

Avant-propos

« Je ne mens que pour dire la vérité. »

Proverbe chinois

Cet ouvrage propose une vue d'ensemble des objectifs, des principes et des méthodes de l'analyse spatiale et des systèmes d'information géographique dans le domaine de la santé, et plus particulièrement dans l'étude des maladies infectieuses et des relations santé-environnement. Il est construit comme une introduction pratique à l'analyse spatiale et spatio-temporelle pour l'épidémiologie et la géographie de la santé. Il a pour objectif de décrire en détail les objectifs, les concepts, et la plupart des méthodes et techniques disponibles dans ce domaine, avec une démarche pédagogique illustrée d'exemples concrets. Il est construit comme un ouvrage didactique à destination des étudiants et des professionnels de la santé publique, épidémiologistes, inspecteurs de santé publique, géographes de la santé, spécialistes des rapports santé-environnement, que ce soit en santé humaine ou en santé animale, et qui souhaitent trouver un panorama complet du sujet sans avoir à l'aborder de façon trop mathématique ou statistique. Enfin, il est également construit comme un outil à l'usage de tous ceux qui souhaitent une introduction aux méthodes générales de l'analyse spatiale.

L'analyse spatiale inclut toute technique qui étudie des objets et leurs attributs en utilisant des propriétés topologiques ou géométriques, généralement dans un espace métrique de dimension deux ou trois. Cette définition est très générale et s'applique à de nombreux domaines. L'analyse spatiale n'est pas une discipline récente ; elle est utilisée depuis de nombreuses années en biologie, en botanique, en épidémiologie, pour le traitement d'image, pour l'analyse de réseaux, dans la conception électronique, en chimie, en cosmologie, en climatologie, en hydrologie, en économie, etc. Et bien sûr en géographie, où l'analyse spatiale est définie comme « l'analyse formalisée de la

configuration et des propriétés de l'espace géographique, tel qu'il est produit et vécu par les sociétés humaines » [PUM 97].

En épidémiologie (étude des facteurs influant sur la santé et les maladies dans une population) et en géographie de la santé (analyse géographique du système de santé et de la distribution spatiale des maladies)¹, nous utiliserons le terme d'« analyse spatiale » pour décrire les techniques d'analyse appliquées aux « objets » décrits ou utilisés par l'épidémiologie ou la géographie, dès lors qu'ils sont localisés dans l'espace et que l'analyse utilise cette localisation : les individus, les vecteurs, les réservoirs, les populations, les territoires, l'environnement, qu'il soit naturel, urbain, rural, etc. L'analyse spatiale utilise les relations, topologiques ou géométriques, des individus avec leur environnement et des individus entre eux. Est exclu ce qui se passe « dans » le malade (dans l'organe, dans la cellule, ou au niveau de la biologie de l'agent pathogène). Par exemple, cet ouvrage ne traite pas d'imagerie médicale et des techniques associées de traitement d'image, même si certaines de ces techniques sont parfois très proches de celles que nous décrirons.

Les phénomènes de santé sont rarement distribués de façon aléatoire dans l'espace : un phénomène de santé comporte bien souvent des facteurs de risque liés à des facteurs géographiques, à des facteurs mésologiques et aux relations spatiales entre les individus. Utiliser la localisation est donc fondamental dans l'analyse et la compréhension des phénomènes de santé et de leurs mécanismes : en prenant en compte les relations et interactions spatiales entre les acteurs de la maladie – vue comme un système complexe – l'analyse spatiale permet de mieux identifier et comprendre les mécanismes et les processus qui sous-tendent les phénomènes.

En épidémiologie, l'analyse spatiale offre ainsi des éléments qui permettent de consolider l'épidémiologie « classique » et de nourrir la recherche et le paramétrage de modèles. Elle permet également d'enrichir la démarche analytique en géographie de la santé, dont le corpus méthodologique intègre également tout un ensemble d'approches qualitatives. L'analyse spatiale *descriptive* inclut l'analyse cartographique, la recherche de caractéristiques géométriques et spatio-temporelles, l'analyse de la variabilité spatiale d'une valeur, la recherche d'agrégats, la recherche d'échelles d'analyse et de synthèse, la recherche de corrélations environnementales, etc. L'analyse spatiale *explicative* est essentiellement statistique, avec la recherche de modèles statistiques incluant les relations spatiales entre individus. Nous n'aborderons que très succinctement la modélisation des processus spatiaux, ce sujet dépassant le cadre de cet ouvrage.

