

## Chapitre 4

# Cartographie des accidents de la circulation

*Florence Huguenin-Richard*

La gestion du risque routier au niveau d'une ville passe de plus en plus par une phase de « territorialisation » des accidents de la circulation, en somme par l'établissement d'une étude d'enjeux en matière de sécurité routière – la mise en carte des accidents pouvant permettre d'identifier des lieux dangereux et d'établir après observations sur le terrain un diagnostic de sécurité où des facteurs de risque sont clairement identifiés.

La constitution d'une base de données spatiales – précises et fiables – sur les accidents, son intégration au sein d'un système d'information géographique parfois spécifiquement dédié à la problématique<sup>1</sup>, et plus encore l'utilisation de méthodes cartographiques appropriées sont les garants d'une bonne identification des lieux dangereux.

Ainsi, nous répondrons dans ce chapitre aux deux points suivants :

– le premier concerne la constitution d'une base de données géographiques sur les accidents de la circulation : quelles données intégrées ? Où les trouver ? Comment localiser avec précision les accidents ? Comment représenter sur une carte les accidents ?

– le second s'intéressera aux méthodes d'identification des lieux dangereux : comment estimer le risque routier ? A quelles échelles géographiques observer ce risque ? Quelles méthodes d'analyse et quels modes de représentation cartographique ?

---

<sup>1</sup> Par exemple *Concerto*, un logiciel développé par le SETRA et le CERTU de traitement et d'exploitation du fichier national sur les accidents

#### **4.1. Constitution d'un système d'information géographique sur les accidents**

La première difficulté pour constituer un système d'information géographique sur les accidents réside dans la constitution même de la base de données, qui permettra de cartographier finement les accidents survenus et de les replacer dans leur contexte géographique. La localisation précise des accidents à partir du fichier de recensement existant au niveau national est un délicat problème qui passe bien souvent par une étape de géocodage comme nous allons le voir.

##### **4.1.1. Source des données sur les accidents de la circulation**

En France, chaque accident corporel fait l'objet d'un procès-verbal dressé par la police en milieu urbain, par la gendarmerie en milieu rural et dans les agglomérations de moins de 7000 habitants environ. Les procès-verbaux ont pour première finalité d'établir les responsabilités juridiques. À partir de ces documents, les forces de l'ordre constituent des fichiers codés et rendus anonymes – les Bulletins d'Analyse des Accidents Corporels de la Circulation (BAAC) – qui seront centralisés au SETRA tous les mois et qui viendront alimenter le fichier national de suivi de sécurité routière (les procès-verbaux restant confidentiels). Les accidents matériels ne font pas l'objet d'une collecte systématique.

Un BAAC se compose de quatre niveaux qui renseignent chacun des composantes de l'accident, soit au total environ soixante variables recueillies :

- les caractéristiques de l'événement (localisation, date, heure, luminosité, type de collision, conditions météorologiques, type d'environnement, etc.) ;
- le lieu (catégorie de la voie, type de carrefour, régime de circulation, état de la route, aménagement de l'infrastructure, signalisation, etc.) ;
- les véhicules impliqués (type, catégorie administrative, date de la première mise en circulation, date du dernier contrôle technique, chargement, manœuvre principale avant le choc, etc.) ;
- les usagers impliqués lorsqu'ils ont nécessité un traitement médical (type d'usager, gravité des blessures, place de l'usager dans le véhicule, âge, sexe, catégorie socioprofessionnelle, taux d'alcoolémie, motif du trajet, utilisation ou non des équipements de sécurité, manœuvre du piéton, etc.).

Plusieurs critiques peuvent être portées sur la nature et la qualité de l'information délivrée par les BAAC. En effet, il est reconnu que le recensement des accidents n'est pas exhaustif [FLE 98]. Un certain nombre échappe aux statistiques notamment ceux qui impliquent un seul véhicule ou un usager vulnérable. Par ailleurs, cette information est avant tout orientée vers une analyse épidémiologique du risque. Son mode d'acquisition résultant d'un codage sur ordinateur après lecture

du procès-verbal peut poser problème : certains renseignements portés dans les procès-verbaux manquent parfois d'exactitude, d'autres rubriques ne sont pas toujours bien remplies. Enfin la localisation exacte de l'événement est un point délicat. De manière courante, le département et le code INSEE de la commune sont renseignés. La position précise de l'accident, quant à elle, est exprimée en milieu rural par une évaluation de sa distance au point de repère le plus proche. En milieu urbain, on utilise l'adresse postale la plus proche, ce qui limite la précision de l'information dans les zones faiblement urbanisées. C'est cette dernière variable qui nécessite généralement le plus de corrections avant une exploitation des données, demandant bien souvent une relecture des procès-verbaux.

#### **4.1.2. Nécessaire géocodage des données accidents**

L'opération de géocodage (fonction présente sur l'ensemble des systèmes d'information géographique courants), s'effectue à partir d'un fichier « accidents » corrigé où l'un des attributs du tableau contient une adresse postale et d'un référentiel linéaire composé des rues. Pour chacun des tronçons composant une rue (communément, un tronçon est une portion de rue entre deux carrefours), le nom de la rue est référencé ainsi que les numéros des adresses aux deux extrémités. Chaque objet ponctuel créé sera localisé sur le tronçon, en fonction de son adresse et de la longueur de la rue selon une règle de proportionnalité simple – sous *Concerto* tout du moins. Le système récupérera *a posteriori* les coordonnées géographiques en x et y de chaque accident. A partir de là, il est possible de cartographier un semis de points.

Des problèmes de géocodage peuvent survenir dès lors que certaines adresses portées dans le fichier des accidents ne peuvent pas être reconnues directement par le système de géocodage, car soient absentes du référentiel géographique, soient erronées. C'est pourquoi il est nécessaire de contrôler la qualité des informations recueillies de manière à minimiser le nombre d'accidents qui ne sera pas automatiquement géocodé. Par ailleurs, la représentation cartographique d'un réseau routier étant complexe (certains tronçons peuvent être représentés sur la carte par plusieurs polygones, comme un boulevard composé de deux chaussées distinctes), le logiciel ne pourra pas localiser directement l'accident, ce qui nécessite soit de simplifier le réseau lors de sa cartographie, soit de créer un système d'adressage plus précis.

