

SYSTÈMES D'INFORMATION GÉOGRAPHIQUE EN SANTÉ ENVIRONNEMENT

Florent Occelli

Université de Lille, EA4483 - IMPact de
l'environnement chimique sur la santé
humaine

Faculté de pharmacie

Laboratoire des sciences végétales et fon-
giques EA4483, Lille, France

Les systèmes d'information géographique (SIG) sont de puissants outils d'analyse des liens en santé-environnement. Les principaux enjeux reposent aujourd'hui sur l'accès aux données multidisciplinaires et multidimensionnelles, ainsi que sur leur interopérabilité.

Quelles utilisations des SIG en santé-environnement ?

La dimension géographique exerce très certainement une influence majeure sur les déterminants de santé des populations, tels que la génétique, les caractéristiques socioéconomiques et l'environnement de vie. Elle prend tout son sens dans l'évaluation des liens santé-environnement, puisqu'elle va conditionner le contexte environnemental et sociodémographique de notre lieu de vie, notre proximité aux sources de nuisances, ainsi que la qualité des milieux environnementaux auxquels nous sommes exposés. Dans cette optique, les SIG sont désormais considérés comme des outils incontournables en santé-environnement.

« Un SIG est un ensemble organisé de matériels informatiques, de logiciels, de données géographiques et de personnels capables de saisir, stocker, mettre à jour, manipuler, analyser et présenter toutes formes d'informations géographiquement référencées » [1].

Ces outils permettent non seulement de cartographier des jeux de données géoréférencées et multidisciplinaires (ex : pollution de l'environnement, statistiques démographiques et socioéconomiques, morbidité des populations), mais surtout d'étudier leur interaction statistique dans l'espace-temps. L'un des précurseurs de leur utilisation fut sans doute John Snow, qui en 1854 eut l'idée de représenter sur une carte de Londres la densité de cas de choléra sur plusieurs périodes et les points de distribution d'eau. Il avait alors émis l'hypothèse que cette pathologie se disséminait par l'intermédiaire de l'eau potable. Il a ainsi mis en évidence que les cas se concentraient autour d'une pompe à eau de Broad Street, dans le district de Soho [2]. Depuis la fin du XX^e siècle et l'avènement de l'ère numérique, les SIG se sont développés pour aujourd'hui employer de puissants outils d'analyse spatiale basés sur des concepts de distance, de centralité ou d'interaction spatiale. Il ne s'agit plus aujourd'hui de superposer des données, mais d'étudier leurs corrélations géographiques.

Quelques exemples d'outils utilisés en santé-environnement :

- le géocodage pour localiser des patients à partir de leur adresse de résidence ;
- l'analyse de distance pour calculer la proximité d'un individu aux axes routiers, sites et sols pollués, industries à risque, en tenant compte de la direction des vents dominants ;
- le calcul d'accessibilité pour estimer le temps d'accès aux services de proximité comme les professionnels de santé, les services écosystémiques ;
- la jointure spatiale pour identifier les populations sensibles potentiellement exposées à un danger (ex : présence de crèches ou d'écoles sur d'anciens sites à risque) ;

- l'analyse de tendance pour identifier les points noirs environnementaux (détection de points chauds : zones géographiques présentant une pollution statistiquement atypique) ;
- la cartographie de maladies pour modéliser la distribution spatiale d'incidence des événements de santé sur la base de registres (ex : cancers, maladies cardiovasculaires) ;
- la détection de clusters pour identifier les zones géographiques de sur- et de sousincidence atypique ;
- l'interpolation spatiale pour modéliser les données environnementales en tout point de l'espace à partir d'informations fragmentées ;
- la régression géographique pour étudier les liens géographiques entre indicateurs sanitaires, environnementaux et socioéconomiques et identifier des facteurs de risques des maladies.

La donnée : principale source d'incertitude

La donnée constitue le carburant des SIG. La disponibilité des données environnementales, sanitaires et sociodémographiques sur des territoires communs, à des échelles suffisamment fines et à des temporalités adaptées au temps de latence des maladies est indispensable. La donnée représente la principale source d'incertitude dans l'étude spatialisée des risques sanitaires liés à l'environnement. Cette source d'incertitude peut être double : interne à la donnée et classiquement discutée, telle que la performance de la mesure du paramètre (précision, exactitude, spécificité, sensibilité, reproductibilité, répétabilité) ou externe à la donnée et spécifique à son implémentation dans un SIG, comme l'approche des zones blanches (dépourvues d'informations) ou la propagation des incertitudes dans les analyses conjointes¹. La disponibilité et la bonne qualité de la donnée d'entrée sont donc primordiales car elles déterminent, pour une grande part, le degré de confiance accordé aux résultats issus des analyses. En dehors des études appliquées, la recherche dans ce domaine s'intéresse à la réduction de ces incertitudes pour augmenter la robustesse des conclusions.

