



L'espace compte ! Mesurer les structures spatiales du changement social.

Sébastien Oliveau

► To cite this version:

Sébastien Oliveau. L'espace compte ! Mesurer les structures spatiales du changement social.. Géographie. Université d'Aix-Marseille 1, 2011. <tel-01139493>

HAL Id: tel-01139493

<https://halshs.archives-ouvertes.fr/tel-01139493>

Submitted on 5 Apr 2015

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Mémoire d'habilitation à diriger des recherches

Volume 1

L'espace compte !

Mesurer les structures spatiales du changement social

Jury :

Rodolphe Dodier (rapporteur)

Jean-Paul Ferrier (tuteur)

Christophe Z. Guilmoto

Denise Pumain (rapporteur)

Lena Sanders

Isabelle Thomas (rapporteur)

Université Aix-Marseille 1

UFR des sciences géographiques et de l'aménagement

Pour Jean-Luc.

SOMMAIRE

<i>Remerciements</i>	<i>1</i>
<i>Avant-Propos.....</i>	<i>5</i>
<i>« Space matters »</i>	<i>7</i>
<i>L'espace au cœur des sociétés</i>	<i>13</i>
<i>Voisinages</i>	<i>27</i>
<i>La mesure des structures spatiales.....</i>	<i>49</i>
<i>Les indices locaux</i>	<i>66</i>
<i>Au-delà des indices locaux.....</i>	<i>83</i>
<i>Ouvertures</i>	<i>97</i>
<i>Bibliographie.....</i>	<i>105</i>
<i>Annexe : Deux approches de la vulnérabilité des populations</i>	<i>120</i>
<i>Table des matières.....</i>	<i>124</i>

REMERCIEMENTS

La recherche constitue un beau système de don/contre-don. L'écriture de l'habilitation me permettra en partie de rendre ces dons, même si j'ai conscience que les mots seuls parfois ne suffisent pas. Consacrer 3 pages à des remerciements, ce n'est certes pas raisonnable, mais cela ne me permettra de tout dire pour autant. Que les absents ne se vexent pas, ils ne sont pas oubliés pour autant.

Je dois d'abord remercier Jean-Paul Ferrier, qui a accepté avec enthousiasme de m'accompagner dans le cheminement de cette réflexion géographique (et au passage Colette de m'avoir toujours si bien accueilli chez eux). Sa confiance ~~depuis~~ avant mon premier jour à Aix-en-Provence m'honore. Il retrouvera certainement dans ces lignes des emprunts issus de nos conversations, qu'il me pardonne de ne pas les avoir notés plus explicitement.

Seul Christophe Z. Guilmoto connaît l'étendue de ma dette envers lui, depuis Paris-Pondichéry en 1997-98 jusqu'à aujourd'hui, en passant par le Censur et les LISA, les terrains explorés sont immenses ! Je pense qu'il sera fier du chemin parcouru.

Denise Pumain, la première, m'a poussé à écrire cette habilitation, qu'elle soit ici remerciée pour ces 10 années de travail (et d'encouragement) et toutes les opportunités scientifiques créées (notamment le GDRE S4).

Plus récemment rencontré, Rodolphe Dodier m'a fourni de précieux conseils et accepté de lire ce travail, un peu éloigné de ses préoccupations sans doute, mais qui nous permettrons, je l'espère, de mieux travailler ensemble à Aix.

Alors qu'elle a toujours été disponible, je n'ai jamais pris le temps d'échanger avec Lena Sanders sur ma propre recherche. L'habilitation sera donc cette première fois.

Souvent lue, Isabelle Thomas aurait du être de mon jury de thèse. Cela n'avait pu se faire. Je suis heureux de l'occasion qui m'est donnée aujourd'hui.

Malgré ma résistance à concevoir des variables aléatoires dont nous ne serions que des réalisations, christophe Pouet a toujours repris mes écrits pour les corriger et essayer de m'expliquer les lois qui organisent mes données... Qu'il soit encore remercié pour sa patience et son attention.

J'espère que ces pages confirmeront à Céline Rozenblat qu'elle a bien fait de me faire confiance.

Les rencontres sont trop riches et trop nombreuses pour pouvoir être rappelées ici, certaines le seront au long du texte. Je m'attarderais toute même sur quelques institutions marquantes pour remercier les individus qui les font vivre :

- L'université de Nanterre et le département de géographie qui m'ont formé aux différentes facettes du métier d'enseignant-chercheur. Mes responsabilités d' élu et mon engagement associatif m'ont sensibilisé largement aux conditions de vie étudiantes (Merci à Ludovic Bu de m'avoir emmené dans son aventure). La FAGE a eu un rôle très formateur, que tous ses acteurs en soient remerciés (particulièrement Laurent Dailloux).

- L'Institut Français de Pondichéry et surtout son personnel, au sein du département de Sciences Sociales : Venkata qui m'a ouvert des portes de sa culture, Ramanujam, et bien sûr Sattia. Je n'oublie pas tous les autres, Tiare, Balu, Adi, etc.

- l'AJEI (Association Jeunes Etudes Indiennes) qui continue aujourd'hui son chemin sans moi et qui m'a permis de si nombreuses rencontres.

- L'UMR Géographie-cités qui m'a accompagné tout au long de la thèse et dont j'ai toujours plaisir à retrouver les membres, trop nombreux pour être cités. Leur exigence (émulation) et leur écoute (coopération) sont pour moi deux facteurs de réussite.

- L'UMR Thema, pour l'organisation de Théoquant (que Jean-Christophe Foltête nous pardonne) et la vitalité scientifique que toute l'équipe apporte à la géographie en participant au balayage collectif du champ des possibles (Piombini, 2006)... On ne peut pas citer tout le monde, mais Jean-Philippe Antoni tient une place particulière pour moi. Je n'oublie pas Cécile Tannier, Pierre Frankhauser, Florian Tolle et loulou !

- Je ne peux pas oublier mes collègues de l'UMR ESPACE, mais ne peux pas non plus tous les citer.

- L'Université de Provence, et son personnel. Et au sein de celui-ci plus particulièrement Xavier Lafon, Eric Rostang, Natacha Mauric, Dominique Batoux, Henriette Stoffel, Thierry Rippol, Matthew Graves... Annie-Paule Gauthier, toutes les filles du 1^{er} étage. J'ai une pensée particulière pour Sylvie Milhet qui s'occupe si bien de nos doctorants.

Au sein de l'UFR des sciences géographiques et de l'aménagement : Jocelyne Barnabé, "Mimi" Mireille Gauthier, Géraldine Evrard, Chantal Chaput, Fadila Tahouati, Carole Sevin, Isabelle Ginoux... et la plupart de mes collègues enseignants qui supportent mon caractère parfois trop direct et savent me pardonner. Alexandre Grondeau tient une place à part : son arrivée m'a poussé à m'expliquer sur mes positions épistémologiques et à redécouvrir la géographie radicale (et tant d'autres choses).

Une mention spéciale doit être attribuée à l'équipe de la formation MASS, et plus particulièrement Marie-Christine Roubaud, bien sûr, mais aussi Paul Allard et Julie Raud. Tout ça n'aurait pas eu lieu sans ma rencontre avec Christophe Pouet, dont la patience et la disponibilité n'ont d'égales que celles de Thomas Willer quand il s'agit de m'expliquer les variables "iid" et de reprendre mes approximations statistiques...

C'est l'ensemble des étudiants croisés – le temps d'un cours, mais souvent plus longtemps – qui donne son vrai sens à ce travail. La liste est trop longue, mais je me dois d'en citer plusieurs : Marion Borderon et Camille Perchoux pour leur confiance, Mouhamadou Moustapha Mbacke Ndour et Alphousseyni N'Donky. Il est des promotions d'étudiants que l'on n'oublie pas : la bande des affreux du master SDS 2010-2011 (Joel Querci, Yoann Doignon, Lionel Kieffer, Agnès Couvray, Aymeric Morizot, Guillaume Ner, Séverin Heyraud et j'y ajoute Claire Gago-Chidaine, passagère clandestine) et à travers eux tous ceux qui les ont précédé (notamment Elodie Zubar, Joan Perez, Cyril Pivano, Matthieu Pauchet, Géraldine Savin, Joseph Chauvin, Robin Burgi, Rishi Walker...) en attendant la relève !

Enfin, parents et amis (particulièrement ceux de Fontjoyeuse), qui ne saisissent pas forcément les enjeux de cet exercice et qui se sont demandés pourquoi j'ai passé tant de temps avec mon ordinateur plutôt qu'avec eux, sans jamais m'en tenir rigueur. Tant de temps à travailler plutôt qu'à regarder les enfants grandir.

Merci papa et maman.

J'en profite pour faire un clin d'œil aux nombreux(ses) enseignant(e)s qui m'ont fait devenir ce que je suis (depuis la maternelle jusqu'à l'université en passant par le lycée Les Pierres Vives le collège Maupassant et l'école Jean Destraves). Je n'en nommerais qu'un : Marc Ségard, maître à qui je dois la passion des études, des bandes dessinées et du jeu d'échec.

Merci particulièrement à Isabelle et Philippe Venier (c'est ton tour maintenant !) pour l'oxygène sans cesse apporté, et à travers eux tous les amis poitevins (Kamel & Leila plus particulièrement).

Merci à Frédéric Audard pour son soutien sans faille.

Je sais que Pauline saura me pardonner cette année de sacrifice.

Odile et Léone ont tout fait pour rendre impossible la rédaction de cette HDR : qu'elles soient remerciées de nous rappeler chaque jour que l'essentiel est ailleurs.

Dès la première page d'introduction de son propre mémoire d'habilitation à diriger des recherches, Jean Luc Bonnefoy rappelle la tension qui existe entre l'approche matérielle et l'approche idéelle des organisations spatiales humaines pour conclure qu' « aucun géographe n'est strictement d'un bord ou de l'autre, mais l'adoption de tel ou tel *a priori* donne à voir des réalités différentes, complémentaires. » (Bonnefoy, 2005). Sa volonté de s'intéresser aux habitants¹ l'a amené à repartir de la "brique de base" de cette étude, les individus pris isolément, dans leur liberté de choix et leur capacité à interagir. L'espace envisagé devenait complexe, construit par la multitude des interactions entre ces individus.

Confronté plusieurs années à cette manière de voir critique et radicale, il devenait nécessaire d'expliquer mes propres choix, complémentaires aux siens. Je suis de ceux qui « font grand cas de la spatialisation des phénomènes, de leur autocorrélation spatiale », pour reprendre ses termes. Bien que partageant une même ambition (exigence) de théorisation de la géographie, de rigueur dans l'analyse spatiale, il nous a semblé souvent être ce que l'on pouvait faire de plus opposés en géographie : lui qui centrait son approche sur l'individu, comment se retrouver, moi qui ai toujours regardé les « paquets » d'humains et des « bouts d'espace » ? Intéressé par la liberté des acteurs, il s'opposait à moi qui préfère observer les contraintes qui les enserrant.

On l'aura compris, le point de vue proposé ici sera celui d'une analyse spatiale généralement agrégée, s'intéressant aux groupes sociaux dans ce qu'ils partagent de plus communs, et dans les formes lisibles que ces communautés inconscientes laissent voir. Nous abandonnons donc, temporairement et sans la renier, la science complexe.

L'ensemble de ce travail repose sur des travaux en langue anglaise ou française. On le verra, la dynamique de la géographie n'est pas la même dans ces deux univers et cela impose des allers-retours entre chaque. Cela donne parfois des choses assez artificielles, des sortes de coupure entre ces mondes. Je ne sais pas exactement ce qu'il en a été avant les années 2000, mais on peut imaginer que les contacts étaient plus difficiles qu'aujourd'hui, et ils ne sont

¹ Il est de ce point de vue l'héritier direct de la géographie humaniste de Jean Paul Ferrier.

toujours pas si fréquents, finalement (la distance continue d'opposer sa résistance !). Néanmoins, la littérature, nous y reviendrons, est devenue plus aisément accessible. Il n'en reste pas moins que si la révolution géographique a eu lieu d'abord dans le monde anglophone, elle s'est diffusée en France où elle a été réappropriée localement. La relative indépendance de ces deux aires culturelles a créé des différences qui perdurent. Pour respecter ces différences, mais aussi jeter des ponts, j'ai pris soin de respecter autant que possible les citations et termes originaux en les laissant en anglais et en proposant soit la traduction établie, soit une traduction personnelle.

Depuis l'ouvrage de Peter Hagett² (1965 ; 1973 pour la traduction française) consacré à l'analyse spatiale en géographie humaine, la discipline a connu de grands progrès dans l'intégration des méthodes formelles issues des mathématiques appliquées. L'ouvrage *Spatial processes. Models and applications* de Cliff et Ord en 1981³, ou celui de D.Ripley *Spatial statistics* la même année, témoigne de ce point de vue de l'arrivée à maturité du projet collectif de géographie quantitative initié moins de 20 ans auparavant.

Les français participent de cet effort : en dehors de la géographie avec les travaux des géologues (Matheron, 1965) ou des mathématiciens (Lebart, 1969) mais aussi au sein de la discipline (Brunet, 1967, Sanders & Durand-Dastès, 1985 ; Saint-Julien, 1985, Dauphiné & Voiron, 1988⁴). Les géographes français, partis avec 10 ans de retard (Pumain, Robic, 2002) orientent en parallèle leurs travaux dans des directions plus théoriques (GIP Reclus 1985 ; Pumain *et al.*, 1989 ; Brunet, Dollfus, 1990 ; Sanders, 1992 ; Bailly, Pumain, Ferras, 1992). De la géographie quantitative, pas seulement pour décrire mais pour comprendre et expliquer. L'ambition en France est d'abord théorique et quantitative, comme en témoignent encore aujourd'hui deux colloques bisannuels majeurs⁵ : Théoquant⁶ et dans une moindre mesure Géopoint⁷.

² D'autres ont suivi que nous ne pourrions pas tous citer (par exemple Berry, Marble, 1968 ou Cole, King, 1968). On doit rappeler qu'en Suède les *Lund studies in geography*, avaient déjà publié des travaux à partir de 1949 (dont le fameux *The Propagation of Innovation Waves* d'Hägerstrand en 1952), entament une série C, intitulée *General and mathematical geography* en 1962 (Bunge). L'ouvrage de Peter Hagett reste néanmoins le plus emblématique de la période, particulièrement en France.

³ Cet ouvrage avait été précédé 8 ans avant d'un premier opus consacré uniquement à l'autocorrélation spatiale (Cliff, Ord, 1973). On se rappellera que si Andrew Cliff est un géographe, Keith Ord est un mathématicien.

⁴ On ne parlera pas ici de la géographie physique, qui constitue à mon sens une discipline autonome de la géographie humaine, bien qu'elles se retrouvent sur les questions de méthodes d'une part et sur le thème environnemental d'autre part.

⁵ On peut aussi évoquer l'ECTQG (European Colloquium of Theoretical and Quantitative Geography) où la participation française est toujours importante.

⁶ Le colloque Théoquant est organisé tous les deux ans à Besançon par le laboratoire ThéMA (Théoriser et Modéliser pour Aménager) depuis 1993 (<http://thema.univ-fcomte.fr/theoq/>).

⁷ Le colloque Géopoint est organisé tous les deux ans à Avignon par le groupe Dupont depuis 1976. Il est aujourd'hui appuyé par l'UMR 6012 ESPACE (<http://www.groupe-dupont.org/geopoint.htm>). Il se démarque du colloque Théoquant par l'absence de référence explicite à la théorie (bien que ce fût le thème du premier colloque) et par une prise de distance vis-à-vis de l'approche quantitative (voir par exemple l'avant-propos des actes de 1976).

Un quart de siècle plus tard, ces différences perdurent. Les fers de lance de l'analyse spatiale de langue anglaise sont plus tournés vers une approche statistique et/ou directement appliquée (voir les travaux du CSISS⁸ aux Etats-Unis, du CASA⁹ en Angleterre ou du NCG¹⁰ en Irlande), alors que les français ont plus insisté sur la dimension théorique, qui donne du sens à ces transferts des mathématiques vers les sciences sociales (voir les travaux développés au sein du GDRE S4¹¹ et les publications de Denise Pumain 2003, 2004, ou Claude Grasland, 2009).

On peut envisager ces différences à partir d'une lecture particulière de la situation française. La géographie y est considérée avant tout comme participant des sciences sociales¹², et la dimension littéraire des écrits y est importante depuis toujours (en témoigne le poids encore marqué de l'épreuve de l'agrégation dans la discipline). A ce titre, les tentatives de formalisation mathématique sont souvent regardées avec suspicion et l'on préfère des formes plus intuitives d'explication¹³. Dans le cadre de l'analyse spatiale française, il semblait donc assez raisonnable d'aller vers des approches plus théoriques, qui permettaient de mieux faire passer le message auprès de collègues réticents. Par ailleurs, cela permettait de donner aux géographes des outils de compréhension de ces méthodes afin qu'ils puissent les adopter ou au moins les accepter¹⁴.

On peut aussi y voir une différence plus essentielle, à savoir l'orientation plus appliquée de la géographie de langue anglaise par opposition à l'orientation plus académique de la géographie française, qui cherche avant tout à déchiffrer le monde (Brunet, 1990) et à lui donner du sens.

⁸ Le CSISS (Center for Spatially Integrated Social Science) : <http://www.csiss.org>

⁹ CASA (Center for Advanced Spatial Analysis) : <http://www.casa.ucl.ac.uk>

¹⁰ NCG (National Centre for Geocomputation) : <http://ncg.nuim.ie>

¹¹ GRE S4 (Spatial Simulation for the Social Sciences) : s4.parisgeo.cnrs.fr

¹² Dans les universités, les départements de géographie sont bien souvent et de plus en plus rattachés aux secteurs de lettres et sciences humaines qu'aux sciences dites dures.

¹³ Le sujet est abordé dans Pumain (2004), nous l'avons aussi évoqué aussi (Oliveau 2010).

¹⁴ La lecture critique du *Dictionnaire de la géographie et de l'espace des sociétés* de Jacques Lévy et Michel Lussault par Kevin Cox (2006) souligne bien à quel point une partie des géographes français n'ont toujours pas compris la démarche d'analyse spatiale, ses apports dans et hors de la discipline et plus généralement son projet scientifique.

C'est dans cette perspective particulière que nous souhaitons ré-aborder la question de l'autocorrélation spatiale¹⁵. En effet, le sujet est en grande partie statistiquement maîtrisé¹⁶ mais n'a pas encore connu le succès que l'on pourrait envisager¹⁷. Il convient donc de revisiter cette méthode, non pas uniquement du point de vue statistique, cela a été fait avant nous (et généralement mieux), mais du point de vue de la géographie, en envisageant cette mesure dans ses différentes dimensions, du global au local, et à travers ses apports, sa mise en œuvre, son interprétation, ses potentiels et bien sûr ses limites.

Comme nous pouvons le voir, l'ambition est didactique autant que scientifique, car les enjeux de l'utilisation de cet outil sont majeurs. Il concerne d'abord la validité statistique des mesures faites couramment en géographie (au premier rang desquels la mesure de la corrélation), mais aussi des outils appliqués de façon automatique (lissages spatiaux offerts par les logiciels de cartographie par exemple). Il porte aussi sur l'appréhension des espaces étudiés en portant notre regard non plus sur les ruptures, frontières et autres fragmentations, mais en insistant au contraire sur l'homogénéité et les convergences parfois plus difficiles à percevoir.

La question statistique

Revenons sur ces derniers points et d'abord sur celui de la validité statistique des résultats proposés en sciences sociales. Même les méthodes qui semblent les plus simples en analyse statistique exploratoire (corrélation, régression) reposent sur des hypothèses concernant les variables qui sont relativement fortes et devraient être vérifiées systématiquement (ce que rappelait déjà Peter Gould dans son article de 1970 : « Is statistix inferens the geographical names for a wild goose ? »). Tel est le cas de l'hypothèse d'indépendance des variables, qui n'est pas respectée s'il y a autocorrélation spatiale (Haining, 2003 : 273).

Ainsi en est-il par exemple de l'analyse des résidus d'une régression. La présence d'autocorrélation spatiale va augmenter artificiellement les coefficients et risquer de rendre

¹⁵ L'autocorrélation spatiale est la dépendance entre attributs d'individus statistiques voisins dans l'espace (Charre 1995). Autrement dit : « la corrélation d'une variable avec elle-même (autocorrélation) attribuable à l'organisation géographique des données (spatial) » (Griffith, 1992).

¹⁶ Même s'il reste des progrès à faire, notamment concernant les hypothèses de départ (toutes les questions autour de la normalité des variables) et les tests de significativité des résultats, par exemple.

¹⁷ C'est moins vrai Outre-Atlantique, et cela se répand doucement en France, dans le milieu de l'analyse spatiale, bien sûr (depuis Banos 2001, et plus récemment Zaninetti, 2004,) Le nombre de thèses qui y font explicitement appel augmente : en analyse spatiale (Casanova 2010), mais aussi au-delà en géographie de la santé par exemple (Charreire 2007) ou du développement (Daix, 2010).

statistiquement significative une relation qui ne l'est pas (elle réduit la variance)¹⁸. La présence d'autocorrélation spatiale doit alerter le modélisateur sur l'existence de variable(s) non prise(s) en compte ou de l'influence des valeurs des individus sur celles de leurs voisins (nous y revenons plus en détail page 92 et suivantes).

Au-delà, les lissages spatiaux de type « krigeage » reposent sur l'hypothèse que les données ont une structure spatiale et s'appuie sur cette structure pour construire l'interpolation des valeurs. Avec le développement important de la puissance logicielle et informatique, ces lissages sont aujourd'hui accessibles à tous. Pourtant, si la machine autorise l'utilisateur à effectuer un krigeage, celui-ci devrait être réservé au cas où l'autocorrélation spatiale est identifiée et proposer des paramètres en fonction d'un variogramme spécifique, après une analyse des conditions de stationnarité¹⁹ et d'isotropie²⁰ du phénomène étudié. Ceci n'est pas toujours le cas, loin s'en faut.

La mesure de l'homogénéité.

Le second argument justifiant le recours à des mesures d'autocorrélation spatiale n'est plus mathématique mais théorique. Au-delà du biais statistique que représente la présence d'autocorrélation spatiale, c'est aussi une information d'ordre "purement"²¹ géographique qui nous est donnée : l'espace compte et se rappelle à nous (c'est notamment l'argumentaire de Robert Haining, 2003, 2009 ou de Michael Goodchild, 2009). Des structures spatiales existent, qui peuvent être mesurées certes, mais également interprétées. La présence d'autocorrélation spatiale révèle l'existence de convergences régionales et d'ensembles homogènes. C'est ce que nous dit la théorie des variables régionalisées (Matheron, 1965) à propos de la géologie et qui se traduit aussi dans l'espace social. De ce point de vue, la mesure de l'autocorrélation spatiale s'opposerait en quelque sorte à (ou viendrait en

¹⁸ Smith & Lee (2011) nuancent le propos en rappelant que si les variables explicatives ne sont pas spatialement autocorrélées, alors l'affirmation n'est plus exacte.

¹⁹ Pour le dire rapidement, la stationnarité est la caractéristique d'une variable spatiale dont la distribution est identique quelque soit le positionnement de la fenêtre d'observation (sous-espace) étudié. Autrement dit, une variable spatiale qui ne connaît pas de poches locales où les valeurs se différencient de la moyenne globale.

²⁰ L'isotropie est le fait qu'une variable spatiale ne montre pas d'orientation privilégiée (voir page 48). C'est d'ailleurs généralement l'anisotropie qui est recherchée dans les données. Houlding (2000 : 88 et suivantes) propose un exemple clair.

²¹ Nous utilisons dans le texte deux variétés de guillemets. « » indiquent une citation ou une référence, " " indiquent un relâchement du langage volontaire (abus de langage, termes excessifs).

complément de) l'étude des discontinuités, ruptures et autres frontières, approches quasi consubstantielles à la géographie, et qui perdure²².

C'est ici un de nos *credo* qui s'exprime. Pour des raisons statistiques d'une part, mais surtout pour des raisons géographiques, l'étude de l'autocorrélation spatiale constitue à notre avis un point de départ nécessaire aux travaux en géographie.

²² (Foucher, 1986 redonne indubitablement une actualité et une pertinence au sujet). Dans le cadre plus restreint du courant lié à l'analyse spatiale, on se rappellera qu'il s'agissait déjà du sujet de la thèse complémentaire de Brunet (1967), remis au goût du jour 30 ans plus tard (Gay, 1995 ; Brunet et al. 1997).

Il y a 25 ans, Lewis Branscomb, ancien directeur de la recherche pour la société IBM répondait à un exercice de science fiction pour la revue *American Scientist*. Il s'y livrait à une projection de ce que pourrait être la recherche 20 ans après en 2006. Dans son exposé, quelques lignes sont à retenir :

« By 2000, it was recognized that modern geography is the integrated view of man and his planet, the bringing together of ecology, the study of human habitats, geomorphology, social anthropology, and economics-in short, all the tools necessary to understand how human beings should view their fragile planetary home. Once again geography became a popular course of study in school, particularly since students were no longer required to memorize state capitals and map features. People carried such information with them in their pocket data banks. »

On notera d'abord que l'information géographique dans la poche est aujourd'hui une réalité via les applications de téléphone mobiles de type *google map* et autres²³. Même si il aura fallu attendre un peu plus tard que 2000 pour avoir réellement dans la poche ces informations, on peut considérer que le tournant du 21^{ème} siècle a été marqué par l'explosion de la diffusion de l'information géographique, autant du point de vue quantitatif que qualitatif. Le plus important dans la remarque de Branscomb tient surtout au rôle central qu'il envisageait pour la géographie, science d'interfaces entre les autres sciences et au cœur des problématiques se rapportant à l'humain.

Les géographes ont longtemps manqué d'ambition théorique, préférant agir et se frotter au terrain, plutôt que de réfléchir à ce qui constituait les fondements épistémologiques de la discipline. Cette faiblesse intellectuelle constituait dans le même temps une force, gardant la discipline unie autour de ses pratiques, à défaut de se définir un objet scientifique commun. L'hétérogénéité des recherches regroupées sous le terme de géographie en témoigne encore aujourd'hui. Néanmoins, deux faits majeurs ont amené la géographie à évoluer. La première remonte à la fin des années 60, lorsque la révolution quantitative a atteint la géographie française. La seconde est liée au développement de la géomatique puis de l'Internet, particulièrement à partir du milieu des années 1990. Si la première est bien documentée

²³ On remarquera quand même que les prémisses étaient déjà visibles en 1986 : « 3615iti » sur le minitel français permettait dès 1987 de définir des itinéraires...

(Pumain, Robic, 2002 pour le contexte français ; Morrill, 1984, Haggett, 2008 ou Haining 2009 pour le contexte anglophone), on ne réalise peut être pas encore complètement l'impact qu'a eu la seconde. En effet, si les premiers logiciels de cartographie sur ordinateur sont contemporains de la révolution quantitative, il faudra attendre les années 1990 pour que se répandent leur usage au-delà des cercles restreints des laboratoires de cartographie. Il paraît néanmoins pertinent de retracer en grandes lignes l'histoire de la cartographie automatique jusqu'à aujourd'hui, car elle éclairera sans doute la situation présente.

La naissance de la cartographie automatique sur ordinateur peut être assimilée au développement du logiciel SYMAP à partir de 1965 dans le laboratoire de *computer graphics*²⁴ à Harvard autour de Howard Fisher (Chrisman, 2005). Il faudra juste 10 ans pour qu'ait lieu la première conférence américaine sur la cartographie assistée par ordinateur (Elliott, 1976). Les ordinateurs de l'époque ne sont pas encore micro et encore moins personnels et les investissements nécessaires trop importants. Les premiers micro-ordinateurs apparaissent vers 1975, mais on fait habituellement remonter l'ordinateur personnel à la commercialisation de l'Apple II en 1977. Les premiers PC (*Personal Computer* d'IBM) datent eux de 1981²⁵.

Il n'en demeure pas moins que les éléments nécessaires à toute l'évolution de la cartographie vers les Systèmes d'Information Géographique (SIG) sont déjà présents et que des expérimentations²⁶ de logiciels de cartographie verront le jour ici et là dans les années 1980²⁷. Dès 1982, ESRI lance la première le logiciel ARC/INFO (programmé en FORTRAN) et quatre ans plus tard Mapinfo commercialise le premier SIG pour PC. C'est néanmoins au tournant des années 1990 que les SIG se développent réellement.

Aujourd'hui, chacun peut disposer d'un SIG sur son ordinateur, et son utilisation va bien au-delà des géographes spécialisés. La décennie 2000 a vu la démocratisation rapide de cet outil, rendant la gestion de l'information géographique accessible aux autres disciplines. On peut néanmoins remarquer que le corpus scientifique qui s'était développé en parallèle dans la discipline (sémiologie graphique et analyse spatiale notamment) peine à suivre la diffusion.

²⁴ Il prendra le nom de *computer graphics and spatial analysis* en 1968. Pour la petite histoire, on se rappellera que Brian Berry a soutenu le projet dès le départ, il sera d'ailleurs à la tête de ce laboratoire de 1975 à 1981.

²⁵ Nous avons eu la chance d'avoir un directeur d'école primaire qui monta un club informatique à base de ZX81 (Sinclair), nous amenant à mettre les doigts sur un clavier (à membrane) avant l'âge de 10 ans... que M. Laglantine soit ici remercié.

²⁶ Expérimentations au sens où peu ont survécu et ont pu se répandre.

²⁷ L'article de Philippe Waniez (2010) qui porte sur son logiciel Philcarto propose une contextualisation française intéressante.

On n'a jamais tant produit de cartes, mais elles n'ont peut être jamais été aussi mal réalisées et souvent inutiles. « L'intelligibilité supplémentaire apportée par la cartographie des résultats », pour reprendre l'expression de Denise Pumain et Marie-Claire Robic (2002), n'est pas toujours évidente...

Autrement dit, nous sommes aujourd'hui dans une situation quelque peu paradoxale. Jamais l'information géographique n'a été aussi disponible, mais l'entreprise de théorisation de la géographie a du mal à se faire entendre²⁸. Le développement des outils informatiques, dont on aurait pu penser qu'il favoriserait les approches de type théorique et quantitatif, a surtout favorisé la diffusion des outils de description cartographique, dans lesquels les tenants d'une géographie qu'on qualifiera de non réformée s'engouffrent. Les SIG ont tous les aspects du sérieux scientifique : objectivisation, informatisation, reproductibilité avec une touche de communication en plus puisqu'il permet de faire de jolies cartes. On en trouve donc dans tous les projets et l'on prétend faire avec de l'analyse spatiale. Mais l'analyse spatiale repose sur un positionnement épistémologique que beaucoup ont du mal à accepter. Dessiner des « tampons » pour délimiter des zones, ce n'est pas faire de l'analyse spatiale...

Pourtant, les théories et les méthodes d'une analyse plus poussée sont largement à la disposition des chercheurs aujourd'hui. Décrire des formes, mesurer des structures, modéliser des relations sont à la portée de tous, souvent à travers des logiciels plutôt conviviaux, souvent gratuits, mais rarement intégrés²⁹. L'analyse factorielle, les classifications et les règles de sémiologie graphique sont enseignées dès la licence. Le corpus théorique a été vulgarisé depuis longtemps (mais en partie seulement) : Racine & Reymond (1973) ; Sanders, Durand-Dastès (1985), Dauphiné, Voiron (1988), Bailly *et al.* (1992), Charre (1995), Gay (1995), Derycke *et al.* (1996), Pumain, Saint Julien (1997), Pumain, Saint Julien (2001), Sanders (2001)³⁰... Cependant, de nombreux classiques n'ont jamais été traduits en français. Que l'on pense simplement à Christaller (1933), Lösch (1940) ou Hägerstrand (1953, 1967 pour la version anglaise) par exemple, et Haggett (1973) fait alors figure de grande exception. Nous avons sans doute insuffisamment œuvré et il manque encore de nombreuses

²⁸ Rhind *et al.* en faisaient déjà le titre d'une communication en 1988 : « The analysis of geographical data: data rich, technology adequate, theory poor ».

²⁹ Longley et Batty le faisaient déjà remarquer il y a 15 ans de cela. Les logiciels de SIG ne sont pas développés par des personnes ayant des cursus de géographie, ce qui n'aide pas à l'intégration des outils dont aurait besoin l'analyse spatiale. (Longley, Batty, 1996 : 3)

³⁰ Nous n'avons pas cherché à être exhaustif.

publications pour rendre ce champ plus accessible³¹. Le monde de langue anglaise est bien mieux fourni et plus divers (mais les forces plus nombreuses aussi...).

Dans le même temps, les disciplines hors de la géographie ont investi le champ spatial de manière souvent moins frileuse (c'est aussi le constat fait par Robert Haining, 2009). Au premier rang, il faut sans doute citer l'économie, à travers l'économie géographique, mais surtout l'économétrie spatiale et d'autres approches qui intègrent aisément ces formalismes. L'écologie a aussi largement intégré l'espace comme un « nouveau paradigme » (Legendre, 1993). Au sein des sciences sociales, la collaboration avec historiens et surtout archéologues est ancienne. Elle peine à émerger avec les sociologues mais connaît des développements rapides chez les démographes. On peut penser que les disciplines déjà tournées vers la quantification aient moins de mal à intégrer les deux nouvelles dimensions qu'apporte l'espace³².

Petit à petit, l'espace prend sa place d'acteur majeur de nos sociétés, et les géographes doivent saisir cette opportunité de se placer au premier plan sur les questions de société contemporaines. En effet, la géographie et ses méthodologies possèdent tous les atouts pour se placer au cœur des débats politiques actuels. En se situant dans la perspective d'un espace³³ qui contraint nos sociétés, on remet au centre de nos réflexions les possibilités et les capacités des individus et des groupes humains à agir. La révélation des contraintes est la première démarche pour les combattre.

Nous incluons donc nos travaux dans cette perspective en nous impliquant dans un champ qui nous est cher, celui des inégalités sociales et de leur expression spatiale. Nous sommes assez proche de ce point de vue de la démarche radicale de géographes américains comme David Harvey (1996), même si on peut regretter qu'il ait abandonné le formalisme de ses premiers travaux³⁴.

³¹ Nous n'incluons volontairement pas les colloques dans le travail de vulgarisation. Il s'agit pour nous de lieux d'échanges où les participants s'auto-sélectionnent en fonction de leurs intérêts existant. Autrement dit, il y a un biais de sélection, puisque les néophytes s'auto-excluent la plupart du temps.

³² Certes l'espace est constitué de trois dimensions, mais l'altitude est souvent négligée, pour des questions de complexité de calcul, et du peu de plus value que sa prise en compte entraîne.

³³ A mon sens, l'espace est la structure basique. Le territoire ou l'environnement sont des espaces qualifiés pour décrire une caractéristique particulière qu'on lui attache (identité culturelle, biologique, etc.)

³⁴ Kevin Cox a suivi le même chemin. William Bunge a eu une trajectoire assez similaire, mais plus dramatique (Goodchild, 2008).

Poverty mapping et spatial poverty traps

Avec la diffusion des systèmes d'information géographiques, le milieu des années 1990 marque un regain d'intérêt pour la prise en compte de l'espace dans les analyses en sciences sociales (Bailey & Gatrell, 1995; Arlinghaus, 1996; Martin, 1996). C'est à cette époque qu'apparaît au sein de la Banque Mondiale l'idée que cartographier la pauvreté constituerait un apport au ciblage des populations à aider en priorité.

Lors du colloque « including the poor » en 1989, Datt et Ravallion (1993) envisagent l'apport du ciblage (*targeting*) des populations pour la réduction de la pauvreté entre les régions indiennes. Leur conclusion est mitigée et contredit en partie les résultats plus clairs obtenus précédemment en Indonésie (Ravallion, 1993). Néanmoins, ce dernier continuera ses travaux (au Bangladesh) dans cette direction et arrivera à la conclusion que l'on peut définir des espaces pauvres, qui ne sont pas simplement la concentration des pauvres dans l'espace (Ravallion & Wodon, 1997).

L'idée de Ravallion est que les zones les plus pauvres ne se caractérisent pas seulement par la concentration de populations pauvres mais que ces populations restent plus pauvres aussi parce qu'elles occupent ces espaces. Le concept de « piège spatial à pauvreté » (*spatial poverty traps*) est exposé en 1997 dans une publication avec Jalan (Jalan & Ravallion, 1997).

Par ailleurs, on constate aussi que les populations les plus pauvres habitent dans les zones les plus dangereuses (que le risque soit "naturel" ou sanitaire) et que les populations les plus riches résident dans des endroits plus sûrs (Lall & Deichmann, 2009³⁵). Les vulnérabilités sont multiples et enferment plus encore les populations dans leur précarité lorsque des catastrophes surviennent³⁶.

La cartographie de la pauvreté (*poverty mapping*) se développe donc pour différents pays tout au long de la décennie 2000, appuyée par les efforts méthodologiques des économistes (Deichmann, 1999, Davies, 2003). Alors que les études se multiplient chez les économistes (voir par exemple la bonne synthèse éditée par Bedi *et al.*, 2007), l'exercice ne semble pas avoir percolé dans la géographie du développement, où il pourrait être fécond et renouveler

³⁵ Nous arrivons aux mêmes conclusions concernant le lien entre paludisme et pauvreté à Dakar (Couvray *et al.*, 2011 ; Borderon *et al.*, 2011).

³⁶ Après s'être beaucoup concentrée sur les risques, la géographie s'intéresse aujourd'hui aux vulnérabilités, mais la dimension sociale est encore largement ignorée au profit des dimensions écologiques et infrastructurelles (le numéro à venir de la revue *GeoJournal* en témoigne : Matisziw & Grubestic, 2011).

les questionnements.³⁷ La cartographie de la pauvreté concentre en effet plusieurs questions méthodologiques, que l'on retrouve ailleurs.

Ainsi, si chacun sait ce qu'est la pauvreté, sa définition pose rapidement problème. Il n'y pas de consensus chez les économistes. On s'accorde néanmoins à définir la pauvreté selon un « seuil » (souvent monétaire ou monétarisé) en dessous duquel la population sera déclarée pauvre. Ce seuil peut être défini de façon absolue ou relative, voire connaître plusieurs définitions (haute et basse par exemple). On utilise ainsi souvent le seuil de revenu de 1\$ par jour et par habitant (en parité de pouvoir d'achat) pour définir de façon absolue la pauvreté à l'échelle mondiale. On trouve aussi parfois le seuil de 2\$.

Evidemment, cette solution, qui permet les comparaisons à petite échelle n'est pas toujours pertinente. Ce seuil de 1\$/habitant/jour n'est, par exemple, pas à même de refléter la situation dans les pays développés. C'est pourquoi on a aussi recours à des mesures relatives de la pauvreté, définie en fonction du niveau de vie de la région concernée. Ainsi, les statisticiens d'Eurostat « utilisent en général un seuil à 60 % de la médiane des niveaux de vie » (définition extraite du site de l'INSEE). Cette seconde approche est toute aussi artificielle que la première, puisque l'on pourrait tout aussi bien choisir 40 ou 50% (ce qui réduirait le nombre de pauvres). Pour illustrer le propos, on peut rappeler que l'INSEE estime que le seuil de pauvreté en 2008 est à 949€ de revenus mensuels³⁸ (seuil à 60%) ou à 791 euros (seuil à 50%). On est loin des 1\$ ou 2\$ quotidiens définis par la Banque Mondiale.