1. Des définitions précises seront données au chapitre 1.

Dans le domaine de la santé, l'analyse spatiale n'est pas seulement utilisée pour des études en épidémiologie ou en géographie. Elle intervient également au niveau des politiques publiques, avec le développement de nouvelles applications en santé publique : systèmes d'alerte, systèmes de gestion de crises, systèmes de prévention et d'analyse de risques, préparation de campagne de vaccination, enquêtes et sondages.

Dans cet ouvrage, nous présentons les concepts généraux sur lesquels s'appuie la démarche de l'analyse spatiale et cherchons à expliquer et à éclairer les principes utilisés dans les méthodes d'analyse. L'accent sera mis également sur l'utilisation pratique de ces méthodes : de nombreux exemples concrets seront utilisés tout au long de cet ouvrage. Ils sont tirés de données réelles collectées au cours de travaux de recherche. Ces exemples couvrent des situations souvent rencontrées dans la pratique.

L'utilisation de l'analyse spatiale dans le domaine de la santé s'est fortement accrue ces dernières années avec le développement de la géomatique et des systèmes d'information géographique (SIG). L'analyse spatiale en santé bénéficie, comme dans d'autres domaines d'application, de la diffusion des SIG, du développement de leurs fonctionnalités techniques et de la disponibilité grandissante de données géographiques, même si la qualité de ces données est encore souvent insuffisante.

Il est difficile, sinon impossible, de ne pas utiliser un SIG pour gérer, transformer, manipuler, analyser, représenter l'information spatialisée. Une annexe² est consacrée à la description des SIG et des techniques de manipulation de données localisées qu'il faut connaître pour développer un projet d'analyse spatiale en épidémiologie, et qui permettent de mettre en œuvre les techniques d'analyse présentées dans cet ouvrage. Le logiciel utilisé pour l'illustrer (SavGIS) et des exemples complets sont disponibles en téléchargement sur le site Internet de SavGIS³.

Pour finir, qu'il me soit permis de remercier tous ceux qui ont contribué à la genèse de cet ouvrage. Tout d'abord Jean-Paul Gonzalez, médecin, virologue, qui a su inspirer, stimuler et diriger sans pareil, avec un enthousiasme permanent ; Florent Demoraes, géographe, qui a permis de renforcer, consolider et compléter ces réflexions ; Bernard Lortic, ingénieur, dont l'exigence restera sans égale ; tous les collègues, étudiants, doctorants, stagiaires qui ont contribué directement ou indirectement à l'amélioration de cet ouvrage, et en particulier Nitin Tripathi, José Tupiza, Somsakun Maneerat, Julie Vallée, Jothiganesh Sundaram, Tania Serrano. Que tous trouvent ici le témoignage de ma sincère gratitude.

2. Disponible en téléchargement à l'adresse : http://www.iste.co.uk/fr/Souris_Annexe_française_MEP.zip.

3. www.savgis.org.

Logiciels et bases de données

Afin d'aider le lecteur à utiliser les méthodes présentées dans cet ouvrage, nous indiquerons au fil du texte comment les mettre en œuvre avec quelques logiciels que nous avons sélectionnés. Nous présentons également quelques jeux de données qui peuvent être téléchargés et utilisés pour reproduire les exemples illustrant cet ouvrage.

I.1. Logiciels

Nous avons sélectionné quelques logiciels qui permettent de mettre en œuvre la plupart des méthodes présentées dans cet ouvrage : des systèmes d'information géographique généralistes (QGIS, ArcGIS, SavGIS) ou des logiciels plus spécifiques (R, GeoDA, SatScan, GWR4). Après la description des méthodes, nous indiquerons très succinctement pour chaque logiciel quelles sont les procédures à utiliser et comment y accéder, lorsqu'elles sont disponibles, sans plus de détails. Le lecteur pourra si nécessaire se référer aux manuels d'utilisation des logiciels.

I.1.1. QGIS



Quantum GIS (QGIS¹) est un système d'information géographique gratuit et *open source*. Il fonctionne sous Linux, Unix, Mac OSX, Windows et Android, et prend en charge de nombreux formats (vectoriels et matriciels) pour les données et bases de

1. Informations détaillées, documentation, téléchargements, tutoriels sont disponibles sur <http://www.qgis.org>.

données. QGIS offre un nombre sans cesse croissant de fonctionnalités fournies par des fonctions de base et des plug-ins.

I.1.2. ArcGIS



Le système d'information géographique ArcGIS² est un produit commercial de Environmental Systems Research Institute (ESRI). Logiciel très complet, il comprend un grand nombre de fonctionnalités. Le système fournit une infrastructure permettant de rendre les cartes et l'information géographique disponibles et partagées pour une entreprise, une communauté et sur le Web.