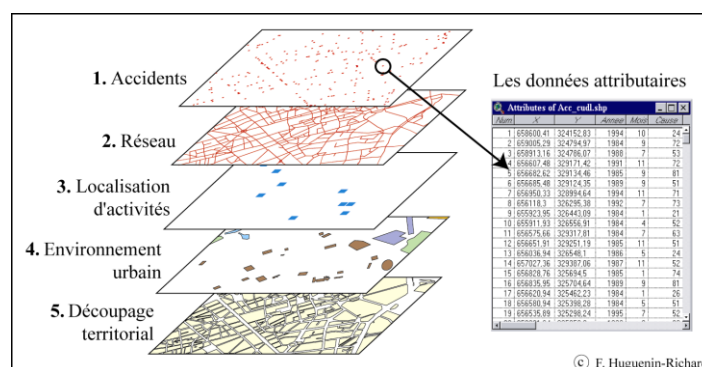
#### **4.1.3. Constitution d'une base de données sur le risque routier**

La description du contexte environnemental dans lequel l'accident s'inscrit nécessite d'enrichir cette base de données avec des informations sur le réseau de

circulation (géométrie, caractéristiques), sur l'aménagement urbain en général (localisation d'éléments comme les écoles, les stations de transport en commun, les espaces verts, le bâti...), et de mesures sur l'exposition au risque d'accidents (telles que le trafic routier, la population habitante, les flux piétonniers). Ce qui représente au final une importante masse de données assez hétérogènes et pas toujours évidentes à obtenir !

La structuration des données en différentes couches thématiques revient implicitement à construire un modèle de type cartographique (cf. Figure 1). Nous proposons l'exemple d'une base de données composée de cinq thèmes principaux :

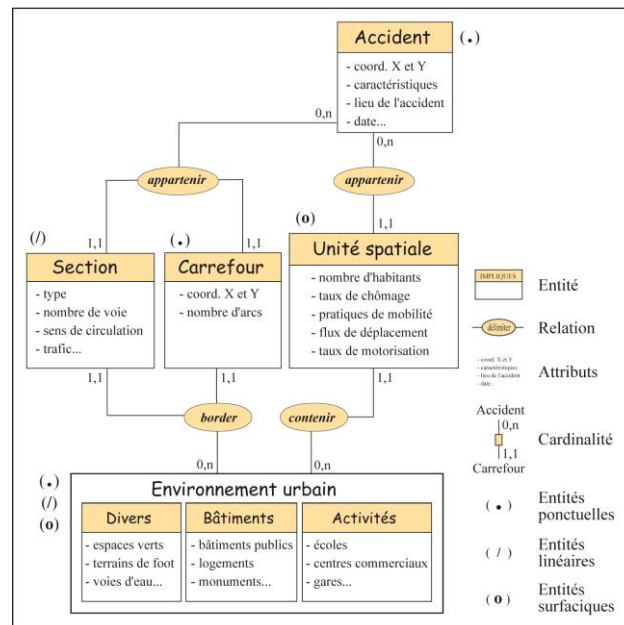
- les accidents ;
- le réseau de circulation ;
- la localisation d'activités génératrices de flux (écoles, hypermarchés, gares) ;
- l'environnement urbain (parcs et espaces verts, monuments, trame du bâti) ;
- des découpages territoriaux auxquels sont associés des données sur les habitants, la mobilité, les flux, etc.



**Figure 1.** Présentation du modèle cartographique

Cet agencement relève d'un modèle de base de données relationnelle (cf. Figure 2). Dans ce cas, chaque thème est associé à un fichier alphanumérique où tous les objets géographiques référencés sur le plan correspondent à un enregistrement, c'est-à-dire à une ligne dans la table. Par conséquent, l'information est consignée dans deux champs distincts :

- la « géobase » qui localise et donne la forme géométrique de chaque objet ;
- les fichiers de données attributaires.



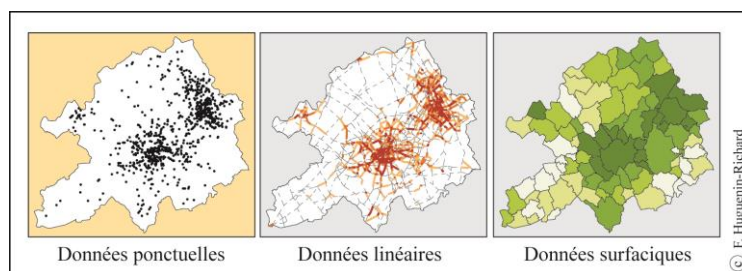
**Figure 2.** Schéma conceptuel de la base de données géographiques

#### 4.1.4. Changer d'échelles pour observer le risque routier

Si la constitution d'une telle base de données constitue un travail long et délicat, elle permet au final une large exploitation en analyse spatiale du fait de la particularité et de la richesse de ces données. Partant de données ponctuelles (le semis de points des accidents), c'est-à-dire désagrégées, il est possible de mener des études à différentes échelles – du local au plus global – suivant que l'on agrège ou pas les données accidents à d'autres entités spatiales telles qu'un tronçon de route, un quartier... (cf. Figure 3).

L'accident est un événement local mais qui s'inscrit dans les multiples niveaux d'organisation d'une ville : celui de la route ou de la rue, celui du quartier, celui du plan de déplacement urbain, etc. De ce fait, les causes de répartition des accidents sur un territoire relèvent probablement de facteurs qui ressortent de ces échelles différentes. À un niveau local, un tronçon peut être dangereux du fait de la proximité d'une école ; au niveau global, le danger sur un tronçon équivalent peut être expliqué par un report de trafic automobile. En réponse, l'analyse du risque routier doit être menée en tenant compte de cette réalité multi-échelle. Il faut donc pouvoir naviguer d'une échelle à l'autre pour saisir les différents niveaux de structure du risque routier, ce que les systèmes d'information géographique permettent aisément

par des fonctions de sélection et d'agrégation spatiale. La configuration des lieux dangereux dépendra de fait de l'échelle à laquelle est menée l'analyse et du type d'analyse (globale ou locale).



**Figure 3.** Type de données graphique et statistique

#### 4.1.5. Définition des lieux à risque

A chaque niveau d'appréhension du risque correspond un type d'unité spatiale élémentaire (cf. Tableau 1). Ces unités spatiales sont la définition des « lieux d'accidents ». De formes et de tailles différentes, il peut s'agir d'entités surfaciques, linéaires ou ponctuelles. Ainsi à petite échelle, l'unité spatiale peut être la commune, le tronçon d'un grand axe de communication ; à grande échelle, elle devient le quartier, la section de voie, le carrefour, la maille d'une grille. Une attention particulière devra être portée au choix de la taille et de la forme des unités spatiales qui induisent des contraintes au niveau des méthodes statistiques employées. En sécurité routière, les unités les plus utilisées sont le carrefour et le tronçon de voie. Même si tous les niveaux d'agrégation sont en principe envisageables, le choix de l'échelle de l'étude et des individus statistiques associés est souvent conditionné par la disponibilité des données sur l'exposition au risque (population, taux de motorisation, flux, trafic automobile, etc.).