Les définitions relatives, si elles ne sont pas parfaites, semblent néanmoins bien plus adaptées à l'étude comparative de la pauvreté et surtout des inégalités face à la pauvreté. Notamment parce que les pays du Sud, les plus concernés par le sujet³⁹, ne sont pas homogènes non plus et connaissent de grandes différences entre eux et au sein de leur territoire (que l'on envisage seulement les différences villes-campagnes, par exemple).

Un autre élément de discussion important concerne le niveau auquel on appréhende la pauvreté. Si les économistes montrent une nette préférence pour des estimations au niveau des individus ou des ménages, les données ne sont pas toujours accessibles et il est plus facile de définir des indicateurs à des niveaux agrégés. Ainsi, il est particulièrement difficile d'évaluer

³⁷ On notera d'ailleurs que le SEDAC (Socioeconomic Data and Applications Center) a investi ce champ depuis 2004, où il trouve de quoi exploiter ses bases de données géographiques (données censitaires et imagerie satellitaire) : <http://sedac.ciesin.columbia.edu/povmap/>

³⁸ Personnes vivant en France métropolitaine dans un ménage dont le revenu déclaré à l'administration fiscale est positif ou nul et dont la personne de référence n'est pas étudiante.

³⁹ On ne doit pas oublier que la pauvreté dans les pays les moins avancés a un impact immédiat sur la vie des individus (survie), là où il est différé dans les pays du Nord (baisse de l'espérance et de la qualité de vie).

avec précision les revenus d'un ménage. La mesure du PIB d'une région est quant à lui bien plus facile. Comme souvent, on a donc une opposition entre le niveau de résolution et le coût général de l'information⁴⁰ : plus l'information est désagrégée plus elle est difficile à obtenir (et plus elle est hétérogène⁴¹).

On le voit, l'exercice de cartographie de la pauvreté entraîne nombre de questions que l'on retrouve, de manière plus ou moins marquée, en étudiant d'autres problématiques. Au cœur des questionnements reviennent des questions de définition des indicateurs, de résolution de l'information (de l'individu au groupe) et d'échelle de l'analyse (qui justifie les méthodes absolues ou relatives). C'est typiquement les questions auxquelles sont confrontés les chercheurs qui travaillent sur le développement, quelque soit la ou les dimensions selon la(es)quelle(s) on l'aborde (économique, sociale, culturelle) et l'angle choisi (individuel ou collectif). Plus généralement, toutes les analyses des changements sociaux (au sens large) butent sur ces questionnements d'évaluation des indices, de résolution et d'échelle d'analyse. Si cela était nécessaire, voici donc justifiée l'utilité d'une approche quantitative et spatiale des sociétés et des comportements des individus.

Nous nous sommes prêtés à l'exercice et aux questions méthodologiques qu'il soulève dès 2003 dans le cadre du projet EMIS (<http://www.demographie.net/emis/>). Celui-ci s'étant achevé alors qu'une bonne partie de ses membres venaient d'achever leur thèse et repartaient vers des horizons différents, il n'y a malheureusement pas eu de valorisation directe sur la pauvreté⁴². Nous avons toutefois eu l'occasion de présenter une partie des réflexions issues de ce projet dans un séminaire du Center for International Earth Science Information Network (Oliveau, 2005b). Nous y avons présenté l'approche suivie par le projet (agrégée au niveau de regroupements de villages) et les options méthodologiques suivies (algèbre d'indicateurs cartographiés) en la comparant à la méthodologie de type synthétique que nous avons utilisée dans notre travail de doctorat. Le détail des deux approches est présenté en annexe (page 120).

⁴⁰ On peut estimer le coût de l'information de multiples façon : en termes de temps, de travail ou de moyens pour l'obtenir, etc.

⁴¹ On constate ainsi un effet de lissage avec l'agrégation des données, qui diminue les différences mesurées. Kendall illustre dès 1939 (pp. 41-42) le fait que changer le mode d'agrégation peut changer la valeur des mesures.

⁴² Mais les valorisations indirectes ont été très nombreuses, à commencer par le soutien apporté aux différentes thèses en cours, des co-publications liées au travail commun (mais sur d'autres sujets) et plus largement une dynamique intellectuelle impulsée. Un rapport final a été produit qui retrace l'expérience (<http://www.demographie.net/emis/plan.htm>)

Dimensions spatiales des comportements sociaux

La cartographie des phénomènes sociaux (pauvreté comprise) exhibe des formes et des régularités que l'œil reconnaît aisément comme n'étant pas le fruit du hasard. Cet ordre sous-jacent, ces structures lisibles, pour reprendre les mots-clefs de Roger Brunet sont à mesurer, et surtout à interroger. Il faut « décoder les messages qu'involontairement nous transmet l'espace. » (Brunet, 1990 : 76).

Cette cartographie, nous l'avons évoqué, apporte une « intelligibilité supplémentaire » aux phénomènes étudiés. Nous ne nous en sommes pas privé lors de nos travaux en interdisciplinarité, particulièrement avec des démographes (mais aussi avec des écologues). La description de la spatialité des phénomènes ne manque pas de poser de nouvelles questions, notamment sur la nature des liens que plusieurs phénomènes peuvent avoir entre eux (corrélation). On peut en effet penser qu'un certain nombre de phénomènes sociaux soient corrélés parce qu'ils ont un lien (causal ou non, ce n'est pas la question). C'est la première explication donnée lorsque l'on envisage une corrélation (et c'est étymologiquement et statistiquement exact). On peut aussi envisager que le lien entre les deux phénomènes soient leur localisation et qu'il n'y ait d'autre lien entre eux que la concomitance de leur site. Bien sûr l'argument est extrême. La coïncidence de positionnement est difficile à envisager si elle est suffisamment importante pour entraîner une corrélation statistique... Néanmoins, à bien y réfléchir, la réflexion n'est pas si absurde qu'elle ne pourrait paraître. L'expérience montre plusieurs phénomènes corrélés dont on sait qu'ils "cachent" un troisième phénomène, qui est l'explication de la corrélation entre les deux premiers.

L'outillage théorique et méthodologique de la géographie doit alors être appelé en renfort pour deux raisons. D'abord parce que l'exercice cartographique nécessite de suivre quelques règles pour que la carte reste fidèle aux données⁴³. D'ailleurs, même fidèle aux données, la carte reste statistiquement une forme de simplification (discrétisation) et doit être maniée avec précaution. La lecture de carte est toujours un exercice périlleux (« dangereux » dira même Dominique Badariotti, 2004) qui peut nous laisser voir des relations qui n'existeraient pas ou seraient bien plus faibles. A l'inverse, la carte peut tout aussi bien nous cacher des relations qui n'apparaîtraient pas suffisamment clairement. Comme le rappelait Jean-Claude Müller (1977) : « [...] il n'existe aucun critère intuitif et absolu dans la comparaison des cartes [...] ».

⁴³ Voir l'article d'Isabelle Thomas (2001) qui renvoie vers la bibliographie classique.

Ensuite, l'outillage théorique de la géographie viendra épauler les commentaires fournis en replaçant les formes observées dans un contexte plus large, faisant appel par exemple à la diffusion spatiale et/ou aux interactions. C'est certainement là qu'est l'apport de la géographie aux sciences sociales. Rappeler que les événements observés se situent dans l'espace et que l'espace contraint les individus, ce qui a une influence directe sur leur possibilité d'interagir avec les autres individus et modulent les capacités et les vitesses de diffusion des phénomènes.

Dans le cadre de notre collaboration avec C.Z.Guilmoto⁴⁴ au sein du South Indian Fertility Project (SIFP), nous nous sommes intéressés très tôt au lien entre démographie et processus spatiaux. Cela constituait d'ailleurs le sujet de notre mémoire de D.E.A., qui reprenait les travaux démographiques sur la baisse de la fécondité au Tamil Nadu sous l'angle de leur interprétation géographique (Oliveau, 1999). L'enjeu était de démontrer l'intérêt d'un regard géographique sur des sujets qui constituaient habituellement la "chasse-gardée" des démographes. Parmi les apports de la géographie dans ce cadre précis, nous avons pu montrer, contrairement aux idées reçues, que la baisse de la fécondité était plus importante dans certaines poches rurales du Tamil Nadu qu'au Kérala (Oliveau, 2005). Cela laissait supposer qu'il pourrait y avoir plusieurs zones de départ de la baisse de la fécondité dans le Sud de l'Inde et non pas simplement une diffusion depuis le Kerala. Par ailleurs, la cartographie fine mise en place pour l'occasion a aussi permis de mettre en évidence les voies préférentielles de diffusion de la baisse de la fécondité (axes de communication) et les zones de résistance majeures (reliefs isolés).

L'exemple le plus marquant reste sans doute celui de la diffusion des inégalités de sex-ratio à la naissance en Inde. Tout au long du 20^{ème} siècle le taux de féminité baisse, illustrant les nombreuses discriminations dont les femmes sont victimes. A partir des années 1980, la sélection des fœtus et les infanticides deviennent tellement importants en Inde qu'il devient difficile de nier l'existence du phénomène. Au cours des années 1990, le sujet devient une préoccupation majeure et plusieurs démographes se penchent sur la question. S'appuyant sur la base de données du SIFP, C.Z.Guilmoto adopte pour sa part une position nouvelle en entamant une cartographie fine du phénomène (Guilmoto, 2000). Partant de cette base de données, nous avons reconstruit pour les années 1961, 1971, 1981 et 1991 les données de sex-ratio à l'échelle des 2000 villages (représentant 5 millions d'habitants) des districts de

⁴⁴ Nous ne dirons jamais assez ce que notre travail doit à celui de C.Z.Guilmoto, particulièrement à travers notre association à son projet de cartographie fine du Sud de l'Inde (<http://demographie.net/sifp/>).

Dharmapuri et Salem (les plus touchés par le phénomène dans le Sud de l'Inde). Ce travail de cartographie nous a permis de montrer la diffusion spatiale du phénomène depuis un groupe de villages situés à l'ouest de la ville de Salem (Oliveau, Vella, 2004 ; Vella, Oliveau, 2005). Si l'on fait souvent l'hypothèse de la diffusion spatiale d'un phénomène, il faut des données à différentes dates pour pouvoir la montrer. Or, si les données spatiales sont déjà coûteuses à mettre en place, les données spatio-temporelles (administratives qui plus est) le sont bien plus.

***Linking census through time*⁴⁵**

Au coût du géoréférencement, se rajoute en effet celui du lien des données à travers le temps. Or, les données administratives de type censitaire évoluent rapidement. Dans les pays du Sud encore plus, où les recensements ne font pas l'objet d'une réflexion très poussée quant à leur utilisation dans des perspectives transversales et fines. On distinguera deux grandes catégories de problèmes pour lier les recensements dans le temps. Le premier est d'ordre géométrique, le second d'ordre administratif.

Géométriquement d'abord, les limites administratives varient généralement d'un recensement à l'autre, souvent séparés d'au moins une décennie. Le peuplement évolue, et les frontières des espaces habités changent, amenant l'administration censitaire à les suivre, ou du moins à s'adapter pour mieux les prendre en compte. A cela s'ajoute d'autres considérations, qui peuvent relever du politique, cherchant à changer la catégorisation des espaces considérés et jouant donc avec leur délimitation. On observe ainsi dans le cas indien des changements fréquents de limite de communes pour faire varier leur classification en ville ou village, selon les besoins.

Autant que les changements de contours à proprement parler, les changements géométriques observés sont bien souvent des recompositions à partir de l'existant. Les regroupements simples (les unités A et B ne deviennent qu'une seule unité) ne posent pas de problèmes majeurs, puisqu'ils ne font que diminuer la résolution spatiale. Par contre, on peut envisager trois cas de figures de modifications géométriques plus ennuyeux pour le suivi spatio-temporel des populations, constitués par des transferts et/ou création de territoires (figure 1).

⁴⁵ Titre d'un article de David Martin *et al.* (2002), qui reprend cette thématique à propos des recensements au Royaume-Uni en 1971, 1981 et 1991, et qui montre bien que ces problématiques ne sont pas propres aux pays du Sud...

Type 1 :

Transfert simple
(de A vers B)

Type 2 :

Transferts complexes
(de A vers C et de C vers B)

Type 3 :

Transferts et création
(Transfert de A et B pour la
création de C)

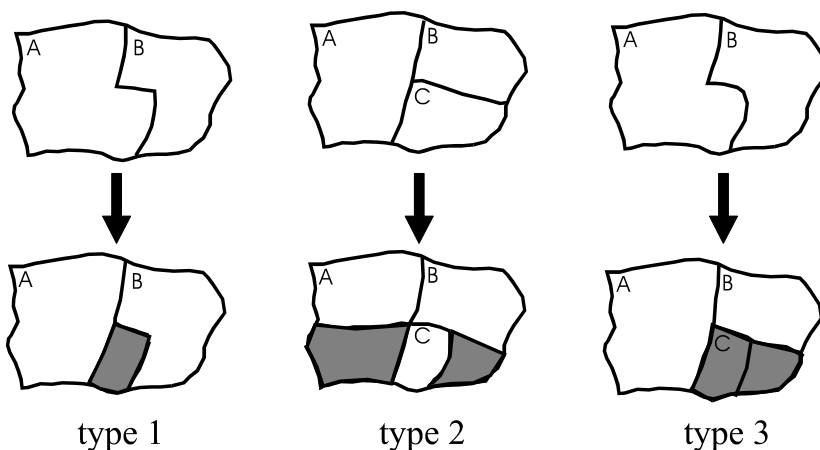


**figure 1 : 3 types de modifications géométriques
(Guilmoto, 1989 : 60)**

A ces considérations s'ajoutent des problèmes administratifs. Souvent liée aux changements de géométrie, la toponymie pose parfois de gros problèmes. Lorsque A et B fusionnent, ils peuvent très bien devenir C ! Par ailleurs, les noms des lieux peuvent changer aussi par simple décision administrative. Il en est ainsi régulièrement. En Inde, Madras est devenue Chennai en 1996, à l'instar de nombreuses autres villes dans le pays. Lorsque les villes sont importantes, il est facile de suivre ces évolutions. Lorsqu'elles sont plus petites, il devient difficile de les suivre⁴⁶. Par ailleurs, outre ces changements de toponymie, peuvent s'ajouter des problèmes de transcription des noms vernaculaires dans les alphabets utilisés par le recensement. Le même nom en tamoul connaît des transcriptions en alphabet latin qui varie au cours du temps... Enfin, il faut rappeler que les recensements ne disposent pas forcément de l'information qu'ils sont censés construire eux-mêmes ! Nous l'avons vécu à deux reprises en Inde. Pendant la thèse nous avons cherché ce que recouvraient les termes « *Municipal Township* » et « *Panchayat Township* » inscrits en face de certaines communes, mais personne n'a pu expliciter ces termes (Oliveau, 2004 : 72). A la même époque, nous avons cherché à localiser des villages pour la base de données SIFP : présents dans les bases de données officielles, une dizaine d'entre eux n'apparaissaient sur aucune carte et n'ont jamais été retrouvés⁴⁷... et la question s'est d'ailleurs reposée pour des villes dépassant parfois les 20.000 habitants pour le recensement de 2001 !

⁴⁶ De façon anecdotique, mais qui illustre bien ce problème, je n'ai pas réussi à dater le changement de nom de la commune française de Chanteloup (78) en Chanteloup-les-Vignes... (merci à Roland Courtot d'avoir fait cette remarque).

⁴⁷ Malgré les recherches poussées au sein de l'administration censitaire par Sattia Vingadassamy, assistant du projet et expert dans ce domaine.

Pour finir, la définition des objets et des individus recensés peut elle aussi varier dans le temps et entraîner une difficulté supplémentaire pour le suivi de la population. Les exemples sont légions, depuis la classification des activités (voir Paulus, 2004, pp. 157-166) jusqu'au genre des individus : une dépêche de l'AFP du 09 janvier 2011 annonçait que le *Central Bureau of Statistics* népalais allait intégrer un troisième genre dans le recensement général de la population (ce que les indiens seraient en passe de faire aussi), il y en aura donc vraisemblablement trois dans les prochains résultats, là où il n'y en avait que deux auparavant... Par ailleurs, le sens donné aux catégories peut lui aussi varier dans le temps. Le statut d'agriculteur en 1950 n'est plus le même aujourd'hui.

On le voit, le changement social est difficile à saisir autant pour des questions de fond que pour des raisons plus formelles. Il nous faudra donc envisager de mettre en œuvre des méthodologies robustes pour pouvoir assurer ce travail. Parmi celles-ci, nous avons pris le parti de nous appuyer sur des jeux de données toujours importants, qui limitent l'importance que pourraient avoir dans les résultats des erreurs ponctuelles. Par ailleurs, nous mettons en œuvre des statistiques robustes, de préférence déjà éprouvées, et que nous remettons aussi à l'épreuve, de manière pragmatique.

Il s'agit alors pour nous de regarder les changements sociaux "de haut", avec suffisamment de recul et un nombre d'individus observés assez important pour pouvoir dégager des tendances et considérer leur transcription dans l'espace. Nous pensons qu'il est nécessaire, sans être suffisant, de connaître les cadres sociaux et spatiaux dans lesquels les individus vivent pour saisir la liberté dont ceux-ci disposent.

Qui se ressemble s'assemble ?

Cette question de la liberté des "acteurs" traverse le champ des sciences sociales. Le reproche souvent fait aux approches quantitatives repose justement sur le fait qu'elles enfermeraient les individus dans des trajectoires façonnées par les modèles. La recherche de régularités oblitérerait la marge d'action individuelle. Il n'en est rien. La recherche de régularités permet de ne pas recommencer éternellement à faire les mêmes constats, et ce faisant à proposer des explications plus générales qui ne soient pas *ad hoc*. Dans le même temps, une fois les régularités établies, nous pouvons mettre l'accent sur les exceptions (il n'y a pas d'exception sans règle...).

Un exemple central dans notre propos est sans nul doute la première loi de la géographie, proposée par Waldo Tobler (1970). Rappelons au passage que l’auteur utilise cette loi comme une pirouette intellectuelle dans un article dont l’ambition était tout autre que de fonder l’analyse spatiale⁴⁸... Que dit cette première loi de la géographie ? « Everything is related to everything else, but near things are more related than distant thing », ce qui peut être traduit en français par : « tout est en relation avec tout le reste, mais les choses proches le sont plus que les choses distantes »⁴⁹.

Contrairement à ce qui lui est souvent (et injustement) reproché, cette loi n’a pas de valeur prédictive mais descriptive⁵⁰. Elle observe une régularité que la culture populaire française à résumer sous forme proverbiale : « qui se ressemble s’assemble ». Une fois le constat fait, et il me semble nécessaire de le faire pour entamer l’activité géographique, on peut envisager deux explications qui amènent à ce résultat.

La première invoque directement la liberté des individus. Puisque les individus sont libres, ils vont chercher à se rapprocher d’autres individus qui leur ressemblent (ce que les économistes appellent gracieusement la maximisation des utilités individuelles). Des processus d’agrégation et de ségrégation se mettent alors en marche qui amène les individus à différencier les valeurs des espaces qu’ils fréquentent ou non. Il en résulte donc que si vous prenez un individu au hasard, vous aurez plus de chances de trouver un individu qui lui ressemble en tirant un autre individu pas trop loin du premier qu’en tirant un individu plus éloigné.

La seconde invoque au contraire l’aliénation des individus à leur environnement. Ainsi, un individu dépend directement des autres individus qu’il fréquente et adopte les caractéristiques de ces individus. Puisque les interactions sont plus fréquentes avec les gens proches qu’avec les gens éloignés, alors, les individus proches se ressemblent juste parce qu’ils sont proches et s’influencent. Il ne s’agit pas d’une proposition tautologique. Les interactions maximum d’un individu sont bien avec sa famille et ensuite les gens qui partagent son activité principale. Or il y a bien un mimétisme minimum qui s’opère, sans quoi la vie sociale devient difficile.

⁴⁸ Un échange intéressant entre géographes américains est proposé dans le numéro 2 du volume 94 des annales de l’association des géographes américains (AAAG) en 2004. On notera particulièrement les propositions de Miller (2004) et Goodchild (2004), ainsi que la réponse générale de Tobler (2004).

⁴⁹ Je n’ai jamais trouvé de traduction en français qui soit aussi élégante que la version originale. C’est un vrai handicap, puisque Tobler (2004) insiste par ailleurs sur la nécessité de la simplicité dans les formulations scientifiques et que l’on y arrive mal ici avec la langue française.

⁵⁰ Et n’avait d’ailleurs la première fois pas de valeur autrement qu’elliptique pour son auteur !

Ces deux explications sont valides et ne s'excluent pas l'une l'autre. On envisage même assez bien que nos vies se construisent en mêlant libertés et contraintes. D'ailleurs, de récentes simulations sur la ségrégation spatiale en milieu urbain montre que la liberté totale et la contrainte absolue n'arrivent ni l'une ni l'autre à reproduire les situations observées. Il faut un mélange de liberté (pouvoir changer de lieu quand notre voisinage nous déplaît) et de tolérance (accepter que notre voisinage ne soit pas exactement tel qu'on le souhaitait) pour obtenir des résultats cohérents (Moreno *et al.*, 2009).

Avant d'envisager les explications, nous pouvons déjà essayer d'établir de manière plus posée ce que l'on entend par « plus en relation ». L'idée communément admise est celle de la ressemblance, même si on pourrait en fait la lire comme l'interaction. Pour notre part, nous resterons sur le sens premier donné par Tobler de ressemblance. Un des premiers objectifs de la géographie pourrait alors être de regarder les ressemblances, autrement dit l'homogénéité des espaces. C'était en quelque sorte ce qui sous-tendait l'approche des études régionales, qui insistaient sur les traits communs aux régions. Néanmoins, c'est jusqu'à présent bien plus les questions de dissemblances qui ont été étudiées à travers les questions de discontinuités et de ruptures (Gay, 1995 ; Brunet *et al.*, 1997 ; Grasland, 1997 ; François, 1998). Plus généralement, les questions de barrières et de frontières ont connu un intérêt marqué ces dernières décennies en France. Pourtant, à bien y regarder, ressemblance et dissemblance ne sont que les deux faces d'une même pièce. L'un ne peut que difficilement se concevoir sans l'autre.

Pour notre part, nous préférons insister sur l'homogénéité des espaces plutôt que sur les ruptures observées. Comme le rappelle Roger Brunet, pourtant opposé aux mesures d'autocorrélation spatiale, « c'est par l'origine, ou le centre, qu'il faut aborder les systèmes spatiaux, et non par leurs limites; car aux limites, où le système est faible, n'importe quel accident mineur suffit pour marquer la rupture et en masquer la raison profonde; mais il n'est que l'occasion, pas la cause » (Brunet, 1967 :301). Ainsi, il conviendrait d'étudier d'abord ce qui fait le système spatial (son homogénéité générale) pour ensuite en définir plus finement les contours (les discontinuités).

Dans cette perspective, nous allons donc chercher à voir dans quelle mesure des espaces voisins se ressemblent ou non. Il nous faudra d'abord revenir sur la qualification de voisins, pour envisager ensuite des méthodes permettant d'évaluer les degrés de ressemblances.

Si l'ambition de la géographie est la mise en évidence et l'explication des processus qui organisent notre espace, celles-ci reposent classiquement sur la description des formes observées. En tant que préalable nécessaire au développement de théories, la qualité de la description reste un enjeu majeur de la discipline. De ce point de vue, et replacée dans le contexte de l'analyse spatiale, l'étude des ressemblances / dissemblances des espaces est sans nul doute un élément central de la description, complémentaire aux études sur les hiérarchies et les échelles. On sait que la définition de l'échelle influe sur la perception des phénomènes et qu'un phénomène visible à un niveau peut ne plus l'être à un autre (Mathian, Piron, 2001 : 64). Il en est de même des voisinages qui peuvent varier et montrer (ou cacher) des relations entre individus voisins (Bailey, Gatrell, 1995 : 35).

Il convient donc de s'attarder sur cette notion préalable à nos travaux, le voisinage. Cette notion, qui peut paraître simple de prime abord, s'avère en fait difficile à formaliser. De plus, cette formalisation repose sur des questions géométriques, qui restent peu explorées en géographie, particulièrement en regard des efforts faits pour le développement des statistiques (Bavaud, 1998).

Les formes de voisinage

La manière de percevoir et la manière de concevoir le voisinage sont extrêmement variables d'un individu à l'autre. Pour s'en convaincre, il suffit de faire un test simple dans une salle de cours -à savoir : demander à un étudiant qui est son voisin- pour s'apercevoir que la perception du voisinage est souvent étroite (l'individu assis à côté de moi à ma gauche et/ou à ma droite) alors que sa conception est souvent plus large. Dès que l'on amène cet étudiant à réfléchir à ce qui fait le voisinage, les limites évidentes (la juxtaposition immédiate) disparaissent et la notion de limite de voisinage devient plus difficile à définir : devant et derrière, les étudiants sont-ils mes voisins ? Le voisin de mon voisin est-il mon voisin ?

De manière plus formelle, il est d'usage d'identifier trois grands modes de voisinage qui reposent sur trois définitions distinctes : la contiguïté, la distance et la proximité⁵¹. Chaque définition a évidemment une logique, des contraintes et des conséquences sur lesquelles il faudra revenir. Il n'y a pas de choix évident ou parfait (Bailey, Gatrell, 1995 ; Bavaud, 1998)

⁵¹ Le chapitre de Dubin (2009) constitue une solide introduction à ces questions.

et la sélection d'un voisinage plutôt que d'un autre restera posée, même si certaines solutions peuvent paraître préférables ou plus logiques dans certains contextes.

Contiguïté

La contiguïté repose sur une définition topologique du voisinage⁵². Elle envisage les objets géographiques comme des surfaces juxtaposées dont la liaison produit le voisinage. Que deux objets aient un point ou une arête commune, et ils seront voisins. On peut par ailleurs étendre cette acception aux voisins des voisins : si un premier objet partage un point ou une arête commune avec un deuxième objet qui lui-même partage une arête ou un point avec un troisième objet, alors le premier et le troisième objet pourront être considérés comme voisins contigus. On précisera alors qu'ils le sont selon un niveau que l'on nommera « d'ordre 2 ».

Le voisinage par contiguïté peut donc prendre une multitude de formes que l'on doit préciser. Tout d'abord la contiguïté peut se faire par le partage d'un point ou d'une arête (figure 2). La première forme de contiguïté sera nommée contiguïté nodale, puisqu'un nœud commun suffit à définir le voisinage. On trouve dans l'informatique le vocabulaire de voisinage de Moore. Luc Anselin (1988 : 19) parle quant à lui de contiguïté de la reine (*queen contiguity*) en référence au déplacement de cette pièce sur un jeu d'échecs. Bouchard (1976 : 536) parle de « définition "diagonale" de contiguïté », mais cela laisse finalement une ambiguïté, particulièrement aujourd'hui où une partie des auteurs utilisent la métaphore échiquéenne pour décrire les contiguïtés.

La seconde forme de contiguïté sera nommée contiguïté latérale (terme présent dans Bouchard, 1976), puisqu'il faut une arête commune (deux points consécutifs) pour être voisins, autrement dit dans ce cas être adjacents. C'est une définition restrictive de la contiguïté. Dans l'informatique, ce voisinage est appelé de « Von Neumann ». Luc Anselin parle de contiguïté de la tour (*Rook contiguity*).

On peut noter qu'il existe théoriquement une troisième forme de contiguïté, qui correspondrait à la différence entre les deux premières, c'est-à-dire celle de deux objets partageant un point seulement. De façon plus concrète, cela correspond au déplacement du fou sur le jeu d'échec et la dénomination de « *bishop contiguity* » existe (Anselin, 1988). On voit mal cependant ce qui pourrait justifier l'utilisation de ce type de voisinage, et c'est sans surprise qu'il est largement absent de la littérature. Nous ne l'envisagerons donc pas ici.

⁵² Miller (2000) propose d'aller vers la notion plus large de « connectivité ». Pour le détail de la connectivité et un renouvellement de l'approche topologique entre voisins, voir Egenhofer & Herring (1994).

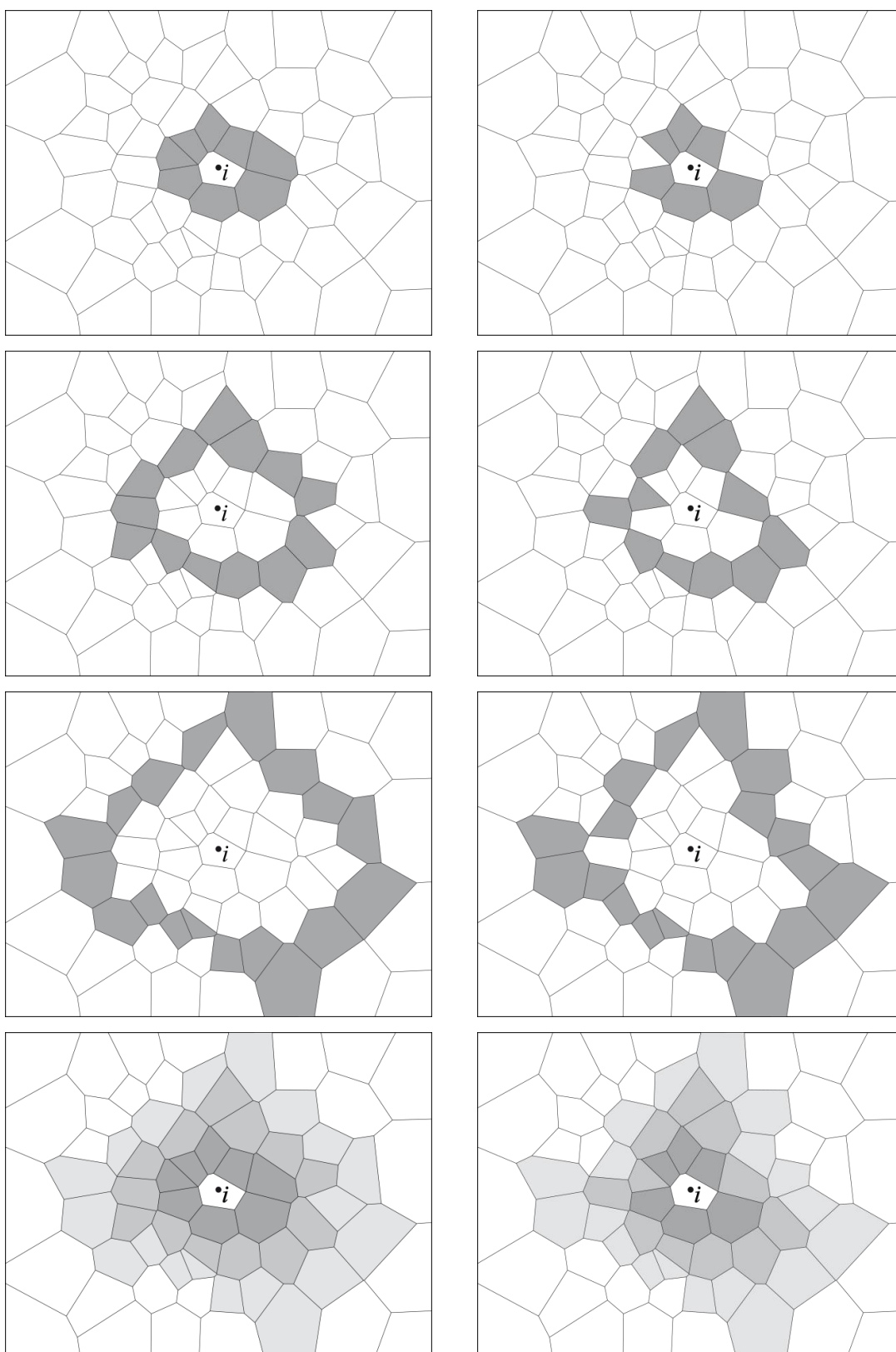


figure 2 : types et ordres de contiguïtés
 (à gauche : contiguïté nodale ; à droite : contiguïté latérale
 de haut en bas : ordre de contiguïté 1,2 3 et tous confondus)

Pour comprendre les implications des différentes formes de contiguïtés sur les mesures d'autocorrélation spatiale, il semble pertinent de s'arrêter un instant sur les cas de figures particuliers que sont les trames géométriques régulières. Le triangle équilatéral, le carré et l'hexagone régulier étant les seuls polygones réguliers qui permettent de remplir complètement un espace, nous allons revenir sur les propriétés de ces différents pavages.

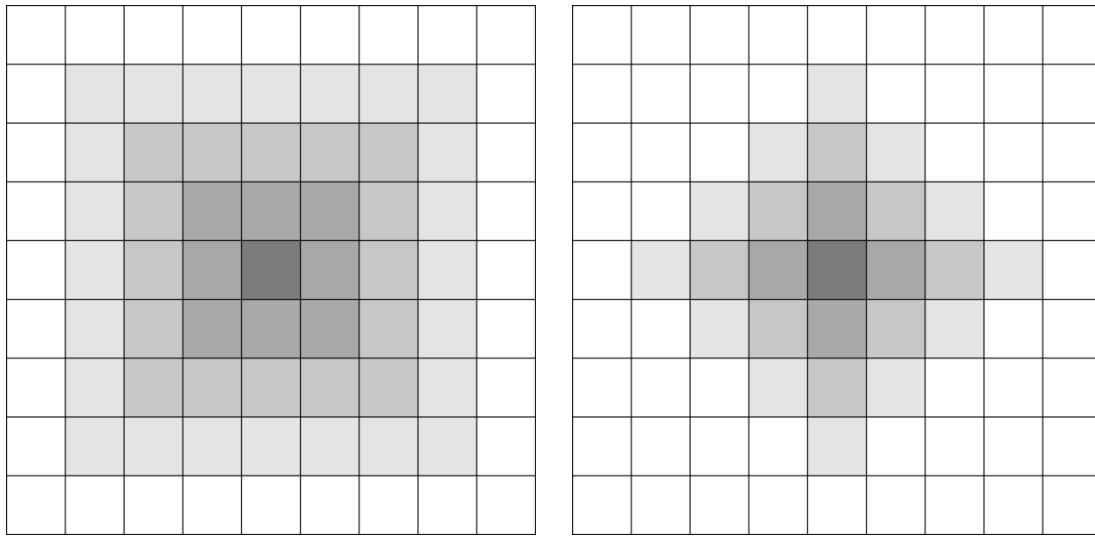


figure 3 : cas particulier de contiguïté : le damier
(à gauche : nodale ; à droite : latérale)

Le plan en damier est un cas simple d'organisation spatiale, que l'on rencontre notamment dans les premiers automates cellulaires. Dans ce type de configuration, la différence entre les contiguïtés latérales et nodales sont très fortes et vont croissant avec l'augmentation de l'ordre de contiguïté.

Pour une contiguïté latérale dans un plan en damier, la formule de récurrence pour calculer le nombre de voisins est : $x_{i+1} = x_i + 4i$ (avec i le niveau d'ordre de contiguïté)

Pour une contiguïté nodale dans un plan en damier, la formule de récurrence pour calculer le nombre de voisins est : $x_{i+1} = x_i + (4 + 4)i$ (avec i le niveau d'ordre de contiguïté)

Le nombre de voisins évolue donc de façon linéaire avec un facteur 4 dans le premier cas et avec un facteur 8 dans le second. La figure 6 résume bien la différence induite par ces deux formes de contiguïtés, et ce dès les premiers ordres de voisinage.

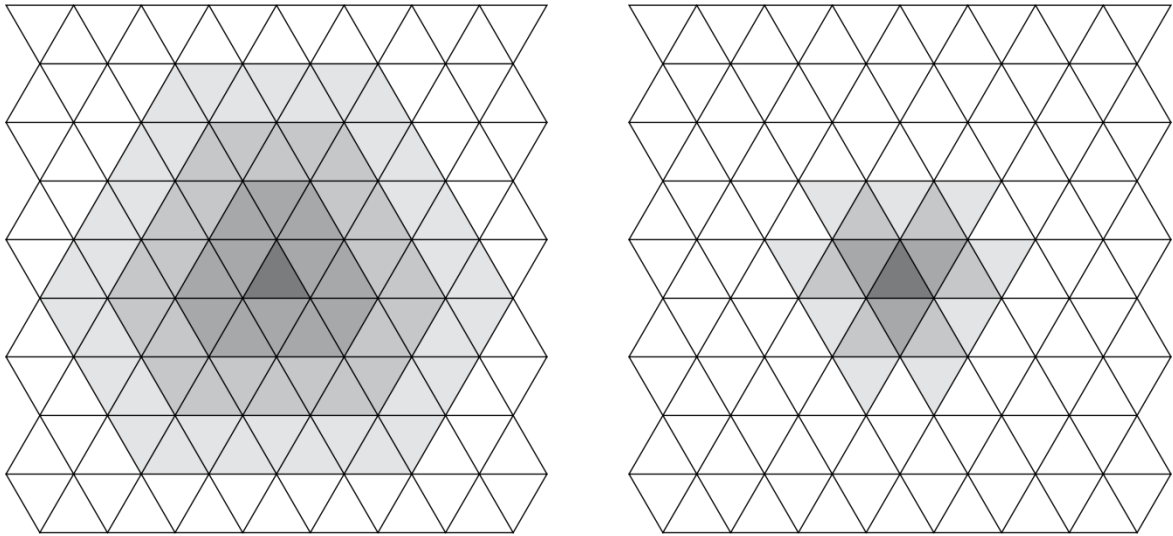


**figure 4 : cas particulier de contiguïté : les triangles
(à gauche : nodale ; à droite : latérale)**

Le cas du pavage en triangle est pédagogique plus que pratique : cette forme ne répond pas à une logique de découpage réel ou modélisé et qui puisse s'interpréter. En outre, il se trouve être assez difficile à mettre en œuvre... Néanmoins, il nous intéresse ici car il permet d'insister sur l'influence du choix du mode de contiguïté sur les résultats. En effet, comme dans le cas du pavage carré, la différence entre les deux modes de contiguïtés est marquée.

Pour une contiguïté latérale dans un plan en triangles équilatéraux, la formule de récurrence pour calculer le nombre de voisins est : $x_{i+1} = x_i + 3i$ (avec i le niveau d'ordre de contiguïté)

Pour une contiguïté nodale dans un plan en triangles équilatéraux, la formule de récurrence pour calculer le nombre de voisins est : $x_{i+1} = x_i + (3 + 3^2)i$ (avec i le niveau d'ordre de contiguïté)

La différence entre les contiguïtés latérale et nodale (facteur 3 versus facteur 12) est encore plus marquée que dans le cas du pavage carré.

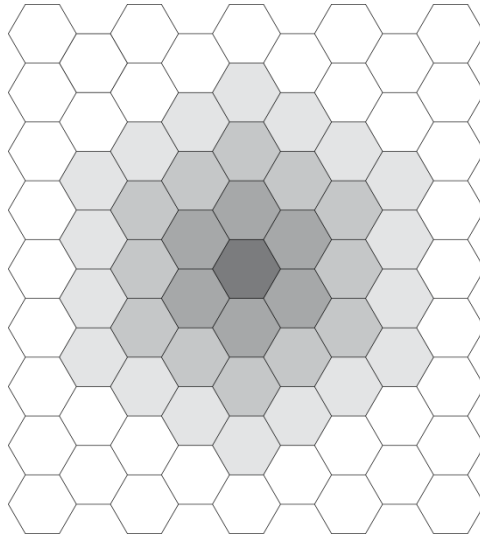


figure 5 : cas particulier de contiguïté : les hexagones

Célèbre en géographie depuis les travaux de Christaller (1933) et ceux de Lösch (1943, 1954), le plan en hexagones réguliers possède un certain nombre de propriétés qu'on ne rappellera pas ici⁵³. Parmi ses propriétés, on doit mentionner que sa configuration fait disparaître la différence entre contiguïté latérale et nodale. Plus exactement il n'y a pas dans ce cas de contiguïté nodale stricto-sensu, puisque la contiguïté est systématiquement latérale. Autrement dit, c'est le pavage le plus simple du point de vue mathématique, puisque sa topologie ne lui confère qu'un type de contiguïté (latéral) et comprenant au plus 6 voisins (contre 8 voisins au maximum pour le carré et 12 pour le triangle).

