Cybergeo : European Journal of Geography* [En ligne], Systèmes, Modélisation, Géostatistiques, article 414.
- BELL W. (1954): "A probability model for the measurement of ecological segregation", *Social Forces*, 32 (4), 357-364.
- CARLSON, S. M. (1992), "Trends in race/sex occupational inequality: conceptual and measurement issues", *Social Problems*, vol. 39, 269-290.
- DUNCAN O.D. et DUNCAN B. (1955a): "A Methodological Analysis of Segregation Indexes", *American Sociological Review*, 41, 210-217.
- DUNCAN, O.D., CUZZOERT, R.P. and DUNCAN, B. (1961): "Problems in analyzing areal data", *Statistical geography*, Glencoe, Illinois: The free press of Glencoe
- FITOUSSI J-P., LAURENT E., MAURICE J. (2004) : "Ségrégation urbaine et intégration sociale". Conseil d'Analyse Economique. La Documentation Française, Paris.
- GEARY R. C. (1954): "The contiguity ratio and statistical mapping", *The Incorporated Statistician*, 5, 115-145.
- GINI C. (1921): "Measurement of inequality of income", *Economic Journal*, 31, 22-43.
- GORARD, S. (2000): "Education and Social Justice", Cardiff: University of Wales Press
- ISARD W. (1960): "Methods of regional analysis: an introduction to regional science". The MIT Press, Cambridge.
- JAYET H. (2001): "Econométrie et données spatiales", *Cahiers d'économie et sociologie rurales*, 58-59
- MASSEY D.S., DENTON N.A. (1988): "The dimensions of residential segregation", *Social forces*, 67 (2), 281-315.
- MORAN P. A. P. (1950): "A test for serial dependence of residuals", *Biometrika*, 37, 178-181.
- MORGAN B. S. (1975): "The Segregation of socioeconomic groups in urban areas: a comparative analysis", *Urban Studies*, 12, pp. 47-60.
- PIELOU, E.C. (1966): "Shannon's formula as a measure of species diversity: its use and misuse", *Am. Nat.*, 100, 463–465
- MORILL R. (1991): "On the measure of geographic segregation", *Geography research forum*, 11, 25-36.

- REARDON, S. F. (1998): "Measures of racial diversity and segregation in multigroup and hierarchical structured Populations", Presented at the annual meeting of the Eastern Sociological Society, Philadelphia, PA.
- REARDON S. F. et FIREBAUGH G. (2002): "Measures of multigroup segregation", *Sociological Methodology*, 32, 33-67.
- REARDON S. F., O'SULLIVAN D. (2004): "Measures of spatial segregation", *Sociological Methodology*, 34, 121-162.
- SAKODA, J. N. (1981): "A generalized Index of dissimilarity", *Demography*, 18, 245-250.
- SHANNON, C.E. (1948): "A mathematical theory of communication", *Bell System Technical Journal*, 27, 379-423 and 623-656.
- SIMPSON E. H. (1949): "Measurement of diversity", *Nature*, 163:688.
- STEARNS, L.B., LOGAN, J.R. (1986) "Measuring trends in segregation: three dimensions, three measures", *Urban affairs quarterly*, 22, 124-150.
- STIGLITZ J., SEN A., FITOUSSI J-P. (2009) : "Mesure des performances économiques et du progrès social réexaminée – Réflexions et vue d'ensemble des questions abordées", [www.stiglitz-sen-fitoussi.fr](http://www.stiglitz-sen-fitoussi.fr)
- THEIL H. (1972). "Statistical Decomposition Analysis". North-Holland, Amsterdam.
- THEIL H., FINEZZA A. J. (1971). "A note on the measurement of racial integration of schools by means of informational concepts". *Journal of Mathematical Sociology*, 1 (2), 187-194
- TOBLER W.R. (1970): "A computer movie simulating urban growth in the Detroit region", *Economic geography*, Supplement 46, 234-40.
- UPTON G., FINGLETON B. (1985): "Spatial data analysis by example", New York, Wiley
- WHITE M. J. (1983) "The Measurement of Spatial Segregation", *American Journal of Sociology*, 88, 1008-1019.
- WONG D.W.S. (1993): "Spatial Indices of Segregation", *Urban Studies*, 30 (3), 559-572.
- WONG D.W.S. (1998): "Measuring multiethnic spatial segregation", *Urban Geography*, 19 (1), 77-87.
- WONG D.W.S. (2003): "Implementing spatial segregation measures in GIS", *Computers, Environment, and Urban Systems*, 27, 53-70.
- WONG D.W.S., CHONG W.K. (1998): "Using spatial segregation measures in GIS and statistical modeling packages", *Urban Geography*, 19, 477-485.