Niveau d'observation	Unités spatiales	Type d'objets
Échelle communale	<ul style="list-style-type: none"> <li>▪ une commune</li> <li>▪ une voie</li> </ul>	<ul style="list-style-type: none"> <li>▪ surfacique</li> <li>▪ linéaire</li> </ul>
Échelle d'une ville	<ul style="list-style-type: none"> <li>▪ un quartier</li> <li>▪ un tronçon de route</li> </ul>	<ul style="list-style-type: none"> <li>▪ surfacique</li> <li>▪ linéaire</li> </ul>
Échelle locale	<ul style="list-style-type: none"> <li>▪ une section de route</li> <li>▪ un carrefour</li> <li>▪ un fenêtré circulaire</li> <li>▪ une maille d'un carroyage</li> </ul>	<ul style="list-style-type: none"> <li>▪ linéaire</li> <li>▪ ponctuel</li> <li>▪ surfacique</li> <li>▪ surfacique</li> </ul>

**Tableau 1.** Echelles géographiques et « lieux à risque »

#### **4.1.6. Les notions de risque et d'exposition au risque d'accident**

Le risque d'accident de la circulation est la probabilité qu'une personne soit victime d'une collision au cours d'un déplacement ou sur une période de temps donnée. Sont exposées au risque d'accident toutes les personnes se déplaçant sur la voie publique quelque soit le mode de transport utilisé ; leur vulnérabilité étant conditionnée par leurs propres caractéristiques (jeune ou vieux, piéton ou automobiliste...).

Pour estimer le risque d'accidents d'un individu ou d'un groupe d'individus, on rapporte l'occurrence de réalisation de l'événement que l'on connaît *a posteriori* (le nombre d'accidents observés) à une mesure de l'exposition de l'individu ou du groupe d'individus pris en compte. Toute la difficulté réside dans le choix de cette mesure, qui peut s'effectuer de différentes manières ; le terme d'exposition ayant été maintes fois débattu dans la littérature [CHA 73] [FON 88] [CAR 93]. Il apparaît que deux mesures sont couramment utilisées pour estimer le risque d'accident au niveau individuel : la distance parcourue ou le temps passé sur la route par l'individu ou un groupe d'individus présentant des caractéristiques similaires.

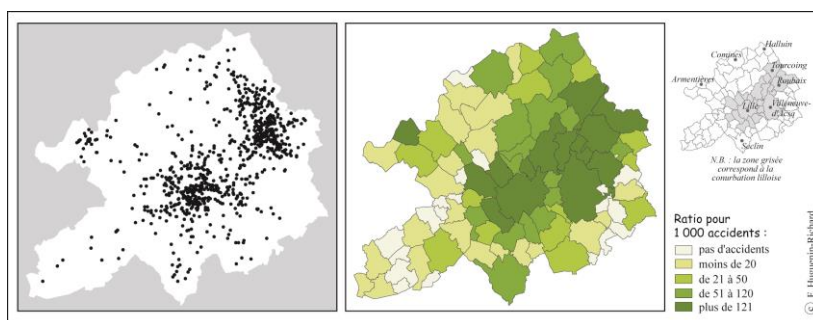
Dans une approche géographique, l'estimation du risque devient collective : ce que l'on cherche à identifier ce sont les lieux présentant un niveau de risque anormal. Des mesures comme la population habitante, le taux de motorisation, la consommation de carburant seront utilisées lors d'études comparatives à l'échelle d'un pays ou d'une région ; des estimations de flux issus d'enquêtes de mobilité, de trafic automobile sur des sections de route permettront une meilleure estimation du risque à des échelles plus fines (au niveau d'un quartier, du réseau de circulation).

En définitive, le choix de la mesure d'exposition au risque dépend du type d'investigation (approche épidémiologique ou géographique), de l'échelle de l'étude et plus prosaïquement de la disponibilité et de la pertinence des informations requises (par exemple, la mesure de l'exposition au risque d'accidents des piétons correspond aux nombre de piétons circulant dans une rue et non au trafic automobile).

### **4.2. Identifier les lieux dangereux : proposition de méthodes cartographiques**

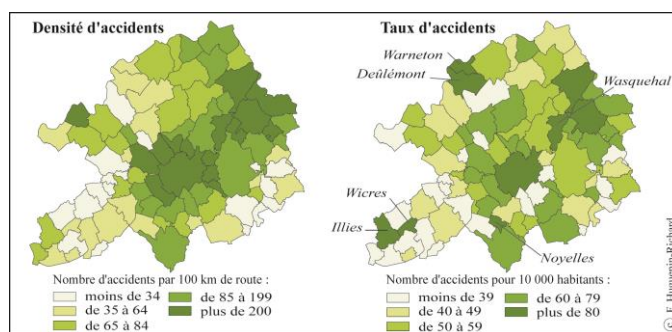
La détection des lieux dangereux sur un territoire donné constitue la seconde étape lors d'une étude d'enjeux de sécurité routière. Une première approche consiste à repérer visuellement sur la carte les zones d'accumulation d'accidents, exprimées de manière brute par leur effectif (cf. Figure 4). On retrouve là toute la littérature de l'accidentologie sur l'identification des « points noirs » dans un réseau de circulation

notamment, correspondant aux zones qui présentent un nombre d'accidents supérieur à un seuil jugé critique [FLE 98].



**Figure 4.** Cartes du semis de points et carte de la proportion d'accidents par commune

A des fins cartographiques, il est fortement recommandé de calculer au moins des densités d'accidents (cf. Figure 6). En effet, cela permet de standardiser les données « accidents » recensées la plupart du temps dans des unités spatiales très hétérogènes. Travailler sur des données brutes aurait alors pour conséquence d'induire des biais statistiques bien connus – l'effet de surface et de taille [HAG 73] [CHA 95] – qui masquent la véritable structure de répartition géographique du phénomène.



**Figure 5.** Cartes des densités d'accidents et du taux d'accidents par commune

Dans le cas où l'on dispose de mesures sur l'exposition au risque fiables et pertinentes, on préférera alors la cartographie de taux de risque (cf. Figure 5)). Cependant, les cartes de taux apparaissent bien souvent très « bruitées » et de fait, délicates à interpréter.