Pour la contiguïté dans un plan hexagonal, la formule de récurrence pour calculer le nombre de voisins est : $x_{i+1} = x_i + 6i$ (avec i le niveau d'ordre de contiguïté)

Comme le montre de façon synthétique la figure 6, le nombre de voisins varie d'abord selon le type de contiguïté (les contiguïtés nodales construisant plus de voisins) et varie ensuite en fonction du pavage, de façon inverse selon qu'il s'agit d'une contiguïté nodale ou latérale.

⁵³ Pour les plus intéressés par les questions de géométrie, Dacey (1965 : 113) rappelle quelques propriétés des pavages hexagonaux et renvoie lui-même à une bibliographie plus précise et notamment aux travaux de Dirichlet.

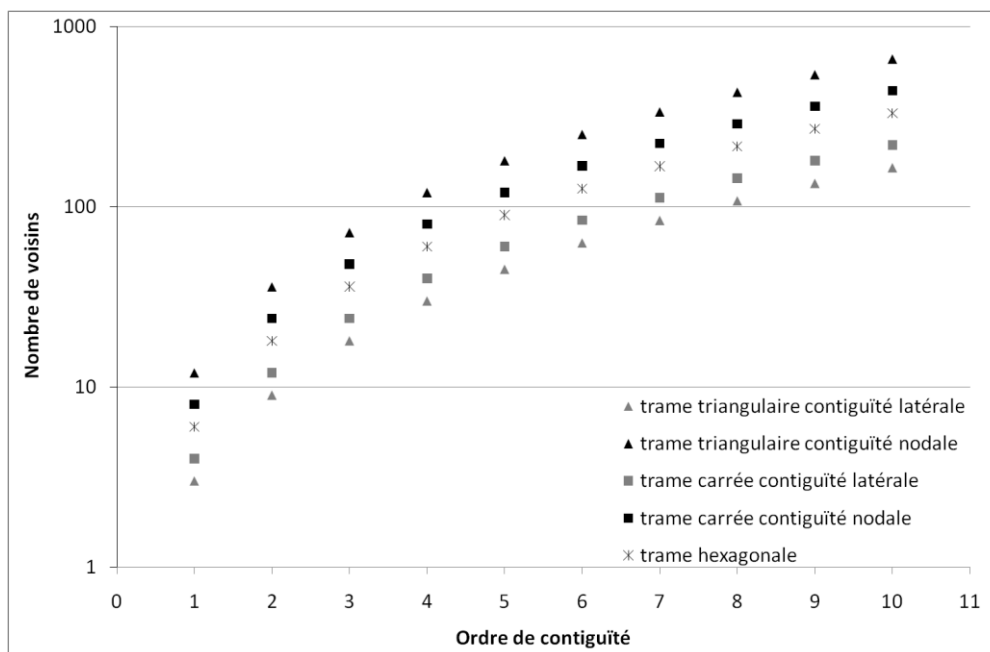


figure 6 : nombre de voisins (échelle logarithmique).

Evidemment, beaucoup d'applications ne se situent pas dans ce contexte, encore que l'on puisse retrouver des trames régulières dans différents endroits, et l'on pensera bien sûr au découpage en townships en Amérique du Nord. D'ailleurs Boots & Tiefelsdorf (2000) soulignent que la trame carrée est celle utilisée pour la télédétection⁵⁴ et que la trame hexagonale ressemble à beaucoup de tessellations empiriques. Si l'on prend une trame plus "classique", les résultats diffèrent alors assez radicalement. Ainsi, pour prendre un exemple plus proche de nous, nous avons produit les mêmes analyses sur les départements français afin de voir comment le nombre de voisins évolue.

La figure 7 présente cet espace, et le tableau 1 (page 40) récapitule les résultats obtenus.

La première remarque concerne l'absence de contiguïté nodale. En effet, il n'y a pas de départements qui ne soient limitrophes que par un point. La seconde, illustrée par la situation de la Corse, marque la première limite de cette approche par la contiguïté, puisque les deux départements de l'île sont certes voisins entre eux, mais n'ont pas d'autre voisin possible.

⁵⁴ Nous avons volontairement laissé de côté la télédétection, bien que ce soit un champ où les mesures de voisinages soient importantes et la littérature sur la question développée (mais de façon très appliquée aux problématiques restreintes de l'image).

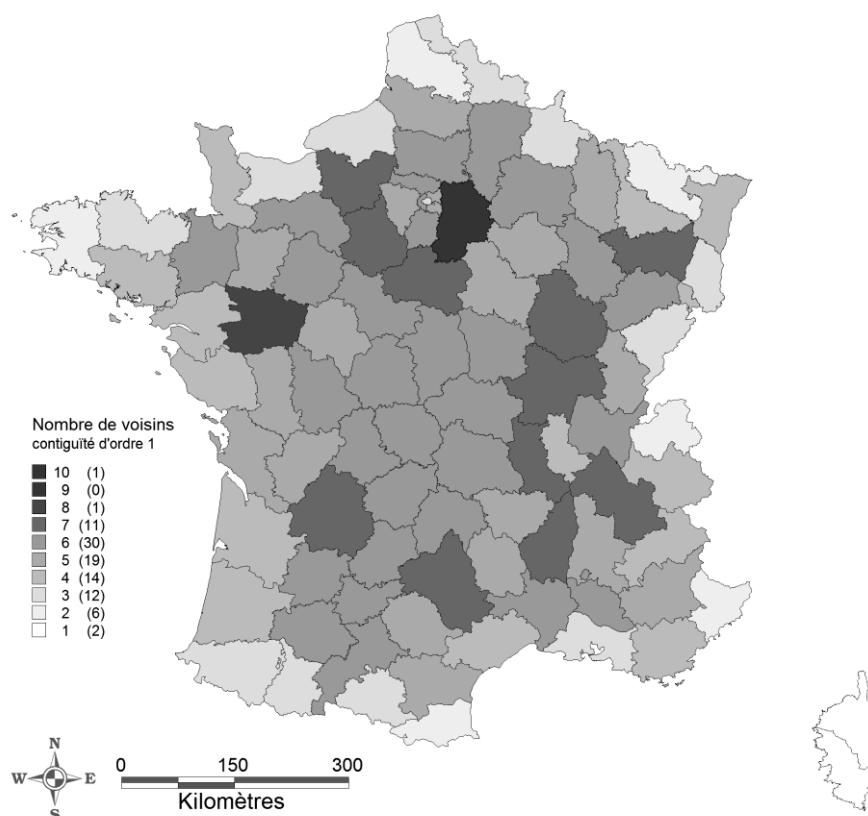


figure 7 : cas pratique de contiguïté, les départements français.
(Les chiffres entre parenthèses indiquent le nombre de départements dans chaque classe)

Si la contiguïté est une forme de voisinage à la fois intuitive et pratique (elle repose sur une mesure topologique simple), elle connaît une limite majeure, soulignée depuis longtemps (Dacey, 1965) mais qui reste d'actualité (Getis, 2009) : elle est invariante selon certaines transformations topologiques. Autrement dit plus simplement : des configurations spatiales différentes seront décrites de la même façon. Nous reviendrons sur ce point particulier plus loin (figure 12, page 45), mais cela justifie déjà l'étude de modes de voisinage autre que la contiguïté.

Par ailleurs, cette conception du voisinage est aussi défailante dans quelques cas particuliers, comme l'existence d'entités géographiques isolées, comme peuvent l'être les îles par exemple, et plus généralement lorsque des discontinuités spatiales existent ("trou" dans la couverture spatiale analysée).

Accessibilité

Si la contiguïté apparaîtrait souvent intuitivement comme une bonne définition du voisinage, la notion d'accessibilité peut paraître, après réflexion, plus pertinente. L'accessibilité renvoie en effet au premier concept de la géographie⁵⁵ : la distance. Distance, dérivé du latin *distantia*, renvoie au préfixe « dis » (qui exprime la séparation) et au verbe « stare », (se tenir, être debout). La distance exprime donc ce qui n'est pas nous et qui se tient séparé de nous. C'est une définition large et pertinente du voisinage : elle nécessite de connaître cet "autre" et de mesurer ce qui nous en sépare. En mesurant ce qui nous en sépare, la distance nous rappelle aussi que nous y sommes lié : la distance relie autant qu'elle sépare (Hall, 1966 ; Paelinck, 2009 : 426). Les voisins sont séparés mais liés par construction.

Que l'on soit dans le domaine strictement mathématique ou que l'on choisisse de s'aventurer dans les représentations individuelles, il existe une multitude de manière de définir la distance (voir la thèse de De Smith, 2003). La définition mathématique de la distance se prête évidemment bien à la formalisation et à la modélisation des voisinages. On rappellera donc en préambule la définition donnée par Huriot et Perreux (1990 : 200) :

« Soit un ensemble quelconque de lieux ; une fonction réelle d définie sur L est une **fonction distance**, ou **métrique**, si et seulement si, elle satisfait les 4 conditions suivantes quels que soient a, b, c appartenant à L

- (c1) non négativité $d(a, b) \geq 0$
- (c2) identité $d(a, b) = 0 \Leftrightarrow a = b$
- (c3) symétrie $d(a, b) = d(b, a)$
- (c4) inégalité triangulaire $d(a, b) \leq d(a, c) + d(b, c)$

Le nombre réel $d(a, b)$ est appelé **distance** de a à b ».

La fonction distance est alors définie par :

$$d_p(A, B) = \left[(x_{AB})^p + (y_{AB})^p \right]^{1/p}, \text{ avec } x_{AB} = |x_A - x_B| ; y_{AB} = |y_A - y_B| ; \text{ et } p \geq 1. »$$

Les termes x_a et x_b sont les longitudes des points a et b (dont on veut calculer la distance), et y_a et y_b leurs latitudes. Le coefficient p est supérieur ou égal à 1. Pour « $p=1$ », on obtient une distance rectangulaire (ou rectilinéaire), qui est aussi appelée « distance de Manhattan ».

⁵⁵ Grataloup présentait la distance comme « l'attribut premier d'un système spatial » (1996, 105), nous allons un peu plus loin en tirant les conséquences de ce positionnement.

La « distance à vol d'oiseau »⁵⁶ est l'appellation populaire de la distance euclidienne. C'est la distance la plus intuitive : le plus court chemin entre deux points. Le coefficient p est alors égal à 2.

On le voit, la définition proposée n'est déjà pas univoque et laisse une infinité de possibilités. On peut rajouter à celles-ci des modulations ou des déformations⁵⁷ en fonction du coût, du temps, des représentations mentales, etc. et l'on comprend alors que si la distance se conçoit aisément, sa formalisation va nécessiter une redéfinition précise dans chaque contexte. Nous en resterons cependant ici à une conception classique de la distance mathématique, qui a pour elle de se modéliser et de pouvoir intégrer des programmes informatiques facilement. Nous utiliserons ici la distance euclidienne, autrement dit la ligne droite entre deux points, qui constitue un compromis pertinent (Peeters et Thomas, 1997 ; Berroir, 1998 ; Oliveau 2004).

Si le voisinage par contiguïté nécessite absolument une trame polygonale, le voisinage par la distance peut s'effectuer aussi bien sur des polygones que sur un semis de points. Cette première possibilité (les polygones) pose plusieurs problèmes qui expliquent sa moindre utilisation, notamment celui de savoir comment on mesure la distance entre deux polygones. On peut en effet envisager que l'on mesure la distance entre les deux points les plus proches de chacun des polygones (auquel cas deux polygones adjacents ont une distance de 0) ou au contraire en fonction des deux points les plus éloignés. On peut aussi envisager de prendre la moyenne des distances entre points, etc. Par ailleurs la difficulté d'implémentation de ce genre de calcul fait préférer des approches par semis de points. Autrement dit, lorsque l'on dispose d'une trame polygonale et que l'on souhaite utiliser une mesure de voisinage par la distance, l'usage le plus répandu (pour ne pas dire quasi systématique) est d'utiliser un point de référence, bien souvent le centroïde du polygone (que les systèmes d'informations géographiques savent très bien déterminer). Ceci n'est pas sans poser encore des problèmes, puisque les centroïdes ne correspondent pas aux centres de gravité, qui peuvent eux-mêmes se trouver hors des polygones. On pourrait envisager de définir des points *ad hoc* pour chaque mesure (les préfectures pour les départements par exemple ou les centres anciens pour les villes), mais cela ne constitue pas, on le devine une solution satisfaisante.

⁵⁶ On ne résiste pas ici à rappeler la réflexion de Pierre Dumolard (1999) : « a-t-on déjà vu des oiseaux voler tout droit d'un point de départ à un point d'arrivée ? », qui nous invite à ne pas forcément se laisser aller à des modélisations trop simplistes.

⁵⁷ On parle de déformations pour envisager le relâchement des conditions dictées par la définition mathématique, notamment la symétrie et l'inégalité triangulaire, quand les modulations ne seraient que des facteurs ajoutés au calcul de la distance.

La distance va donc définir une zone au sein de laquelle le point *i* (figure 8) va pouvoir interagir avec les autres points. Il y a bien l'idée d'accès aux autres dans un pas de distance donné. L'accessibilité met en avant la notion d'espace pour définir son voisinage, et non plus la topologie. L'interprétation et la communication sont d'ailleurs grandement facilitées, puisque l'on va envisager le voisinage « dans un rayon de », ce qui est plus évocateur que la contiguïté dans une perspective géographique.

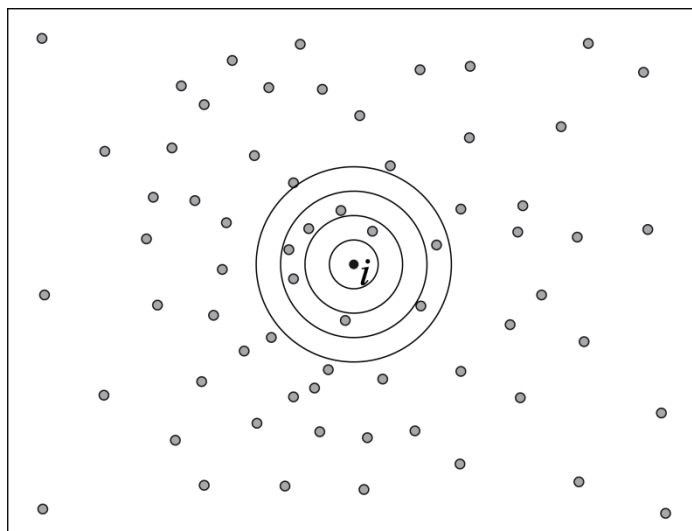


figure 8 : exemple de voisinage défini par la distance

On remarquera cependant que la distance est généralement interprétée de façon absolue : on considère que tous les points dans le rayon de portée envisagé ont la même probabilité d'interaction, et partant, le même poids. On peut cependant envisager de se placer dans un système d'interactions spatiales où la probabilité du lien entre le point d'origine et le point voisin est fonction de la distance (Wong, 2011, effectue différents tests d'autocorrélation spatiale en variant les méthodes de prise en compte de la distance : distance simple, inverse de la distance, carré de l'inverse de la distance). Claude Grasland (2004) oppose dans cette perspective l'analyse spatiale (qui considère l'interaction spatiale en prenant en compte les distances) à la statistique spatiale (qui considère l'autocorrélation spatiale en prenant en compte l'influence de la contiguïté). Nous sommes enclins à le suivre sur cette distinction.

Proximité

La proximité est une troisième manière de définir le voisinage. Il ne s'agit plus là de regarder la contiguïté ou la distance autour d'un objet mais sa proximité. La proximité se

définissant comme la plus courte distance à un objet, on parle souvent de « plus proches voisins » (*k nearest neighbours* en anglais).

On peut dénombrer au moins 3 avantages à cette méthode. Elle est d'abord très simple à mettre en œuvre puisqu'il suffit de mesurer la distance entre les objets, de les ordonner en fonction de cette distance et de décider ensuite combien l'on veut de voisin(s). Le nombre de voisins étant décidé par l'utilisateur, cela crée un réel avantage lors de mesures sur des populations importantes, car il permet de maîtriser le temps de calcul (même si l'accélération continue des vitesses de calcul des ordinateurs rend cette remarque un peu moins pertinente chaque jour). Enfin, cette méthode propose une solution où tous les individus ont le même nombre de voisins, ce qui constitue un atout pour certaines mesures d'autocorrélation spatiale où le nombre de voisins entre en considération (voir les formules des différents indices d'autocorrélation spatiale pages 56 et 60).

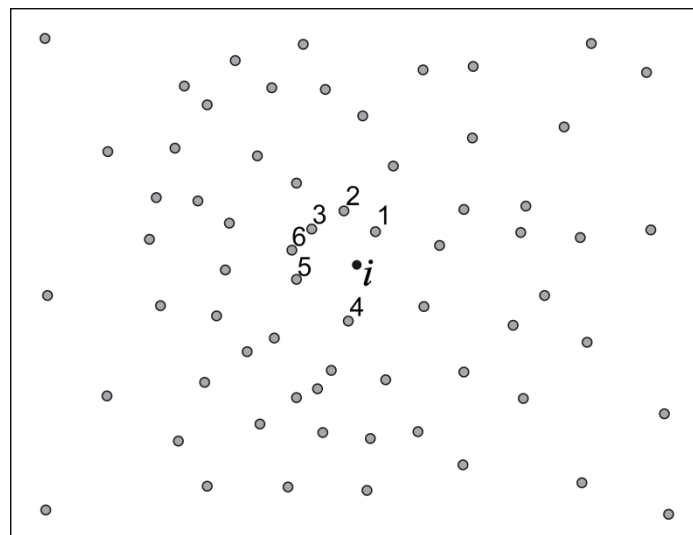


**figure 9 : exemple de voisinage défini par la proximité
6 plus proches voisins du point *i* (centroïdes des polygones de la figure 2)**

Néanmoins ces avantages sont grevés par des inconvénients à ne pas négliger, et qui peuvent être suffisants -à notre avis- pour justifier son rejet dans beaucoup de cas concrets d'utilisation.

Tout d'abord, sur un maillage régulier (voir plus haut), la proximité est prise en défaut : il y a 3 plus proches voisins de niveau 1 pour le triangle, 4 pour le carré, 6 pour l'hexagone. Cela signifie que la maîtrise du nombre de voisins, avantages définis ci-dessus n'est pas strictement garanti. Sur un maillage irrégulier aussi, on peut s'attendre à trouver deux

individus à la même distance de l'individu de référence (des plus proches voisins ex-æquo en quelques sortes...).

Ensuite, et dans le même ordre d'idée, la proximité peut donner des résultats largement contre-intuitifs, voire absurdes. Selon les configurations envisagées, un voisin proche pourrait ne pas être contigu alors qu'un voisin contigu ne serait pas proche. Autrement dit, la méthode de construction de la proximité n'assure pas l'isotropie du voisinage, ce qui peut poser problème (l'exemple de la figure 9 montre une anisotropie marquée pour le point *i* malgré la prise en compte de 6 plus proches voisins).

Enfin, le choix du nombre de voisins à prendre en compte est encore plus difficile à faire que celui d'un ordre de contiguïté ou d'un pas de distance, sauf à répéter les mesures pour différents ordre de proximité. Il paraît néanmoins possible que l'on s'achemine vers des solutions mathématiques (Rogerson, 2010 ; Rogerson & Kedron, 2011) qui donnerait alors un regain d'intérêt à cette méthode dans certains cas de figure.

Au-delà des avantages et inconvénients de la méthode de définition du voisinage par la proximité, il existe une caractéristique particulière de ce mode de voisinage qu'il convient de remarquer. Par construction, un objet A voisin d'un objet B n'inclut pas que l'objet B soit voisin de l'objet A. En effet, A peut être l'objet le plus proche de B sans pour autant que B soit l'objet le plus proche de A. Autrement dit en termes mathématiques, la matrice de voisinage ne sera alors pas symétrique. Cela peut certes questionner, mais on peut aussi l'interpréter comme la traduction de l'influence inégale des individus.

Comment choisir ses voisins ?

Comme le fait remarquer Getis (2009), la question de la définition d'un voisinage optimal pour les mesures d'autocorrélation spatiale a été globalement laissée de côté. Quelques auteurs ont néanmoins tenté d'aborder ce sujet difficile (Bavaud, 1998, Rogerson, 2010). Sans vouloir répondre de manière définitive, Getis apporte un premier élément de typologie en distinguant trois approches possibles, qu'il nomme « théorique » « topologique » et « empirique ». Pour faire bref, on pourrait dire que l'approche théorique renvoie à ce que nous avons appelé accessibilité, ce qui constitue pour l'auteur un progrès par rapport à l'approche topologique, qui regroupe les approches de la contiguïté et de la proximité. Enfin, il envisage qu'une approche empirique soit possible, plus rigoureuse et émanant de la structure même des

données⁵⁸. L'idée est de regarder d'abord comment se construisent les relations spatiales entre les données afin de définir le voisinage à adopter. La voie reste à ouvrir, les travaux sur le sujet étant assez peu nombreux, et les méthodes proposées très (trop ?) lourdes à mettre en œuvre.

Au final, l'approche théorique semble constituer un bon compromis et paraît plus facile à mettre en œuvre. Pour améliorer son choix et mieux maîtriser ses résultats (éviter les mésinterprétations), le chercheur est invité à varier les approches pour regarder les différences par exemple en changeant les pas de distance choisis.

Pour se persuader de l'impact du type de voisinage, et plus généralement en préambule d'analyses reposant sur le voisinage, un exercice simple peut être effectué. Il consiste à regarder le nombre de voisins pour un espace existant, afin de voir quel est l'impact des différentes méthodes sur le nombre de voisins observés. Sans prétention démonstrative, cet exercice est néanmoins éclairant et pallie au moins partiellement l'absence actuelle de méthode pour définir un voisinage optimal.

Méthode	Facteur	Nombre de voisins		Nombre de voisins moyen	Coefficient de variation du nombre de voisins
		minimum	maximum		
Contiguïté latérale	Ordre 1	1 (2)	10	5,0 (5,0)	0,33 (0,31)
Contiguïté nodale	Ordre 1	1 (2)	10	5,0 (5,0)	0,33 (0,31)
Contiguïté nodale	Ordre 2	1 (4)	25	14,1 (14,4)	0,36 (0,33)
Contiguïté nodale	Ordre 3	1 (8)	45	26,3 (26,8)	0,36 (0,33)
Distance	Distance minimale	1 (1)	12	5,2 (5,2)	0,48 (0,46)
Plus proches voisins	Nombre de voisins	égal au nombre choisi			-

tableau 1 : Impact des méthodes sur le nombre de voisins : exemple des départements français (entre parenthèses, résultats sans les deux départements corses).

L'espace décrit comporte des bords et les unités situées en bordure ont donc moins de voisins (voir p. 44) On ne peut donc pas comparer ces résultats avec ceux obtenus précédemment pour les trames régulières.

Pour notre part, nous avons proposé dans notre thèse de doctorat (Oliveau 2004) un récapitulatif des avantages et inconvénients des trois méthodes de prise en compte du

⁵⁸ On retrouve ici l'approche de l'analyse spatiale qui souhaite laisser les données parler d'elles mêmes (« letting the data speak for themselves », Gould, 1981).

voisinage. Nous proposons ici un tableau mis à jour, plus complet, qui reprend cet exercice (tableau 2).

Type de voisinage	Facteur déterminant	Avantages	Inconvénients
Contiguïté nodale	Ordre de contiguïté	- Isotropie assurée par construction - Voisinage défini par la topologie	- Tous les voisins sont égaux, quelque soit leur éloignement réel. - Variabilité du nombre de voisins selon les points. - Espace continu nécessaire - Invariance d'échelle
Contiguïté latérale	Ordre de contiguïté (secondairement : forme géométrique des unités)	- mesure intuitive du voisinage - Voisinage défini par la topologie	- Tous les voisins sont égaux, quelque soit leur éloignement réel. - Variabilité du nombre de voisins selon les points. - espace continu nécessaire - Invariance d'échelle.
Distance	Portée maximum et/ou fonction de la distance choisie	- Isotropie assurée par construction - Voisinage défini par une notion spatiale. - Ajustement possible de la distance en fonction de la connaissance du semis de points. - prise en compte possible d'un espace discontinu	- Très grande variabilité du nombre de voisins selon l'irrégularité du semis. - "aveugle" aux discontinuités.
Plus proches voisins	Nombre de voisins	- Tous les points ont le même nombre de voisins. - Maîtrise du nombre de voisins.	- Pas de prise en compte de l'éloignement des points. - Pas de prise en compte de la contiguïté. - Isotropie non assurée. - Invariance d'échelle.

tableau 2 : Avantages et inconvénients des différents type de définition du voisinage.

Nous pouvons tirer deux conclusions de ces comparaisons. La première porte sur l'absence actuelle de capacité à déterminer une méthode de voisinage optimale (Rogerson 2010, Getis, 2009), ce qui laisse par ailleurs une voie ouverte pour les plus "matheux" des géographes ou pour les plus spatialistes des mathématiciens. Cela nous invite surtout à la plus grande modestie quant aux solutions proposées par la suite. Nous gardons une préférence pour la prise en compte de la distance comme mode de définition du voisinage, parce que ce choix

défend une position "spatialiste" et indique une volonté d'aller vers plus de théorie⁵⁹. A l'inverse, le choix des plus proches voisins reste pour nous le plus difficile à justifier, à moins d'envisager de longues batteries de tests visant à choisir un nombre plus qu'un autre afin d'optimiser les résultats.

La seconde conclusion porte sur le soin à apporter -dans la mesure du possible- à la trame spatiale utilisée. Plus celle-ci sera régulière et tendra vers une forme hexagonale, plus les résultats des différentes formes de voisinage seront semblables et plus la mesure sera donc robuste (au sens où elle devient indépendante du mode de voisinage). On aurait donc intérêt par exemple à envisager des réagréations, lorsque cela est possible, pour obtenir des trames plus adéquates. On doit d'ailleurs faire ici un parallèle avec les questions concernant le MAUP (que nous abordons de façon plus détaillée page 83). On constate ainsi que cette recherche du mode de voisinage optimal théorique rejoint la question des structures spatiales permettant de minimiser les effets d'échelle et de zonage. Les hexagones constituent de ce point de vue une solution qui semble optimale.

Enfin, on se rappellera que quelque soit le choix effectué, il apparaît toujours nécessaire de regarder de manière attentive les impacts des différents choix de voisinage sur les résultats des mesures effectuées avant d'aller plus loin dans les analyses.

Les matrices de voisinage

Une fois défini qui peut être voisin de qui, il convient de formaliser cette relation. La matrice de voisinage met en vis-à-vis les individus en rappelant leur lien.

Matrice binaire

Une matrice de voisinage se présente sous la forme d'un tableau à double entrée où les individus sont présents deux fois, en ligne et en colonne. De cette manière le lien entre deux individus voisins est noté par la valeur « 1 » et l'absence de relation par la valeur « 0 ». Il s'agit donc d'une matrice binaire dont la diagonale est logiquement nulle, puisqu'il est difficile de considérer qu'un individu puisse être voisin avec lui-même. Nous verrons plus

⁵⁹ Un parallèle peut être fait avec les discrétisations faites pour la cartographie. Même si il est parfois plus parlant de faire une discrétisation manuelle pour exprimer de manière concrète des valeurs, il est toujours plus juste de suivre les méthodes issues de la théorie : méthode standard pour une distribution gaussienne par exemple.

loin que cette hypothèse peut néanmoins être envisagée, auquel cas la diagonale prend alors la valeur « 1 ».

Dans le cas de voisinage par contiguïté (latérale ou nodale) et dans le cas d'un voisinage basé sur la distance, la matrice est symétrique⁶⁰. Dans le cas des plus proches voisins, la symétrie n'est pas respectée, puisque comme nous l'avons écrit plus haut, ce mode de voisinage autorise que A soit voisin de B mais que B ne soit pas voisin de A.

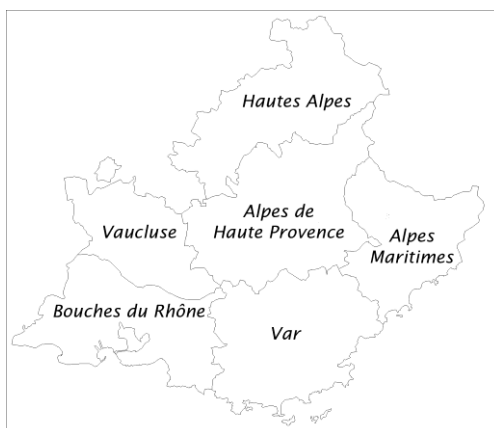


figure 10 : les 6 départements de la région PACA.

	Alpes-de-Haute-Provence	Hautes-Alpes	Alpes-Maritimes	Bouches-du-Rhône	Var	Vaucluse
Alpes-de-Haute-Provence	0	1	1	0	1	1
Hautes-Alpes	1	0	0	0	0	0
Alpes-Maritimes	1	0	0	0	1	0
Bouches-du-Rhône	0	0	0	0	1	1
Var	1	0	1	1	0	1
Vaucluse	1	0	0	1	1	0

figure 11 : matrice de voisinage (contiguïté) pour les départements de PACA.

⁶⁰ La symétrie d'une matrice se lit en fonction de sa diagonale qui sert d'axe de symétrie.

Effets structurels

Comme nous l'avons fait remarquer précédemment, sauf à utiliser la proximité pour définir le voisinage, les individus n'ont pas tous le même nombre de voisins. La figure 11 le montre bien malgré son nombre réduit d'individus. Ce que la matrice ne montre pas, c'est que le nombre des voisins est directement lié à deux éléments : la situation et la morphologie des unités spatiales. Elle ne montre pas non plus un dernier phénomène remarquable : l'invariance d'échelle des mesures de contiguïtés.

Effet de bord

La situation des unités par rapport à l'ensemble de l'espace étudié a un impact direct sur leur nombre de voisins. Ainsi, les unités les plus centrales ont une probabilité plus importante d'avoir plus de voisins que celles situées sur les marges. On le comprend aisément : les unités situées sur les bords auront des parties de leur périmètre sans voisin, alors que celle situées au centre seront littéralement cernées. Le simple exemple des 6 départements de PACA le montre déjà : le Var et les Alpes de Hautes Provence, en situation centrale ont 4 voisins⁶¹ alors que les Hautes Alpes n'en ont qu'un. Cet effet de bord se retrouve systématiquement et il conviendra de le prendre en compte. La figure 7 offre une autre illustration de l'effet de bord.

Morphologie

Le périmètre, la surface et la compacité des unités spatiales jouent aussi un rôle sur le nombre de voisins, mais différemment selon le type de définition choisie. Nous l'avons montré pour les polygones réguliers précédemment, et l'on envisage que cela joue aussi un rôle important pour des polygones irréguliers. Il n'y a pas de règle établie à notre connaissance. Mais une unité avec un périmètre plus important aura une probabilité plus forte d'avoir plus de voisins contigus. De même qu'une unité de grande surface a moins de chance d'avoir des voisins si l'on prend la distance en considération, etc. Autrement dit, et comme nous l'avons déjà remarqué précédemment, il y a un avantage certain à avoir des trames les plus régulières possible pour éviter ou du moins diminuer l'effet de la morphologie des unités spatiales sur la mesure de leur voisinage.

⁶¹ Le Var et le Vaucluse partagent environ 500 mètres de frontières, qui isolent les Bouches-du-Rhône des Alpes-de-Haute-Provence...

Invariance d'échelles

Il y a un dernier élément que nous souhaitons mettre en exergue : l'invariance d'échelle des voisinages définis par la contiguïté (ou la topologie, dans le vocabulaire de Getis). Comme nous le constatons sur la figure 12, le voisinage selon la contiguïté est invariant selon l'échelle choisie. Cela signifie que des configurations spatiales de formes identiques mais de dimensions différentes peuvent avoir la même matrice de voisinage. De même, des configurations spatiales différentes peuvent produire des matrices similaires. On retrouve typiquement cette invariance d'échelle dans les structures fractales, qui caractérisent notamment le bâti urbain (Batty, Longley, 1994, Frankhauser, 1998). Or, cette invariance d'échelle constitue un biais pour la mesure de l'autocorrélation spatiale (Dacey, 1965).

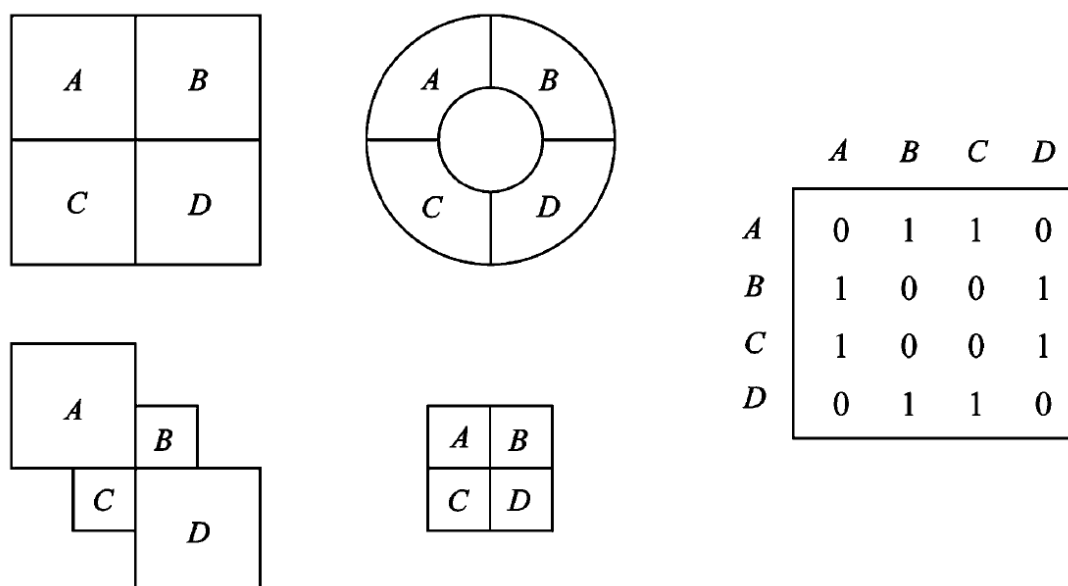


figure 12 : invariance d'échelle du voisinage selon la contiguïté (repris de Getis 2009 d'après Cliff & Ord, 1973).

La figure 12 nous montre quatre configurations spatiales différentes (à gauche) représentées par une même matrice de voisinage (à droite) si l'on prend en compte la contiguïté latérale. En choisissant une contiguïté nodale, trois des quatre configurations gardent un voisinage identique.

Matrice de poids spatiaux

Si la matrice de voisinage donne de façon binaire les relations entre unités spatiales (1 nous sommes voisins, 0 nous ne le sommes pas), nous avons envisagé que les voisinages pourraient

en fait être pondérés, par exemple par une fonction de la distance (voir page 37). Cela consisterait alors à remplacer les 0 et les 1 par la distance mesurée entre les unités spatiales ou toute autre fonction de la distance. Dans le même ordre d'idée, on peut très bien envisager de transformer les matrices pour prendre en compte l'ordre de contiguïté ou des plus proches voisins. La matrice n'est alors plus binaire, mais devient une matrice de poids spatiaux.

Si la matrice de poids spatiaux (que l'on pourra nommer en raccourci matrice de pondération⁶²) offre des possibilités de raffinements infinis, elle n'offre pas forcément beaucoup plus de lisibilité. Certes, la prise en compte du voisinage relatif (distance effective, niveau de contiguïté réel, longueur de périmètre commun, etc.) est un atout intéressant, mais il n'existe pas de procédure pour déterminer la fonction à appliquer à ces voisinages relatifs : doit-on employer la distance, son inverse, le carré de son inverse, une fonction gaussienne ? Et pour la contiguïté, est-ce que l'inverse de l'ordre de contiguïté représente bien la différence entre un voisin direct et le voisin du voisin ? Autant de questions qui restent pour l'instant en suspens.

Nous pouvons même aller encore plus loin en intégrant dans la matrice des éléments extérieurs à la stricte définition du voisinage. Ainsi, le poids donné à chaque individu en fonction de son rapport de voisinage pourrait être lui-même pondéré par une autre variable, comme l'importance de sa population si l'on pense à des unités démographiques. La matrice n'est alors plus strictement spatiale, mais cela n'est pas essentiel.

Standardisation en ligne

Un dernier élément doit être noté avant d'aller plus loin. Comme nous l'avons fait remarquer précédemment à propos des matrices binaires, tous les individus, en fonction de leur nombre de voisins n'auront pas le même poids dans les matrices. Pour pallier ce qui devient rapidement un biais statistique, une mesure élémentaire peut être prise. Elle consiste à diviser le poids de chaque voisin par le poids total de l'individu de référence dans la matrice. Autrement dit à appliquer une standardisation en ligne de la matrice. De cette manière la somme des poids de chaque ligne est égale à 1, les individus reprennent donc tous la même importance.

Cela permettra aussi de comparer des matrices issues de définitions de voisinages différentes, puisque les changements observés ne seront plus fonction du nombre de voisins

⁶² Bavaud (1998), qui parle de « spatial weights matrix » nous rappelle que le vocabulaire est varié en anglais : « contiguity », « connectivity », « adjacency » ou « association » matrix.

mais des relations établies entre eux. Enfin, cela simplifiera l'écriture d'un certain nombre d'indices, puisque la somme des paires de voisins pondérés est égale à la somme des individus (voir page 57).

	Matrice binaire						Matrice standardisée en ligne					
	Alpes-de-Haute-Provence	Hautes-Alpes	Alpes-Maritimes	Bouches-du-Rhône	Var	Vaucluse	Alpes-de-Haute-Provence	Hautes-Alpes	Alpes-Maritimes	Bouches-du-Rhône	Var	Vaucluse
Alpes-de-Haute-Provence	0	1	1	0	1	1	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$
Hautes-Alpes	1	0	0	0	0	0	1	0	0	0	0	0
Alpes-Maritimes	1	0	0	0	1	0	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0
Bouches-du-Rhône	0	0	0	0	1	1	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
Var	1	0	1	1	0	1	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$
Vaucluse	1	0	0	1	1	0	$\frac{1}{3}$	0	0	$\frac{1}{3}$	$\frac{1}{3}$	0

figure 13 : standardisation en ligne de la matrice de voisinage (contiguïté) pour les départements de PACA.

On le voit, la standardisation en ligne redonne un poids égal à tous les individus, notamment ceux situés aux marges des espaces étudiés. Du point de vue de l'importance des individus, l'effet de bord est donc annulé. Néanmoins, il reste un effet de bord qui concerne la variabilité des résultats que l'on peut obtenir. Les individus sur le bord des espaces étudiés ayant moins de voisins auront donc des résultats qui varieront plus et seront moins significatifs, puisqu'il n'y aura pas d'effet de lissage qu'apporte un nombre plus important de voisins, mais l'effet reste mineur.