## ANNEXES : CALCUL DES INDICATEURS

### A.1. LES INDICES DE SEGREGATION SPATIALE

#### Les indices d'égalité

L'indice de ségrégation de Duncan (IS) :  $IS^k = \sum_{i=1}^n \left[ \frac{t_i |p_i^k - P^k|}{2TP^k(1-P^k)} \right]$

L'indice de ségrégation de Morill :  $ISMorill^k = IS^k - \frac{\sum_{i=1}^n \sum_{j=1}^n |c_{ij}(p_i^k - p_j^k)|}{\sum_{i=1}^n \sum_{j=1}^n c_{ij}}$

L'indice de Wong :  $ISWong^k = IS^k - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} |p_i^k - p_j^k| \frac{\frac{Per_i}{A_i} + \frac{Per_j}{A_j}}{2 \max\left(\frac{Per}{A}\right)}$ ,  $w_{ij} = \frac{f_{ij}}{\sum_i f_{ij}}$

L'indice de ségrégation de Gorard :  $ISGorard^k = \frac{1}{2} \sum_{i=1}^n \left| \frac{x_i^k}{X^k} - \frac{t_i}{T} \right|$

L'indice de ségrégation de Gini (G) :  $G^k = \sum_{i=1}^n \sum_{j=1}^n \left[ \frac{t_i t_j |p_i^k - p_j^k|}{2T^2 P^k (1-P^k)} \right]$

L'indice de dissimilarité de Duncan (ID) :  $ID^{k_1, k_2} = \frac{1}{2} \sum_{i=1}^n \left| \frac{x_i^{k_1}}{X^{k_1}} - \frac{x_i^{k_2}}{X^{k_2}} \right|$

#### Les indices d'exposition

L'indice d'isolement (xPx) :  $xPx^k = \sum_{i=1}^n \left( \frac{x_i^k}{X^k} \frac{x_i^k}{t_i} \right)$

L'indice d'isolement ajusté (Eta2) :  $Eta2^k = \frac{xPx^k - P^k}{1 - P^k}$

L'indice d'interaction (xPy) :  $xPy^{k_1, k_2} = \sum_{i=1}^n \left( \frac{x_i^{k_1}}{X^{k_1}} \frac{x_i^{k_2}}{t_i} \right)$

#### Les indices de concentration

L'indice Delta :  $\Delta^k = \frac{1}{2} \sum_{i=1}^n \left| \frac{x_i^k}{X^k} - \frac{A_i}{A} \right|$

L'indice de concentration absolue :  $ACO^k = 1 - \frac{\sum_{i=1}^n (x_i^k A_i / X^k) - \sum_{i=1}^{n1} (t_i A_i / T_1)}{\sum_{i=n2}^n (t_i A_i / T_2) - \sum_{i=1}^{n1} (t_i A_i / T_1)}$

L'indice de concentration relative :  $RCO^{k_1, k_2} = \frac{\left[ \sum_{i=1}^n (x_i^{k_1} A_i / X^{k_1}) / \sum_{i=1}^n (x_i^{k_2} A_i / X^{k_2}) \right] - 1}{\left[ \sum_{i=1}^{n1} (t_i A_i / T_1) / \sum_{i=n2}^n (t_i A_i / T_2) \right] - 1}$

### Les indices de regroupement

L'indice de regroupement absolu :  $ACL^k = \frac{\left[ \sum_{i=1}^n \frac{x_i^k}{X^k} \sum_{j=1}^n (c_{ij} x_j^k) \right] - \left[ \frac{X^k}{n^2} \sum_{i=1}^n \sum_{j=1}^n c_{ij} \right]}{\left[ \sum_{i=1}^n \frac{x_i^k}{X^k} \sum_{j=1}^n (c_{ij} t_j^k) \right] - \left[ \frac{X^k}{n^2} \sum_{i=1}^n \sum_{j=1}^n c_{ij} \right]}$

La proximité moyenne entre les membres d'une catégorie k :  $Pxx^k = \frac{1}{(X^k)^2} \sum_{i=1}^n \sum_{j=1}^n x_i^k x_j^k f(d_{ij})$

L'indice de regroupement relatif :  $RCL^{k_1, k_2} = \frac{Pxx^{k_1}}{Pxx^{k_2} - 1}$

### La centralisation

Proportion du groupe localisé dans l'unité spatiale centrale :  $Pcc^k = \frac{Xcc^k}{X^k}$

L'indice de centralisation absolu :  $ACE^k = \left( \sum_{i=2}^n XC_{i-1}^k AC_i \right) - \left( \sum_{i=2}^n XC_i^k AC_{i-1} \right)$