I.1.3. SavGIS



SavGIS³ est un système d'information géographique gratuit, fonctionnant sous Windows. Issu de la recherche, ce logiciel complet et puissant est en constante évolution pour proposer des solutions innovantes dans le traitement de l'information localisée, avec de nombreux développements concernant l'analyse spatiale et la modélisation pour l'épidémiologie. Outre sa gratuité, les atouts de ce logiciel sont nombreux : rigueur dans la gestion des données, partage des données, puissance d'analyse, fonctions avancées d'analyse spatiale, fonctions d'analyse statistique.

I.1.4. R



R (logiciel libre et *open source*) est un langage de programmation pour l'analyse statistique des données, ainsi qu'un environnement d'analyse de données et de visualisation graphique. Les scientifiques et les chercheurs ont mis en œuvre un grand nombre de procédures spécialisées pour une grande variété d'applications intégrées

2. On trouvera de plus amples informations sur le logiciel ArcGIS sur <https://www.arcgis.com>.

3. Informations détaillées sur <http://www.savgis.org>.

directement dans R. R-GIS.net est un site web qui vise à discuter de la manipulation des données spatiales et l'analyse dans R. Plusieurs « packages »⁴ sont disponibles pour les opérations liées aux données spatialisées : *sp*, *spdep*, etc.

1.1.5. GeoDa



GeoDa⁵ est un logiciel d'analyse spatiale gratuit et *open source*, développé depuis 2003 par l'université de l'État d'Arizona (États-Unis). GeoDa est un outil logiciel focalisé sur l'analyse spatiale et les modèles spatiaux. Le programme fournit une interface graphique conviviale pour les méthodes d'analyse des données spatiales exploratoires (ESDA), telles que les statistiques d'autocorrélation spatiale pour les données agrégées et l'analyse de régression spatiale de base pour les données ponctuelles et zonales.

1.1.6. SaTScan



SaTScan⁶ est un logiciel gratuit qui analyse des données spatiales à l'aide des statistiques spatiales, temporelles ou spatio-temporelles. L'objet principal de SaTScan est la détection des agrégats et la mise en place de systèmes d'alerte ou de détection précoce des maladies. Le logiciel peut également être utilisé pour des problèmes similaires dans d'autres domaines scientifiques.

1.1.7. GWR4

La régression pondérée géographiquement (GWR) est une technique d'analyse spatiale qui tient compte de variables qui présentent de l'autocorrélation spatiale et des variations locales. Cette technique de régression permet de modéliser les relations locales entre les variables explicatives et la variable à expliquer. Plusieurs logiciels

4. Plus d'informations sur <http://r-gis.net> et dans l'ouvrage framebook.org/r-et-espace.

5. Plus d'informations sur <http://geodacenter.github.io/>.

6. Plus d'informations sur <https://www.satscan.org/>.

permettent d'exécuter des régressions pondérées géographiquement (ArcGIS, SpaceStat, SAM, les *packages* *spgwr*, *gwr* ou *GWmodel* de R), mais *GWR4*⁷ est une application Windows autonome.

I.1.8. GAMA



GAMA⁸ est une plateforme libre de modélisation et de simulation à base d'agents développée depuis 2007, et plus particulièrement dédiée à la simulation de phénomènes spatialisés. GAMA offre un certain nombre de fonctionnalités avancées : une gestion poussée des données géographiques ; un ensemble de structures et de commandes facilitant la définition de modèles multiniveaux ; des outils automatisés d'aide à l'exploration de modèles permettant de définir des plans d'expérience et de les exécuter sur des ressources de calcul haute performance (*cluster*, grille) ; un système de plug-ins qui permet d'étendre le langage GAML pour des besoins spécifiques ; des ponts et des possibilités de couplage avec d'autres outils utilisés dans le domaine de la modélisation des systèmes complexes.

I.2. Données pour les exemples

Les méthodes présentées dans cet ouvrage seront illustrées par des exemples tirés de situations et de bases de données réelles. Les données liées à ces exemples sont disponibles sous forme de fichiers Excel pour les données non localisées, de fichiers au format Shapefile pour les données géolocalisées, ou de bases de données géographiques complètes directement exploitables avec le logiciel SavGIS⁹.

7. Plus d'informations sur GWR4 sur <http://gwr.maynoothuniversity.ie/gwr4-software/>.

8. <http://gama-plaform.org>.

9. Ces fichiers, ainsi que les bases de données SavGIS, peuvent être téléchargés à l'adresse www.savgis.org.