L'observation et la recherche de régularités dans l'espace amènent à se poser la question de l'existence de structures spatiales, puis de leur mesure. Cette mesure des structures spatiales peut prendre différentes voies (morphologie mathématique, fractales, etc.), et nous avons choisi pour notre part de nous concentrer sur celle qui permet de prendre en compte les dimensions sociales des données en plus de leurs caractéristiques strictement spatiales. Pour être plus exact, nous avons pris le chemin qui cherchait à croiser l'espace et les statistiques.

Ce croisement de la statistique et de l'espace a été rendu possible par la convergence de deux courants issus des mathématiques, la statistique spatiale d'une part et la géostatistique d'autre part. La statistique spatiale apparaît au milieu du 20^{ème} siècle, lorsque des mathématiciens cherchent à mettre en évidence les perturbations créées par la position spatiale relative des objets entre eux. Il est d'usage de considérer Moran (1948, 1950) et Geary, (1954) et dans une moindre mesure⁶³ Krishna-Iyer (1949) comme les trois premiers statisticiens à formaliser cette relation entre valeur d'une variable pour un individu et valeurs de la variable de ses voisins. La géostatistique apparaît une dizaine d'années plus tard. On considère que l'ouvrage de Georges Matheron *Les variables régionalisées et leur estimation* (1965) marque la naissance de la géostatistique. La géostatistique s'intéresse à la prédiction des valeurs d'une variable spatialisée en fonction des valeurs déjà connues de cette variable. Elle a permis de grands progrès dans la mesure des structures spatiales, leur prise en compte et leur représentation graphique et cartographique (krigeage). Sans intervenir au départ dans le transfert des statistiques spatiales à la géographie, la géostatistique en est aujourd'hui difficilement dissociable⁶⁴. En 1988, dans l'introduction de son ouvrage sur l'économétrie spatiale, Anselin fait justement remarquer que la distinction entre les deux disciplines n'est pas claire⁶⁵ et que pour éviter les confusions, il parlera plutôt d'économétrie spatiale.

⁶³ La *join count statistic* proposée par Krishna-Iyer a connu des développements dans les années 1960, et sert de référence à de nombreux travaux. Il constitue une voie intéressante pour les mesures de « colocalisation » de variables qualitatives -en anglais *colocation*- (Dacey, 1965 ; Leslie, Kronenfeld, 2011), autrement dit pour la mesure de l'autocorrélation spatiale de variables qualitatives.

⁶⁴ « The field of geostatistics includes concepts such as spatial dependence, correlation range, support, and nugget effects. All of these terms are of direct relevance in the context of analyzing spatial variation and spatial autocorrelation, and allowing for area-size effects (support) in spatial analysis » (Haining, 2009 : 370).

⁶⁵ Fotheringham (2009), s'appuyant sur Goovaerts (1997), rappelle que les deux disciplines ont évolué en parallèle sur des sujets communs, développant des vocabulaires qui leur sont propres.

En géographie, l'article de référence est celui de Cliff et Ord paru en 1969⁶⁶, et plus encore l'ouvrage qui suivra en 1973, simplement intitulé « *spatial autocorrelation* ». Il introduit une version profondément remaniée des statistiques de Moran et Geary, qui a fait dire à Getis qu'elle aurait dû être renommée « Cliff et Ord's *r* »⁶⁷ (Getis 2009). Ces travaux ne doivent pas faire oublier l'intense activité autour de ces questions dans la géographie de langue anglaise, et notamment les travaux précurseurs de Michael F. Dacey (1965, 1969), entouré d'un certain nombre d'autres doctorants acteurs de la révolution quantitative, autour de William Garrison⁶⁸ : Brian J.L. Berry, Duane F. Marble, Richard L. Morrill, John D. Nystuen, Robert C. Mayfield, William W. Bunge, Robert C. Mayfield, Ronald Boyce, Waldo R. Tobler, Arthur Getis, John Kolars⁶⁹, (voir Morrill, 1984 et Haggett, 2008 pour le détail). De même, on peut envisager l'ensemble de ces travaux dans la perspective de l'analyse exploratoire des données (EDA, Tukey, 1977), ce que nous avons déjà fait ailleurs (Oliveau, 2004 : 169 et suivantes).

C'est donc à partir de mesures de l'autocorrélation spatiale que nous envisagerons les structures spatiales des phénomènes sociaux. Pour cela, il nous a semblé important de revenir sur les mesures de ces structures en nous concentrant sur les indices les plus connus et leurs utilisations possibles. De nombreux ouvrages et articles reviennent de manière approfondie sur différents points évoqués ici⁷⁰, mais il n'en existe pas encore à notre connaissance qui reprenne de manière simple et pédagogique l'ensemble de ce corpus.

Les indicateurs globaux

Pour bien comprendre ce que saisissent les différents indices proposés et l'utilisation que l'on peut en faire en géographie, il nous a paru nécessaire de présenter les indices les plus connus (et donc les plus répandus), afin de mettre en évidence la manière dont ils s'articulent,

⁶⁶ La revue *Geographical Analysis* a proposé un numéro spécial en 2009 autour de cet article. Cliff et Ord s'y prêtent notamment à la reconstitution du contexte de la rédaction de ce papier, ce qui a eu pour nous un intérêt épistémologique évident.

⁶⁷ « *r* » est la lettre attribuée à l'indice de Moran dans les premiers articles de Cliff & Ord (par exemple 1970).

⁶⁸ William Garrison, au département de géographie de l'université de Washington est à l'origine du tournant quantitativiste (à travers ses cours et ses orientations) pris par les géographes cités après (voir aussi Johnston *et al.*, 2008, Barnes, 2004, Tobler 2002, Berry, 1993).

⁶⁹ Contrairement aux autres, il n'a pas soutenu sa thèse à l'université de Washington (<http://depts.washington.edu/geog/ma-and-ph-d-theses/>).

⁷⁰ Cliff, Ord, (1973, 1981); Clark, (1979); Cressie (1981); Ripley, (1981, 1998); Griffith, (1987); Anselin, (1988); Haining, (1990, 2003); Isaaks, Srivastava, (1990); Bailey, Gatrell, (1995); Longley, Batty, (1996); Arlinghaus, (1996); Burrough, Mc Donnell, (1998); Fotheringham, *et al.*, (2000) et bien sûr des revues, particulièrement *Geographical analysis* ou le *Journal of geographical systems*.

c'est-à-dire comment l'un vient compléter, remplacer, affiner, l'autre. La présentation sera donc volontairement partielle et partielle. Les méthodes présentées s'appuient toutes sur des mesures de variance et de covariance. D'après Decroly et Grasland (1996), cette direction spécifique des travaux a certainement reposé sur leur "facilité" de mise en œuvre et sur le fait qu'elle nécessitait moins de calcul. Les deux auteurs suggèrent que d'autres voies n'ont ainsi pas été explorées, notamment celles reposant sur des métriques de type Chi².

Autrement dit, l'ensemble des indices proposés ici vont chercher à savoir si la distribution statistique observée est biaisée par la proximité des individus entre eux. Pour cela ils s'appuieront généralement sur l'écart observé des valeurs des individus avec les valeurs de ses voisins en comparaison de l'écart des valeurs des individus avec l'ensemble des individus (principe directeur des mesures de variance et de covariance).

Quelques exemples de structures spatiales

Pour donner un peu de corps à la notion de structure spatiale, nous pouvons nous appuyer sur des exemples artificiels, sortes d'archétype des structures que l'on peut rencontrer. Dans un article récent, David Wong (2011) s'est prêté à cette modélisation, et nous reprenons ici les 6 cas de figures proposés⁷¹, qui reflètent bien les configurations observables (figure 14).

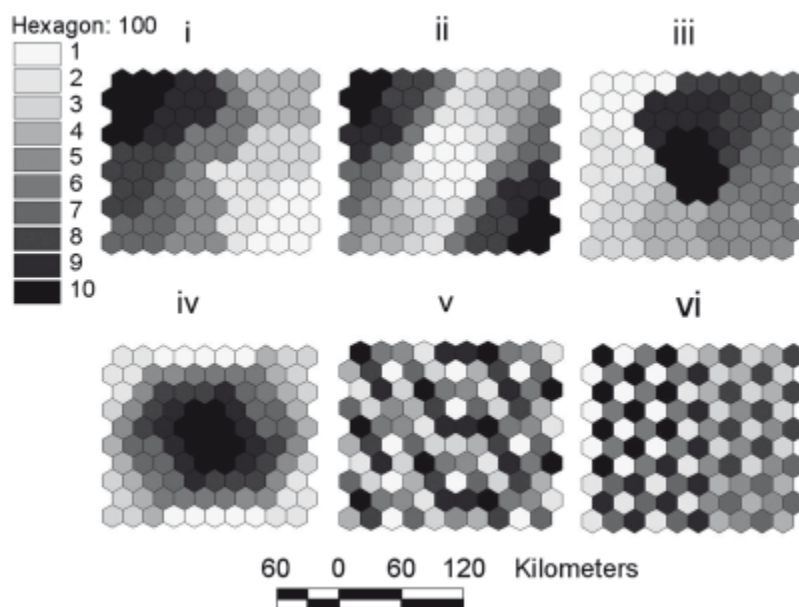


**figure 14 : 6 archétypes de structures spatiales
(Wong, 2011 : 332)**

⁷¹ On pourrait envisager d'ailleurs de construire d'autres cas de figures à partir des mêmes données. Wong, d'après les données qu'il nous a transmises (ce n'est pas indiqué dans l'article) a d'ailleurs testé au moins deux autres cas : un regroupement à chaque angle et 2 regroupements entre centre et angle.

Wong identifie 6 configurations spatiales qui correspondent à la présence sur l'espace étudié de : (i) tendance – *trend* ; (ii) regroupements à deux angles - *cluster 2 corners* ; (iii) regroupement local - *cluster local* ; (iv) regroupement centré - *cluster-center* ; (v) disposition aléatoire – *even* ; (vi) voisins opposés - *opposite neighbours*.

L'auteur mesure pour ces 6 modèles le niveau d'autocorrélation spatiale. Celle-ci est fortement positive et va en diminuant du modèle (i) au modèle (iv), elle est nulle pour le modèle (v) et modérément négative pour le modèle (vi). Nous proposons page 77 une exploration plus détaillée de ces structures spatiales

Notations

La statistique utilise des notations qui lui sont propres et qui facilitent les échanges. Il convient néanmoins de préciser leur sens. Nous devons ainsi rappeler quelques usages propres aux statistiques spatiales. Nous nous situons dans un univers statistique où l'espace entre en jeu. De ce fait, les lettres x et y sont réservées au positionnement des individus dans le plan (ou longitude et latitude si on envisage le positionnement terrestre). Par conséquent, la variable sera notée Z et la valeur qu'elle prend pour un individu Z_i . Puisque tout individu i peut avoir un voisin, celui-ci sera indiqué par la lettre j et la matrice de pondération W_{ij} (*weights*, poids en anglais). Enfin n correspond au nombre total d'individus et m au nombre total de paires de voisins. Enfin, sauf indication contraire, les sommes sont écrites de façon simplifiée $\sum_i z$ et non pas $\sum_{i=1}^n z$.

Pour exemple, et puisque nous aurons l'occasion d'en parler, la variance de la variable Z s'écrit donc comme indiqué dans l'équation 1.

$$\text{équation 1 (variance)} : \frac{\sum_i (z_i - \bar{z})^2}{n}$$

Semivariance

Le premier indice développé pour tenter de mesurer les effets de structure spatiale est la semivariance, ainsi appelée car elle ressemble à la variance, mais que le dénominateur n'est pas n mais $2m$ (voir équation 2).

$$\text{équation 2 (semivariance)} : \frac{\sum_i \sum_j (z_i - z_j)^2}{2m}$$

La semivariance ne peut pas s'interpréter directement, mais en comparaison avec d'autres mesures. On l'utilise en faisant varier les niveaux de voisinage. S'il y a autocorrélation spatiale, alors la semivariance sera plus faible dans les voisinages proches⁷² et plus élevés dans les voisinages lointains.

On utilise aussi la semivariance pour comparer les structures spatiales de deux variables dans le même espace ou de la même variable dans deux espaces distincts (voir Dauphiné, Voiron-Canicio 1988). On trouvera donc la semivariance souvent représentée dans un variogramme⁷³, diagramme servant à représenter la semivariance en fonction de différentes valeurs de voisinage.

Covariance spatiale

La covariance est un outil de statistique descriptive bien connu. Elle permet de mesurer dans quelle proportion les valeurs d'une variable varient comme les valeurs d'une seconde variable. Adaptée à la problématique qui nous intéresse, il suffit de remplacer les valeurs de la seconde variable par les valeurs de la même variable dans le voisinage. On obtient alors une mesure de la covariance spatiale (équation 3).

$$\text{équation 3 (covariance)} : \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{m}$$

Lorsque la covariance est non nulle, cela signifie que la valeur d'un point est liée à la valeur de ses voisins (il y a donc autocorrélation spatiale). Si la covariance est positive, alors l'autocorrélation spatiale est positive elle aussi, c'est-à-dire que les voisins ont des valeurs qui évoluent dans le même sens. Si la covariance est négative cela signifie au contraire que l'autocorrélation spatiale est négative, autrement dit que les valeurs des points évoluent en sens inverse de celles de leurs voisins. On notera que la nullité de la covariance ne garantit pas forcément l'absence d'autocorrélation spatiale.

La covariance spatiale, malgré sa simplicité, n'est que rarement utilisée, pour les mêmes raisons que la covariance traditionnelle. D'abord parce qu'elle n'est pas statistiquement très robuste (comme le soulignent Dauphiné & Voiron Canicio, 1998 : 34), ensuite parce qu'elle a

⁷² On parlera de « voisinage proche » pour parler de distance faible, de contiguïté d'ordre réduit ou de nombre de plus proches voisins faibles, en opposition aux « voisinages lointains », c'est-à-dire avec des distances plus importantes, des ordres de contiguïtés plus élevés ou un nombre de plus proches voisins plus élevé.

⁷³ On devrait parler plus exactement de semivariogramme, mais l'usage est répandu de ne pas indiquer « semi ».

été remplacée par des indices plus performants et plus faciles à commenter (hormis le « 0 », la covariance n'a pas de borne qui permette de l'interpréter).

Isotropie ?

La mesure de l'autocorrélation spatiale à travers des indices comme la semivariance et sa représentation à travers des variogrammes ne tient pas compte de l'existence possible d'une orientation spatiale particulière. Il considère que l'espace observé est isotrope. Or, il peut arriver que l'espace ne soit pas isotrope. C'est fréquemment le cas en géographie physique et en géologie, mais l'on peut envisager aussi que l'espace humain soit anisotrope (orientation préférentielle liée à des axes de communications, effets de vallée, etc.) La figure 14 montre deux exemples d'espaces anisotropes avec les organisations i et ii.

Il peut alors être intéressant de mesurer la variance dans différentes directions, de manière indépendante. On peut donc produire des variogrammes anisotropes où l'on superposera les courbes correspondant à la variographie effectuée selon différentes orientations (par exemple et de façon simple nord-sud versus est-ouest).

A propos du krigeage

Le krigeage⁷⁴ est une méthode d'interpolation spatiale mise en place par Matheron et nommée ainsi en référence aux travaux de Danie Krige. C'est un estimateur exact qui cherche à minimiser la variance de l'estimation produite en s'appuyant sur la structure spatiale des données connues⁷⁵. Pour cela, le krigeage s'appuie sur un variogramme, dit variogramme expérimental, qui permettra de définir un variogramme théorique à même de modéliser le comportement spatial de la variable pour les zones où les valeurs ne sont pas connues. Cette étape de modélisation est primordiale pour la bonne qualité de l'interpolation.

De même, le choix de la méthode de krigeage joue un rôle majeur sur la qualité de l'estimation fournie. Il existe ainsi plusieurs méthodes de krigeage, parmi lesquelles nous ne

⁷⁴ Voir Clark (1979), Cressie (1981) ou Bailey & Gatrell (1995). La présentation la plus pédagogique, même si elle n'est pas la plus complète, reste pour moi celle de Burrough et Mc Donnell (1998 : 132-161). Les cours de Denis Marcotte (école polytechnique de Montréal), en français, ont été d'une grande aide : <http://geo.polymtl.ca/~marcotte/g1q3401geo.html>. L'apprentissage s'est aussi largement déroulé à travers la pratique au sein du *South India Fertility Project* (<http://demographie.net/sifp/>).

⁷⁵ On doit rappeler que le krigeage est dit « transitif », c'est-à-dire que le fait de rajouter un point dont la valeur correspond à l'estimation faite par le krigeage ne change pas le krigeage (mais diminue la variance de son estimation).

citerons que les trois plus connues. Le krigeage simple est le plus facile à calculer, mais nécessite de connaître la moyenne de la variable et surtout repose sur la nécessité d'être face à une variable stationnaire (ce qui constitue une contrainte forte⁷⁶). Il est donc peu utilisé, avantageusement remplacé par le krigeage ordinaire qui ne se réfère pas à la moyenne générale et autorise un relâchement de la condition de stationnarité⁷⁷. On trouve aussi le krigeage universel qui intègre dans sa modélisation la présence d'une tendance spatiale (on parle aussi de dérive en géostatistique). On notera par ailleurs que les krigeages peuvent être effectués à partir de points ou de blocs de points.

Quelque soit la méthode de krigeage, elles partagent des caractéristiques qui les rendent en moyenne supérieures aux autres formes d'interpolation spatiale, à savoir : la prise en compte de la position relative des points connus pour l'estimation, de l'effet d'écran, de l'effet de pépité et de l'anisotropie.

Le krigeage estime les valeurs autour de chaque point connu à partir du variogramme théorique défini par l'utilisateur. Il prend en compte pour cela la position des autres points connus autour et leur donne un poids relatif en fonction de leur positionnement. Par exemple, 4 points situés à la même distance et régulièrement autour du premier point (un au nord, un à l'est, un au sud et un à l'ouest) se verront donner chacun 25% d'influence dans le calcul. Si on retire un point, par exemple celui du nord, alors les autres points verront leur poids changer : le point sud gardera 25% d'influence mais les deux autres points se "partageront" l'influence totale, soit 37,5% chacun. De la même façon, la valeur des points les plus proches est plus prise en compte que celle des points éloignés. Si deux points sont dans la même direction mais que l'un est placé entre le second et le point connu, il aura un poids très important au détriment de l'autre. C'est ce que l'on appelle « l'effet d'écran ». Cela a pour conséquence de ne pas donner à deux points dans la même direction plus de poids qu'un seul point dans une direction différente.

« L'effet de pépité » est ainsi nommé en rapport avec les fouilles géologiques : vous pouvez trouver une pépité sans que cela indique un filon. Autrement dit, bien que l'interpolation repose sur des valeurs connues, on peut envisager que celles-ci soient localement atypiques et que l'on ne souhaite pas les prendre en compte. L'effet de pépité permet alors de recalculer les valeurs connues à partir des points voisins, à condition de

⁷⁶ Getis & Ord (1996 : 271) rappellent que ce genre de données ne se trouve quasiment jamais en géographie.

⁷⁷ La moyenne utilisée pour effectuer l'estimation est calculée dans une fenêtre déterminée (au lieu de prendre la moyenne générale de l'échantillon), ce qui s'avère localement plus exact qu'une moyenne estimée pour l'ensemble de l'échantillon (Haining, 2003 : 331)

respecter les valeurs de la variance. L'effet de pépité est une sorte de "tolérance" pour l'estimation qui a un effet de lissage assez fort sur les données. Lorsque l'on décide d'affecter un effet de pépité au krigeage, l'effet d'écran est diminué.

Enfin, il est possible de prendre en compte l'anisotropie des variables dans la modélisation. Pour cela, on envisage des variogrammes théoriques différents selon les directions, en fonction des variogrammes expérimentaux observés.

Les logiciels qui permettent de mettre en place des interpolations par krigeage sont aujourd'hui nombreux (parmi les plus anciens et toujours en évolution, on trouve Surfer⁷⁸ et GS⁷⁹). Les SIG intègrent l'option (c'est le cas d'ArcGIS via le module geostatistical analysis, Vertical Mapper ajoute cette possibilité à Mapinfo).

Encore une fois, le développement et la facilité d'accès de ces méthodologies entraînent les utilisateurs à « appuyer sur le bouton » sans trop réfléchir. Or le krigeage est une technique d'interpolation qui fait des hypothèses fortes sur la distribution spatiale des phénomènes, et il convient auparavant de vérifier que les conditions sont réunies, et l'étude de l'autocorrélation spatiale est un préalable obligatoire. La modélisation du variogramme expérimental (donc la bonne connaissance de la structure spatiale du phénomène) et le choix de la méthode de krigeage sont deux éléments qui ne peuvent être négligés.

L'indice de Moran

L'indice de Moran est aujourd'hui l'indicateur le plus robuste statistiquement (Cliff, Ord, 1981 : 54-56), c'est pourquoi nous avons choisi d'appuyer notre travail sur sa mesure. Néanmoins, après l'avoir présenté et montré comment on pouvait l'utiliser, nous reviendrons sur d'autres indices proches et qui forment la famille des statistiques « gamma ».

Présentation

L'indice de Moran tire son origine d'un article de Pat Moran daté de 1950. Néanmoins, c'est sa réécriture proposée par Cliff et Ord (1969)⁸⁰ qui a permis sa diffusion, et c'est sous cette forme nouvelle qu'il est connu. L'indice de Moran, aussi appelé *I* de Moran, rapporte la covariance spatiale à la variance d'une variable (équation 4). La division de la covariance

⁷⁸ Développé par Golden Software : <http://www.goldensoftware.com>

⁷⁹ Développé par Gamma Design Software : <http://www.gammadesign.com>

⁸⁰ Et pour être totalement exact, il s'agit de la proposition d'Andrew Cliff dans sa thèse de doctorat (Cliff, 1969).

spatiale par la variance a pour effet de borner l'indice entre -1 et +1⁸¹. L'indice s'interprète de la façon suivante : la valeur 0 indique l'absence d'autocorrélation spatiale⁸², et plus l'indice tend vers 1, plus il y a d'autocorrélation spatiale. Le signe indique si l'autocorrélation spatiale est positive ou négative.

C'est un indice intéressant car il détecte bien les phénomènes d'attraction et de répulsion. *I* est grand lorsqu'on a de sites où les valeurs de la variable sont toutes en excès ou toutes en défaut par rapport à la moyenne générale (ce que l'on peut nommer attraction). A l'inverse, *I* se rapprochera de -1 lorsqu'on aura des sites où les valeurs de la variable sont divergentes entre voisins mais éloignées de la moyenne générale (ce que l'on nommera répulsion locale).

$$I \text{ de Moran} = \frac{\text{COV}}{\text{var}} \text{ soit } \text{cov} \div \text{var}$$

$$I \text{ de Moran} = \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{m} \div \frac{\sum_i (z_i - \bar{z})^2}{n}$$

$$I \text{ de Moran} = \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{m} \times \frac{n}{\sum_i (z_i - \bar{z})^2}$$

équation 4 (I de Moran) : $\frac{n}{m} \times \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$

C'est ici qu'intervient l'effet de la standardisation en ligne de la matrice de poids spatiaux évoquée précédemment (page 46). Dans une matrice de poids spatiaux standardisée en ligne, la somme des voisins pondérés est égale à la somme des individus. Ce que l'on peut écrire : $m=n$. On peut alors simplifier l'équation du *I* de Moran (équation 5). C'est cette formule qui sera systématiquement utilisée ici (la standardisation en ligne sera donc systématiquement mise en œuvre).

équation 5 (I de Moran après standardisation en ligne) : $\frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$

⁸¹ En fait, le bornage n'est pas absolu. Dans certains cas de figures, rares, la valeur 1 peut être dépassée.

⁸² On parle de 0 par facilité, mais il s'agit de $\frac{-1}{n-1}$ (qui tend donc très rapidement vers 0).

Sa forme et son interprétation le rapprochent donc du coefficient de corrélation linéaire, ce qui a sans doute aidé à le populariser. On peut ainsi interpréter la valeur de l'indice de Moran comme la part de variance explicable par le voisinage (c'est ce que fait Guilmoto dans Guilmoto, Oliveau, *et al.*, 2004 : 94). Un indice de Moran de 0,25 attribuerait ainsi 25 % de la variance aux valeurs dans le voisinage. Si l'indice de Moran ressemble au coefficient de corrélation linéaire, on ne peut cependant pas comparer leurs résultats. En effet, le I de Moran a tendance à sous-estimer la corrélation par rapport à un coefficient de corrélation linéaire.

Pour le comprendre, il faut envisager que dans le cas du I de Moran, la seconde variable du coefficient de corrélation linéaire est en fait la variable Z mais pour les voisins des individus (on la notera Z_j). Or cette variable a un écart-type plus faible que Z_i , puisqu'il s'agit de la moyenne pondérée de la valeur des voisins (il y a donc un lissage des valeurs qui se fait). Mais, dans le cas de l'indice de Moran, on ne prend pas en compte cet écart-type, mais celui de la variable d'origine⁸³. Le dénominateur a donc une valeur plus forte qu'il n'aurait dans le cas d'une corrélation linéaire et tend donc à sous-estimer la corrélation.

Comme pour la semivariance, on peut produire des mesures à différents niveaux de voisinage, ce qui permettra de produire un diagramme, appelé corrélogramme⁸⁴, montrant comment la structure spatiale évolue en fonction de la distance, du niveau de contiguïté, etc. Le corrélogramme permet aussi la comparaison d'une variable avec d'autres variables et/ou pour différents espaces.

On se rappellera par contre que si la matrice de voisinage est modifiée, la partie de l'équation se référant à Z_j va changer aussi, c'est-à-dire que la valeur de $(Z_j - \bar{Z})$ sera transformée. En augmentant le nombre de voisins, on diminuera d'autant les écarts de Z_j à $\bar{Z}$ (effet de lissage par agrégation). Plus le nombre de voisins pris en compte sera important, et plus l'autocorrélation spatiale sera faible, ce qui explique en partie (mais pour une petite partie seulement) la décroissance observée sur le corrélogramme.

⁸³ Au lieu d'avoir comme dénominateur la multiplication des deux écarts-type, on a la variance de la variable (pour mémoire, il s'agit du carré de l'écart-type de la variable).

⁸⁴ On parle de corrélogramme pour la représentation d'indices mesurant l'autocorrélation spatiale et de variogramme lorsqu'il s'agit de représenter la variance ou la semivariance. La courbe représentée sur un corrélogramme descend lorsque la différence entre voisins augmente alors que celle du variogramme monte.

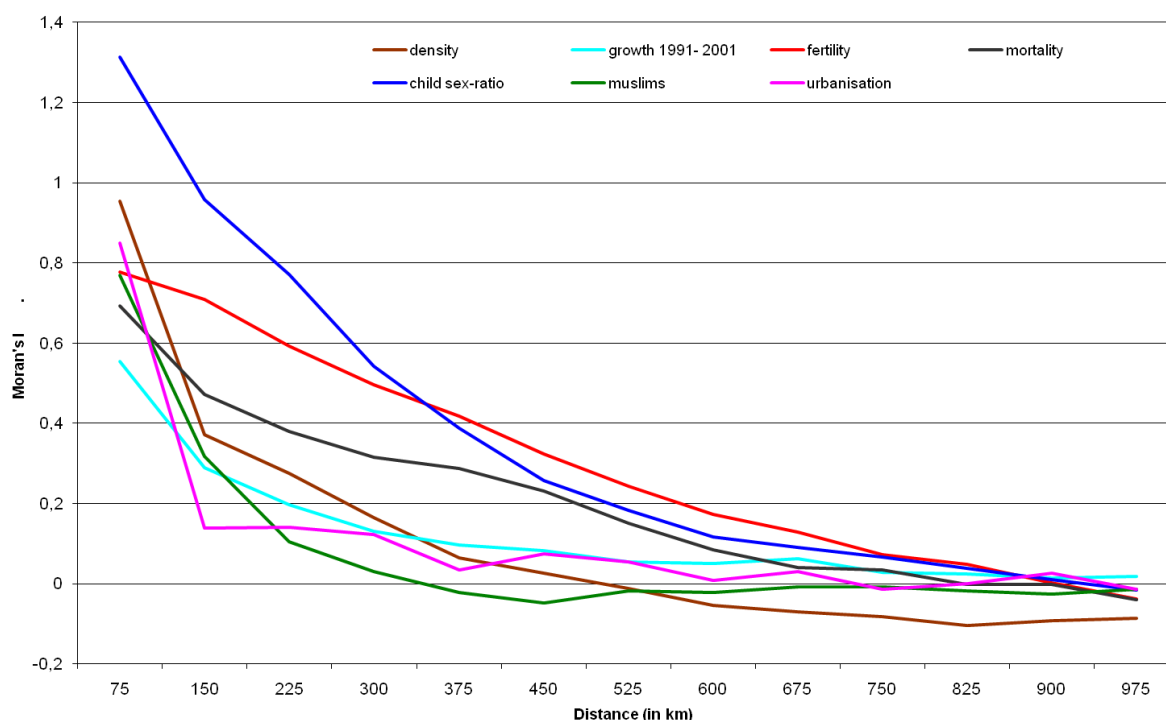


figure 15 : corrélogrammes

La figure 15 propose la superposition de corrélogrammes pour 7 indicateurs sociodémographiques à l'échelle des districts indiens en 2001⁸⁵. Il s'agit de comparer les structures spatiales de ces différents phénomènes. Pour lire le corrélogramme, on partira des valeurs dans un voisinage proche (gauche de la courbe) pour aller vers la droite. Les points des courbes indiquent les niveaux d'autocorrélation spatiale à différents pas de distances, les pentes des courbes nous donnent des indications sur la structure spatiale des phénomènes dans l'espace. On considère, sans s'appuyer sur aucune démonstration mathématique, que la valeur 0,2 constitue pour le I de Moran une valeur seuil en dessous de laquelle il n'y a plus d'autocorrélation spatiale.

On s'intéressera donc d'abord aux valeurs dans des voisinages proches. On voit que tous les indicateurs n'ont pas des structures spatiales aussi marquées. Le sex-ratio enfantin (*child sex-ratio*) a le niveau d'autocorrélation spatiale le plus important (on remarque d'ailleurs que le I de Moran dépasse 1, ce qui est très rare), et la croissance démographique entre 1991 et 2001 (*growth 1991-2001*) le niveau le plus faible.

⁸⁵ Le détail des indicateurs et les analyses détaillées des indicateurs sont disponibles dans Oliveau, Guilmoto, 2005. La matrice de voisinage repose sur des pas de distance de 75 km. Les mesures ont été produites avec le logiciel Variowin (Pannatier, 1996).

Le deuxième niveau de lecture concerne les pentes des courbes représentées. On y voit par exemple que si le pourcentage de musulmans par districts (*muslims*) a le même niveau d'autocorrélation spatiale que le taux de fécondité (*fertility*) dans un rayon de 75 km, leur structure spatiale est en fait très différente. Les musulmans semblent concentrés dans certains voisinages de l'ordre de 150 km, alors que les ressemblances de comportements reproductifs s'étalent sur plus de 400 km. Plus généralement, la fécondité présente la structure spatiale la plus étendue (à 400km, le *I* de Moran est toujours supérieure à 0,4 et surpasse le niveau d'autocorrélation spatiale du sex-ratio enfantin), même si elle n'est localement pas la plus forte (elle est ainsi inférieure au taux d'urbanisation -*urbanisation*- ou à la densité -*density* pour un voisinage de 75 km).

Le nuage de points de Moran

On peut interpréter le *I* de Moran comme l'estimation par les moindres carrés ordinaires du coefficient de régression linéaire de Z_j sur Z_i ⁸⁶. Dans ce cas, il est possible de produire une représentation graphique de l'indice de Moran, qu'Anselin a nommé « nuage de points de Moran » (*Moran scatterplot*⁸⁷). Il s'agit avant tout d'un mode de visualisation des données spatiales et pas encore d'une désagrégation d'un indice global. Le nuage de points de Moran fait partie des fonctionnalités intégrées dans la version 1.50 du logiciel SpaceStat (Anselin, 1992, 1994) et du logiciel GeoDA (Anselin, 2003).

Luc Anselin, pour élaborer le nuage de points de Moran, part du constat que si les techniques d'exploration des données s'étaient bien développées depuis les travaux de Tukey⁸⁸, il n'en était pas de même pour les données spatiales. L'analyse exploratoire des données (EDA⁸⁹) nécessitait une mise à jour importante pour intégrer la dimension spatiale et devenir l'analyse exploratoire des données spatiales (ESDA⁹⁰). Plusieurs travaux ont vu le jour à la même époque, mais le nuage de points de Moran tient une place à part, puisqu'il

⁸⁶ Dray (2011) utilise cette propriété pour montrer les limites inhérentes à la mesure globale du *I* de Moran, à savoir son incapacité à déceler de l'autocorrélation spatiale globalement lorsque qu'autocorrélations spatiales positive et négative coexistent.

⁸⁷ Luc Anselin publie deux *working papers* en 1993 (Anselin, 1993a, 1993b), le premier sera repris dans un ouvrage collectif en 1996 (Anselin, 1996), il propose une présentation complète de l'outil. Le second fera l'objet d'une publication dans *Geographical analysis* (Anselin, 1995).

⁸⁸ Voir par exemple Monmonnier, 1989.

⁸⁹ *Exploratory Data Analysis*.

⁹⁰ *Exploratory Spatial Data Analysis*. Pour un bref aperçu en français, voir Banos (2001).

permet de "regarder l'intérieur" du I de Moran, c'est-à-dire de comprendre comment celui-ci est composé.

Le nuage de points de Moran se construit simplement. Il s'agit de projeter sur un graphique les valeurs d'une variable Z pour l'ensemble des points i (en abscisse) et les valeurs dans le voisinage⁹¹ j de ces points i (en ordonnée). On obtient ainsi un nuage de points qui permet de voir l'autocorrélation spatiale de la variable. Comme pour une corrélation linéaire classique, la forme du nuage de points et son orientation nous donnent l'intensité de la relation (ici autocorrélation spatiale). Une petite étape précède l'élaboration du nuage de points de Moran. Elle consiste à centrer-réduire la variable (à la standardiser, comme dirait les anglophones). Cette opération permet la comparaison des nuages entre eux et surtout élimine les effets d'échelle liés aux différences d'ordre de grandeur d'une variable à l'autre.

Une fois centrées-réduites et représentées sous forme d'un nuage de points on peut voir aussi si certains points se comportent comme des valeurs atypiques (*outliers*). On parlera alors de valeurs spatialement atypiques (*spatial outliers*) puisque leur caractéristique est d'avoir des valeurs très différentes de l'ensemble de l'échantillon ou/et des valeurs des points qui les entourent. Enfin, un certain nombre de points peuvent avoir des valeurs éloignées de l'ensemble de la population mais proches de celles de leurs voisins. Plutôt que de les envisager comme des valeurs spatialement atypiques on parlera alors de poches d'hétérogénéité locale (les valeurs sont localement homogènes, mais se présentent comme hétérogènes par rapport à la distribution globale).

⁹¹ Le voisinage est toujours défini par une matrice W_{ij} , voir page 49.

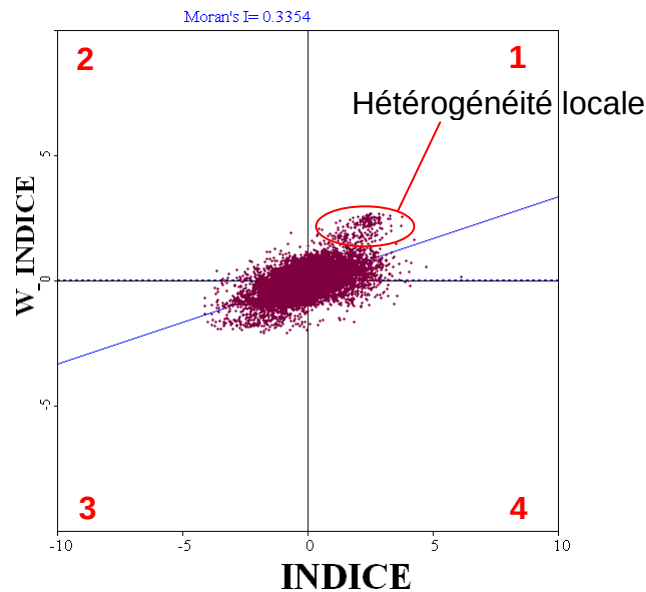


figure 16 : nuage de points de Moran pour l'indice de modernisation dans un voisinage de 15 km (Oliveau, 2004)

Si l'on considère que le nuage de points représente la régression entre les valeurs en « i » (INDICE) et en « j » (W_INDICE), alors le I de Moran est déduit du coefficient de la pente de la droite de régression (en bleu sur la figure), ici 0,3354⁹² (Griffith, 1996, Dray, 2011). Les quadrants (numérotés de 1 à 4 en rouge) nous indiquent logiquement les types d'association spatiale en présence.

- Dans le premier quadrant se trouvent les points dont les valeurs sont supérieures à la moyenne et dont le voisinage a aussi des valeurs supérieures à la moyenne. Il y a donc une participation positive à l'autocorrélation spatiale globale.
- Dans le deuxième quadrant se trouvent les points dont les valeurs sont supérieures à la moyenne et dont le voisinage a des valeurs inférieures à la moyenne. Il y a donc une participation négative à l'autocorrélation spatiale globale.
- Dans le troisième quadrant se trouvent les points dont les valeurs sont inférieures à la moyenne et dont le voisinage a aussi des valeurs inférieures à la moyenne. Il y a donc une participation positive à l'autocorrélation spatiale globale.
- Dans le quatrième quadrant se trouvent les points dont les valeurs sont inférieures à la moyenne et dont le voisinage a des valeurs supérieures à la moyenne. Il y a donc une participation négative à l'autocorrélation spatiale globale.

⁹² On ne garde habituellement que deux décimales, ce qui est amplement suffisant. Le logiciel GeoDA en donne 4 de façon systématique.

On visualise ainsi de quelle manière chaque point participe à la mesure globale : les points situés dans les quadrants 1 et 3 auront tendance à rendre l'autocorrélation spatiale positive et ceux des quadrants 2 et 4 à la rendre négative. On parle souvent de « *spatial outliers* » (valeurs spatialement atypiques) pour désigner ces derniers lorsque l'autocorrélation spatiale globale est positive, puisqu'ils ont des valeurs opposées à celles de leurs voisins

On comprendra aussi le poids que peuvent avoir des individus ayant des valeurs extrêmes, notamment des poches d'hétérogénéité locale, particulièrement lorsque l'échantillon n'est pas très important. Dans l'exemple présenté, les points formant la poche d'hétérogénéité spatiale identifiée (points entourés en rouge sur la figure) n'influent que marginalement sur la valeur du I global, du fait du nombre total de points très important (16085 individus).

Pour illustrer le poids que peuvent avoir quelques valeurs spatialement atypiques, nous nous intéresserons au sex-ratio des enfants à l'échelle des districts indiens (au nombre de 591 ici). Le phénomène a été largement étudié, notamment dans sa dimension spatiale (Guilmoto, 2008). La figure 17 nous montre le nuage de points de Moran pour le sex-ratio des enfants de moins de 6 ans en Inde en 2001 (les données nous ont été fournies par C.Z. Guilmoto).