Adapté aux SIG, le concept de réutilisation de données produites dans un autre cadre (limitant ainsi le coût de production de nouvelles données) permet de répondre à cette attente. Dans le domaine de la santé, un article paru en 2017 montre par exemple la possibilité de réutiliser les données issues du Programme médicalisé des systèmes d'information (PMSI) pour générer des cartographies nationales d'incidence d'événements médicaux à fine échelle sur le territoire national [3]. Ces cartographies étaient auparavant conditionnées par la présence de registres de santé, dont l'étendue spatiotemporelle reste souvent limitée, comme pour les cancers.

L'établissement de conventions ou partenariats entre producteurs de données et équipes de recherche permet également de répondre à l'objectif de rendre disponible une donnée fiable et pertinente. En ce sens, le SIG représente un support adéquat pour le montage de projets multidisciplinaires. Cependant, dans de nombreux cas, le verrou de la mise en relation des données issues de multiples sources et de natures diverses est à lever : il est rarement possible de directement intégrer les données de santé, les données socioéconomiques et les données environnementales au sein d'une même analyse sans traitement préalable. D'un côté, les données caractérisant l'état de santé ou le statut socioéconomique des populations (ex : taux d'incidence d'un événement de santé ou taux de chômage) sont de nature agrégée. Elles caractérisent un groupe de population, représenté par une unité spatiale administrative (données latticielles). Il est rare de pouvoir disposer de la géolocalisation individuelle de patients, du fait de la confidentialité de cette information. D'un autre côté, les données environnementales de pollution sont de nature géostatistique, c'est-à-dire non agrégée. Les modélisations de pollution des milieux sont bâties soit sur des modèles déterministes à partir de plusieurs paramètres d'entrée tels que les sources d'émissions, la topographie, la météo et la dynamique des fluides (ex : modèles de dispersion atmosphérique), soit sur l'interpolation spatiale de mesures réelles de terrain à partir d'un échantillon de points (ex : krigeage). Ces modélisations prédisent les concentrations selon une maille régulière, allant de quelques mètres à quelques kilomètres de côté. Il faut

1 L'incertitude interne de chaque donnée/variable d'entrée contribue à la l'incertitude totale dans une analyse combinant plusieurs données/variables.

ensuite agréger ces données en calculant une valeur moyenne pour chaque unité administrative (figure 1). Assurer l'interopérabilité statistique entre les différentes bases de données fait aujourd'hui l'objet de nombreux travaux de recherche afin d'optimiser les méthodologies.

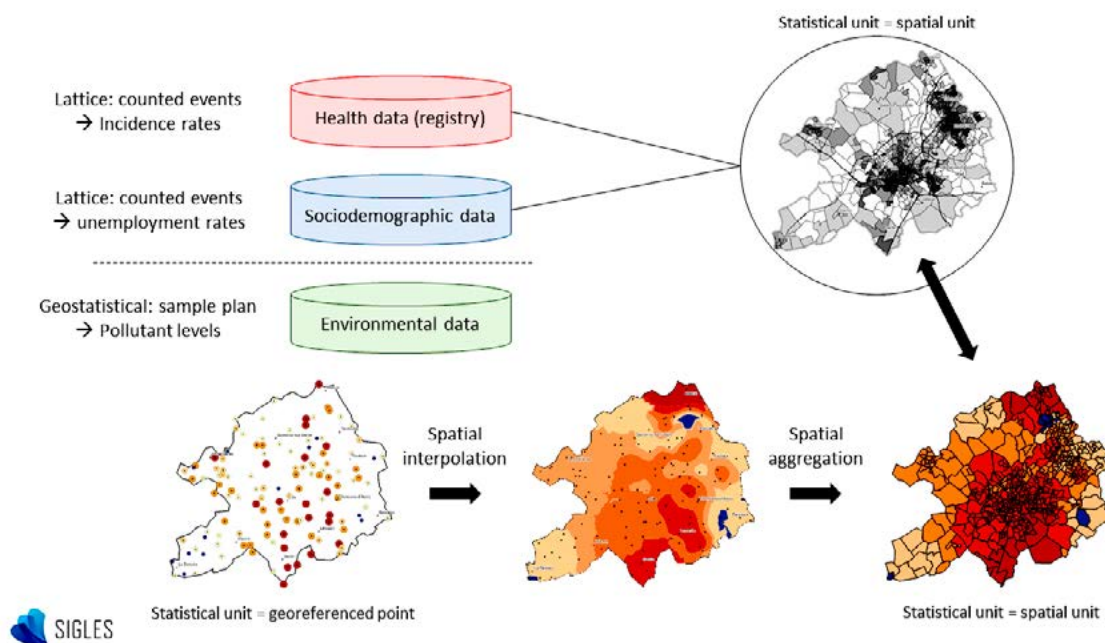


Figure 1. Interopérabilité spatiale des données sanitaires, sociodémographiques et environnementales (Source : SIGLES)