L'indice de centralisation relatif :  $RCE^{k_1, k_2} = \left( \sum_{i=2}^n XC_{i-1}^{k_1} XC_i^{k_2} \right) - \left( \sum_{i=2}^n XC_i^{k_1} XC_{i-1}^{k_2} \right)$

### Les indices multigroupes

L'indice de Shannon-Wiener:  $H_{SW} = - \sum_{k=1}^K P^k \ln P^k$

L'indice de Pielou:  $J_{Pielou} = \frac{H_{SW}}{\ln K}$

L'indice d'interaction de Simpson :  $I_{Simpson} = \sum_{k=1}^K P^k (1 - P^k)$

$$\text{L'indice de Gini multigroupe : } G_{multi} = \frac{1}{2T^2 I_{Simpson}} \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^n (t_i t_j | p_i^k - p_j^k |)$$

$$\text{L'indice de dissimilarité multigroupe : } ID_{multi} = \frac{1}{2TI_{Simpson}} \sum_{k=1}^K \sum_{i=1}^n (t_i | p_i^k - P^k |)$$

### La ségrégation locale

$$\text{Le quotient de localisation : } QL_i^k = \frac{x_i^k / t_i}{X^k / T}$$

$$\text{L'indice d'entropie locale : } H_{PielouLocal}^i = - \frac{\sum_{k=1}^{K_i} P_i^k \ln P_i^k}{\ln K_i}$$

où  $i$  = unité spatiale ;  $n$  = nombre d'unités spatiales ;  $k$  = catégorie de population ;  $p_i^k$  = proportion du groupe  $k$  dans la population totale de l'unité spatiale  $i$  ;  $P^k$  = proportion de la catégorie  $k$  dans la population totale de la zone d'étude ;  $x_i^k$  = population du groupe  $k$  dans l'unité spatiale  $i$  ;  $X^k$  = population totale de la catégorie  $k$  dans la zone d'étude ;  $t_i$  = population totale dans l'unité spatiale  $i$  ;  $T$  = population totale de la zone d'étude ;  $c_{ij}$  = les éléments binaires de la matrice de contiguïté ;  $Per_i$  = périmètre de l'unité spatiale  $i$  ;  $f_{ij}$  = longueur de frontière commune entre les unités spatiales  $i$  et  $j$  ;  $A_i$  = aire de l'unité spatiale  $i$  ;  $AC_i$  = la proportion cumulée de la superficie de l'unité spatiale  $i$  ;  $A$  = la superficie totale de la zone d'étude ;  $n1$  = numéro de l'unité spatiale quand la somme cumulée de la proportion totale des unités spatiales est égale à la somme de la population du groupe  $k$  dans la zone (somme de 1 à  $n1$ , les unités spatiales étant ordonnées par leur superficie) ;  $n2$  = numéro de l'unité spatiale quand la somme cumulée de la proportion totale des unités spatiales est égale à la somme de la population du groupe  $k$  dans la zone (somme de  $n2$  à  $n$ ) ;  $T1, T2$  = population totale dans les unités spatiales de 1 à  $n1$  (de  $n2$  à  $n$ ) ;  $\max(Per/A)$  = rapport maximum entre le périmètre et l'aire des unités spatiales  $j$  ;  $f(d_{ij})$  = une fonction choisie pour l'impact de la distance sur la proximité (fonction linéaire, quadratique, exponentielle, etc...). Dans OASIS nous avons utilisé la forme linéaire :  $f(d_{ij}) = d_{ij}$  ;  $X_{cc}^k$  la population du groupe  $k$  dans l'unité spatiale centrale  $cc$  ;  $XC_i^k$  = la proportion cumulée du groupe  $k$  dans l'unité spatiale  $i$  (les unités spatiales sont triées selon la distance du centre)



## A.2. LES INDICES D'AUTOCORRELATION SPATIALE

$$\text{L'indice de Moran : } I_{Moran} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$E(I_{Moran}) = -\frac{1}{n-1}$$

$$V(I_{Moran}) = E(I_{Moran}^2) - (E(I_{Moran}))^2$$

$$\text{Ainsi, la valeur critique z-score est égale à : } z_I = \frac{I_{Moran} - E(I_{Moran})}{\sqrt{V(I_{Moran})}}$$

$$\text{L'indice de Geary : } C_{Geary} = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{2 \left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\text{L'indice de Moran local : } I_i = \frac{(n-1)(x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

où  $i, j$  = unité spatiale ;  $n$  = nombre d'unités spatiales ;  $x_i$  est la valeur de la variable dans l'unité  $i$  ;  $\bar{x}$  est la moyenne de  $x$  ; et  $w_{ij}$  sont les éléments de la matrice d'interactions spatiales, définie sous la forme de la contiguïté, les distances ou les frontières communes.