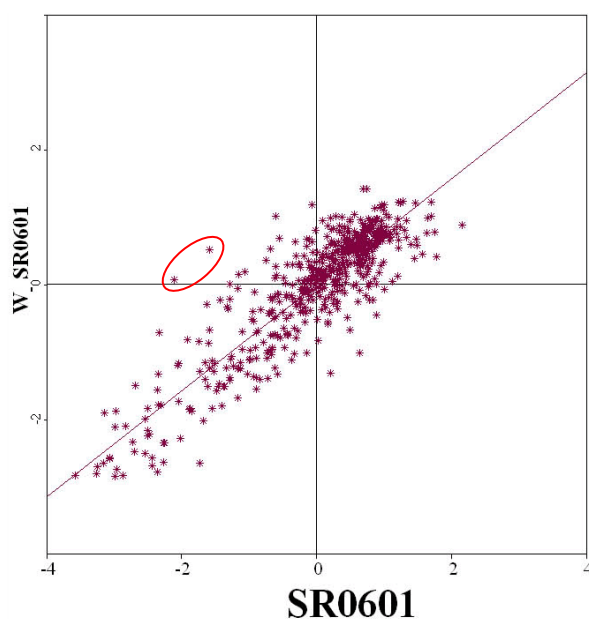


figure 17 : nuage de points de Moran pour le sex-ratio des enfants en Inde (2001)

Le nuage de points est très étiré et montre une régularité assez remarquable. Le I de Moran calculé est de 0,776⁹³. En regardant plus en détail, on voit bien sûr quelques valeurs spatialement atypiques. Nous avons choisi de refaire les mesures en retirant du calcul deux

⁹³ Pour la démonstration que nous souhaitons faire la troisième décimale est exceptionnellement nécessaire.

d'entre elles (entourées en rouge sur la figure). Le I de Moran passe alors à 0,787. Cela signifie que ces deux points (0,3% de la population) influent pour 1,4% de la valeur de l'indice de Moran, ce qui est loin d'être négligeable. Comme l'a dit Griffith (2006), l'analyse des espaces participant de l'autocorrélation spatiale négative dans les structures globales n'a pas reçu l'attention méritée.

Le coefficient de Geary et les statistiques gamma

Le I de Moran, nous l'avons déjà fait remarquer, entre dans la catégorie des indices statistiques reposant sur des mesures de la variance. Il en existe plusieurs autres, qui sont souvent assez proches. Parmi ceux là, on trouve notamment le coefficient (ou ratio) de Geary, appelé en raccourci c de Geary. Le c de Geary est lui aussi le fruit de l'actualisation par Cliff et Ord (1969) d'un indice plus ancien, datant du milieu du 20^{ème} siècle (Geary, 1954).

$$\text{équation 6 (c de Geary)} : \frac{n-1}{2m} \times \frac{\sum_i \sum_j w_{ij} (z_i - z_j)^2}{\sum_i (z_i - \bar{z})^2}$$

Comme le montre l'équation 6, le coefficient de Geary repose lui aussi sur la variance mais compare les écarts des individus à leurs voisins, et non pas les écarts des individus et de leurs voisins à la moyenne. Plus simplement, on retiendra qu'il s'agit du rapport de la semivariance sur la variance. L'indice évolue entre 0 (autocorrélation spatiale positive) et 2⁹⁴ (autocorrélation spatiale négative) en passant par 1, qui marque l'absence d'autocorrélation spatiale. Les résultats sont donc aussi faciles à interpréter que ceux du I de Moran, bien que moins intuitifs, puisque lorsque l'autocorrélation spatiale augmente, l'indice diminue. On notera que les deux indices, si ils ne sont pas exactement redondants, sont statistiquement liés (Cliff, Ord, 1970 : 273), mais ne varient pas exactement de la même manière (Griffith, 1987, 44). Il peut ainsi être utile de les comparer, mais bien souvent le seul I de Moran suffit et l'on ne croise pas souvent le c de Geary dans la littérature (et plus rarement encore sans le I de Moran).

Cette ressemblance entre les indices a amené Cliff et Ord⁹⁵ (1981 : 23) à envisager que les indices globaux d'autocorrélation spatiale étaient bien souvent des cas particuliers de ce qu'ils

⁹⁴ Comme pour le I de Moran, l'indice n'est pas strictement borné. Griffith (1987 : 36) fait remarquer que l'indice peut exceptionnellement dépasser 2.

⁹⁵ Ils s'appuyaient notamment sur les articles de Mantel (1967) et de Hubert *et al.* (1981).

nomment les statistiques de produits croisés (*cross product statistics*)⁹⁶. De façon plus générale encore, les statistiques de produits croisés s'apparentent aux statistiques gamma (Anselin, 1995 : 97⁹⁷) dont la particularité est de mesurer les similarités (ou dissimilarités) entre deux (ou plus) matrices. L'équation 7 montre comment on peut résumer l'ensemble de ces statistiques gamma (Γ).

$$\text{équation 7 (statistique gamma)} : \Gamma = \sum_i \sum_j w_{ij} v_{ij}$$

Dans cette formule, « w_{ij} » est une matrice de poids spatiaux alors que « v_{ij} » représente l'association statistique entre les valeurs des points i et de celles de leurs voisins j . Autrement dit, il s'agit toujours de comparer les proximités statistiques et spatiales. Bien entendu, tout se mesure pour l'ensemble des points ($\sum_i$) et de leurs voisins ($\sum_j$).

En donnant $v_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$, on retrouve le I de Moran ;

le c de Geary pour $v_{ij} = \frac{(z_i - z_j)^2}{\sum_i (z_i - \bar{z})^2} \times \frac{n-1}{2m}$;

ou bien encore la semivariance avec $v_{ij} = \frac{(z_i - z_j)^2}{2m}$.

⁹⁶ Getis (1991) présente en détail les statistiques de produits croisés. C'est pour lui l'occasion de rappeler le lien entre interaction et autocorrélation spatiales.

⁹⁷ Cette forme simplifiée des indices globaux d'autocorrélation spatiale est utilisée par Cliff et Ord (1981 : chapitre 2) pour mettre au point des tests statistiques de significativité.

L'ensemble des statistiques présentées ont un défaut commun très tôt souligné : elles produisent des mesures globales, qui sont aveugles à la présence possible de phénomènes locaux d'autocorrélation spatiale⁹⁸. Elles peuvent indiquer l'absence d'autocorrélation spatiale à une échelle globale, alors que des études portant sur certains sous-espaces pourraient montrer des niveaux d'autocorrélation spatiale (positive ou négative) remarquables⁹⁹. Ainsi, si l'on mesurerait l'autocorrélation spatiale des seules poches d'hétérogénéité locale, on trouverait alors que l'indépendance statistique des observations n'est pas garantie, ce qui repose évidemment les problèmes statistiques déjà soulignés précédemment. De plus, la stationnarité spatiale de la variable ne peut pas être évaluée ainsi (et ce, quelque soit le niveau global de l'autocorrélation spatiale).

La première solution a été de varier les indices utilisés (semivariance, I de Moran, c de Geary, etc.), de développer des méthodes de contrôle de ces indices (variogramme, corrélogramme, isotrope et anisotropes, nuage de points de Moran). Une seconde proposition consistait à développer des méthodes de lissages spatiaux (à partir de krigeages sophistiqués notamment). Néanmoins, il semble que la méthode la plus riche de résultats a été de mettre au point des techniques d'exploration spécifiques pour les données spatiales qui permettent réellement de "voir" à l'intérieur des variables. Le nuage de points de Moran, évoqué précédemment, constitue une voie, mais la production s'est tournée d'abord vers des indices locaux capables de mettre au jour ces sous-structures au sein des structures globales.

Au-delà des contraintes statistiques (indépendance et stationnarité), on peut envisager au moins deux autres raisons qui ont poussé les travaux dans la direction des indices locaux. La première a été de comprendre comment des mesures globales étaient affectées/construites par des mesures locales. Si le nuage de points de Moran joue un rôle important de ce point de vue, les recherches de Getis sur la statistique G_i l'ont précédé (Getis, 1991 ; Getis & Ord, 1992 ; voir plus bas). La deuxième raison qui a poussé au développement de ces indices locaux est la volonté de pouvoir tester pour l'ensemble des points ceux qui ont une influence plus importante sur le voisinage, l'idée étant qu'il existe des regroupements de valeurs (*clusters*).

⁹⁸ Même s'il est vrai que lorsqu'une forte autocorrélation spatiale globale est identifiée, cela importe moins (Getis & Ord, 1996 :276).

⁹⁹ C'est bien cette limite qui a permis à Roger Brunet de développer son objection à l'utilisation de l'autocorrélation spatiale (Brunet & Dollfus, 1990 : 87), en s'appuyant sur le fait que les variables n'étaient jamais ni totalement autocorrélées spatialement, ni absolument pas autocorrélées.

Cette préoccupation occupe d'ailleurs largement le champ de l'épidémiologie spatiale depuis 40 ans (depuis Mantel, 1967, jusqu'à Rogerson aujourd'hui).

Des associations spatiales aux statistiques G

Les indices locaux vont donc d'abord chercher à identifier localement des regroupements de valeurs proches, des « associations spatiales », la notion d'autocorrélation spatiale étant mise un peu de côté. Pourtant, que l'on recherche des clusters, des associations spatiales ou des mesures locales de l'autocorrélation spatiale, méthodes et démarches sont identiques. Seul le vocabulaire change, l'idée de l'association spatiale étant de rappeler l'association statistique de deux individus voisins. Lorsque des associations spatiales sont identifiées, on les nomme aussi souvent *hot-spots*¹⁰⁰. L'idée sous-jacente est de repérer les « poches spatiales de non-stationnarité » des variables, souligner l'hétérogénéité des variables (Fotheringham, 2009).

Getis & Ord (1996 : 276) rappellent d'ailleurs que, si l'autocorrélation spatiale est forte, la recherche de poches locales est moins importante, et qu'elles alors ont un rôle généralement moins marqué (notamment dans le cas de tendances spatiales). Par contre, il peut exister des poches locales alors que l'autocorrélation spatiale globale est nulle. Ce type d'analyse, on le comprend, est donc fondamental pour s'assurer des conditions dans lesquelles on va produire ses résultats description ou modélisation, comme pour le krigeage par exemple.

Fort du constat de la nécessité d'une approche locale, Getis (1991) développe une nouvelle mesure, la famille des statistiques G , qui permet de détecter les « poches locales de dépendance ». Ces statistiques sont conçues pour compléter l'information générale d'un indice comme le I de Moran. L'équation 8 donne la formule de la première statistique proposée, la statistique G_i .

$$\text{équation 8 (statistique } G_i) : G_i = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}$$

Cette première statistique rapporte les valeurs de la variable X dans le voisinage de i ¹⁰¹ (x_i n'étant pas compris) à la somme des valeurs de X , hormis x_i . Si l'on admet que i soit son

¹⁰⁰ « Points chauds » en français, c'est-à-dire des « lieu[x] où il se passe quelque chose » pour reprendre la définition du *Robert*.

¹⁰¹ Comme toujours défini par la matrice de poids spatiaux W_{ij} .

propre voisin¹⁰² (que j puisse comprendre i), on définit alors la statistique G_i^* . Les deux statistiques sont donc assez proches. Getis & Ord (1992 : 194) rappellent eux-mêmes que « G_i , and G_i^* typically convey much the same information »... A partir de cette statistique locale, Getis & Ord envisagent la définition d'un indice global, nommé « G » (équation 9).

$$\text{équation 9 (la statistique G)} : G = \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i \sum_j x_i x_j}$$

La statistique G est d'une forme proche du I de Moran et fait partie des statistiques de produits croisés¹⁰³. N'étant pas construite à partir de la covariance mais de la somme des valeurs, elle ne propose néanmoins pas exactement les mêmes résultats. Si la détection des concentrations de hautes valeurs est identique pour les deux indices, ils diffèrent fortement lorsque l'on est en présence de concentrations de valeurs faibles. Le I de Moran reste dans ce cas positif, alors qu'il devient fortement négatif pour G . De la même manière, lorsque l'on est en présence de valeurs faibles dans un voisinage de valeurs fortes, le I de Moran réagit beaucoup plus.

Association de valeurs	Valeur de G	Valeur de I
Fortes et fortes	Très positive	Très positive
Fortes et moyennes	Moyennement positive	Moyennement positive
Moyennes et moyennes	Nulle	Nulle
Aléatoires	Nulle	Nulle
Fortes et faibles	Moyennement négative	Très négative
Moyennes et faibles	Moyennement négative ¹⁰⁴	Moyennement négative
Faibles et faibles	Très négative	Très positive

tableau 3 : association de valeurs et mesure de I et G
(d'après Getis et Ord, 1992)

¹⁰² De prime abord, concevoir que l'individu soit son propre voisin n'a pas particulièrement de sens. Néanmoins, dans une perspective de modélisation, on peut envisager d'intégrer la valeur de i . Voir page 87.

¹⁰³ si l'on considère que $v_{ij} = \frac{x_i x_j}{\sum_i \sum_j x_i x_j}$, on retrouve bien l'équation 7.

¹⁰⁴ La valeur prise par G est plus négative pour l'association de valeurs « moyennes et faibles » que pour l'association de valeurs « fortes et faibles ».

Le tableau montre bien l'apport des statistiques G_i et G à la mesure des associations spatiales. En effet, la statistique G_i va permettre de cartographier les *clusters*, puisque chaque individu se voit associer une valeur, et on pourra dans le même temps obtenir une information sur la manière dont les statistiques globales reflètent les structures observées, en comparant les statistiques G et I .

Les LISA

Si Arthur Getis a véritablement ouvert la voie des statistiques spatiales locales, c'est à Luc Anselin que l'on doit la plus grande avancée. Anselin, en s'appuyant sur les statistiques gamma, met sur pied la désagrégation des indices globaux. L'idée est toujours la même : une statistique globale n'indiquant pas la présence d'autocorrélation spatiale ne garantit pas forcément une situation d'indépendance statistique localement. Le nuage de points de Moran permet déjà une première visualisation, et la statistique locale G permet de cartographier les clusters locaux. Néanmoins, si l'on pouvait remplacer la statistique G par un indice global désagrégé, on gagnerait en cohérence.

Contrairement à Getis qui propose un indice local qui peut être agrégé en un indice global, Anselin (1995¹⁰⁵) présente une version locale des différents indicateurs globaux : les indicateurs locaux d'association spatiale (*Local Indicators of Spatial Association* qui deviendront "célèbres" sous l'acronyme LISA). Il en propose une définition volontairement très large : « un indicateur local d'association spatiale est n'importe quelle statistique qui réponde à deux critères :

- a) Le LISA donne pour chaque observation une indication sur l'étendue d'un regroupement spatial significatif de valeurs similaires autour de cette observation.
- b) La somme des LISA pour toutes les observations est proportionnelle à un indicateur global d'association spatiale » (traduction libre).

L'idée est finalement assez simple. Puisque les statistiques gamma sont la somme pour tous les individus i de la mesure de la similarité (ou dissimilarité) avec leur voisin, on peut définir un indice local à partir de l'indice global en retirant la somme des i de l'équation (en introduisant un facteur de proportionnalité γ si nécessaire). Autrement dit, un gamma global est la somme pondérée des gammas locaux. Ce qui, en langage mathématique, donne l'équation 10.

¹⁰⁵ Comme nous l'avons mentionné précédemment, cette publication a fait l'objet d'un *working paper* identique 2 ans auparavant (Anselin, 1993).

$$\text{équation 10 (gamma global / gamma local)} : \sum_i \Gamma_i = \gamma \Gamma$$

Le I de Moran Local

Nous l'avons déjà dit, le I de Moran est une statistique gamma connue et largement utilisée. C'est donc assez logiquement que Luc Anselin développe son premier exemple de LISA à partir de cet indice. Si l'on utilise la forme simplifiée du I de Moran (équation 5), sa transformation en LISA se fait de la manière suivante (équation 11).

$$\text{équation 11 (I de Moran local)} : I_i = \frac{\sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$$

avec $\sum_i I_i = \gamma \cdot I$ et $\gamma = n$

L'introduction du facteur de proportionnalité γ permet ici d'apprécier la valeur d'un gamma local en fonction du gamma global. Ainsi, il est aisé de mettre en évidence des valeurs atypiques, en comparant la valeur du gamma local à celle de la valeur de la moyenne des gammas locaux (c'est-à-dire le gamma global divisé par le nombre d'individus « i »). Cela permet aussi de recalculer des gammas "intermédiaires", c'est-à-dire des mesures artificielles du I de Moran reflétant la force des regroupements observés localement (voir page 89).

Significativité

Lorsque l'on dispose d'un indicateur, il convient de savoir si les résultats sont significatifs, et dans quelle mesure. Cela vaut aussi bien pour les indices globaux que pour les locaux. Nous n'avons pas abordé le sujet plus tôt car il est relativement compliqué et souhaitons le traiter globalement.

Pour faire vite, la question est de savoir qu'elle est la probabilité que l'indice mesuré soit le résultat du hasard. Si l'on connaît la distribution de notre indice, on peut savoir si la valeur observée est probable ou non, ce que l'on traduira par plus ou moins significative (plus la valeur est probable, moins elle est significative). Par exemple, si la distribution de l'indice est de type normal, la probabilité la plus forte est que l'indice obtenu soit proche de la moyenne. S'il s'en éloigne fortement, alors on pourra penser que le résultat obtenu est significatif, car probablement pas dû au hasard. Dans le cas de la statistique G , par exemple, la distribution est connue (Getis & Ord, 1992).

Par contre, de nombreuses statistiques n'ont pas de distribution connue ou demande des conditions restrictives pour être connues (Anselin, 1995). On utilise alors la variable étudiée pour construire une pseudo-distribution de l'indice. Il s'agit de garder la valeur de i mais de déplacer aléatoirement¹⁰⁶ les autres valeurs dans l'espace. En recommençant l'opération plusieurs fois, on obtient alors une distribution empirique de l'indice pour la distribution de la variable étudiée. On peut alors comparer la valeur observée de l'indice à cette pseudo-distribution, comme si elle était la vraie distribution. Cette opération s'appelle *pseudo-randomization* ou *permutation*. Plus le nombre de permutations est important, plus on peut s'assurer de la significativité des résultats¹⁰⁷.

Cartographie

Depuis le G de Getis, et avec les LISA (dont le I de Moran local), nous sommes capable de proposer une mesure individuelle de la similarité statistique des individus avec leurs voisins. Nous allons donc pouvoir cartographier ces indicateurs locaux pour montrer les structures des espaces considérés et enrichir la cartographie traditionnelle des variables étudiées. Par ailleurs, puisque nous savons pour chaque point estimer la significativité de l'indice local, nous serons aussi en mesure de cartographier cet élément.

Pour illustrer notre propos, nous reprendrons la cartographie effectuée pendant notre doctorat (Oliveau, 2004). Pas par facilité, les cartes ont été refaites pour l'occasion¹⁰⁸, mais parce qu'il est le fruit d'un travail minutieux, et que l'on maîtrise cette variable mieux qu'une autre¹⁰⁹. En outre, le lecteur pourra se reporter à notre thèse de doctorat pour les détails. Pour mémoire la variable cartographiée est un indice de modernisation, construit grâce à une analyse factorielle effectuée sur les données censitaires indiennes de 1991. Les données sont à l'échelle des villages tamouls (et des villages du territoire de Pondichéry compris dans cet espace). L'indice présenté est la version débarrassée de l'effet d'enclavement et de centralité propre des villages. Nous retiendrons que la variable présente les dimensions socio-économiques des villages pondérée par les différentiels d'accessibilité à la ville et les

¹⁰⁶ Pour s'assurer du caractère aléatoire de l'opération, on utilise généralement un processus de type Monte Carlo.

¹⁰⁷ Cette question de significativité des résultats du I de Moran est particulièrement explorée par Michael Tiefelsdorf (voir par exemple Tiefelsdorf, 1998 ; Boot, Tiefelsdorf, 2000).

¹⁰⁸ 7 ans après, des fichiers bien rangés permettent de reprendre ses données...

¹⁰⁹ Un autre exemple est disponible à partir de l'étude du sex-ratio en Inde du Sud dans une des parties que nous avons rédigé du rapport EMIS (Guilmoto, 2005 : chapitre 3.4).

différentiels de population des villages. Autrement dit, les structures spatiales qui subsistent ne sont pas le fait des caractéristiques socio-démo-économiques ni de l'influence urbaine¹¹⁰.

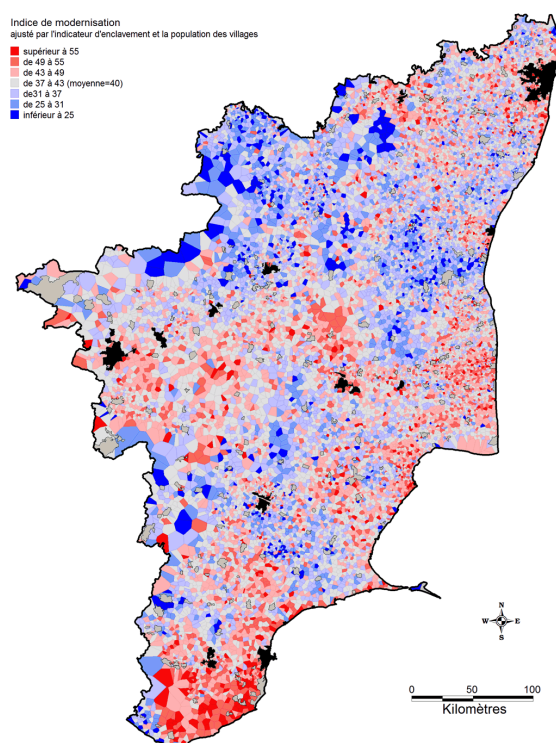


figure 18 : indice de modernisation ajustée

Si l'œil décerne bien des zones qui semblent se détacher, on voit aussi avec cette carte la limite de la lecture cartographique. A l'exception des espaces où les villages sont les plus grands, et où, par ailleurs, les phénomènes semblent plus marqués (au nord-ouest et à l'extrême sud), il est difficile de dire ce qu'il en est vraiment. Le delta de la Cauvery (à l'est au centre) avec ses nombreux villages ne permet pas de se faire une idée claire du niveau de structuration spatiale. On pense toujours voir des tendances, mais la vue est un sens trompeur et ne donne pas, quoi qu'il arrive, la force des relations. La cartographie des LISA donne cette indication de la force de relations entre voisins.

¹¹⁰ Sur le rôle spécifique de l'influence urbaine, nous renvoyons le lecteur aux pages 190 et suivantes de notre thèse de doctorat. On y mesure le rôle structurant des villes tamoules.

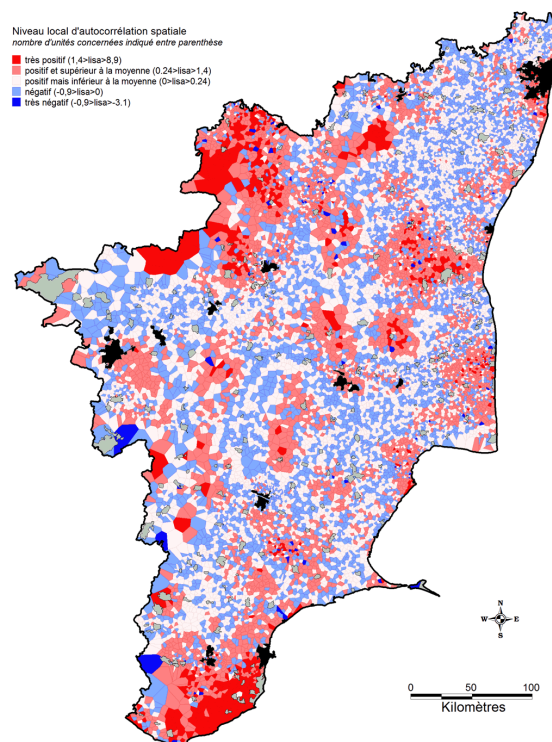


figure 19 : I de Moran local de l'indice de modernisation ajustée¹¹¹

La cartographie des LISA pose problème et n'a jamais été évoquée à ma connaissance (les LISA sont rarement cartographiées en l'état). En effet, même si la distribution tend à être normale (surtout avec 16085 individus), la méthode standard n'est pas forcément pertinente, puisque la valeur « 0 » a un sens fort. Luc Anselin, qui ne parle pas de cartographie, rappelle dans son article que l'on peut utiliser les repères "classiques" de la statistique pour identifier les valeurs extrêmes (plus de 2 écarts-types au dessus ou au dessous de la moyenne).

La moyenne est ici de 0,24. Il s'agit, par construction, de la valeur du I de Moran global. Nous avons donc fait le choix d'utiliser cette limite de 2 écarts-types (soit 1,15) pour définir les deux classes extrêmes. Ensuite, nous avons séparé les deux classes autour de la moyenne, mais extrait de la classe inférieure les valeurs positives. Il en résulte 5 classes qui nous laissent lire l'intensité locale des relations entre unités. On retrouve beaucoup plus d'associations positives que d'associations négatives, ce qui explique la valeur positive du I global.

¹¹¹ Tous les résultats ont été produits avec le logiciel GeoDA, logiciel d'ESDA, et ensuite exportés pour être cartographiés avec le SIG Mapinfo.

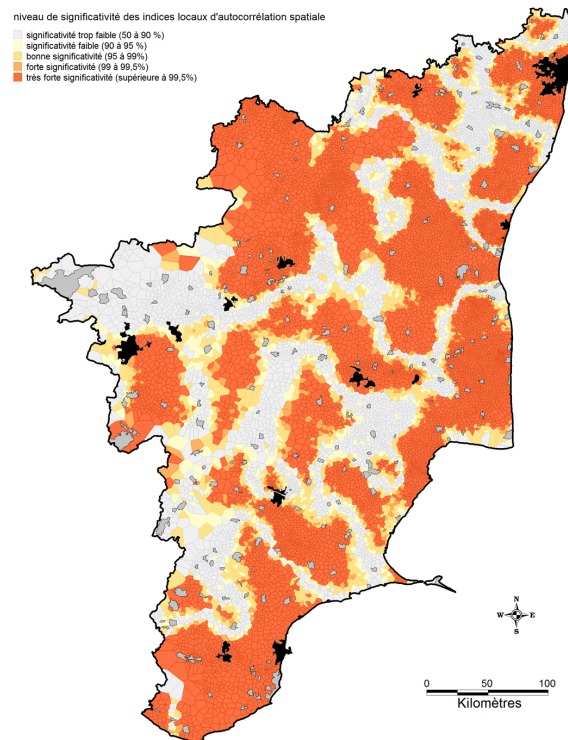


figure 20 : significativité des I de Moran locaux

La carte de la significativité des LISA est intéressante à plusieurs titres. D’abord elle nous indique clairement les espaces où il se passe des choses qui ne sont pas dues au hasard. Autrement dit, et en complément de la carte des LISA, elle permet d’éliminer les zones où la relation entre voisins n’est pas significative.

Cette première remarque faite, cela nous amène à re-réfléchir le sens de l’autocorrélation spatiale. En effet, les zones qui ne sont pas significatives se caractérisent par l’absence de valeurs fortes de LISA (par définition). Pourtant les individus peuvent avoir des valeurs proches les unes des autres. La non significativité marque avant tout la proximité des valeurs individuelles avec la moyenne globale : les individus concernés peuvent très bien se ressembler localement¹¹², ce que nous rappelait justement le tableau 3 page 68. Comme nous l’avons souligné à la suite de Griffith (2006), il semblerait néanmoins convenable de regarder plus en détail ces espaces où les valeurs ne sont pas significatives pour savoir ce qu’il s’y passe réellement.

Enfin, une dernière remarque s’impose concernant la forme que prend la répartition des niveaux de significativité. Il s’organise lui-même en poches dont les franges sont les moins

¹¹² En poussant le raisonnement jusqu’à l’absurde, on peut défendre cette position en rappelant que si tous les individus ont la même valeur, il n’y a plus d’autocorrélation spatiale, puisque l’espace ne compte plus.

significatives. Il n'y a donc pas de rupture nette, mais un continuum, entre les espaces significatifs et ceux qui ne le sont pas. Ces limites varient d'ailleurs (mais de façon marginale) lorsque l'on relance l'estimation de la significativité (puisque'il s'agit d'une procédure aléatoire, les résultats de deux séries de tirages consécutifs ne sont pas identiques).

Regroupements d'associations spatiales

Puisque nous avons désormais à disposition une information sur la valeur des individus, le sens de la relation avec leurs voisins et le niveau de significativité de cette relation, on peut donc proposer une typologie des individus. L'usage définit 5 groupes principaux, mais rien n'empêche d'en définir plus ou autrement. Nous reprenons ici les résultats produits avec le logiciel GeoDA¹¹³, qui sont les plus largement répandus

On distinguera d'une part les associations spatiales positives des associations spatiales négatives, puis les associations de valeurs fortes des associations de valeurs faibles et on exclura finalement les individus dont les valeurs de LISA ne sont pas significatives.

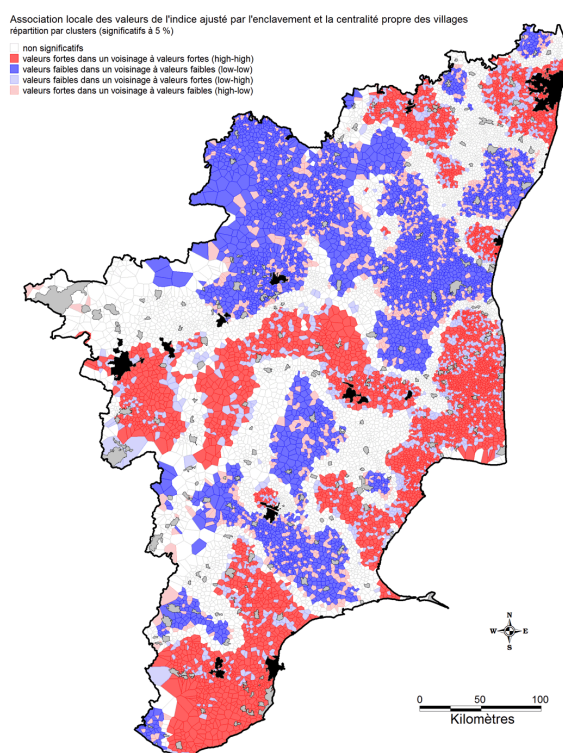


figure 21 : clusters de l'indice de modernisation ajustée

¹¹³ Le logiciel GeoDA a été développé par Luc Anselin après avoir produit un module pour Arc View : « SpaceStat » (maintenant développé et distribué par BioMedWare). GeoDA est distribué gratuitement. Luc Anselin a développé en parallèle de sa recherche les outils pour mettre en œuvre ses méthodes, ce qui a largement contribué à les diffuser tous.

La carte ci-dessus nous montre le résultat de cette catégorisation. Elle identifie en rouge les valeurs fortes et en bleu les valeurs faibles. Les couleurs foncées renvoient à des associations spatiales locales positives et les couleurs claires à des associations spatiales négatives. Les LISA non significatifs sont en blanc. On peut proposer un tableau récapitulatif des catégories :

		Valeurs des individus	
		Faibles	Fortes
Valeurs du voisinage	Fortes	Associations spatiales négatives	Associations spatiales positives
	Faibles	Associations spatiales positives	Associations spatiales négatives

tableau 4 : associations spatiales

On retrouve dans la littérature le terme de *clusters*, qualifié par le type d'association : *high-high* (association spatiale positive de valeurs fortes), *low-high* (association spatiale négatives : valeurs faibles dans un voisinage de valeurs fortes), *low-low* (association spatiale positive de valeurs faibles), *high-low* (association spatiale négative : valeurs fortes dans un voisinage de valeurs faibles).

« Moran scatterplot is not a LISA »

Le tableau 4 nous rappelle le nuage de points de Moran, puisqu'ils se construisent de manière similaire. Les deux éléments sont intimement liés, puisque les quadrants du nuage de points permettent de retrouver les associations spatiales. Cependant, comme il le rappelle lui-même, « le nuage de points de Moran n'est pas un LISA », puisqu'il ne permet pas d'identifier les associations spatiales¹¹⁴. Le nuage de points ne permet pas en effet de présumer de la significativité des individus, ni de leur proximité spatiale. Le nuage de points

¹¹⁴ « Note that the Moran scatterplot is not a LISA [...], since no indication of significant local spatial clustering is obtained » (Anselin, 1995 : 105).

ne permet donc pas d'identifier des poches d'hétérogénéité spatiale, tout au plus montre-t-il les valeurs spatialement atypiques.

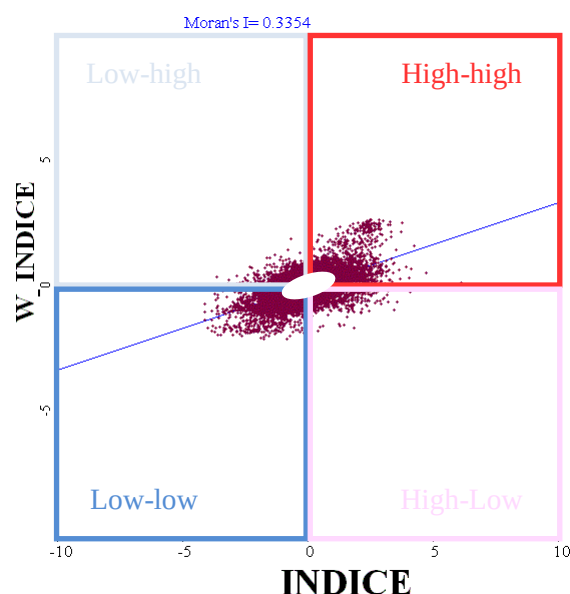


figure 22 : nuage de points de Moran et associations spatiales

Le quadrant 1 identifie les clusters *high-high*, le 2 les *low-high*, le 3 les *low-low* et le 4 les *high-low*. L'interactivité des logiciels d'exploration de données spatiales autorise le « balayage géographique » (*geographic brushing*, Monmonnier, 1989), qui permet la visualisation sur la carte des entités identifiées sur le nuage de points. Celui-ci apparaît alors comme un complément intéressant à l'exploration des structures spatiales.

Retour sur les statistiques G

Si le nuage de points de Moran a toujours son utilité, qu'en est-il de la statistique G ? La

statistique G_i est assurément une forme de gamma local : $\Gamma_i = \sum_j w_{ij} v_{ij}$ avec $v_{ij} = \frac{x_j}{\sum_j x_j}$,

mais ne constitue pas pour Anselin un LISA, puisque la statistique G n'est pas une statistique globale d'association spatiale (le tableau 3 montre bien que les associations de valeurs faibles ne sont pas identifiées à l'échelle globale). Par ailleurs, Getis et Ord avaient déjà fait remarquer que leur indice global n'apportait rien de plus que le I de Moran (1992 : 198¹¹⁵) et ils montrent même que le I_i et le G_i^* sont statistiquement liés (Ord & Getis, 1995 : 289). Avec l'apparition des LISA, et le lien entre indice de Moran global et indices de Moran locaux, la

¹¹⁵ Pour une comparaison détaillée des statistiques G_i , G_i^* et des LISA, voir Bao & Henry (1996).

famille des statistiques G semble donc être devenue superflue. Elle reste pourtant utile dans les développements que propose Getis autour des filtrages spatiaux (voir page 92). D'ailleurs, le logiciel GeoDA l'intègre depuis sa version 0.9.9.11 (qui date du premier semestre 2011). Cette récente innovation a pour nous une vertu : celle de nous permettre la comparaison entre les différentes statistiques locales, et de nous confirmer la très grande similarité de leurs résultats.

Retour sur les archétypes de structures spatiales

En nous appuyant sur les données fournies par David Wong, nous avons pu reprendre l'idée des structures spatiales archétypiques évoquées précédemment (page 51). La variable se compose de 10 groupes de 10 individus dont les valeurs s'échelonnent régulièrement de 1 à 10. Moyenne et médiane sont donc égales à 0,55 et l'écart-type est de 2,87.

La répartition spatiale des individus change pour chaque configuration, c'est donc bien uniquement la structure spatiale qui change, la distribution statistique restant identique. Le voisinage utilisé, comme dans l'article, est de 60 km¹¹⁶. Pour chacune d'elles nous cartographions la variable, les regroupements locaux d'associations spatiales et nous proposons le nuage de points de Moran correspondant (au dessus duquel est indiqué la valeur du I de Moran global)¹¹⁷. Sur le nuage de points de Moran, on se rappellera que les valeurs des individus sont en abscisses, et que l'on a donc une gradation de 1 à 10 en allant de gauche à droite. A défaut d'interactivité, cette indication permettra au lecteur de mieux voir à partir du graphique la relation entre le nuage de points et la situation des individus sur les cartes.

¹¹⁶ Nos résultats diffèrent légèrement en termes de valeurs (pas d'interprétation), puisque nous utilisons la distance là où il utilise des fonctions de la distance.

¹¹⁷ Cartographie de la variable avec Philcarto, des clusters avec GeoDA. Tous les traitements ont été effectués avec GeoDA, 10.000 permutations ont été effectuées systématiquement.

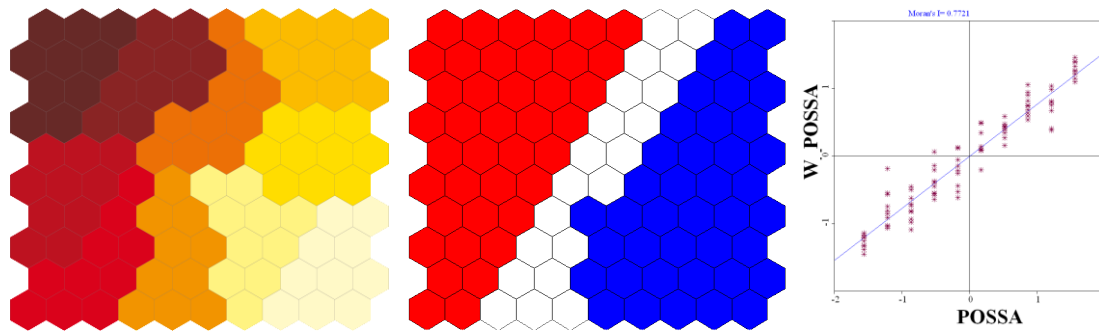


figure 23 : archétype de structure spatiale, tendance.
Données Wong (2011)

Le premier cas de figure envisagé est la présence d'une tendance spatiale. En présence d'une tendance spatiale, la mesure de l'autocorrélation spatiale est évidemment très forte (I de Moran = 0,77). Le nuage de points de Moran marque une remarquable régularité, et seules les valeurs pour les individus au contact des deux poches d'autocorrélation spatiale positive ne sont pas significatives. On notera que ces valeurs sont les valeurs moyennes de la distribution, qui ne sont (par construction) jamais significatives.