Afin de répondre à ces problèmes d'ordre méthodologique (l'absence de mesures suffisamment précises sur l'exposition au risque et la détection de lieux dangereux qui ne le sont peut-être pas), nous proposons des méthodes cartographiques issues de l'analyse spatiale. Elles apparaissent être un recours intéressant pour améliorer le diagnostic. Il s'agit des estimés empirico-bayésiens reposant sur un principe de lissage spatial, et de la recherche par fenêtre circulaire mobile de concentrations locales anormales d'accidents dans leur semis de points.

#### **4.2.1. Cartographie de taux de risque relatif par les estimés empiriques bayésiens**

Les méthodes empiriques bayésiennes permettent de lisser spatialement les indices de risque tels que la densité ou un taux d'accidents, pris par un ensemble d'unités spatiales dont les valeurs d'exposition au danger sont très hétérogènes. Elles calculent une nouvelle estimation appelée risque relatif. L'application de ces méthodes est prise en compte en épidémiologie, notamment pour cartographier l'incidence ou la mortalité par maladie [COL 97] [MOL 91] [RIC 92].

L'idée sous-jacente est de lisser les différences spatiales d'un indice de risque en intégrant localement la situation globale observée par l'ensemble des unités spatiales de la zone étudiée. Ces méthodes sont ainsi à la croisée des techniques plus simples à mettre en œuvre : le quotient de localisation (Pumain, Huguenin-Richard) et le lissage spatial, mais moins appropriées à la problématique.

Lors du calcul d'un niveau de risque par des méthodes que nous qualifierons de classique (ratio, densité ou taux d'accidents), certaines unités spatiales enregistrent parfois des valeurs excessives ou aberrantes compte tenu de la faible importance de la mesure de l'exposition au danger sur laquelle repose l'estimation (population résidente, longueur de réseau). La pertinence des résultats peut être mise en doute. Prenons comme exemple celui de la commune de Warneton (cf. Figure 5), peuplée de 179 habitants, qui enregistre quatre accidents entre 1993 et 1997 mais qui présente un taux de risque observé par rapport à la population de 22 ‰, alors qu'il n'est que de 10 ‰ à Lille (17 2142 habitants, 1 844 accidents).

Lorsqu'un excès de risque est mesuré à partir d'un nombre important de cas, la fiabilité de l'information est renforcée, étant donné que la variance de cette valeur est moins importante par rapport à l'ensemble des valeurs prises dans la zone étudiée. De ce fait, l'interprétation d'une carte de risque est liée à la lecture :

- de l'information sur la mesure de l'exposition au danger ;
- des valeurs prises par les unités spatiales environnantes ;
- et de la situation observée sur l'ensemble du territoire.

Le principe fondamental des méthodes bayésiennes appliquées à l'estimation des risques relatifs est le suivant : « *plus une information individuelle est stable, c'est-à-dire plus la variance de l'estimateur de risque associé est faible, plus cette information est crédible* » [COL 97]. En revanche, moins une information individuelle est stable, moins elle est fiable. Sa modification par rapport aux valeurs de risque estimées dans d'autres unités spatiales, qui selon une hypothèse faite *a priori* présentent une exposition au risque similaire, est envisageable et recommandée.

#### 4.2.1.1. Quelques éléments sur les modèles probabilistes bayésiens

Dans une démarche classique d'estimation du risque, seul le nombre de cas observés est considéré. Une démarche bayésienne suppose que les observations résultent de deux mécanismes qui peuvent être modélisés ensemble [COL 97] :

- en premier lieu, les valeurs de risque relatif fluctuent en fonction des valeurs prises par les mesures d'exposition au risque, modélisés par une loi de paramètres ( $\gamma$ ,  $\theta$ ), appelée loi *a priori* ;
- le nombre de cas observés est la réalisation d'une variable aléatoire de Poisson  $O$ , dont l'occurrence est modélisée par une loi conditionnelle  $F(O/p)$  ;
- enfin, l'estimation bayésienne des risques relatifs est basée sur l'espérance de la loi  $p(p/O)$ , appelée loi *a posteriori*.

Le lien entre ces différentes lois est réalisé en s'appuyant sur le théorème de Bayes relatif aux probabilités conditionnelles.

Une première méthode consiste à utiliser une estimation du paramètre ( $\theta$ ) obtenue directement à partir de l'information connue par la distribution de  $O$ . En fait, il s'agit d'utiliser l'information disponible (le nombre de cas observés) afin d'obtenir une estimation du paramètre de la loi *a priori*. Cette démarche est appelée méthode bayésienne empirique. La connaissance *a priori* se fait uniquement à partir des observations. En l'absence de connaissance du paramètre de la loi *a priori*, la seconde approche consiste à tenir compte de l'incertitude existant sur ce paramètre en faisant intervenir une nouvelle distribution  $q(\theta ; \zeta)$ , appelée loi *hyper a priori*. L'estimation des risques relatifs devient alors beaucoup plus complexe, mais présente l'avantage d'obtenir des résultats par intervalle. Ces méthodes sont appelées méthodes bayésiennes « pures ».

#### 4.2.1.2. Présentation de l'estimateur des risques relatifs par une méthode bayésienne empirique

Pour chaque unité spatiale de la zone étudiée, on suppose que l'on connaisse une distribution *a priori* de moyenne ( $\gamma_i$ ) et de variance ( $\theta_i$ ). L'estimation du risque,

appelée risque relatif ( $RR_i$ ), fondée sur une combinaison de la distribution *a priori* et du risque observé se calcule comme suit. Soit  $i$  une unité géographique :

$$RR_i = w_i r_i + (1 - w_i) \gamma_i$$

où :

- $r_i$  est le risque observé par un estimateur classique ;
- $\gamma_i$  le risque moyen dans les unités spatiales dont on suppose que  $i$  ne diffère pas ;
- $w_i$  un système de pondération propre à l'individu  $i$ .