De la multiexposition à l'exposome des territoires

Initialement mis en œuvre pour simplement associer la densité de population à la présence d'une seule source de nuisance, les SIG en santé-environnement sont aujourd'hui utilisés en routine pour générer des indicateurs complexes spatialisés à partir de bases de données recensant de multiples sources de nuisances ou de surveillance de la contamination des milieux. Dans le cadre de la mesure de l'exposition des populations aux facteurs de risques environnementaux, on peut alors explorer des dimensions telles que la proximité aux industries à risques, aux sites et sols pollués, aux axes routiers ou la modélisation des concentrations environnementales dans l'air, l'eau et les sols. L'intérêt de ces indicateurs se trouve dans la notion spatiale et intégrée de la pollution. Cela répond aux attentes actuelles de la majeure partie des approches scientifiques en santé-environnement, pour lesquelles la mesure de l'exposition et des risques associés fait l'objet d'une interprétation substance par substance. Bien qu'ils ne pourront jamais prétendre évaluer l'exposition réelle des populations, ces indicateurs demandent à être optimisés dans la prise en compte du continuum source-vecteur-cible. Le couplage des modèles multimédia² avec des modèles toxicocinétiques³ et des approches spatialisées constituent une piste pertinente [4].

De plus, l'homme se situe dans un contexte récurrent de pollutions complexes, du fait notamment de la multiplicité des activités anthropiques émettrices. Il est quotidiennement exposé à son environnement d'une manière globale, et donc à l'ensemble des contaminants environnementaux. Les effets de ces mélanges sont difficilement identifiables en toxicologie ou écotoxicologie et l'additivité des risques reste encore peu utilisée dans les méthodes d'évaluation des risques sanitaires. À partir de multiples bases de données géoréférencées, les approches spatiales développées actuellement apportent de

² Modélisation du comportement des polluants dans l'environnement et des voies d'exposition pour le calcul des doses d'exposition.

³ Modélisation du devenir des polluants dans l'organisme, par exemple les modèles pharmacocinétiques physiologiques (PBPK).

nouveaux indicateurs combinant différentes dimensions du risque, permettant l'identification des zones de cumul d'exposition. En ce sens, même s'il est encore difficile de générer un indicateur qui caractérise la qualité globale de l'environnement, tous polluants confondus, des indices de pollution intégrant plusieurs contaminants de manière concomitante peuvent être développés [5].

Bien qu'il ne soit pas encore formalisé, le concept d'exposome territorial a récemment été avancé. Il peut être approché comme la vision globale, intégrée et multidimensionnelle de l'exposition des populations aux facteurs de risques, au sein d'un territoire. L'évolution des SIG est donc naturellement de relever ce défi autour de la caractérisation (quantitative et qualitative) des territoires pour enrichir les connaissances sur les liens entre santé et environnement.

Conclusion

Encore trop peu mises en œuvre dans la recherche en santé-environnement, les techniques de spatialisation des données ont tout intérêt à être développées pour répondre aux enjeux des Plans nationaux santé environnement, tels que l'identification de facteurs de risques des maladies, l'évaluation de l'exposition des populations et la hiérarchisation des inégalités environnementales et sociales de santé (inégalités territoriales). Les SIG ont en effet un rôle à jouer dans la description spatiale et spatio-temporelle des maladies chroniques, notamment non-transmissibles, mais surtout dans la recherche de pistes étiologiques environnementales.

Le transfert et la vulgarisation des données et des résultats auprès du grand public nécessitent également d'être accentués, avec par exemple le développement de cartographies web (www.sirsepaca.org) et d'histoires racontées avec des cartes. Cet enjeu s'inscrit parfaitement dans le cadre de la Loi pour une république numérique de 2016, puisque ces outils permettent de diffuser les données sous la forme de cartographies interactives, idéalement accompagnées d'un descriptif et d'une analyse critique afin de limiter les biais de lecture et d'interprétation des cartes.

Liens d'intérêt en rapport avec le texte publié : **aucun**

Références

1. De Blomac F, Gal R, Hubert M, Richard D, Tourret C. *Arc/Info, concepts et applications en géomatique*. Paris : Hermès, 1994 ; 256 p.
2. Snow J. *On the Mode of Communication of Cholera*. London : Churchill 2nd ed., 1855 ; 162 p.
3. Ghenassia A, Beuscart J-B, Ficheur G, et al. A generic method for improving the spatial interoperability of medical and ecological databases. *Int J Health Geogr* 2017 ; 16(1) : 36.
4. Walker E, Leclerc M, Rey JF, Beaudouin R, Soubeyrand S, Messéan A. A spatio-temporal exposure-hazard model for assessing biological risk and impact : spatio-temporal exposure-hazard model for assessing biological risk and impact. *Risk Anal* 2017.
5. Burgass MJ, Halpern BS, Nicholson E, Milner-Gulland EJ. Navigating uncertainty in environmental composite indicators. *Ecol Indic* 2017 ; 75 : 268-78.

Pour en savoir plus

Cartographies web

www.sirsepaca.org