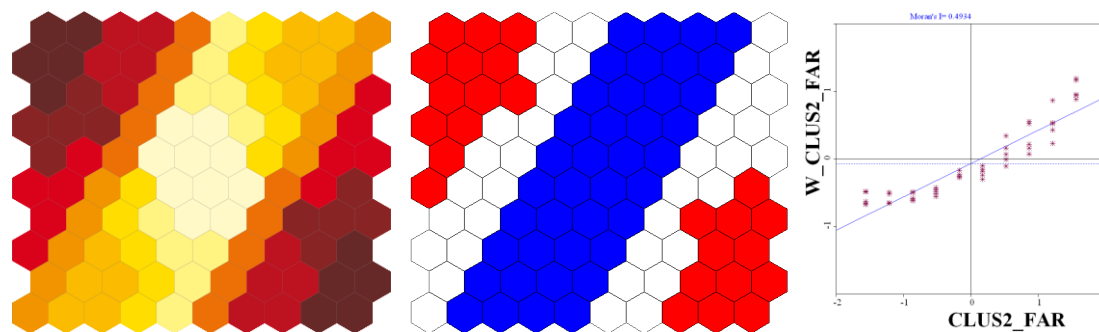


figure 24 : archétype de structure spatiale, regroupement à deux angles.
Données Wong (2011)

Le modèle suivant envisage deux regroupements situés à des angles opposés. L'autocorrélation spatiale globale est encore marquée (I de Moran = 0,49). On constate une forte autocorrélation spatiale locale pour chacun de ces deux regroupements, mais aussi pour la bande centrale qui constitue un regroupement de valeurs faibles, très bien identifiées par les I de Moran locaux, mais ne jouant pas pour autant un rôle très marqué dans la mesure globale.

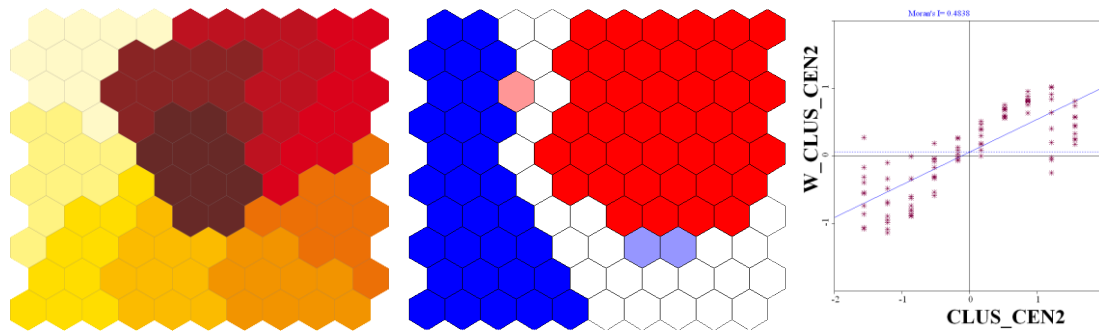


figure 25 : archétype de structure spatiale, regroupement local.
Données Wong (2011)

Wong présente ici un modèle qui illustre le cas d'un regroupement de valeurs situées au centre et entouré d'autres regroupements, mais moins marqués. Si la valeur du I de Moran est proche de celui de la figure précédente (I de Moran = 0,48), on voit que la structure spatiale est assez différente (que ce soit en regardant le nuage de points ou la cartographie des *clusters*).

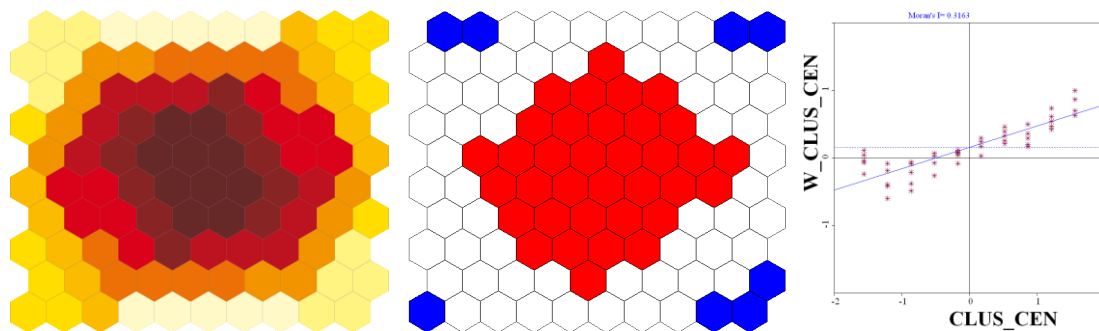


figure 26 : archétype de structure spatiale, regroupement centré.
Données Wong (2011)

La figure 26 montre une organisation spatiale concentrique autour d'un regroupement. On y voit une structure spatiale forte, mais peu caractéristique de l'autocorrélation spatiale. Il en ressort un I de Moran finalement assez faible (0,32)¹¹⁸. Nous avons déjà fait la remarque (Oliveau, 2004 : 189 et suivantes) que ces structures aréolaires, assez typiques des polarisations urbaines, marquaient fortement l'espace sans constituer l'élément essentiel des effets de régionalisation.

¹¹⁸ On le voit sur le nuage de points, ce I de Moran faible est aussi le résultat de l'absence d'autocorrélation spatiale dans les valeurs faibles (cluster *low-low*).

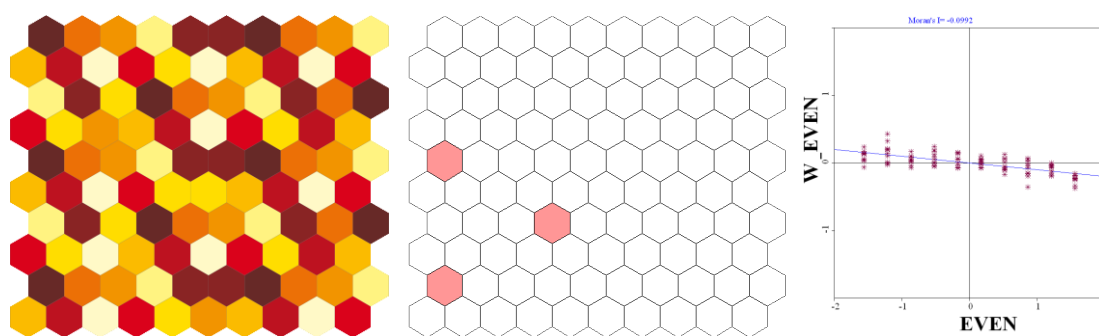


figure 27 : archétype de structure spatiale, disposition aléatoire.
Données Wong (2011)

La distribution spatiale aléatoire donne les résultats attendus : le I de Moran est presque nul (I de Moran= 0,10), et on ne peut pas détecter de regroupements locaux significatifs. Malgré tout, on peut identifier des variables spatialement atypiques. Leur niveau de significativité est assez faible : ces valeurs spatialement atypiques apparaissent et disparaissent au gré des simulations.

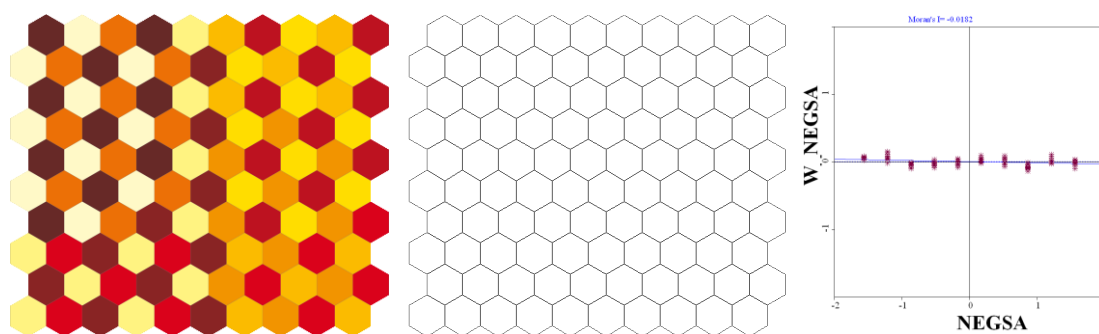


figure 28 : Archétype de structure spatiale, voisins opposés.
Données Wong (2011)

Le dernier cas de figure pourrait présenter une autocorrélation spatiale négative, mais donne une autocorrélation spatiale nulle (I de Moran= -0,02)¹¹⁹. Il nous permet d'aborder ici une caractéristique spécifique de l'autocorrélation spatiale négative, que nous avons déjà évoquée. Elle est difficile à saisir, et ce pour deux raisons liées. La première, la plus importante, concerne la définition du voisinage. Si l'on envisage un voisinage défini par une contiguïté d'ordre 1, alors l'autocorrélation spatiale devient clairement négative (voir figure 29 ; I de Moran= 0,62). On le constate ici, la mesure de l'autocorrélation spatiale négative est beaucoup plus dépendante de la définition du voisinage que l'autocorrélation spatiale positive.

¹¹⁹ C'est l'occasion de rappeler qu'une autocorrélation spatiale nulle peut parfois cacher des structures spatiales à d'autres échelles (voisinage ou résolution).

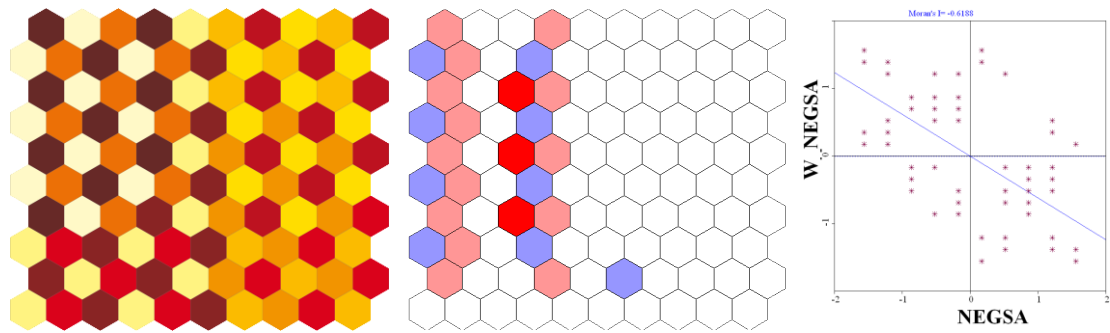


**figure 29 : archétype de structure spatiale, voisins opposés
(voisinage par contiguïté d'ordre 1).
Données Wong (2011)**

La seconde raison concerne l'autocorrélation spatiale négative elle-même, dont nous avons déjà dit qu'elle était difficile à saisir. Ainsi, outre sa sensibilité à la définition du voisinage, elle est aussi beaucoup plus sensible à la présence de valeurs moyennes. En effet, dans le carroyage spécifique nécessaire à l'existence de l'autocorrélation spatiale, les valeurs moyennes agissent de manière plus marquée que dans les regroupements de valeurs similaires : le fait de disséminer dans l'espace des valeurs moyennes tend à ramener l'ensemble des mesures vers la moyenne mais les regroupements de valeurs proches y sont moins sensibles.

Maintenant que la méthode est maîtrisée, comment s'en servir dans des études en géographie, au-delà (même si c'est déjà suffisant en soi) de repérer les structures spatiales fines d'un phénomène social ? Rapidement, on envisage de comparer des phénomènes entre eux sur un même espace, ce que nous avons fait pour différents indicateurs démographiques en Inde (Oliveau, Guilmoto, 2005), on peut aussi proposer de comparer les situations dans différents espaces, ce que nous avons fait en comparant les sex-ratios en Inde et en Chine (Guilmoto Oliveau, 2005).

Puis, la méthodologie nous rappelle, de nouvelles questions apparaissent. Que se passe-t-il lorsque l'on change de niveau, d'échelle ? Le(a)quel(le) choisir pour travailler ? Qu'entraînent les passages de l'un à l'autre sur les perceptions que l'on a des phénomènes et les mesures que l'on en fait ?

Jeux d'échelles

Avant d'entamer nos réflexions à propos des échelles et des niveaux, il nous a semblé prudent de retourner voir les définitions de ces mots. Nous en aurons besoin de plusieurs pour décrire la réalité des jeux que nous allons proposer au lecteur. Nous envisagions qu'il y eût une différence et que la subtilité du langage devait être éclaircie, alors que l'habitude (et l'enseignement) nous amène parfois à ne pas être suffisamment précis.

Nous nous sommes donc référés aux deux "essentiels" de la géographie que constituent *les Mots de la Géographie* (Brunet, et al. 1992) et *Hypergeo* (Elissalde, non daté¹²⁰). Le second ne comporte que l'entrée « échelle » (rédigée par Christian Grataloup). De la lecture de ces trois articles, il ressort que l'échelle est l'ensemble des niveaux (mais on peut dire aussi échelons) qui la composent... On ne change donc pas d'échelle, mais de niveau, bien que l'échelle puisse être comprise dans son sens cartographique de rapport entre deux dimensions, auquel cas le changement de niveau constitue un changement d'échelle. Ne nous voilà pas beaucoup plus avancé.

¹²⁰ Voilà un merveilleux outil, que l'on retrouve régulièrement dans les travaux de nos étudiants que l'on ne sait comment citer, alors qu'elle le mérite. *Hypergéô* est une « une encyclopédie électronique, en accès libre, consacrée à l'épistémologie de la géographie ». Un copyright sur le site indique 2004, mais on ne peut pas en savoir plus.

Nous parlerons donc de « résolution spatiale » (fine ou grossière) ou de « niveau » pour identifier le maillage utilisé (commune, département, région etc.) et de « voisinage »¹²¹ (proche ou lointain) pour parler des différences de contiguïtés (ordres) ou de distance (portée). Il s'agit bien là pourtant de deux dimensions scalaires différentes qui jouent sur les mêmes espaces.

La première remarque concernera le voisinage. Il semble que lorsque l'on augmente la portée (accessibilité) ou l'ordre (contiguïté), voire le terme « k » (proximité), le niveau global d'autocorrélation spatiale tend à diminuer. En effet, en changeant le voisinage, on ne change qu'une partie de l'indicateur, celui qui se réfère à j . Dans le cas du I de Moran, par exemple, seule change la partie $(z_j - \bar{z})$. Or on sait par ailleurs que si le nombre d'individus augmente, leur moyenne va se rapprocher de la moyenne globale de la population. Autrement dit $(z_j - \bar{z})$ va diminuer. Le numérateur du I de Moran tendra donc à diminuer alors que le dénominateur restera constant. Le I de Moran tend donc à diminuer de façon automatique avec l'augmentation du voisinage, ce que montrent d'ailleurs fort bien les corrélogrammes. Il y a donc dans la décroissance des pentes des corrélogrammes (voir figure 15, page 59) une part liée à la construction de l'indice lui-même, et pas uniquement à la structure spatiale intrinsèque de la variable.

Pour illustrer cette caractéristique, nous reprendrons les données sur le sex-ratio des enfants utilisées dans Guilmoto & Oliveau (2007), en nous concentrant uniquement sur les données indiennes. Pour cela nous présenterons d'abord la cartographie à l'échelle indienne (figure 30), puis trois cartes (figure 31) des clusters d'associations spatiales pour les ordres de contiguïtés 1, 2, 3. La valeur du I de Moran est respectivement de 0,69 ; 0,62 ; 0,57.

La structure spatiale globale, très forte, cache en fait deux phénomènes. Le premier se caractérise par la concentration de forts déséquilibres en grandes poches dans la moitié nord-ouest du pays¹²², le second par l'existence de poches d'équilibre du sex-ratio, autocorrélées elles-aussi, mais qui semblent moins structurées. Lorsque l'on augmente le voisinage, l'autocorrélation spatiale globale diminue, mais le nombre d'unités significatives augmente. De plus, le paysage se marque par une opposition plus marquée entre moitié nord et ouest et moitié sud et est.

¹²¹ Nous ne précisons pas que le voisinage est spatial, puisqu'il l'est étymologiquement, tout comme la proximité, d'ailleurs.

¹²² Ces grandes poches, tendent d'ailleurs à se rejoindre. On ne peut le voir ici, mais la cartographie du phénomène sur plusieurs années le montrerait par ailleurs.

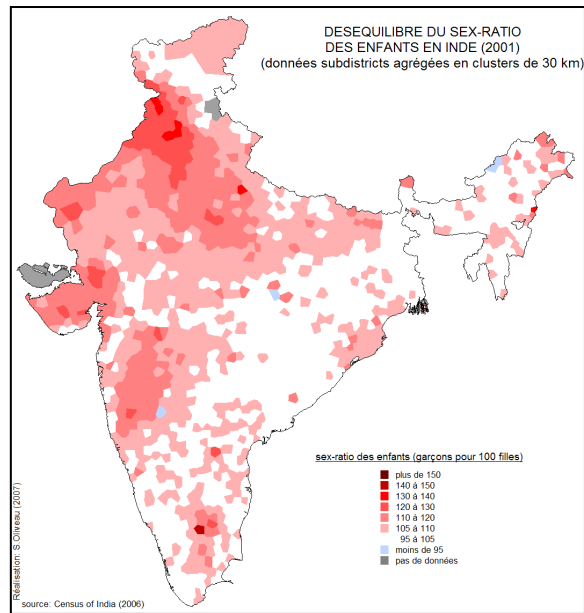


figure 30 : le sex-ratio des enfants en Inde (1991)
(Guilmoto, Oliveau, 2007)

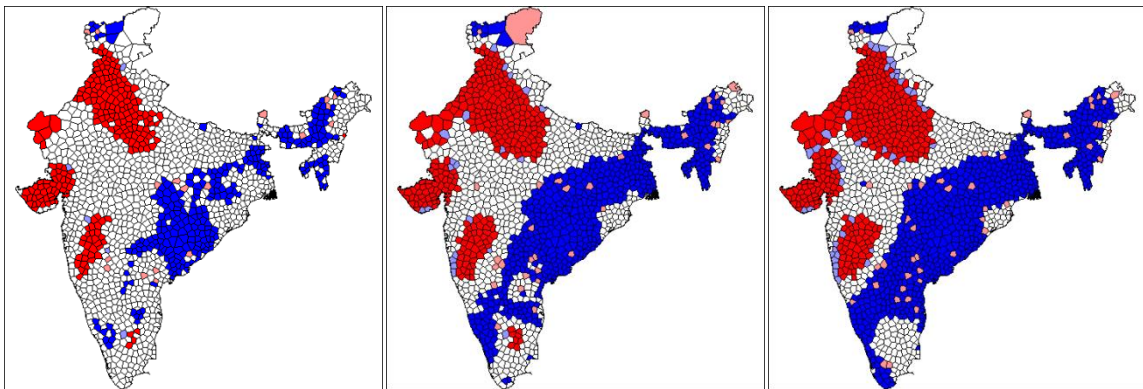


figure 31 : cartographie des clusters pour le sex-ratio des enfants en Inde (1991)
de gauche à droite :contiguïté nodale d'ordre 1, 2 et 3 (Guilmoto, Oliveau, 2007)

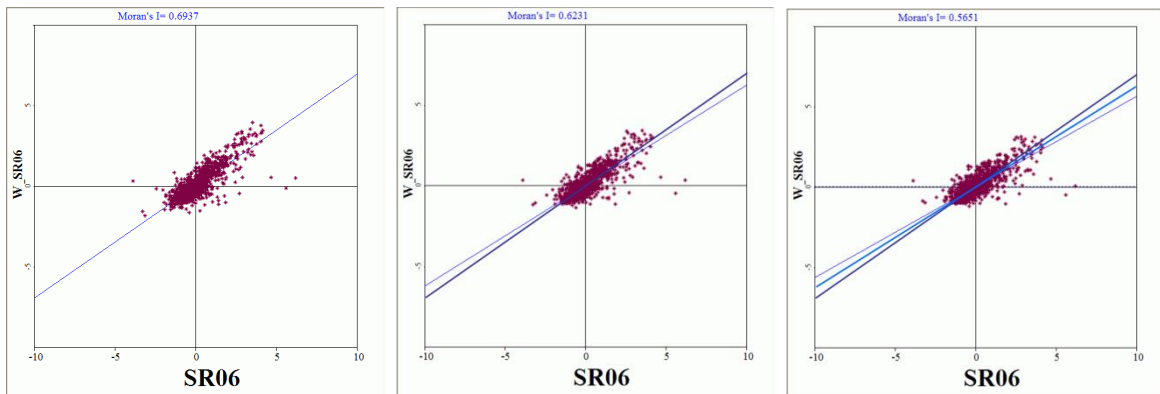


figure 32 : nuage de points de Moran pour le sex-ratio des enfants en Inde (1991)
de gauche à droite :contiguïté nodale d'ordre 1, 2 et 3 (Guilmoto, Oliveau, 2007)

La figure 32 compare les nuages de points de Moran correspondant aux trois cartes de la figure 31. On a conservé les droites des contigüités inférieures pour visualiser le changement. On y voit la stabilité globale de la forme du nuage alors que le voisinage change, mais on perçoit une sorte de tassement vers le centre, notamment en ce qui concerne les valeurs les plus extrêmes du premier quadrant (regroupements *high-high*). Les valeurs spatialement atypiques restent les mêmes. On voit aussi très bien la droite de régression se rapprocher de l'horizontale au fur et à mesure que le voisinage augmente.

Dans l'étude des structures spatiales des phénomènes sociaux, on essaye généralement d'avoir accès à une information ayant la résolution la plus fine possible, sachant que l'on peut toujours l'agréger pour aller vers des niveaux supérieurs, alors que l'inverse n'est pas possible. Par ailleurs, la résolution influe directement sur le voisinage que l'on peut explorer : une résolution fine permet d'explorer des grandes variétés de voisinages, qui se réduisent au fur et à mesure que l'on agrège.

Cet aspect des choses a déjà été évoqué (Oliveau, 2010 ; Guilmoto, 2005), nous les reprendrons donc succinctement. Le premier constat concerne le changement de résolution spatiale. Il faut garder en tête que tous ces indicateurs mesurent avant tout l'hétérogénéité spatiale des variables. Or, lorsque l'on agrège des entités, spatiales, les valeurs tendent à se rapprocher de la moyenne générale, autrement dit, l'hétérogénéité diminue, comme en témoigne d'un point de vue strictement statistique la baisse de la variance de ces variables. Arbia l'a bien résumé sous la forme d'une seconde loi de la géographie¹²³ (Arbia *et al.*, 1996) : « tout est en relation avec tout le reste, mais les choses observées à une résolution grossière le sont plus que les choses observées à une résolution fine »¹²⁴.

A défaut de le démontrer (mais Reynolds, 1998 l'a fait très bien), nous l'avons souligné à partir de différents exemples. Le premier concerne la fécondité¹²⁵ en Inde du Sud¹²⁶ et montre comment le *I* de Moran augmente au fur et à mesure que l'on change de niveau (figure 33).

¹²³ Cette seconde loi est bien sûr à prendre comme un clin d'œil à la première.

¹²⁴ « Everything is related to everything else, but things observed at a coarse spatial resolution are more related than things observed at a finer resolution. »

¹²⁵ Plus exactement un indicateur indirect de la fécondité, le rapport enfants-femmes qui permet de travailler à des échelles fines malgré le manque d'information quant à la structure par âge.

¹²⁶ Les données sont celles du SIFP et ont comme source le recensement de la population de 1991.

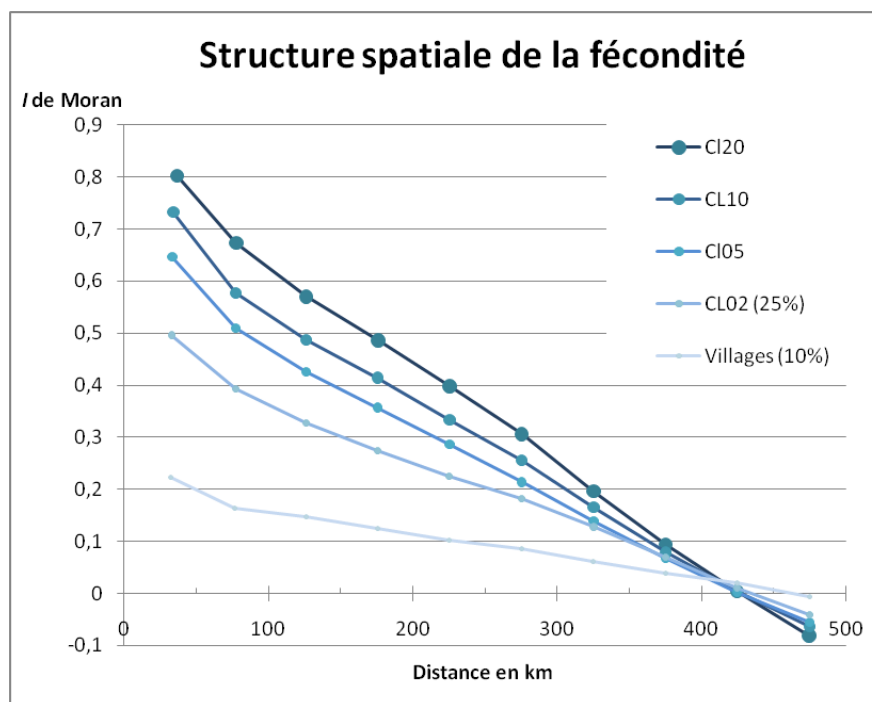


figure 33 : mesure de l'effet d'agrégation sur l'autocorrélation spatiale globale rapport enfants-femme en Inde du Sud (d'après Guilmoto, 2005)

En partant des villages, nous agrégeons les unités selon des regroupements (clusters) de différentes tailles : clusters de 2 km, 5 km, 10 km, 20 km¹²⁷. Le fait de passer des villages aux clusters de 2 km, évacue une part importante du bruit statistique induit par la petite taille démographique des unités, il en résulte une homogénéité bien plus grande des unités et une augmentation importante du *I* de Moran. Cela nous rappelle qu'il faut parfois se méfier des approches à des résolutions spatiales fines pour des indicateurs qui nécessitent des effectifs importants pour être stabilisés et le recours nécessaire aux regroupements et aux approches agrégées pour mesurer les phénomènes démographiques (voir aussi Guilmoto, Oliveau, 2007).

Ces jeux d'échelles ne sont donc pas qu'anecdotiques. Ils influent directement sur notre perception des phénomènes. Cela nous rappelle que réfléchir aux unités spatiales et aux voisinages utilisés est fondamental et qu'il n'y a pas de "recette toute faite". On soulignera par ailleurs que les comparaisons restent sujettes à précaution particulièrement lorsque les résolutions spatiales des unités spatiales observées ne sont pas similaires (il devient alors assez difficile de comparer des mesures d'autocorrélation spatiale).

¹²⁷ La méthodologie de ces regroupements est expliquée dans Guilmoto *et al.*, 2002.

La question du MAUP

Ces réflexions sur les échelles se rapportent à une question plus large connue sous le terme de MAUP (*Modifiable Areal Unit Problem*). Il n'y a pas de traduction établie et stable en français pour le terme MAUP (De Ruffray *et al.* font le même constat en 2011), la grande majorité de la bibliographie étant en anglais et les francophones n'ayant pas fait trop d'effort de traduction (c'est souvent le cas face à des acronymes...). La première traduction (Openshaw, 1981) proposait « problème de l'agrégation spatiale ». On peut proposer « problème des surfaces aréolaires variables » ou « problème des unités spatiales modifiables », voire « problème des unités géographiques modifiables ». Finalement, cela ne constitue peut être qu'un problème de « variabilité des maillages » (comme le fait Denise Pumain dans l'article « maillage » d'*Hypergeo*).

Le MAUP est un sujet récurrent en analyse spatiale, et nous allons donc y revenir rapidement¹²⁸. C'est Stan Openshaw (et Peter Taylor) qui baptisent le MAUP dans un article de 1979 : « a million or so correlation coefficients : three experiments on the Modifiable Areal Unit Problem ». L'idée, développée en détail par la suite (Openshaw, 1984¹²⁹), est que les mesures statistiques faites sur des espaces sont directement dépendantes de la résolution et de la forme des mailles utilisées. On distingue donc habituellement ces deux effets sous les termes d'effet d'échelle d'une part et d'effet de zonage d'autre part.

L'effet d'échelle insiste sur les variations que subissent les données lorsque l'on change de niveau d'observation, notamment l'effet de lissage statistique qui apparaît lors de l'agrégation des données. La disparition des détails au fur et à mesure que l'on « monte » dans les niveaux géographiques et son apport à la compréhension des phénomènes et de leurs structures ont déjà été bien soulignés (en français, voir Mathian, Piron, 1999).

L'effet de zonage souligne le rôle des formes des découpages territoriaux sur les résultats. Ainsi, lors du processus d'agrégation le nombre de possibilités pour associer les mailles entre elles et en former de nouvelles est d'autant plus important que le nombre de mailles est grand. Le nombre de combinaisons possibles augmente donc rapidement et chaque solution entraînera des changements de valeurs pour les mailles nouvellement obtenues. L'effet de zonage est caractéristique des découpages administratifs (particulièrement électoraux) et vient s'ajouter à l'effet d'échelle.

¹²⁸ On consultera sur le MAUP le chapitre de Wong (2009) et le numéro spécial de la revue *Geographical systems* (Wong & Amrhein, 1996b).

¹²⁹ En français : Openshaw, 1981.

Puisque le MAUP a des effets sur les mesures, il a des implications sur les statistiques. Ainsi, Gehlke & Bielh (1934 : 170) sans se référer au MAUP remarquent déjà que les « variations dans la valeur du coefficient de corrélation semblent conditionnées par les changements dans la taille des mailles utilisées, avec une valeur du r plus petite lorsque les unités sont plus petites. Des manières différentes de regrouper les unités ont une influence considérable sur le r , comme l'ont les différences de surfaces. »¹³⁰. Ce qui vaut pour la corrélation linéaire vaut pour l'autocorrélation spatiale, comme le montre Reynolds (1998) à propos du I de Moran ou Amrhein & Reynolds (1996) pour la statistique G_i .

Si la question est ancienne on voit que le problème reste d'actualité. L'effet de zonage reste peu étudié et mériterait certainement d'être exploré plus à fond. L'effet d'échelle est plus facile à mesurer mais mériterait d'être lui aussi spécifiquement étudié pour que l'on puisse envisager de mieux le prendre en compte dans les mesures (Reynolds, 1998).

Par ailleurs, une autre question se pose, celle de savoir si le MAUP, plutôt que de constituer un « problème » ne serait pas en fait constitutif de la donnée spatiale, et, à ce titre, largement autant une information qu'un problème (Grasland & Madelin, 2006 ; Grasland *et al.* 2007). En effet, une fois alerté de l'existence de ce phénomène de variabilité due à l'agrégation, son exploration constitue une nouvelle source d'informations, comme nous l'avons déjà souligné précédemment.

Des gammas intermédiaires ? Une nouvelle utilisation des LISA.

L'exploration des échelles spatiales des phénomènes sociaux (niveaux et résolutions) peut être complétée par une interprétation des indices d'autocorrélation spatiale à des échelles intermédiaires entre les LISA et les statistiques globales, que nous nommerons donc « gammas intermédiaires »¹³¹.

Les LISA, nous l'avons rappelé page 69, sont proportionnels aux indices globaux. Suivant cette relation, il est donc possible en connaissant les valeurs des LISA pour tous les individus de recalculer la valeur de l'indice global. Il serait donc aussi possible de calculer des indices intermédiaires, qui seraient la somme des indices locaux pour un espace donné.

¹³⁰ « Variations in the size of the correlation coefficient seem conditioned upon changes in the size of the unit used, with a smaller value of r associated with the smallest unit rather than with the largest. Various ways of grouping have considerable influence on the r , as well as has the size of the area ».

¹³¹ A notre connaissance, il n'y a pas encore eu d'autre utilisation similaire des LISA.

Il faut bien garder en tête néanmoins qu'il ne s'agit pas de recalculer un indice global pour la zone étudiée, ce qui donnerait des résultats différents, mais de faire la somme des indices locaux. Autrement dit, cette somme se réfère à la moyenne de l'ensemble des valeurs et non pas du sous-espace envisagé.

Pourquoi souhaitons-nous mesurer des gammas intermédiaires plutôt que de recalculer un indice local à l'échelle du sous-espace envisagé ? La mesure d'un gamma intermédiaire évalue le niveau d'autocorrélation spatiale dans une région donnée par rapport à l'ensemble de l'espace étudié. Il permet donc de mettre en exergue les espaces qui participent le plus à l'autocorrélation spatiale globale, c'est-à-dire les régions les plus structurantes à l'échelle globale. Nous avons développé cette idée dans le cadre de l'analyse des inégalités spatiales de la modernisation au Tamil Nadu, complétant ainsi l'analyse précédemment développée (Oliveau, 2004, 2006). A partir de l'analyse des LISA (cartographie présentée page 75), nous avons isolé chaque regroupement d'associations spatiales et mesuré pour chacun d'eux un gamma intermédiaire. La figure 34 présente ce résultat.

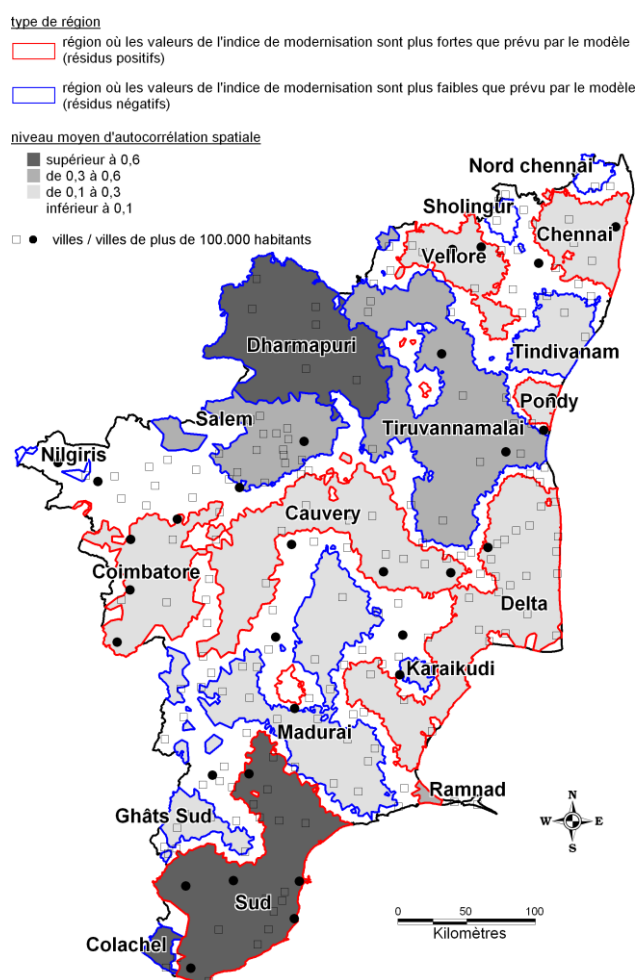


figure 34 : cartographie des gammas intermédiaires.

La cartographie obtenue vient remettre en perspective la cartographie précédente des clusters en hiérarchisant les espaces en fonction de la valeur moyenne des LISA observés, c'est-à-dire leur participation au I de Moran global. La carte nous montre ainsi que les régions les plus influentes pour la détermination du I de Moran (0,24) sont le Sud (avec un gamma intermédiaire de 0,79) et Dharmapuri (gamma intermédiaire de 0,63). On notera que le record est à Colachel (0,92).

Une approche bivariable de l'autocorrélation spatiale

Nous nous sommes concentré sur la mesure de l'autocorrélation spatiale et avons volontairement laissé de côté les autres dimensions des variables sociales. Il peut pourtant être intéressant d'explorer les structures spatiales des données sociales en tenant compte parallèlement d'autres dimensions : sociale, écologique, temporelle, etc. On peut envisager de regarder le lien entre pauvreté et couvert forestier dans les pays du Sud (Sunderlin *et al.*, 2007), d'évaluer la concentration spatiale d'entreprises dans le temps (Felkner & Townsend, à paraître) ou encore de faire l'hypothèse que la mortalité dépende de la pollution alentour (Hu & Rao, 2009).

La mise en œuvre, là encore, n'est pas très difficile¹³². Si l'on se réfère, par exemple, à la définition du I de Moran, on pourrait très bien remplacer la valeur des voisins pour la variable (Z_j) par la valeur des voisins pour une autre variable (U_j). On obtient alors un LISA bivariable (on trouve parfois l'acronyme « BiLISA ») dont la formule est donnée ci-dessous :

$$\text{équation 12 : LISA bivariable (BiLISA) : } I_i = \frac{\sum_j w_{ij} (z_i - \bar{z})(u_j - \bar{u})}{\sum_i (z_i - \bar{z})^2}$$

avec U une seconde variable dont les valeurs u_j sont prises dans le voisinage j de i .
(d'après Anselin *et al.*, 2002)

On peut tout aussi bien considérer que la variable U représente la variable Z à un temps antérieur (« t-1 »), ce qui nous permet d'introduire le temps et d'obtenir une analyse proprement spatio-temporelle (même si elle reste assez frustrée, puisqu'elle ne permet d'analyser qu'un seul pas de temps). Néanmoins, on peut ensuite multiplier les BiLISA

¹³² A condition de ne pas oublier de standardiser les variables préalablement.

temporels à différents pas de temps pour observer les changements... On le voit, les applications potentielles sont nombreuses.

Dans le cadre particulier des LISA bivariés, on peut se dire que l'exclusion de i de la définition de son voisinage est à discuter. En effet, il peut être intéressant de prendre en considération la possibilité que j comprenne la valeur de i comme cela a été évoqué précédemment à propos de la statistique G^*_i (voir page 67) : la valeur de l'individu à un temps antérieur est corrélée à la valeur au temps observé.

La régression spatiale

Après avoir mesuré les structures spatiales, on sera évidemment intéressé à les intégrer dans des modèles, c'est d'ailleurs l'objectif premier de l'approche économétrique, qui a fortement contribué au développement des statistiques spatiales. Nous disposons pour cela de différents modèles de régressions spatiales (voir Anselin, 1988)¹³³.

On distingue actuellement deux approches principales dans la mise en place de régressions incluant des dimensions spatiales identifiées. La première est celle des modèles spatiaux autorégressifs et la seconde celle des modèles à filtrage spatial¹³⁴. Les premiers intègrent l'autocorrélation spatiale dans les modèles alors que les seconds séparent les effets spatiaux des variables (d'où la notion de filtrage spatial) : ils suppriment l'autocorrélation spatiale pour rendre les variables spatialement indépendantes (Getis & Griffith, 2002). Pour notre part, nous en resterons ici aux modèles spatiaux autorégressifs pour deux raisons. La première est liée à la plus grande maîtrise de ce champ¹³⁵. La seconde repose sur l'idée d'incorporer l'espace dans le modèle, comme partie intégrante des phénomènes, là où le filtrage spatial cherche à l'extraire des données, comme un biais. D'une manière qui peut sembler paradoxale, cette deuxième position est celle des économètres alors que les filtrages sont développés par des géographes...

La régression spatiale fonctionne sur le même principe que la régression aspatiale. Pour mémoire, l'idée est d'exprimer un phénomène en fonction d'autres phénomènes, auquel on ajoute un terme d'erreurs (couramment qualifié de « résidu »). On l'exprime souvent sous

¹³³ L'objectif ici n'est pas de rentrer en détail dans l'économétrie spatiale mais de proposer un aperçu très rapide des possibilités offertes et des solutions proposées accessibles.

¹³⁴ Nous faisons volontairement l'impasse sur la régression géographiquement pondérée (*geographically weighted regression*) de Fotheringham *et al.* (1996). <http://ncg.nuim.ie/ncg/GWR/>

¹³⁵ Nous cherchons avant tout à exploiter des méthodologies éprouvées, répandues et reproductibles. Si les filtres spatiaux semblent un peu plus efficaces statistiquement, ils restent plus difficiles à mettre en œuvre.

cette forme : [Données= modèle + erreurs]. Ce que l'équation 13 présente de façon plus formelle.