La fonction  $w_i$  dépend de la population exposée au risque dans chaque unité spatiale ( $n_i$ ) et de la variance de la distribution *a priori* ( $\phi_i$ ). Elle reflète la confiance que l'on accorde aux connaissances *a priori* du risque. Le principe est que lorsque la population exposée est grande,  $w_i$  tend vers 1 : par conséquent, l'estimation bayésienne empirique du risque est proche de l'estimation classique du risque. Dans ce cas, on a davantage confiance dans le risque observé. En revanche, lorsque la mesure d'exposition au risque est petite, l'estimation classique est d'autant plus instable. L'estimateur bayésien empirique du risque relatif influencé par la valeur moyenne ( $\gamma_i$ ) apporte une meilleure information.  $w_i$  vaut :

$$w_i = \frac{\phi_i}{\left( \phi_i + \frac{\gamma_i}{n_i} \right)}$$

L'estimation de ( $\gamma_i$ ) et ( $\phi_i$ ) représente un gros travail de calibrage. Une solution plus simple consiste à utiliser l'information disponible (le nombre de cas observés) pour paramétrer les deux estimateurs de la loi *a priori* par application de la méthode des moments :

- $\hat{\gamma}_i$  est une estimation du paramètre ( $\gamma_i$ ), qui correspond à un risque moyen où  $y_i$  est le nombre de cas observés et  $n_i$  la population exposée au risque :

$$\hat{\gamma} = \frac{\sum y_i}{\sum n_i}$$

- $\hat{\phi}_i$  est une estimation du paramètre ( $\phi_i$ ), calculée comme suit :

$$\hat{\phi} = \frac{\sum n_i (r_i - \hat{\gamma})^2}{\sum n_i} - \frac{\hat{\gamma}}{\bar{n}}$$

– de ce fait, la valeur du poids  $w_i$  est également approchée par  $\hat{w}_i$  :

$$\hat{w}_i = \frac{\hat{\phi}}{\left( \hat{\phi} + \frac{\hat{\gamma}}{n_i} \right)}$$

– finalement, le risque relatif dans l'unité spatiale  $i$  est estimé par la formule suivante :

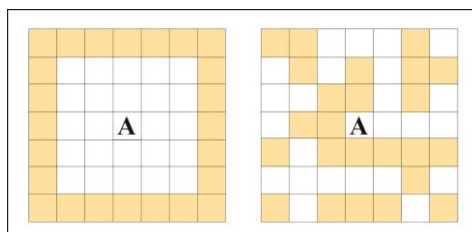
$$RR_i = \hat{\gamma} + \frac{\hat{\phi}(r_i - \hat{\gamma})}{\left( \hat{\phi} + \frac{\hat{\gamma}}{n_i} \right)}$$

Deux hypothèses peuvent être faites *a priori* sur la similarité entre unités géographiques :

– *a priori* l'exposition au risque ne varie pas de manière systématique parmi l'ensemble des unités géographiques. La variation des risques est donc modélisée par une distribution de probabilité globale ou lissage global ;

– en intégrant une information spatiale, on postule que l'exposition au risque d'une unité spatiale donnée ne varie pas systématiquement par rapport aux unités proches géographiquement. Ce lissage par méthode empirique bayésienne est de type local. Les paramètres  $(\gamma_i)$  et  $(\phi_i)$  de la loi *a priori* sont calculés localement en fonction du voisinage de chaque unité spatiale.

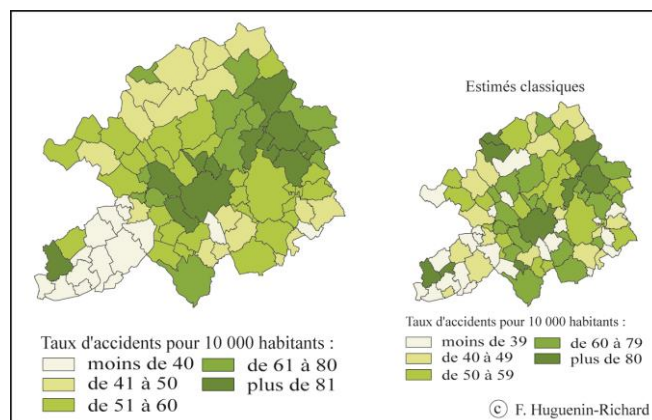
Le choix d'un lissage global ou d'un lissage local conditionne directement les résultats. Colonna [COL 97] montre comment la prise en compte de la position géographique des individus affecte les résultats lors d'un lissage. Les deux cartes de la figure 6 représentent deux répartitions spatiales différentes d'individus à faible et à fort risque d'accident. Un lissage spatial global appliqué pour ces deux structures conduit à une estimation bayésienne empirique du risque relatif en A identique dans les deux cas. En effet avec un lissage global, A est influencée de la même manière par l'ensemble des individus quelle que soit leur répartition spatiale. En revanche, un lissage spatial local fournit des résultats différents selon la localisation géographique des individus les uns par rapport aux autres. Dans le premier cas, l'estimation bayésienne empirique tend vers une valeur de risque relatif faible en A. Dans le second, l'estimation bayésienne empirique tend vers une valeur de risque relatif fort en A.



**Figure 6.** Intérêt d'un lissage local selon la situation des individus

#### 4.2.1.3. Application à l'estimation du risque d'accident par commune

Ce principe de lissage spatial a été appliqué pour calculer un taux relatif de risque d'accidents à partir de la distribution *a priori* du nombre d'habitants. De manière générale, les résultats obtenus par le lissage spatial global sont assez proches de ceux obtenus par le calcul classique du taux. En effet, la technique du lissage global faisant partager la distribution totale de la variable de référence par l'ensemble des individus, conduit à une répartition plus hétérogène des estimations de risque relatif sur le territoire. La distribution de cette variable présente en fait une très forte dispersion statistique, qui affecte les résultats. Warneton est toujours estimée comme très dangereuse. En revanche, l'estimation bayésienne empirique par lissage local conduit à la production de cartes de risque relatif d'accidents beaucoup moins « bruitées » (cf. Figure 7).

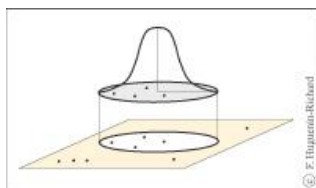


**Figure 7.** Estimations empiriques bayésiennes du risque relatif d'accidents par lissage local en fonction du nombre d'habitants par commune

Finalement, les méthodes empiriques bayésiennes ne renseignent pas directement sur la dangerosité des sites étudiés, mais elles montrent la confiance que l'on peut accorder à l'estimation d'un risque. Lorsqu'une commune est détectée comme dangereuse par un estimateur classique, sa comparaison avec l'estimé de risque relatif calculé à partir des mêmes paramètres permet de juger de la fiabilité de l'observation. Par exemple, sur la carte des taux d'accidents, différentes communes enregistrent un fort risque : Lille, Roubaix, Tourcoing, Wasquehal, Warneton, Illies, Wicres, Noyelles, Deülémont. Certaines d'entre elles sont de très petites communes (Warneton, Wicres). Avec les estimés de risque relatif, Wicres n'apparaît plus comme une commune dangereuse ; Warneton le reste mais dans une moindre mesure uniquement.

#### 4.2.2. *La recherche de concentrations locales anormales d'accidents dans leur semis de points*

La manière la plus simple de cartographier les accidents est de les représenter à travers un semis de points (cf. Figure 3). Chacun d'entre eux identifie la position exacte d'un ou de plusieurs événements. Ce type de document de conception aisée est toutefois difficile à lire et à interpréter en raison d'une saturation de l'information. Cependant, l'avantage du semis de points – se référant à toute la collection des accidents dans un référentiel spatial unique (coordonnées en X et Y) – est que la même base d'information peut être traitée à tous les niveaux d'échelle. Ainsi, l'intégration peut se faire par commune, quartier ou tronçon, mais aussi par d'autres objets spatiaux moins familiers tels que des fenêtres circulaires mobiles (cf. Figure 8). À partir de ces dernières, des méthodes spécifiques peuvent être conduites : elles apporteront des solutions opérationnelles pour diagnostiquer les lieux à risque à travers une prise en compte individuelle des accidents.



**Figure 8.** *Une fenêtre circulaire mobile tri-dimensionnelle<sup>2</sup>*

La méthode que l'on envisage maintenant est issue de la « machine d'analyse géographique » développée au début des années quatre-vingt par

<sup>2</sup> A ce sujet, de plus amples explications sur le site suivant : <http://web.univ-pau.fr/~banos/banos-densités.htm>

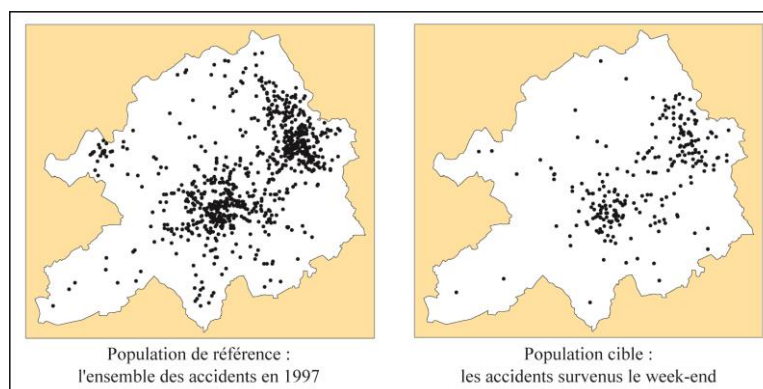
S. Openshaw *et al.* [OPE 87]. Elle intègre un très grand nombre de calculs locaux et originaux et teste de nombreuses hypothèses spatiales. Cette « machine » a largement été diffusée : [GOU 92], [MAR 91], [OPE 95], [THO 91], [BAN 00] et [BAN 01].

La recherche de concentrations locales d'accidents, aussi appelées par l'auteur « points chauds », a pour objectif de détecter des agrégats spatiaux anormaux et excessifs d'événements dans le semis de points, agrégats impossibles à identifier par simple lecture. Cette méthode offre un mode opératoire de diagnostic [BAN 99]

#### 4.2.2.1. Principe de la méthode<sup>3</sup>

Le principe consiste à comparer localement la distribution d'une sous-population d'événements ponctuels (la population cible) avec la distribution spatiale d'une population de référence (la population à risque). On estime que la répartition de la population cible ne devrait pas être très différente de celle de la distribution de référence, sauf si des problèmes de sécurité existent. La méthode propose de tester cette hypothèse par l'application d'un test statistique. Il repose sur la loi de Poisson et permet de mesurer si les écarts sont significatifs.

La figure 9 montre la répartition d'une population à risque (l'ensemble des accidents survenus sur une période de temps donnée) et celle d'une sous-population extraite par requête thématique (la population cible). Les distributions spatiales semblent obéir aux mêmes règles de répartition.



<sup>3</sup> A ce sujet, de plus amples explications sur le site suivant : <http://web.univ-pau.fr/~banos/banos-gam.html>

**Figure 9.** Exemple d'une population de référence et d'une population cible

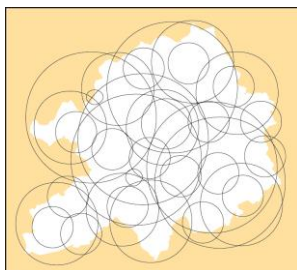
La méthode initiée par S. Openshaw a été spécifiquement adaptée à un semis de points par S. Fotheringham [FOT 96]. L'algorithme original, qui intégrait des informations agrégées à des entités surfaciques relatives à la population à risque (par exemple la population résidente par unité administrative), a été optimisé par ces chercheurs en utilisant une nouvelle technique de fenêtrage. Les résultats restent très proches de ceux obtenus par S. Openshaw.

#### 4.2.2.2. Technique de fenêtrage aléatoire

La zone étudiée est couverte de fenêtres mobiles circulaires avec les règles suivantes :

- le nombre de fenêtres est fixé par l'utilisateur ;
- la localisation est définie au hasard dans la zone d'étude (tirage avec remise) ;
- le rayon, variable, est choisi au hasard dans un intervalle fixé au préalable par l'utilisateur.

Le parcours non systématique assure une bonne couverture de l'espace avec un nombre restreint de fenêtres (Fig. 10). Le paramétrage est en contrepartie plus délicat, puisqu'il est nécessaire d'ajuster conjointement le nombre de fenêtres et leur rayon minimal et maximal.



**Figure 10.** Parcours de la zone d'étude par la technique de fenêtrage aléatoire

#### 4.2.2.3. Présentation de l'algorithme

Dans un premier temps (Fig. 11), on calcule l'occurrence moyenne d'un événement parmi la population cible ( $i$ ) dans la zone d'étude :

$$P_i = \frac{\text{effectifs de la population cible}}{\text{effectifs de la population de référence}}$$