équation 13 (modèle de régression linéaire classique) :

$$Y = \beta.X + \varepsilon$$

avec : Y la variable à expliquer¹³⁶,
 X la variable indépendante (explicative),
 β le paramètre
 ε un terme d'erreurs.

On peut alors introduire la dimension spatiale de deux manières dans la régression. On peut d'abord considérer que l'autocorrélation spatiale observée est le résultat de l'influence des valeurs voisines sur la variable (« effet spatial » pour reprendre la terminologie de Baller *et al.* 2002). On peut ensuite envisager que l'autocorrélation spatiale est due à une variable qui n'est pas prise en compte dans le modèle et dont l'influence est par conséquent présente dans le terme d'erreurs (« perturbation spatiale »).

On utilisera dans le premier cas un modèle dit à variable endogène spatialement décalée (*spatial lag model*), en raccourci modèle à variable décalée. [Données = modèle spatial + erreurs]. Dans le second cas, on se référera à un modèle dit à erreurs spatialement autocorrélées (*spatial error model*) en raccourci modèle à erreurs spatiales [Données = modèle + erreurs spatiales]. La figure 35 illustre les différences entre ces différents modèles.

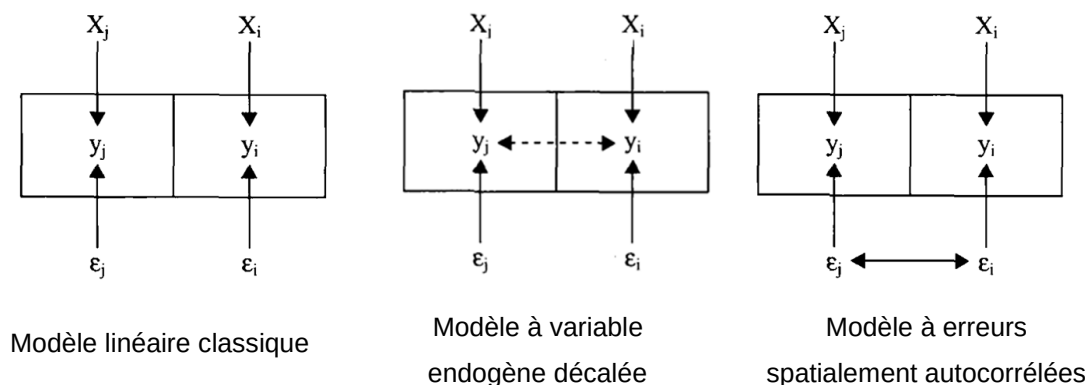


figure 35 : modèles de régression
D'après Baller et al. (2001)¹³⁷

¹³⁶ Les coordonnées géographiques n'apparaissent pas en économétrie, puisque l'espace est représenté via des matrices de pondérations. Il est d'usage d'utiliser X et Y (et non plus Z) pour les variables étudiées.

¹³⁷ L'illustration proposée par Baller *et al.* est plus complète. Il s'agit d'une simplification proposée par Ronald Briggs dans un cours sur l'autocorrélation spatiale à l'université du Henan (Chine). www.utdallas.edu/~briggs

On peut bien sûr présenter ces modèles de façon plus formelle :

équation 14 (modèle de régression à variable endogène décalée) :

$$Y = \rho WY + \beta.X + \varepsilon$$

W la matrice de pondération spatiale et ρ un paramètre.

L'équation 14 montre la prise en compte de l'autocorrélation spatiale de la variable Y à travers la mise en place de la matrice de poids spatiaux W , auquel on adjoint un paramètre ρ .

équation 15 (modèle de régression à erreurs spatialement autocorrélées) :

$$Y = \beta.X + \rho W\varepsilon + \xi$$

Pour faciliter la lecture, l'équation 15 est présentée sous une forme semblable à l'équation 14. Ici, c'est le terme d'erreur ε qui est autocorrélé. On ajoute un élément « ξ » qui correspond au bruit résiduel (c'est une variable iid - indépendante et identiquement distribuée).

Le modèle à variable endogène décalée ne permet de contrôler que l'autocorrélation spatiale de la variable et peut donc laisser la place à de l'autocorrélation spatiale dans les résidus. Le modèle à erreurs spatiales contrôle directement les résidus et est donc plus robuste statistiquement, mais sa capacité explicative s'en trouve réduite d'autant. Il existe des tests pour départager l'utilisation préférentielle de l'un ou de l'autre (tests fondés sur les multiplicateurs de Lagrange - *Lagrange Multiplier*¹³⁸). On s'oriente aujourd'hui vers des modèles de régression spatiale mixte (que décrit la figure 36) qui intègrent les deux approches (voir l'équation 16 pour la formalisation).

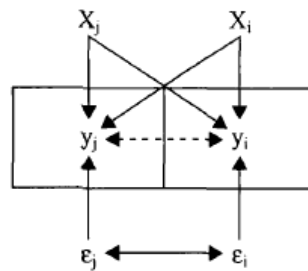


figure 36 : modèles de régression spatiale mixte
D'après Baller et al. (2001)

équation 16 (modèle de régression spatiale mixte) :

$$Y = \rho WY + \beta.X + \lambda W\varepsilon + \xi$$

¹³⁸ Ces tests se retrouvent dans la littérature (Anselin *et al.*, 1996) et les logiciels sous la dénomination LMlag (pour les modèles à variable décalée) et ou LMerr (pour les modèles à erreur spatiale). Il en existe des versions robustes RLMlag et RLMerr, mais ceci nous éloigne de nos intérêts présents.

On l'imagine, les applications sont nombreuses et les économistes, particulièrement ceux de l'économétrie spatiale et des sciences régionales, s'en sont emparés depuis longtemps. De même, les épidémiologues intègrent désormais la dimension spatiale dans leurs travaux, tout comme les démographes. Les géographes, surtout en France, semblent à la traîne. Les revues françaises ne montrent pas de signe de production, alors que l'espace est roi dans ces approches. Cruel paradoxe. Les géographes ont pourtant un rôle important à jouer, et pas simplement sur l'évolution méthodologique, mais sur l'interprétation des phénomènes.

A deux niveaux, l'expertise géographique peut intervenir de manière très pertinente. La plupart des travaux reviennent sur la difficulté de mettre en place une matrice de pondération adéquate. Or, cette matrice de pondération n'est jamais que la formalisation d'un voisinage : on en revient au questionnement initial de notre travail. Les géographes restent les plus à même, parce que les plus sensibilisés dans leur parcours, à donner du sens au voisinage. Au-delà de la mesure, il y a l'interprétation que l'on en donne, et les niveaux de voisinage pertinents pour une l'étude d'une variable renvoient directement à sa structuration spatiale, autrement dit aux interactions qui l'ont construite.

De même, la réflexion autour du rôle que peut jouer l'espace dans les sociétés amène les chercheurs à introduire des variables venant de champs très différents, à croiser toujours plus les disciplines. Le reproche longtemps fait aux géographes devient alors une qualité. Qui peut mieux que les géographes intégrer dimensions sociales et environnementales dans des approches qui mettent en jeu la spatialité des sociétés ?

Retour aux sources

Au cœur du fonctionnement des sociétés, il y a le rapport de l'homme à l'espace, et c'est ce que les géographes essayent de montrer. Nous avons pris le parti de souligner que cet espace contraint l'activité humaine. Cette contrainte s'exprime sous la forme d'une résistance aux interactions, la fameuse « rugosité de l'espace » (Helle, 1993), qui n'est autre que la concrétisation de la « tyrannie de la distance » (Bairoch, 1985 : 33). Cette résistance aux interactions entraîne une différenciation des espaces, par des phénomènes de regroupements (et dans une moindre mesure d'exclusion). L'espace différencié présente à l'observateur des formes visibles, que l'on sera tenté de décrypter. Au-delà de les lire, on a cherché à mesurer ces formes. La mesure permet l'objectivation d'une partie du phénomène, condition nécessaire à la pratique scientifique.

Il faut pourtant mettre en garde contre les mesures. Utiliser aveuglément, elles ne permettent pas de remplir leur tâche initiale. Il faut donc apprendre à les maîtriser, c'est-à-dire comprendre pourquoi elles sont construites, comment elles fonctionnent (ce qu'elles montrent et ce qu'elles cachent) et qu'elles sont les éléments qui les influencent. Alors, mais alors seulement, elles deviennent de précieuses auxiliaires pour le chercheur.

Distance et mesure

Dans ce volume, nous avons donc souhaité revenir sur les mesures de l'autocorrélation spatiale en insistant plus particulièrement sur les dimensions multiscalaires de certains indices et les utilisations que l'on pouvait en faire, de la description à la modélisation des phénomènes sociaux. Ces mesures reposent avant tout sur des matrices de voisinage, qui constituent le "cœur" de ces mesures et conditionnent les résultats obtenus. Il nous a donc semblé primordial de revenir sur les différentes définitions du voisinage, souligner leurs différences et montrer dans quelle mesure les particularités de chacun influent les perceptions que l'on en retirera. Nous avons pu montrer dans quelques cas extrêmes que la définition du voisinage peut radicalement changer la mesure, la faisant passer du négatif au positif... Cela n'invalidé en rien la mesure mais nous rappelle avec quelle prudence il faut toujours agir. La multiplication des angles de vue est un élément essentiel de la description, qui permettra ensuite de proposer des modélisations pertinentes. Car, au bout de ce travail, c'est bien la

recherche de modèles qui nous fait avancer. C'est par la modélisation que l'homme comprend le monde et le dit.

Pour notre part, nous avons fait le choix d'utiliser des modèles utilisant un formalisme de type mathématique, nous nous en expliquons dans le volume 2. Nous nous sommes donc penché sur les indices permettant de mesurer l'autocorrélation spatiale et avons re-parcouru le chemin depuis leur mise en place jusqu'à leur utilisation : des années 50 à aujourd'hui, du global au local. Avant de démarrer, nous avons souhaité revenir sur l'élément essentiel et pourtant trop négligé que constitue le voisinage. Chemin faisant, nous nous sommes arrêtés pour répondre à nos propres problématiques et creuser l'échelon intermédiaire entre mesure globale et observations locales, tentant d'apporter notre petite pierre à l'édifice.

Cette idée de se reposer sur des mesures d'autocorrélation spatiale pour dessiner des régions avait été développée dès la fin du doctorat, de façon pratique. Elle repose plus largement aujourd'hui pour nous la question de savoir comment on souhaite définir nos objets. Devons-nous nous appuyer sur les dissimilarités, chercher les frontières, ou au contraire insister sur la cohésion et l'homogénéité, quitte à laisser des zones de floues, des marges, entre chaque ? Cette seconde option, que nous avons choisie pour l'instant, s'intéresse donc plus aux phénomènes d'agglomération, de diffusion que de rupture, il offre à voir une vision plus "souple" du territoire, construit comme un continuum. Cette vision continue de l'espace permet de répondre à la critique souvent faite aux approches quantitatives de « mettre les choses dans des boîtes ». Ici, les boîtes ont des contours incertains. Sans pour autant intégrer le domaine de la logique floue¹³⁹, nous ne sommes pas insensibles à ces approches, même si nous avons conscience qu'elles peuvent être déroutantes pour personnes extérieures à la modélisation : on cherche à mettre en place des modèles formels qui produisent des résultats incertains !

Une prochaine étape, du point de vue méthodologique, serait plus certainement de réfléchir aux interactions qui sous-tendent la production de ces formes spatiales que l'on a étudiées, c'est du moins ce que nous avait suggéré la lecture de l'article de Getis (1991). Mais ce n'est peut être plus à travers la statistique spatiale que nous aborderons ces questions, nous y reviendrons plus loin.

Au cœur des méthodologies présentées se trouve la question du voisinage, et nous avons particulièrement insisté sur le rôle spécifique de la distance pour le définir. Ce rôle de la distance comme « attribut premier d'un système spatial » (Grataloup, 1996 : 105), nous

¹³⁹ Nous renvoyons ici le lecteur à l'HDR de Sophie de Ruffray (2007), dont c'est le sujet.

l'avons déjà souligné dans nos travaux de doctorat. Il traverse l'ensemble de notre recherche. C'est pourquoi nous proposons dans le volume 2 d'en faire la clef de voûte de l'épistémologie moderne de notre discipline.

Espace et population

En replaçant la distance et l'espace au cœur de la discipline géographique, on ouvre ainsi la porte aux autres disciplines des sciences sociales. Par l'espace (de sa représentation à sa modélisation), le géographe propose une vision nouvelle (ou au moins renouvelée) aux objets partagés avec les disciplines connexes qui mettent quant à elle l'accent sur d'autres dimensions spécifiques de nos sociétés (temps, ressources, structures sociales, représentations, etc.). Il se place aussi souvent en médiateur, la carte constituant alors un moyen du dialogue interdisciplinaire.

Dans un monde que nous savons complexe (nous y reviendrons), nous ne pensons pas qu'un individu seul puisse prendre en compte les multiples dimensions évoquées. L'interdisciplinarité s'impose à tous. Pour qu'elle soit efficace, le chercheur doit être solide dans sa discipline et se former dans une (ou des) autres(s). Nous avons fait le choix de la démographie, par opportunité et par goût (on y trouve de beaux effets de structure !).

Cette ouverture vers la démographie rentre dans le champ de la maintenant traditionnelle géographie de la population. Alors que les liens entre population et espace sont au centre des nombreuses réflexions et que la démographie a commencé à investir le champ spatial, cette sous-discipline semble en perte de vitesse en France. On peut envisager plusieurs causes, aussi bien structurelles que conjoncturelles. Nous y voyons 3 éléments principaux. La première est due à l'intérêt général de la géographie de la population qui fait que nombreux sont les géographes qui en participent sans pour autant s'y identifier. La géographie de la population, à force d'être évidente (on la trouve comme un enseignement incontournable dans de nombreux cursus de licence) n'a pas été pensée¹⁴⁰. La deuxième tient peut être au manque d'investissement dans la dimension proprement spatiale des phénomènes ou du moins dans la formalisation de ceux-ci¹⁴¹. Le dernier élément, presque ironique, pourrait être un effet démographique de non-renouvellement des générations.

¹⁴⁰ Sans épistémologie, point de salut ?

¹⁴¹ Sans accuser Pierre Georges, la sous-discipline qu'il a contribué à faire émerger n'a peut être pas su se détacher de sa conception vidalienne de la géographie.

Il semble pourtant y avoir une place réelle pour la géographie de la population. Nous l'avons évoquée en début de volume, la cartographie de la pauvreté, et plus généralement la géographie des inégalités (qui insisterait notamment sur les vulnérabilités sociales des espaces) offrent un objet qui permettrait de reprendre une place dans la discipline dans les sciences sociales et plus largement dans le débat public. Les inégalités observées ont des origines et des transcriptions spatiales, les géographes sont donc à même de tenir des discours importants et de ne pas laisser la place aux seuls sociologues et économistes.

Perspectives

Dans cette perspective d'étude des inégalités spatiales que pouvons-nous faire aujourd'hui ? Il nous semble que nous pouvons identifier trois axes de travail. Le premier est de diversifier nos terrains pour multiplier les contextualisations de notre approche. En effet, de la théorie à la pratique il y a une marche conséquente à franchir. Nous avons déjà éprouvé ces méthodes dans le contexte indien à différentes résolutions (villages, sous-districts, districts) et échelles (Tamil Nadu, Inde du Sud, Inde, Inde-Chine). Nous avons commencé depuis 3 ans à explorer de nouvelles échelles à Dakar : résolution fine en milieu intra-urbain¹⁴².

Le programme DemoMed¹⁴³ nous offrira l'opportunité d'explorer avec Yoann Doignon les rives de la méditerranée. DemoMed est un projet d'observatoire démographique de la méditerranée, développé par Isabelle Blöss (Widmer), démographe à l'université de Provence. Nous avons profité de son offre pour intégrer l'équipe et tenter de mettre en valeur les dimensions spatiales des phénomènes démographiques. Vue la diversité des contextes, cela nous permettra de mettre à l'épreuve nos méthodologies. Outre la mesure de l'autocorrélation spatiale se posent des questions en rapport avec le MAUP et surtout des questions pratiques concernant l'harmonisation des maillages administratifs¹⁴⁴. Plus globalement, cela nous donnera l'occasion de mettre en opposition effet de frontière et continuum spatial et nous amènera à observer plus en détail des effets d'hétérogénéité spatiale locale (dans les Balkans par exemple).

¹⁴² Avec l'appui de plusieurs étudiants : Alphousseyni N'Donky, Marion Borderon, Agnès Couvray.

¹⁴³ Le programme DemoMed est présenté rapidement dans le Volume 2. Il réunit des démographes et géographes de plusieurs pays.

¹⁴⁴ On rejoindra les préoccupations de recherche de l'UMS RIATE (<http://www.ums-riate.fr>).

Par ailleurs, nous travaillons actuellement avec des démographes brésiliens (université de Campinas) qui souhaitent mettre en place une recherche sur les liens entre pauvreté et environnement au Brésil. Ce projet me permettra d'aborder un élément laissé de côté jusqu'ici mais qui met en relation les dimensions démographiques avec l'environnement dans un cadre spatial qui servira d'interface.

Le deuxième axe sera la vulgarisation de ces méthodes, dont nous avons déjà dit qu'elle nous semblait nécessaire. A travers l'enseignement d'une part, mais le public touché est restreint. A travers des publications d'autre part. Des publications qui illustrent par l'exemple, bien sûr, mais la prochaine étape sera plus sûrement la publication d'un manuel, qui mêlerait théorie (mais légèrement, les bibliothèques en sont déjà pourvues) et surtout cas pratiques permettant d'illustrer le propos, et "dédramatisant" ces approches.

Le troisième axe est moins clairement établi. Il est constitué de nouvelles dimensions à explorer. La mode est aux flux et aux réseaux. Il faut reconnaître que cette perspective est plaisante, puisqu'elle met l'accent sur les interactions. Du point de vue de l'autocorrélation spatiale et de sa mesure, l'application aux réseaux n'est pas encore très développée (Voir Peeters & Thomas, 2009). Une première approche simple consisterait à appliquer aux réseaux les modes de définitions de voisinage définis précédemment. Les nœuds sont comme des points et les liens qui les unissent permettent de définir des distances ou des topologies, comme on le souhaite. En se plaçant dans le cadre de la théorie des graphes, la description du graphe à l'aide d'une matrice permet d'utiliser les statistiques gamma. Néanmoins, on peut rapidement envisager des situations un peu plus compliquées où les liens sont valués, ce qui complique un peu la définition de la matrice (Peeters & Thomas, 2009 reviennent de façon plus détaillée sur ces questions). Une question nous intéresse notamment, celle de l'effet de la forme du réseau par rapport à l'effet du nombre de voisins dans les mesures¹⁴⁵. A priori, il n'existe cependant pas de raison de ne pas appliquer ces indices aux réseaux. Si cette ouverture sur les réseaux doit se concrétiser, il se fera certainement avec C.Rozenblat, en reprenant la suite de son article de 2010. Les données sont déjà explorées, bien maîtrisées, et les indices pourraient être comparés.

¹⁴⁵ Cette question est liée à la communication présentée par Arnaud Banos à Besançon en 2009 (réunion du GDRE S4 autour du groupe de travail MODUS), et qui s'est traduit par une publication (Banos, 2012).

Quoi qu'il en soit, une partie de ces perspectives devront s'envisager dans le cadre d'un rapprochement avec des statisticiens¹⁴⁶ pour pouvoir aller plus loin. Il me semble en effet inutile de réinventer la roue alors qu'ils l'ont bien tournée depuis longtemps... De plus, les statisticiens sont demandeurs de base de données "réelles" et de cas de figure concrets. Ils sont donc prêts à répondre à nos problèmes ! Enfin, l'exemple montre que ces alliances sont souvent porteuses de fruits : Arthur Cliff & Keith Ord ; Arthur Getis & Keith Ord pour prendre les plus célèbres. Dominique Peeters & Isabelle Thomas pour la francophonie ; Pierre Frankhauser ou Lena Sanders (ces deux derniers ayant une double formation en statistiques et en géographie).

Du groupe aux individus ?

Les structures décrites proviennent de l'interaction d'individus, dont les comportements, une fois agrégés montrent des régularités. Cette approche agrégée est nécessaire pour pouvoir saisir les mouvements de fond que connaissent nos sociétés. On ne peut en effet pas se placer au niveau des individus pour comprendre la plupart des phénomènes démographiques. La fécondité n'a de sens qu'à l'échelle d'une population, ainsi en est-il de la mortalité, etc. Pourtant naître, avoir un enfant, mourir sont autant d'actions issues de comportements individuels...qui ne devraient donc pas tout à la liberté humaine ? Si l'on est capable de montrer des tendances générales, c'est bien qu'un certain nombre de personnes se comportent de manière similaire. C'est d'ailleurs sur cette hypothèse que nous avons construit notre travail.

Diffusion spatiale

On peut envisager deux scénarios possibles, qui ne s'opposent pas et travaillent même certainement de conserve. Le premier propose que les individus qui ont les mêmes comportements tendent à se regrouper. On retrouve donc au même endroit des personnes qui ont les mêmes comportements. Cela suppose tout de même deux hypothèses un peu fortes. La première que les individus soient libres de se déplacer comme ils le souhaitent, et l'on sait que ce n'est pas le cas. La seconde que les individus soient totalement informés des caractéristiques des individus qu'ils vont rejoindre, ce qui est aussi difficilement imaginable.

¹⁴⁶ Comme nous l'évoquons dans le volume 2, nous avons commencé ce rapprochement auprès de certains membres du LATP (Laboratoire d'Analyse, Topologie, Probabilités UMR 6632).

Enfin, cela aurait pour conséquence que les individus se ressembleraient pour tout, et que l'on aurait donc pour tous les éléments comparés la même distribution spatiale, or il n'en est rien. On doit donc envisager plutôt un second scénario qui voudrait que les comportements se transmettent entre voisins. C'est évidemment l'idée de la diffusion spatiale des innovations, le changement de normes sociales étant considéré comme une innovation¹⁴⁷. Les formes spatiales des agrégats observés ne seraient donc pas le simple résultat du mouvement des individus, mais aussi celui des idées.

Pour expliquer les formes observées, nous pourrions donc essayer de descendre à l'échelle des personnes eux-mêmes et tenter de modéliser leurs comportements en intégrant non pas uniquement leurs caractéristiques individuelles, mais aussi leurs interactions avec les autres personnes, socialement ou spatialement proches. Repartir d'une vision centrée sur l'individu pour retrouver les comportements de groupe¹⁴⁸. Autrement dit, faire émerger à un niveau supérieur des phénomènes construits à un niveau inférieur. Cette approche dite individus-centrée a déjà quelques années (Bousquet, 1999, Bonnefoy, 2005).

Simulation et approche individus-centrée

La difficulté réside dans la mise en œuvre pratique de ces réflexions. Les données individuelles sont plus rares que les données collectives, et plus difficiles à utiliser. Par ailleurs, on doit se placer sur des temporalités longues et suivre des individus qui ont plusieurs fâcheuses tendances : ils changent, se déplacent, etc. Par ailleurs, plus que de les suivre, on aimerait savoir ce qui se passerait si tel ou tel élément venait à changer... Cela seul nous permettrait de démontrer les causalités et de dépasser la description des corrélations. Nous atteignons là une limite des sciences sociales : il n'est pas possible de mener des expériences sur les sociétés humaines. Pas tant pour des questions éthiques que pour des raisons pratiques : le sujet est transformé par l'expérience. Or, chaque sujet étant unique, l'expérience ne peut donc être reproduite. Elle n'acquiert donc aucune valeur de démonstration.

Ce problème est d'ailleurs rencontré dans d'autres sciences où les expérimentations sont possibles mais trop chères. C'est pour ces raisons (difficulté et coût) que les chercheurs ont

¹⁴⁷ Nous avons construit notre réflexion sur la diffusion spatiale des innovations à partir de 4 sources principales : Hägerstrand (1967) Rogers (1962) et Saint Julien (1985) bien sûr, mais aussi les travaux de Daudé (2003, Langlois, Daudé, 2007).

¹⁴⁸ Nous devons beaucoup ici à Jean-Luc Bonnefoy, plus encore par nos interactions que par sa production. Nous renvoyons néanmoins le lecteur à sa production, particulièrement Bonnefoy 2001, 2003, 2005).

développé des stratégies de simulation. La dernière de ces stratégies, le développement des systèmes multi-agents (Ferber, 1999, Wooldridge , 2002) offre des perspectives riches.

Grâce aux systèmes multi-agents (SMA) il devient possible et relativement aisé de construire des modèles simulant de multiples dimensions de l'activité humaine et notamment d'intégrer les notions de choix, contraintes, interaction, le tout dans un environnement qui peut être spatial. On peut partir d'environnements virtuels pour tenter de reconstituer les dynamiques en action. L'important n'est plus alors de reconstituer des modèles réalistes, mais plutôt d'essayer de comprendre comment évoluent les phénomènes en fonction des hypothèses posées. Ce sont alors les processus, autant que les résultats, qui sont mis au premier plan. Nous nous sommes récemment prêtés à l'exercice avec Yoann Doignon (Doignon, Oliveau, 2011), pour voir comment l'espace constituait un facteur essentiel de la diffusion. Les résultats nous montrent que l'on arrive à reproduire à partir de comportements individuels les dynamiques observées à un niveau agrégé (dans ce cas la diffusion de la baisse de la fécondité et ses différentes conséquences).

Le plus intéressant dans cette approche est la dimension stochastique possible de ces simulations. Non déterministe, ces modèles permettent d'intégrer le hasard dans le déroulement des processus. C'est évidemment une grande force, même si cela pose des problèmes notamment en ce qui concerne l'évaluation des résultats. Néanmoins, on peut envisager de multiplier les expériences suffisamment pour produire des statistiques de probabilités et rendre ces simulations utiles pour la mise en place de scénarios.

Faut-il conclure ?

Nous sommes conscients que les voies proposées ne seront peut être pas toutes explorées, par manque de temps notamment. Elles constituent plutôt des guides pour saisir des opportunités et entraîner des étudiants sur ces voies. Cela ne constitue donc pas une conclusion, mais bien une ouverture. Mais c'est bien là le but de l'habilitation à diriger des recherches : non pas tant conclure (sur) son travail que d'ouvrir de nouvelles perspectives.

BIBLIOGRAPHIE

Dans la mesure du possible, nous avons indiqué où trouver les références sur Internet. Nous utilisons en priorité le DOI des fichiers (Digital Object Identifier), identifiant unique des documents en ligne (que l'on peut comparer aux ISSN) et qui permettent de retrouver les documents même si leur localisation change (à partir du site web <http://dx.doi.org>). Pour les articles du portail Jstor, nous renvoyons à l'adresse permanente (*stable URL*). Lorsque les revues n'éditent pas de DOI, comme c'est le cas pour le portail Revues.org, nous renvoyons à l'adresse de l'article dans la revue. Enfin, si les documents ne sont pas disponibles selon un de ces 3 critères, nous renvoyons à l'adresse où nous l'avons trouvé, le cas échéant, conscient que leur localisation peut changer avec le temps.

- AMRHEIN, C.G., REYNOLDS, H., (1996), « Using spatial statistics to assess aggregation effects », *Geographical systems*, Vol. 3, pp. 143-158.
- ANSELIN, L., (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, 284 p.
- ANSELIN, L., (1993b), « Local indicators of spatial association - LISA », *research paper 9331*, Regional Research Institute, West Virginia University, 26 p.
<http://rri.wvu.edu/pdffiles/wp9330.pdf>
- ANSELIN, L., (1995), « Local indicators of spatial association - LISA », *Geographical Analysis*, Vol. 27, n°2, pp. 93-115.
doi : 10.1111/j.1538-4632.1995.tb00338.x
- ANSELIN, L., (1996), « The Moran scatterplot as an ESDA tool to assess local instability in spatial association », in FISCHER, Manfred, SCHOLTEN, Henk K., UNWIN, David, *Spatial analytical perspectives on GIS*, Taylor & Francis, London, pp. 111-125.
- ANSELIN, L., (2003), *GeoDa 0.9 User's Guide*, Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics and CMISS, University of Illinois, 125 p.
- ANSELIN, L., BERA, A.K., FLORAX, R., YOON, M.J., (1996), « Simple diagnostic tests for spatial dependence », *Regional Science and Urban Economics*, Vol. 26, n°1, pp. 71-104.
doi : 10.1016/0166-0462(95)02111-6
- ANSELIN, L., SYABRI, I., KHO, Y., (2006), « *GeoDa: An Introduction to Spatial Data Analysis* », *Geographical Analysis*, n°38, pp. 5-22.
doi : 10.1111/j.0016-7363.2005.00671.x
- ANSELIN, L., SYABRI, I., SMIRNOV, O., (2002), « Visualizing multivariate spatial correlation with dynamically linked windows », in Anselin, L., Rey, S., (ed.), *New Tools for Spatial Data Analysis: Proceedings of the Specialist Meeting*, Center for Spatially Integrated Social Science (CMISS), University of California, Santa Barbara, CD-ROM.
- ANSELIN, L., (1992), « Spacestat : A Program for the Analysis of Spatial Data », *Technical Software Series S-92-1*, National Center for Geographic Information and Analysis, University of California, Santa Barbara, Californie.

- ANSELIN, L., (1993a), « The Moran scatterplot as an ESDA tool to assess local instability in spatial association », Communication au GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, 1-5 Décembre 1993; *research paper 9330*, Regional Research Institute, West Virginia University, 23 p.
<http://rri.wvu.edu/pdffiles/wp9330.pdf>
- ANSELIN, L., (1994), « Spacestat : Version 1.50: revision notes », *research paper 9428*, Regional Research Institute, West Virginia University, 30 p.
<http://rri.wvu.edu/pdffiles/wp9428.pdf>
- ARLINGHAUS, S.L., (Ed.), (1996), *Practical Handbook of Spatial Statistics*, CRC Press, New York, 307 p.
- BADARIOTTI, D., (2004), « De la comparaison des cartes électorales », *Cybergeo*, article n°267.
<http://cybergeo.revues.org/3670>
- BAILEY, T.C., GATRELL, A.C., (1995), *Interactive spatial data analysis*, Longman, Essex, 413 p.
- BAILLY, A., FERRAS, R., PUMAIN, D., (1992), *Encyclopédie de géographie*, Economica, Paris, 1132 p.
- BAIROCH, P., (1985, 2^{ème} éd. corrigée), *De Jéricho à Mexico. Villes et économie dans l'histoire*, Arcades, Gallimard, Paris, 706 p.
- BALLER, R.D., ANSELIN, L., MESSNER, S.F., DEANE, G., HAWKINS, D.F., (2001), « Structural covariates of U.S. county homicide rates. Incorporating spatial effects », *Criminology*, Vol. 39, pp. 561-588.
doi : 10.1111/j.1745-9125.2001.tb00933.x
- BANOS, A., (2001) « A propos de l'analyse spatiale exploratoire des données », *Cybergeo*, Systèmes, Modélisation, Géostatistiques, article 197, mis en ligne le 18 octobre 2001.
<http://www.cybergeo.eu/index4056.html>
- BANOS, A., (2012, accepté), « Network effects in Schelling's model of segregation: new evidences from agent-based simulation », *Environment and Planning B*.
- BARNES, T.J., (2004), "A Paper Related to Everything but More Related to Local Things", *Annals of the Association of American Geographers*, Vol. 94, No. 2, pp.278-283.
- BATTY, M., LONGLEY, P., (1994), *Fractal cities: a geometry of form and function*, Academic Press, San Diego (Californie), 314 p.
<http://www.fractalcities.org/>
- BEDI, T., COUDOUÉL, A., SIMLER, K., (2007), *More than a pretty picture : using poverty maps to design better policies and interventions*, Banque Mondiale, 286 p.
<http://go.worldbank.org/P6S3FQPOU0>
- BERRY, B.J.L., (1993), « Geography's quantitative revolution initial conditions, 1954-60: A personal memoir », *Urban Geography*, Vol. 14, pp. 434-41.
- BIVAND, R., (2009), « Applying Measures of Spatial Autocorrelation: Computation and Simulation », *Geographical analysis*, pp. 375-384.
- BOCQUET-APPEL, J.P., COURGEAU, D., PUMAIN, D., (Ed.), (1996), *Spatial Analysis of Biodemographic Data*, INED/John Libbey Eurotext, Montrouge, 367 p.

- BONNEFOY J.-L., BOUSQUET, F., ROUCHIER, J., (2001), « Modélisation d'une interaction individus, espace et société par les systèmes multi-agents : pâture en forêt virtuelle », *L'Espace géographique*, Vol. 30, n°1, pp. 13-25.
- BONNEFOY, J.-L., (2003), « From households to urban structures: space representations as engine of dynamics in multi-agent simulations », *Cybergeo : European Journal of Geography*.
<http://cybergeo.revues.org/1627>.
- BONNEFOY, J.L., (2005), *Etude de géographie théorique et expérimentale*, Habilitation à Diriger des Recherches de l'Université de Provence, non publié.
- BOOTS, B., TIEFELSDORF, M., (2000), « Global and local spatial autocorrelation in bounded regular tessellations », *Geographical systems*, Vol. 2, pp. 319-38.
- BORDERON, M., OLIVEAU, S., LALOU, R., (2011), « Modelling The Malaria infection in Dakar, Senegal, From urban ecology to social geography », 17th European Colloquium on Quantitative and Theoretical Geography, 2-5 septembre 2011, Athènes.
- BOUSQUET, F., GAUTIER, D., (1999) « Comparaison de deux approches de modélisation des dynamiques spatiales par simulation multi-agents : les approches "spatiale" et "acteurs" », *Cybergeo*.
<http://cybergeo.revues.org/2389>
- BRANSCOMB, L.M., (1986), « Science in 2006 », *American Scientist*, November-December.
<http://www.americanscientist.org/issues/feature/science-in-2006/1>
- BRUNET, R., (1967), *Les Phénomènes de discontinuité en géographie*, Edition du CNRS, coll. Mémoires et documents du centre de recherches et documentation cartographiques et géographiques, vol. 7, 117 p.
- BRUNET, R., DOLLFUS, O., (Ed.), (1990), *Mondes nouveaux*, Géographie Universelle, Belin / GIP RECLUS, Paris - Montpellier, 551 p.
- BRUNET, R., FERRAS, R., THERY, H., (Ed.), (1992, 1997 3^{ème} éd.), *Les mots de la géographie, dictionnaire critique*, GIP RECLUS / La Documentation française, Montpellier / Paris, 518 p.
- BRUNET, R., FRANÇOIS, J.-C., GRASLAND, C., (1997), « la discontinuité en géographie: origines et problèmes de recherche », *L'Espace Géographique*, n°4, pp. 297-308
- BUNGE, W., (1962), *Theoretical geography. Lund studies in Geography*, Série C: Mathematical and general geography, 210 p.
- BURROUGH, P.A., MC DONNELL, R., (1998), *Principles of Geographical Information Systems*, Oxford University Press, Oxford, 333 p.
- CASANOVA, L., (2010), *Les dynamiques du foncier à bâtir comme marqueurs du devenir des territoires de Provence intérieure, littorale et préalpine. Éléments de prospective spatiale pour l'action territoriale*, thèse de doctorat en géographie, soutenue à l'université d'Avignon, 441 p.
<http://tel.archives-ouvertes.fr/tel-00555793/fr/>
- CHARRE, J., (1995), *Statistique et territoire*, Montpellier, coll. Espaces modes d'emploi, GIP RECLUS, Montpellier, 120 p.
- CHARREIRE, H., (2007), *Santé périnatale et territoire urbain : analyse géographique des inégalités sociales de santé en Seine-Saint-Denis*, thèse de doctorat en géographie, soutenue à l'université Paris Est-Créteil, Val de Marne,

- CHAUDHURI, S., JALAN, J., SURYAHADI, A., (2001), *Assessing Household Vulnerability to Poverty: a Methodology and Estimates for Indonesia*, Discussion Paper n°0102.52, Columbia University, 35 p.
<http://www.columbia.edu/cu/economics/discaprp/DP0102-42.pdf>
- CHRISMAN, N., (2006), *Charting the unknown; how computer mapping at Harvard became GIS*, ESRI Press, DVD, 218 p.
- CHRISTALLER, W., (1933), *Die zentralen Orte in Süddeutschland*, Gustav Fischer, Iena, (traduction partielle par BASKIN, Charlisle W., (1966) *Central Places in Southern Germany*, Prentice Hall, Englewood Cliffs).
- CLARK, I., (1979), *Practical geostatistics*, Applied Science Publishers, Barking, Essex, 129 p.
- CLIFF, A.D., (1969), *Some measures of spatial association in areal data*, Thèse de doctorat de l'université de Bristol, non publiée.
- CLIFF, A.D., ORD, K.J., (1969), « The Problem of Spatial Autocorrelation », in Scott, A.J., (ed.), *Papers in Regional Science, Volume 1, Studies in Regional Science*, Pion, Londres, pp. 25-55.
- CLIFF, A.D., ORD, K.J., (1970), « Spatial Autocorrelation: A review of existing and new measures with applications », *Economic Geography*, pp. 269-292.
- CLIFF, A.D., ORD, K.J., (1973), *Spatial Autocorrelation*, Pion, Londres, 266 p..
- CLIFF, A.D., ORD, K.J., (1981), *Spatial processes. Models and applications*, Pion, Londres, 266 p.
- CLIFF, A.D., ORD, K.J., (2009), « What were we thinking? », *Geographical analysis*, pp. 351-363.
- COLE, J.P., KING, C.A.M., (1968), *Quantitative Geography*, Glasgow University Press, Glasgow.
- COUVRAY, A., OLIVEAU, S., LALOU, R., (2011), « Quelle relation entre risque sanitaire et pauvreté ? Paludisme et vulnérabilité économique à Dakar », communication au 10ème colloque Théoquant, 23-25 mars 2011, Besançon.
- COX, K., (2006) « Une lecture anglophone et marxiste », *EspacesTemps.net*, Il paraît.
<http://espacestemp.net/document2115.html>
- CRESSIE, N., (1991), *Statistics for Spatial Data*, Wiley, New York, 928 p.
- DACEY, M.F., (1965), « The Geometry of Central Place Theory », *Geografiska Annaler. Series B, Human Geography*, Vol. 47, No. 2, pp. 111-124.
- DACEY, M.F., (1969) « A Review on Measures of Contiguity for Two and K-Color Maps », in BERRY, B.J.L., MARBLE, D.F., *Spatial Analysis : A Reader in Statistical Geography*, Englewood Cliffs, Prentice-Hall, pp. 479-490.
- DAIX, N., (2010), *Internal rural migration and marginality. The case of Agusan del Sur, Philippines*, thèse de doctorat à la faculté de sciences de Namur (FNUDP), 301 p.
- DATT, G., RAVALLION, M., (1993), « Regional Disparities, Targeting, and Poverty in India », in LIPTON, M., VAN DER GAAG, J., (eds.), *Including the poor*, Banque Mondiale, pp. 91-114.

- DAUDE, E., (2002), *Modélisation de la diffusion d'innovation par la simulation multi-agents. L'exemple d'une innovation en milieu rural*, thèse, université d'Avignon et des pays du Vaucluse, 328 p.
<http://tel.archives-ouvertes.fr/tel-00440100/fr/>
- DAUPHINE, A., VOIRON-CANONICIO, C., (1988), *Variogrammes et structures spatiales*, coll. Reclus Modes d'emploi, n°12, GIP RECLUS, Montpellier, 56 p.
- DECROLY, J.-M., GRASLAND, C., (1996), « organisation spatiale et organisation territoriale des comportements démographiques : une approche subjective », in Bocquet-APPEL, J.-P., COURGEAU, D., PUMAIN, D., (Ed.), *Spatial Analysis of Biodemographic Data*, INED/John Libbey Eurotext, Montrouge, pp. 131-156.
- DEICHMANN, Uwe, (1999), *Geographic aspects of inequality and poverty*, texte pour le site Internet de la banque mondiale, non publié, 13 p.
http://siteresources.worldbank.org/INTPGI/Resources/Pro-Poor-Growth/5319_povmap.pdf
- DE RUFFRAY, S., (2007), *L'imprécision et l'incertitude en géographie. L'apport de la logique floue aux problématiques de régionalisation*, Mémoire d'habilitation à diriger des recherches, Université de Paris VII, Volumes 1, 2 et 3, 588 pages.
- DE RUFFRAY, S., HAMEZ, G., GRASLAND, C., LAMBERT, N., HAMM, A., GALLET-MORON, E., (2011), *Enjeux des territoires frontaliers à l'échelle nationale Vers la détermination et la délimitation de pôles transfrontaliers*, rapport pour la DATAR, 34 p.
http://www.ums-riate.fr/documents/rapportDATAR_mai2011.pdf
- DERYCKE, P.-H., HURIOT, J.-M., PUMAIN, D., (coord.), (1996), *Penser la ville, Théories et modèles*, Anthropos, coll. Villes, Paris, 335 p.
- DOIGNON, Y., OLIVEAU, S., (2011), « Comprendre la baisse de la fécondité en Inde : apport de la modélisation individu-centrée », communication au 10^{ème} colloque Théoquant, 23-25 mars 2011, Besançon.
- DRAY, S., (2011), « A New Perspective about Moran's Coefficient: Spatial Autocorrelation as a Linear Regression Problem », *Geographical Analysis*, Vol. 43, pp. 127-141.
- DUBIN, R., (2009), « Spatial Weights », in FOTHERINGHAM, A.S., ROGERSON, P., *The SAGE handbook of spatial analysis*, Sage, pp. 124-158.
- DUMOLARD, P., (1999), « Accessibilité et diffusion spatiale », *L'Espace Géographique*, n°3, pp. 205-214.
- EGENHOFER MJ, MARK DM, HERRING JR (eds), (1994), *The 9-intersection: Formalism and its Use for Natural-language Spatial Predicates*. National Center for Geographic Information and Analysis Technical Report 94-1, 71 p.
http://www.ncgia.ucsb.edu/Publications/Tech_Reports/94/94-1.PDF
- ELISSALDE, B., (coord.), (non daté), *Hypergeo*.
<http://www.hypergeo.eu>
- ELLIOTT, M., (ed.), (1976), *Auto-carto I*, proceedings of the International Conference on Automation in Cartography, 09-12 décembre 1974, Reston (Virginie), American Congress on Surveying and Mapping, 318 p.
<http://mapcontext.com/autocarto/proceedings/auto-carto-1/pdf/auto-carto-1.pdf>
- FELKNER, J.S., TOWNSEND, R.M., (à paraître), « The Geographic Concentration of Enterprise in Developing Countries », *Quarterly Journal of Economics*, non paginé.
- FERBER J. (1995), *Les systèmes multi-agents. Vers une intelligence collective*, InterEditions, Paris, 522 p.

- FINGLETON, B., (2009), « Spatial Autoregression », *Geographical analysis*, pp. 385-391.
- FORTIN, M.-J., DALE, M.R.T., (2009), « Spatial Autocorrelation in Ecological Studies: A Legacy of Solutions and Myths », *Geographical analysis*, pp. 392-397.
- FOTHERINGHAM, A.S., ROGERSON, P. (2009), *The SAGE handbook of spatial analysis*, Sage, 513 p.
- FOTHERINGHAM, S.A., (2009), « "The Problem of Spatial Autocorrelation" and Local Spatial Statistics », *Geographical analysis*, Vol. 41, n°4, pp. 398-403.
doi : 10.1111/j.1538-4632.2009.00767.x
- FOTHERINGHAM, S.A., BRUNDSON, C., CHARLTON, M., (1996), « Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity », *Geographical analysis*, Vol. 28, n°4, pp.281-298.
doi : 10.1111/j.1538-4632.1996.tb00936.x
- FOTHERINGHAM, S.A., BRUNDSON, C., CHARLTON, M., (2000), *Quantitative geography : perspectives on spatial data analysis*, Sage Publications, Londres, 270 p.
- FOUCHER, M. (1986), *L'invention des frontières*, Paris, Fondation pour les Études de Défense Nationale, 326 p.
- FRANÇOIS, J.-C., (1998), « Discontinuités territoriales et mise en évidence de systèmes spatiaux dans l'espace des collèges de l'agglomération parisienne », *L'Espace géographique*, n°1, pp. 63-75.
- FRANKHAUSER, P., (1998), « The fractal approach. A new tool for the spatial analysis of urban agglomerations », *Population : an English Selection*, Vol. 10, n°1, Special issue *New methodological Approaches in the Social Sciences*, vol. 94, n° 2, pp. 205-240.
<http://www.jstor.org/stable/2998685>
- GAY, J.-C., (1995), *Les discontinuités spatiales*, coll. Poche géographie, Economica, Paris, 112 p.
- GEARY, R.C., (1954), « The contiguity ratio and statistical mapping ». *The incorporated Statistician*, vol.5, n°3, pp.115–145.
- GEHLKE, C. E., BIEHL, K., (1934), « Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material », *Journal of the American Statistical Association*, Vol. 29, No. 185, Supplement: Proceedings of the American Statistical Journal (Mars), pp. 169-170
<http://www.jstor.org/stable/2277827>
- GETIS, A., (1991), « Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach », *Environment and Planning A*, Vol. 23, 1269-77.
- GETIS, A., (2009), « Spatial Weights Matrices », *Geographical analysis*, Vol. 41, n°4, pp. 404-410.
doi : 10.1111/j.1538-4632.2009.00768.x
- GETIS, A., GRIFFITH, D., (2002), « Comparative Spatial Filtering in Regression Analysis », *Geographical analysis*, Vol. 34, n°2, pp. 130-140.
doi : 10.1111/j.1538-4632.2002.tb01080.x
- GETIS, A., ORD, J.K., (1992), « The analysis of spatial association by use of distance statistics », *Geographical analysis*, Vol. 24, No. 3, pp. 189-206.

- GETIS, A., ORD, J.K., (1996), « Local spatial statistics: an overview », in LONGLEY, P., BATTY, M., (éd.), *Spatial Analysis : Modelling in a GIS environment*, Geoinformation International, Cambridge, pp. 261-277.
- GIP RECLUS, (1985), *Pour la géographie universelle. Charte de la rédaction*, coll. Reclus modes d'emploi, GIP RECLUS, Montpellier, 56 p.
- GOODCHILD, M.F., (2004), « The Validity and Usefulness of Laws in Geographic Information Science and Geography », *Annals of the Association of American Geographers*, vol. 94, n° 2, pp. 300-303.
<http://www.jstor.org/stable/3693988>
- GOODCHILD, M.F., (2008) « Theoretical Geography (1962): William Bunge », in HUBBARD, P., KITCHIN, R., VALENTINE, G., (eds.), *Key Texts in Human Geography*, SAGE, Los Angeles, pp. 9-16.
- GOODCHILD, M.F., (2009), « What Problem? Spatial Autocorrelation and Geographic Information Science », *Geographical analysis*, Vol. 41, n°4, pp. 411-417.
doi : 10.1111/j.1538-4632.2009.00769.x
- GOODCHILD, M.F., JANELLE, D.G., (2004), *Spatially Integrated Social Science*, coll. Spatial Information Systems, Oxford University Press, New York, 456 p.
- GOOVAERTS, P., (1997), *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 483 p.
- GRASLAND C., (1997), « L'analyse géographique des discontinuités territoriales : l'exemple de la structure par âge des régions européennes vers 1980 », *L'Espace Géographique*, n°4, pp. 309-326.
- GRASLAND C., (2009), « Spatial analysis of social facts. A tentative theoretical framework derived froms Tobler's first law of geography and Blau multilevel structural theory of society », in BAVAUD, F., MAGER, C., *Handbook of theorical and quantiativ geography*, FGSE, University of Lausanne, Lausanne.
- GRASLAND, C., (2004), « Discontinuités statistiques et discontinuités spatiales - L'exemple des inégalités de richesse par habitant en Italie (1951-1991) », communication aux Journées d'Etude Statistiques de Marseille-Luminy, 38 p.
- GRASLAND, C., MADELIN, M., (2006), *ESPON Project ; 3.4.3*, « The modifiable areal unit problem, final report »
http://www.espon.eu/mmp/online/website/content/projects/261/431/index_EN.html
- GRASLAND, C., MADELIN, M., MATHIAN, H., SANDERS, L., (2007), « La question du MAUP dans une problématique d'aménagement européen : problème ou progrès », Communication au 8^{ème} colloque ThéoQuant, 10-12 janvier 2007.
- GRATALOUP, C., (1996), *Lieux d'Histoire, essai de géohistoire systématique*, coll. Espaces modes d'emploi, GIP RECLUS, Montpellier, 200 p.
- GRIFFITH, D.A. (1996), « Computational simplifications for space-time forecasting within GIS: the neighbourhood spatial forecasting model », in LONGLEY, P., BATTY, M., (éd.), *Spatial Analysis : Modelling in a GIS environment*, Geoinformation International, Cambridge, pp. 247-260.
- GRIFFITH, D.A. (2006), « Hidden Negative Spatial Autocorrelation », *Journal of Geographical Systems*, Vol. 8, n°3, pp. 335-355.
doi : 10.1007/s10109-006-0034-9

- GRIFFITH, D.A., (1987), *Spatial autocorrelation*, resource publications in Geography, Association of American Geographer, Washington, 86 p.
- GRIFFITH, D.A., (1992), « Spatial regression analysis on the PC: spatial statistics using SAS », department of geography and Interdisciplinary statistics Program, Syracuse University.
- GUILMOTO, C.Z. (coord.), (2005), *Espace et mesure en Inde du Sud (EMIS). L'espace de la vulnérabilité en Inde du Sud*, Programme national de recherche «Société de l'information » du CNRS, rapport scientifique, 78 p.
<http://www.demographie.net/emis>
- GUILMOTO, C.Z., (1989), Un siècle de démographie tamoule. Evaluation des sources et analyse de l'évolution de la population du Tamil Nadu de 1871 à 1981, Thèse de doctorat en démographie, Université Paris I, 502 p, microfichée.
- GUILMOTO, C.Z., (2000), « Mapping Girl Child Neglect in Rural Tamil Nadu (1991 Census Data) », in Sabu M., George, P. Phavalam, (ed.), *Female Feticide in Tamilnadu*, Report of the State Level Consultation at Chennai, Society for Integrated Rural development, Madurai, pp. 38-43.
- GUILMOTO, C.Z., OLIVEAU, S., VINGADASSAMY, S. (2002), « Un système d'information géographique en Inde du Sud : Théorie, mise en œuvre et applications thématiques », *Espace, Populations et sociétés*, Lille, pp. 147-163
- HÄGERSTRAND, T., (1967), *Innovation diffusion as a spatial process*, University of Chicago press, Chicago, 334 p., traduit du suédois par Allan Pred. Edition originale: (1953), *innovations förloppet ur korologisk synpunkt*, C.W.K. Gleerup, Lund (Suède).
- HAGGETT, P., (1973), *L'analyse spatiale en géographie*, Armand Colin, Paris, 390 p., (1^{ère} édition : 1965, *Location Analysis in human geography*, Arnold, Londres).
- HAGGETT, P., (2008), « The Local Shape of Revolution: Reflections on Quantitative Geography at Cambridge in the 1950s and 1960s », *Geographical analysis*, pp. 336-352.
- HAINING, R.P., (1990), *Spatial data analysis in the social and environmental sciences*, Cambridge University Press, Cambridge, 409 p.
- HAINING, R.P., (2003), *Spatial data analysis : Theory and practice*, Cambridge University Press, Cambridge, 452 p.
- HAINING, R.P., (2009), « Spatial Autocorrelation and the Quantitative Revolution », *Geographical analysis*, Vol. 41, n°4, pp. 364-374.
doi : 10.1111/j.1538-4632.2009.00763.x
- HARVEY, D., (1996), *Justice, nature and the geography of difference*, Wiley-Blackwell, 468 p.
- HELLE, C., (1993), « Essai de mesure de la rugosité de l'espace : application à l'espace vauclusien », *L'Espace Géographique*, n°4, pp. 346-352.
- HOULDING, Simon W., (2000), *Practical geostatistics: modeling and spatial analysis*, Springer, 160 p.
- HU, Z., RAO, K.R., (2009), « Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data », *Environmental Health*, Vol. 8, n°26, non paginé.
doi:10.1186/1476-069X-8-26
- HUBERT, L.J., GOLLEDGE, R.G., CONSTANZO, C.M., (1981), « Generalized procedures for evaluating spatial autocorrelation », *Geographical Analysis*, n°13, pp.224-233.

- HURIOT, J.-M., PERREUR, J., (1990), « Distances, espaces et représentations », *RERU*, n°2, pp. 197-237.
- ISAAKS, E.H., SRIVASTAVA, R.M., (1990) *An introduction to applied geostatistics*, Oxford University Press, 592 p.
- JOHNSTON, R., FAIRBROTHER, M., HAYES, D., HOARE, T., JONES, K., (2008), « The Cold War and geography's quantitative revolution: Some messy reflections on Barnes' geographical underworld », *Geoforum*, Vol. 39, n°6, pp. 1802-1806
doi :10.1016/j.geoforum.2007.09.006
- KENDALL, M.G., (1939), « The Geographical Distribution of Crop Productivity in England », *Journal of the Royal Statistical Society A*, Vol. 102, pp. 21-48.
Stable URL: <http://www.jstor.org/stable/2980138>
- KRISHNA IYER, P.V.A., (1949), « The First and Second Moments of Some Probability Distributions Arising from Points on a Lattice, and their Applications », *Biometrika*, n°36, pp. 135-41.
- LALL, Somik V., DEICHMANN, Uwe, (2009), *Density and Disasters. Economics of Urban Hazard Risk*, World Bank Policy research working paper, n°5161, 48 p.
http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2009/12/29/000158349_20091229205549/Rendered/PDF/WPS5161.pdf
- LANGLOIS, P., DAUDE, E., (2007), « Concepts et modélisation de la diffusion », *Cybergegeo*, n°. 364, 23p.
<http://www.cybergegeo.eu/index2898.html>
- LE GALLO, J., (2002) « Econométrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire », *Economie et Prévision*, vol. 155, n° 4, pp. 139-158
- LEGENDRE, P., (1993), « Spatial autocorrelation: trouble or new paradigm? », *Ecology*, vol. 74, n° 6, pp. 1659-1973.
- LESLIE, T.F., KRONENFELD, B.J., (2011), « The Colocation Quotient: A New Measure of Spatial Association Between Categorical Subsets of Points », *Geographical Analysis*, Vol. 43, pp. 306-326
- LONGLEY, P., BATTY, M., (éd.), (1996), *Spatial Analysis : Modelling in a GIS environment*, Geoinformation International, Cambridge, 392 p.
- LÖSCH, A., (1954), *The Economics of Location*, Yale University Press, New Haven, (traduction de *Die Räumliche Ordnung der Wirtschaft*, 1943).
- MABY, J., Dir., (2003), *Objets et indicateurs géographiques*, UMR ESPACE, 316 p.
- MANTEL, N., (1967), « The Detection of Disease Clustering and a Generalized Regression Approach », *Cancer Research*, 27, pp. 209-220.
- MARTIN, D., (1996), *Geographic Information Systems. Socioeconomic applications*, Routledge, Londres, 210 p.
- MARTIN, D., DURLING, D., MITCHELL, R., (2002), « Linking censuses through time : problems and solutions », *Area*, n°34.1, pp. 82-91.
- MATHERON, G., (1965), *Les variables régionalisées et leur estimation*, Masson, Paris, 305 p.
- MATHIAN, H., PIRON, M., (2001), « Echelles géographiques et méthodes statistiques multidimensionnelles », in SANDERS, L., *Modèles en analyse spatiale*, Hermès Sciences, Paris, pp. 61-104.

- MATISZIW, T.C., GRUBESIC, T.H., (2011), « Geographic perspectives on vulnerability analysis », *GeoJournal*, 2p.
doi : 10.1007/s10708-011-9420-z
- MILLER, H.J., (2004), « Tobler's first law and spatial analysis », *Annals of the Association of American Geographers*, vol. 94, n° 2, pp. 284-289.
<http://www.jstor.org/stable/3693985>
- MONMONIER, M., (1989), « Geographic Brushing enhancing exploratory analysis of the scatterplot matrix », *Geographical analysis*, Vol. 21, n°1, pp. 81-84.
- MORAN, P.A.P., (1948), « The interpretation of statistical maps », *Journal of the Royal Statistical Society, Series B*, Vol. 10, pp. 243-251.
- MORAN, P.A.P., (1950), « Notes on continuous stochastic phenomena », *Biometrika*, Vol. 37, n°1/2, pp. 17-23.
- MORENO, D., BADARIOTTI, D., BANOS A., (2009), « Graph based automata for urban modelling », in BAVAUD F., MAGER, C., (dir.), *Handbook of Theoretical and Quantitative Geography*, FGSE, University of Lausanne, Lausanne, Switzerland, p.261-309.
- MORRILL, R. L. (1984). « Recollections of the "Quantative Revolution's" Early Years: The University of Washington, 1955–65 », in BILLINGE, M., GREGORY,D., MARTIN, R., (ed.) *Recollections of a Revolution: Geography as Spatial Science*, Macmillan Press, Londres, pp. 57–72.
- MULLER, J.-C., (1977), « Comparaison visuelle des cartes », *L'espace Géographique*, t.6, n°1, pp. 59-72.
- OLIVEAU, S., (1999), Les aspects géographiques d'un phénomène démographique : la baisse de la fécondité au Tamil Nadu, mémoire de DEA de l'université de Paris X Nanterre, non publié, 91 p.
- OLIVEAU, S., (2004), *Modernisation villageoise et distance à la ville en Inde du Sud*, thèse de doctorat en géographie, soutenue à l'université Paris 1, 279 p.
<http://tel.archives-ouvertes.fr/tel-00007661>
- OLIVEAU, S., (2005a), « Mapping out fertility in South India : Methodology and results », in GUILMOTO, C.Z., RAJAN, S.I., (Ed.), *Fertility Transition in South India*, Sage, Delhi, pp. 90-113.
- OLIVEAU, S., (2005b), « Two ways to assess vulnerability at village level in South-Indian context », séminaire au CIESIN (Columbia University), 13 septembre 2005, Palisades, NY.
- OLIVEAU, S., (2006), « Les indicateurs locaux d'association spatiale (LISA) comme méthode de régionalisation : Une application en Inde », *actes du 7^{ème} colloque Théoquant*, 26-28 janvier 2005, Besançon, 10 p.
- OLIVEAU, S., (2010), « Autocorrélation spatiale : leçons du changement d'échelle », *L'Espace Géographique*, n°1, pp. 51-64.
- OLIVEAU, S., GUILMOTO, C.Z., (2005), « Spatial correlation and demography. Exploring India's demographic patterns. », communication à la XXVth IUSSP International Population Conference (Session N°1207: "Spatial demography including modelling"), 18-23 juillet 2005, Tours.

- OLIVEAU, S., VELLA, S., (2004), « Evolution spatio-temporelle du rapport de féminité en Inde : un marqueur pertinent de la discrimination sexuelle », communication au séminaire « Genre, territoire, développement : quels regards géographiques ? », 25-26 Mars 2004, Lyon.
- OPENSHAW, S., (1981), « Le problème de l'agrégation spatiale en géographie », *L'espace Géographique*, n°1, pp. 15-24.
- OPENSHAW, S.O., (1984), *The Modifiable Areal Unit Problem*, Concepts and techniques in Modern Geography 38, Norwich : Geo books, 41 p.
- OPENSHAW, S.O., TAYLOR, P.J., (1979), « A million or so correlation coefficients : Three experiments on the Modifiable Areal Unit Problem », in WRIGLEY, N., (Ed.), *Statistical applications in the spatial science*, Pion Limited, Londres, pp. 127-144.
- ORD, J.K., GETIS, A., (1995), « Local spatial autocorrelation statistics: distributional issues and an application », *Geographical analysis*, Vol. 27 (4), pp. 286-306.
doi : 10.1111/j.1538-4632.1995.tb00912.x
- PACE, K.R., LESAGE, J., ZHU, S., (2009), « Impact of Cliff and Ord on the Housing and Real Estate Literature », *Geographical analysis*, Vol. 41 (4), pp. 418-424.
doi : 10.1111/j.1538-4632.2009.00771.x
- PAELINCK, J.H.P., (2009), « Rook or Queen, But How About Some Other Royal Court Suite Members? », *Geographical analysis*, pp. 425-429.
- PANNATIER, Y., (1996), *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York, NY.
- PATIL, G.P., (2009), « Impacts and Wider Impacts on Statistics », *Geographical analysis*, pp. 430-435.
- PAULUS, F., (2004), *Coévolution dans les systèmes de villes : croissance et spécialisation des aires urbaines françaises de 1950 à 2000*, thèse de doctorat en géographie, soutenue à l'université Paris 1, 405 p.
<http://tel.archives-ouvertes.fr/tel-00008053/fr/>
- PEETERS, D., THOMAS, I., (2009), « Network autocorrelation », *Geographical analysis*, Vol. 41, n°4, pp. 436-443.
doi : 10.1111/j.1538-4632.2009.00773.x
- PEETERS, D., THOMAS, I., (1997), « Distance- l_p et localisations optimales. Simulations sur un semis aléatoire de points. », *Les cahiers scientifiques du transport*, n°31, pp. 55-70.
- PIOMBINI, A., (2006), *Modélisation des choix d'itinéraires pédestres en milieu urbain. Approche géographique et paysagère*, thèse de doctorat en géographie, soutenue à l'université de Franche-Comté, 299 p.
- PUMAIN, D., (2003), « une approche de la complexité en géographie », *Géocarrefour*, vol.78, n°1, pp. 25-31.
- PUMAIN, D., (2004), « L'invention permanente : la hiérarchie urbaine », In BRIDEL, P., (ed.), *L'Invention dans les sciences humaines. Hommage du groupe Raison et rationalités à Giovanni Busino*, Labor et Fides, Genève, pp. 66-86.
- PUMAIN, D., ROBIC, M.-C., (2002), « Le rôle des mathématiques dans une « révolution » théorique et quantitative : la géographie française depuis les années 1970 ». *Revue d'histoire des sciences humaines*, 6, pp. 123-144.
doi : 10.3917/rhsh.006.0123

- PUMAIN, D., SAINT JULIEN, T., (1997), *L'analyse spatiale*, coll. Cursus, Armand Colin, Paris, 167 p.
- PUMAIN, D., SAINT JULIEN, T., (2001), *Les interactions spatiales*, coll. Cursus, Armand Colin, Paris, 191 p.
- PUMAIN, D., SANDERS, L., SAINT JULIEN, T., (1989), *Villes et auto-organisation*, Paris, Economica, 191 p.
- RACINE, J.-B., REYMOND, H., (1973), *L'analyse quantitative en géographie*, Presses Universitaires de France, Paris, 316 p.
- RAVALLION, M., (1993), « Poverty Alleviation Through Regional Targeting: A Case Study for Indonesia », in HOFF, K., BRAVERMAN, A., STIGLITZ, J.E., (eds.), *The Economics of Rural Organization*, Oxford University Press.
- RAVALLION, M., WODON, Q. (1997), *Poor areas, or only poor people?*, Policy research working paper, n°1798, 34 p.
- REYNOLDS, H.D., (1998), *The modifiable area unit problem: Empirical analysis by statistical simulation*, Thèse de doctorat de l'université de Toronto, non publiée, 92 p.
- RHIND, D., OPENSHAW, S., GREEN, N., (1988) « The analysis of geographical data: data rich, technology adequate, theory poor », in RAFANELLI, M., KLENSIN, J.C., SVENSSON, P., (eds.), *SSDBM'1988 Proceedings of the 4th international conference on Statistical and Scientific Database Management*, Springer, Londres, pp. 427-454.
- RICHARDSON, S., GUIHENNEUC-JOUYAUX, C., (2009), « Impact of Cliff and Ord (1969, 1981) on Spatial Epidemiology », *Geographical analysis*, pp. 444-451.
- RIPLEY, B.D., (1981), *spatial statistics*, John Wiley, Chichester.
- RIPLEY, B.D., (1998), *Statistical inference for spatial processes*, Cambridge University Press, Cambridge, 148 p.
- ROGERS, Everett M., (1995, 1ère éd. 1962), *Diffusion of innovations*, Free Press, New York, 518 p.
- ROGERSON, P.A, (2010), « Optimal geographic scales for local spatial statistics », *Statistical Methods in Medical Research* », published online 2 June 2010.
doi : 10.1177/0962280210369039
- ROGERSON, P.A, KEDRON, P., (2011), « Optimal Weights for the Local Moran Statistic », communication au Western Regional Science Association 50th Anniversary Meeting, Monterey, Californie.
- ROZENBLAT, C., (2010), « Opening the Black Box of Agglomeration Economies for Measuring Cities' Competitiveness through International Firm Networks », *Urban Studies*, Vol. 47, pp. 2841-2865.
doi : 10.1177/0042098010377369
- SAINT JULIEN, T., (1985), *La diffusion spatiale des innovations*, coll. Reclus modes d'emploi, GIP RECLUS, Montpellier, 40 p.
- SANDERS, L., (1992), *Système de villes et synergétique*, Paris, Anthropos, coll. Villes, 274 p.
- SANDERS, L., (2001), *Modèles en analyse spatiale*, Hermès Sciences, Paris, 333 p.
- SANDERS, L., DURAND-DASTES, F., (1985), *L'effet régional : les composantes explicatives dans l'analyse spatiale*, coll. Reclus modes d'emploi, GIP RECLUS, Montpellier, 48 p.

- SMITH, T.E., LEE, K.L., (2011), « The effects of spatial autoregressive dependencies on inference in ordinary least squares: a geometric approach », *Journal of geographical systems*, Vol.13, 34 p.
doi : 10.1007/s10109-011-0152-x
- SUNDERLIN, W.D., DEWI, S., PUNTODEWO, A., (2007), *Poverty and Forests: Multi-Country Analysis of Spatial Association and Proposed Policy Solutions*, CIFOR occasional paper n° 47, CIFOR, Bogor, Indonésie, 44 p.
- THOMAS, I., (2001), « Cartographie d'aujourd'hui et de demain : rappels et perspectives », *Cybergeo*.
<http://cybergeo.revues.org/3812>
- TIEFELSDORF, Michael, (1998), « Some Practical Applications of Moran's I's Exact Distribution », *Papers in Regional Science*, Vol. 77, pp. 101-129.
- TOBLER, W.R., (1970), « Computer movie simulating urban growth in the Detroit region », *Economic geography*, Supplement 46, pp. 234-240.
- TOBLER, W.D., (2002), « Ma Vie », in *Geographical voices*, Gould, P., Pitts, E.R., (ed.), Syracuse University Press, pp. 293-322.
- TOBLER, W.D., (2004), « On the First Law of Geography: A Reply », *Annals of the Association of American Geographers*, vol. 94, n° 2, pp. 304-310.
<http://www.jstor.org/stable/3693989>
- TOWNSLEY, M., (2009), « Spatial Autocorrelation and Impacts on Criminology », *Geographical analysis*, Vol. 41, n°4, pp. 452-461.
doi : 10.1111/j.1538-4632.2009.00775.x
- TUKEY, J.W., (1977), *Exploratory Data Analysis*, Reading MA: Addison Wesley publishing Company, 688 p.
- VELLA, S., OLIVEAU, S., (2005), « Spatio-temporal Trends of Female Discrimination in South India : A Case Study in Tamil Nadu, 1961-1991. », in Attané, I., Véron, J., (Ed.), *Gender discriminations among young children in Asia*, Collection Sciences Sociales n° 9, IFP-CEPED, Pondichéry (Inde), pp 67-90.
- VON THÜNEN, J.H., (1826), *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hambourg.
- WANIEZ, P., (2010), « Philcarto : histoire de vie d'un logiciel de cartographie », *Cybergeo : European Journal of Geography*, n°497.
<http://cybergeo.revues.org/23076>.
- WONG, D., (2011), « Exploring Spatial Patterns Using an Expanded Spatial Autocorrelation Framework », *Geographical analysis*, Vol. 43, n°3, pp. 327-338.
doi : 10.1111/j.1538-4632.2011.00816.x
- WONG, D., (2009), « The Modifiable Areal Unit Problem (MAUP) », in FOTHERINGHAM, A.S., ROGERSON, P., *The SAGE handbook of spatial analysis*, Sage, pp. 105-124-158.
- WONG, D., AMRHEIN, C., (1996a), « Research on the MAUP : old wine in a new bottle or real breakthrough ? », *Geographical systems*, Vol. 3, n°2-3, pp. 73-76.
- WONG, D., AMRHEIN, C., (Ed.), (1996b), *Geographical systems*, special issue on « The Modifiable Areal Unit Problem », Vol. 3, n°2-3.
- WOOLDRIDGE, M., 2002, *An Introduction to Multiagent Systems*, New York: John Wiley and Sons, 348 p.

Cette annexe reprend une partie du contenu de la présentation faite dans le cadre d'un séminaire au CIESIN (Columbia University) le 13 septembre 2005 et qui n'a pas fait l'objet de valorisation par la suite. Elle ne s'étendra pas sur la longueur d'un article, reprenant succinctement les éléments clefs de la présentation, mais laissant de côté les questions abordées concernant l'agrégation de données, la mesure des structures spatiales et le détail des indicateurs utilisés. Elle est directement redevable du travail mené par C.Z.Guilmoto et l'équipe du projet EMIS : Christophe Z. Guilmoto,, Olivia Aubriot, Virginie Chasles, Katia Radja, Alain Vaguet, Stéphanie Vella et moi même.

Starting from the definition of vulnerability given by Chaudhuri, Jalan et Suryahadi (2001), «We define vulnerability [...] as the ex-ante risk that a household will, if currently non-poor, fall below the poverty line, or if currently poor, will remain in poverty». This could also be the definition of Poverty traps, which leads to the concept of spatial poverty traps in geographical context.

Lack of data

To assess vulnerability you need information on the household. In developing countries, this information may lack, or be partial. In the Indian case, different sources are at our disposal: NFHS (Demographic and Health Survey for India); National Council for Applied Economic Research Database. But these data are sampled, and not significant below main regions level.

The census option

Census in India is old (1871 for the first one) and regular (every 10 years since then). Quality of the census is good for a developing country. Information is available at different scales (villages and town, sub-districts, districts). But purpose of the census is not to assess vulnerability. We need to adapt our question to the available data.

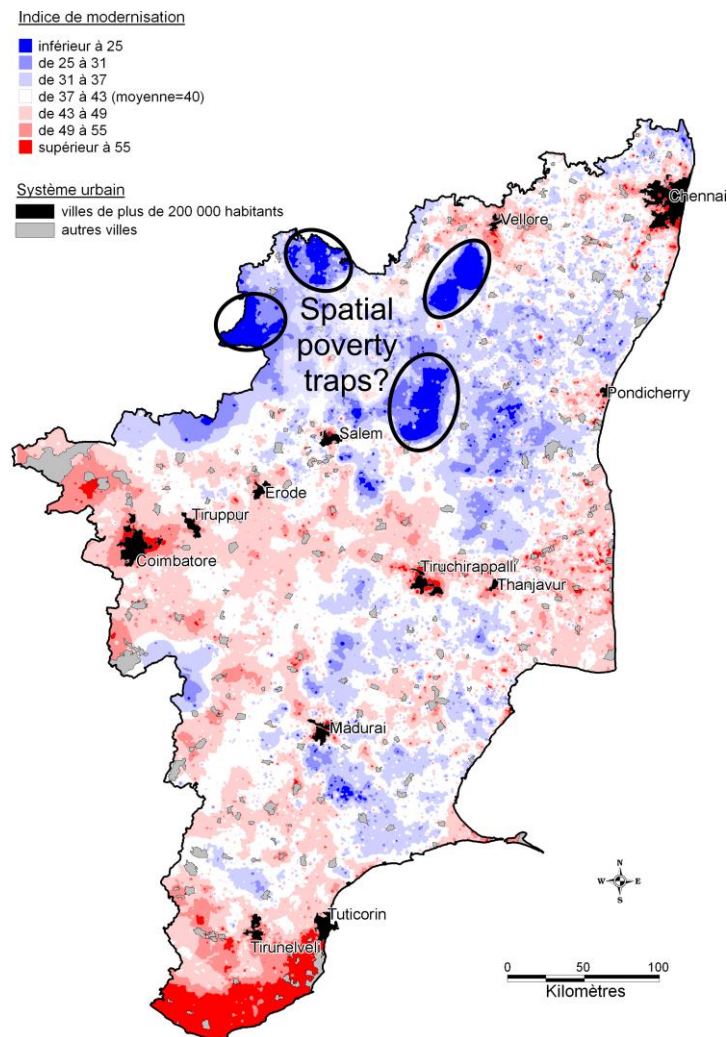
Two ways to assess vulnerability

The synthetic way: Choosing the more meaningful indicators and creating a new index.

The split-up approach: Computing all your indicators independently and doing a new map by overlaying them.

Synthetic Way

Construction of an index synthesizing the chosen census data (1st axis of the principal component analysis). Mapping of the index will show us the vulnerability pockets.



Advantages of the method:

Robust way to treat your information (you know how your index is created and what the map is showing).

Show in one dimension the multiple aspect of the dataset. You can easily use the index in econometric models.

Easier to communicate as your information is reduced.

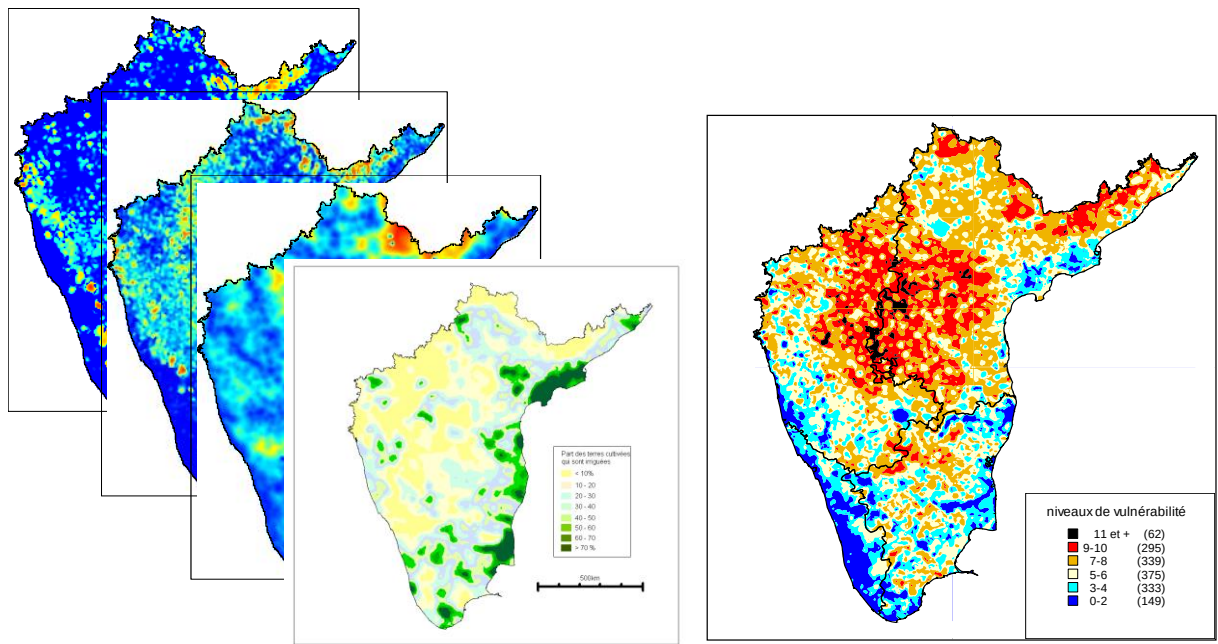
Disadvantages of the method:

Need to select the good data (should perform exploratory data analysis).

You need to redo the principal component analysis if you want to add a new indicator(s).

Split up approach

Mapping all the indicators independently and create a new map by overlaying them.



Clusters of villages painted in black could be defined as spatial poverty traps.

Advantages of the method:

Each variable can be independently studied to know its own characteristics (statistical distribution, spatial pattern, etc.)

You can give (and change) different weight to each indicator according to the importance it has in defining vulnerability.

If you want to create new variables (especially based on space), you just have to map it and overlay it.

If you want to add one more variable, just a new overlay operation to do in the GIS

Disadvantages:

The final scale has no meaning, and is dependent from the number of variables you've put in it.

The weights you can give are not easy to assess (is irrigation two times more important than distance to the city?).

You have redundancy in your results, as your variables are correlated between each others.

TABLE DES MATIERES

<i>Remerciements</i>	1
<i>Avant-Propos</i>	5
« <i>Space matters</i> »	7
La question statistique.....	9
La mesure de l'homogénéité.....	10
<i>L'espace au cœur des sociétés</i>	13
Poverty mapping et spatial poverty traps	17
Dimensions spatiales des comportements sociaux.....	20
Linking census through time	22
Qui se ressemble s'assemble ?	24
<i>Voisinages</i>	27
Les formes de voisinage.....	27
Contiguïté	28
Accessibilité.....	35
Proximité.....	37
Comment choisir ses voisins ?	39
Les matrices de voisinage	42
Matrice binaire	42
Effets structurels	44
Effet de bord	44
Morphologie.....	44
Invariance d'échelles	45
Matrice de poids spatiaux	45
Standardisation en ligne	46
<i>La mesure des structures spatiales</i>	49
Les indicateurs globaux.....	50
Quelques exemples de structures spatiales.....	51
Notations.....	52
Semivariance.....	52
Covariance spatiale	53
Isotropie ?	54
A propos du krigeage	54
L'indice de Moran.....	56

Présentation.....	56
Le nuage de points de Moran.....	60
Le coefficient de Geary et les statistiques gamma	64
<i>Les indices locaux</i>	66
Des associations spatiales aux statistiques G	67
Les LISA	69
Le I de Moran Local	70
Significativité.....	70
Cartographie.....	71
Regroupements d'associations spatiales	75
« Moran scatterplot is not a LISA »	76
Retour sur les statistiques G.....	77
Retour sur les archétypes de structures spatiales	78
<i>Au-delà des indices locaux</i>	83
Jeux d'échelles	83
La question du MAUP	88
Des gammas intermédiaires ? Une nouvelle utilisation des LISA.....	89
Une approche bivariable de l'autocorrélation spatiale.....	91
La régression spatiale.....	92
<i>Ouvertures</i>	97
Retour aux sources	97
Distance et mesure	97
Espace et population	99
Perspectives.....	100
Du groupe aux individus ?	102
Diffusion spatiale.....	102
Simulation et approche individus-centrée.....	103
Faut-il conclure ?.....	104
<i>Bibliographie</i>	105
<i>Annexe : Deux approches de la vulnérabilité des populations</i>	120
<i>Table des matières</i>	124