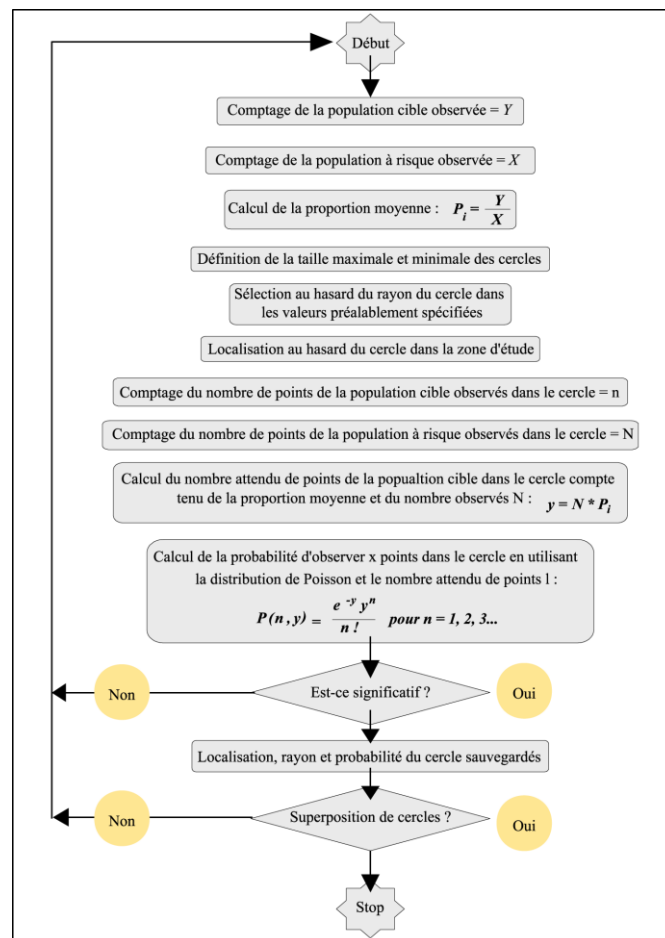
Pour chaque fenêtre, les événements de la population de référence ( $N$ ) et les événements de la population cible ( $n$ ) sont dénombrés. Puis connaissant ( $P_i$ ), on calcule le nombre théorique d'événements de la population cible attendu ( $y$ ), sous l'hypothèse d'une distribution aléatoire :

$$y = N * P_i$$

Puis, l'écart entre le nombre observé ( $n$ ) et le nombre attendu d'événements de la population cible ( $y$ ) est testé par application de la loi de Poisson :

$$P(n, y) = \frac{e^{-y} * y^n}{n!}$$

Enfin, chaque fenêtre qui enregistre un niveau de probabilité inférieur à celui qui est préalablement fixé par l'utilisateur est dessinée sur le semis de points. Les niveaux les plus stables sont ceux pour qui la probabilité est la plus faible. L'accumulation de fenêtres dessinées indique une concentration anormale d'événements de la population cible, c'est-à-dire une zone particulièrement

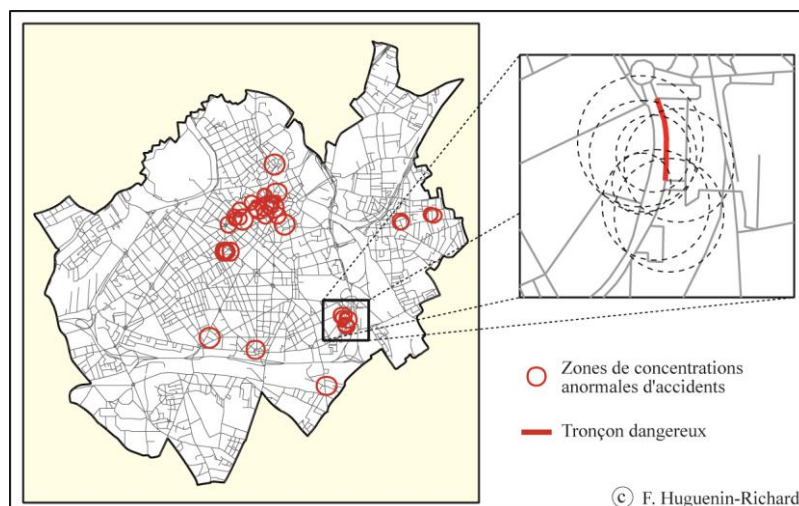


dangereuse pour cette sous-population.

**Figure 11.** Schéma de l'algorithme d'après [FOT 96]

#### 4.2.2.4. Exemple d'application dans la ville de Lille

Cette méthode d'identification cartographique de concentrations anormales d'évènements dans un semis de points a été appliquée au cas des accidents impliquant un piéton adulte dans la ville de Lille. Différents agrégats ont été détectés, figurant en rouge sur la carte 12. En superposant la trame du réseau de circulation à la cette cartographie de « points chauds », il est possible d'identifier avec précision le tronçon de route sur lequel ces accidents anormalement concentrés ont eu lieu.



**Figure 12.** Détection de concentrations spatiales anormales d'accidents de piétons adultes dans la ville de Lille (accidents recensés entre 1993 et 1997)

Nous pourrions multiplier les exemples d'application de cette méthode à partir d'autres thématiques en sécurité routière. Des tests de fiabilité ont été réalisés afin d'être sûr de la validité des résultats obtenus. Nous avons recherché pour cela des concentrations spatiales locales anormales dans des sous-populations d'accidents issues d'un échantillonnage aléatoire de points. La méthode a été répétée sur des populations cibles test afin de vérifier la stabilité des résultats [BAN 99]. Ces contrôles empiriques ont validé la méthode qui maintenant est opérationnelle. Son couplage avec un système d'information géographique permet d'identifier directement au sein des bases de données les objets et les sites concernés.

Cependant, la méthode procédant par agrégation peut conduire à associer, dans la même itération du calcul, des accidents survenus sur des tronçons non contigus. De ce fait, les résultats présentent un biais au stade actuel. D'où la nécessité d'un contrôle par l'utilisateur. L'adaptation de la technique de fenêtrage à l'espace du réseau de circulation permet d'intégrer une contrainte de contiguïté toutefois les temps de calcul impliqués rendent la généralisation de cette approche encore problématique [BAN 01]<sup>4</sup>.

## Conclusion

Au final, il existe de nombreuses méthodes issues du champ de l'analyse spatiale pour cartographier le risque routier et détecter des lieux dangereux. Nous aurions pu aussi présenter l'indice local d'autocorrélation spatiale, qui appliqué à des données « accidents » agrégées au niveau du réseau de circulation, a montré sa pertinence et sa capacité pour identifier des tronçons dangereux sur de grands axes de circulation. On pourra pour cela se référer aux travaux suivants : [HUG 99] et [FLA 01] ; ou encore la méthode des noyaux qui permet de cartographier des densités locales d'accidents [BAN 99] et [HUG 00].

Le choix de la méthode à mettre en œuvre va dépendre de plusieurs facteurs :

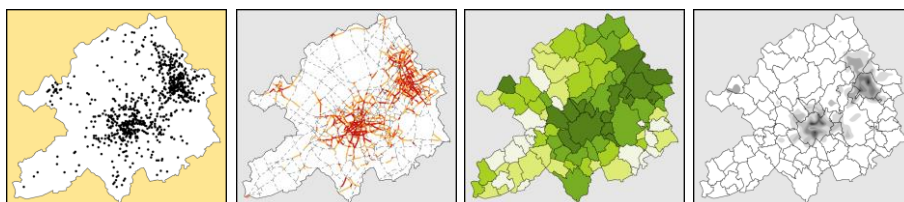
- la précision de la localisation géographique des accidents ;
- la manière dont ils sont représentés au niveau cartographique (sous forme d'un semis de points ou agrégés à des unités spatiales surfaciques ou linéaires, régulières ou irrégulières) ;
- des mesures dont on dispose sur l'exposition au risque, utiles à l'estimation d'un taux de risque ;
- et de l'échelle à laquelle le diagnostic est réalisé.

En effet, il apparaît tout à fait intéressant d'emboîter les échelles d'analyse lors d'études d'enjeux de sécurité routière sur un territoire. Traditionnellement en accidentologie, on cartographie les accidents sur le linéaire des routes. Cela a pour conséquence, entre autre, de produire des documents cartographiques *quasi* illisibles lorsqu'ils sont réalisés à petite échelle (au niveau d'une agglomération par exemple). En emboîtant les échelles cartographiques et en faisant évoluer le niveau d'agrégation spatiale des accidents en fonction de l'échelle (Fig. 13), il est tout à fait envisageable de poser un premier bilan du risque au niveau des communes d'une agglomération ou des quartiers d'une ville avec la méthodes des estimées empirico-bayésiennes de risque relatif, puis d'affiner l'étude d'enjeux dans les communes ou les

---

<sup>4</sup> Des informations complémentaires sont disponibles sur le site suivant :  
<http://web.univ-pau.fr/~banos/poster2.htm>

quartiers identifiés comme dangereux en réalisant des études au niveau du réseau de circulation à l'intérieur de ces zones dangereuses ou à partir du semis de points des accidents.



**Figure 13.** Détection de concentrations spatiales anormales d'accidents de piétons adultes dans la ville de Lille (accidents recensés entre 1993 et 1997)

## Bibliographie

- [BAN 01] BANOS A. ET HUGUENIN-RICHARD F., « Towards integrated spatial surveillance tools of urban risks : space-time analysis of road accidents in Lille (France) », in *Géographie des risques des transports*, Actes du colloque de Besançon, 2001, pp. 281-292.
- [BAN 00] BANOS A. ET HUGUENIN-RICHARD F., « Spatial distribution of road accidents in the vicinity of point sources. Application to child pedestrian accidents », in *Geography and Medicine*, Éditions Elsevier, 2000, pp. 54-64.
- [BAN 99] BANOS A. ET HUGUENIN-RICHARD F., « Méthode d'identification de concentrations locales d'événements dans un semis de points. Application aux accidents de la route », in *Actes du colloque Quatrièmes Rencontres de Théo Quant*, 1999, pp. 245-253.
- [CAR 93] CARRE J.R., DUVAL H., FONTAINE H. ET LASSARRE S., « Les usagers de la route. Une exposition au risque inégale », in *Données Sociales*, 1993, pp. 283-291.
- [CHA 95] CHARRE J., *Statistique et territoire*, GIP Reclus, 1995.
- [CHA 73] CHAPMAN R., « The concept of exposure », in *Accidents Analysis and Prevention*, 1973, Vol. 5, pp. 95-110.

- [COL 97] COLONNA M., « Estimations bayésiennes empiriques de risques relatifs : principes et exemples d'application », in *Revue d'Epidémiologie et de Santé Publique*, Vol. 45, 1997, pp. 42-149.
- [FLA 01] FLAHAUT B., « L'autocorrélation spatiale comme outil géostatistique d'identification des concentrations spatiales des accidents de la route », in *Cybergéo*, 2001.
- [FLE 98] FLEURY D., *Sécurité et urbanisme. La prise en compte de la sécurité routière dans l'aménagement urbain*, Presses des Ponts et Chaussées, 1998.
- [FON 88] FONTAINE H., « Usage de l'automobile et risque d'accidents », in *Recherche Transport et Sécurité*, n° 20, 1988, pp. 5-12.
- [FOT 96] FOTHERINGHAM S. AND ZHAN B., « A comparison of three exploratory methods for cluster detection in spatial point patterns », in *Geographical Analysis*, Vol. 28, n° 3, 1996, pp.200-218.
- [GOU 92] GOULD P., « Epidémiologie et maladie », in *Encyclopédie de géographie*, Economica, 1992, pp. 949-969.
- [HUG 00] HUGUENIN-RICHARD F., *Approche géographique des accidents de la circulation : propositions de modes opératoires de diagnostic. Application au territoire de la métropole lilloise*, Doctorat de géographie, Université de Franche-Comté, 2000.
- [HUG 99] HUGUENIN-RICHARD F., « Identifier les sites routiers dangereux. Application de méthodes d'analyse utilisant la localisation géographique des accidents », in *Revue internationale de géomatique*, Vol. 9, n° 4, 1999, pp. 471-487.
- [MAR 91] MARSHALL R., « A review of methods for the statistical analysis of spatial patterns of disease », in *Journal of the Royal Statistical Society-A*, Vol. 154, n° 3, 1991, pp. 421-441.
- [MOL 91] MOLLIE A. ET RICHARDSON S., « Empirical bayes estimates of cancer mortality rates using spatial models », in *Statistics in Medicine*, Vol. 10, 1991, pp. 95-112.
- [OPE 95] OPENSHAW S., « Developing automated and smart spatial pattern exploration tools for geographical information systems applications », in *The Statistician*, Vol. 44, n° 1, 1995, pp. 3-16.
- [OPE 87] OPENSHAW S., CHARLTON M., WYMER C. AND CRAFT A., « A smart geographical analysis machine for the automated analysis of point data sets », in *International Journal of Geographical Information Systems*, Vol. 1, n° 4, 1987, pp. 335-358.
- [RIC 92] RICHARDSON S., « Modélisation statistique des variations géographiques en épidémiologie », in *Revue d'Epidémiologie et de Santé Publique*, Vol. 40, 1992, pp. 33-45.
- [THO 91] THOMAS R.W., « Quantitative methods : clines, hot spots and cancer clusters », in *Progress in Human Geography*, Vol.15, n° 4, 1991, pp. 445-455.