

# REPRÉSENTATION SURFACIQUE D'ÉVÉNEMENTS PONCTUELS DISCRETS

## COMPARAISON MÉTHODOLOGIQUE À PARTIR DE LA DISTRIBUTION DES ACCIDENTS DE LA ROUTE DANS LA COMMUNAUTÉ URBAINE DE LILLE

Arnaud BANOS

Jérôme BOLOT

THEMA, UPRESA 6049,  
Université de Franche-Comté

### Résumé

*L'identification visuelle de structures spatiales à partir d'un semis de points est une tâche délicate, pour peu que le nombre d'individus soit élevé. Cet article présente et compare deux méthodes permettant d'améliorer significativement la représentation graphique de données spatiales ponctuelles. La première repose sur un carroyage spatial, tandis que la seconde est basée sur l'estimation d'une fonction de densité de Kernel au sein de fenêtres mobiles. Ces deux méthodes sont appliquées aux accidents de route dans la Communauté urbaine de Lille.*

### Summary

*Spatial point pattern analysis of large data sets involves powerful visualization techniques. This paper intends to present two suitable methods: the first one uses regular grid of quadrats superimposed on the event distribution, while the second one is based on Kernel density estimation technique. Both methods are applied to road accidents analysis in the urban community of Lille (France).*

### Mots-Clés

*Accident de la route - Carroyage - Communauté Urbaine de Lille - Événement ponctuel discret - Fonction de densité par noyau - Semis de points*

### Key words

*Kernel density function - Point pattern analysis - Raster display - Road accident*

Les phénomènes géographiques se manifestent souvent sous la forme d'événements ponctuels discrets [4]. Les accidents de la circulation routière en sont un bon exemple [7]. L'analyse de ce type d'information pose alors le problème de sa représentation. La carte par points de la figure 1 est en effet la transcription visuelle la plus simple et la plus courante d'un semis de points. Sa capacité à mettre en évidence les structures spatiales du phénomène étudié est pourtant limitée: en présence d'une information nombreuse, le regard, même entraîné, de l'analyste peine à discerner des structures significatives.

La représentation surfacique lissée de cette information ponctuelle discrète permet de progresser substantiellement dans ce domaine. Deux méthodes appropriées de ce point de vue sont ici présentées et appliquées à la distribution spatiale des accidents de la circulation dans la Communauté Urbaine de Lille<sup>1</sup> en 1996.

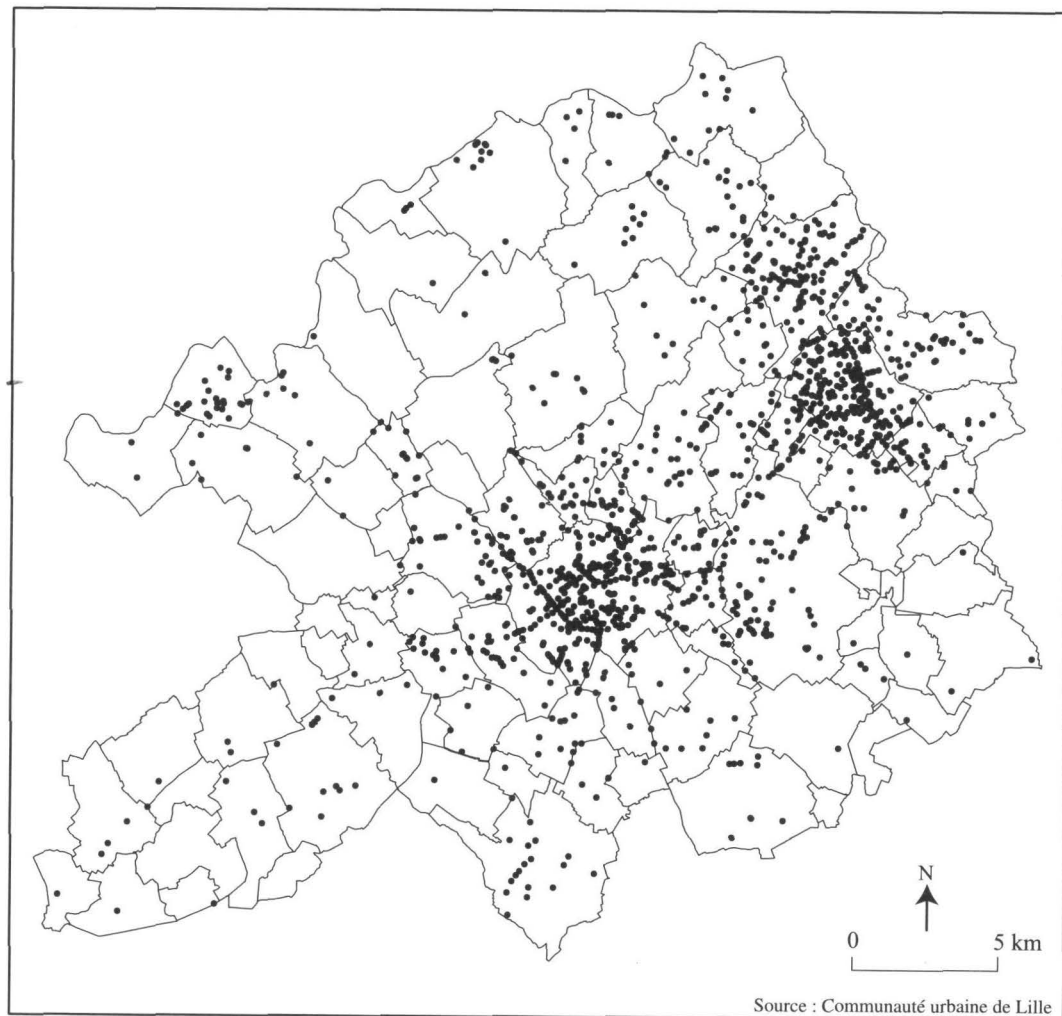


Figure 1 - Les accidents de la circulation dans la Communauté urbaine de Lille en 1996

## 1. Nature du problème

La figure 1 présente l'ensemble des accidents corporels recensés dans la Communauté urbaine de Lille en 1996, soit plus de 1 500. On retrouve évidemment les grandes structures urbaines : Lille, au centre de la communauté urbaine et le bi-pôle Roubaix-Tourcoing, au Nord-Est, constituent les principaux foyers accidentogènes. Il est néanmoins délicat de dépasser ce constat évident. Les zones périurbaines et rurales sont difficiles à qualifier, du point de vue du phénomène étudié, tandis que les pôles urbains sont saturés en information. Autrement dit, cette représentation brute de l'information n'a qu'un intérêt opérationnel limité. Pourtant, l'analyste a besoin d'identifier des structures spatiales significatives. L'idéal serait même qu'il puisse le faire rapidement, à l'issue d'une visualisation brève : ceci supposerait alors de gommer l'information non pertinente (bruit), tout en mettant en valeur les structures spatiales sous-jacentes.

L'agrégation de cette information ponctuelle aux unités spatiales existantes n'a que peu d'intérêt dans cette perspective, puisque cela reviendrait à perdre toute information locale. Par contre, l'idée d'une représentation surfacique de cette information, par nature ponctuelle, est intéressante, à condition de justifier la transition sémantique induite : de quel droit représenter sous forme continue un processus discret ?

Deux arguments peuvent être avancés. En premier lieu, il s'agit d'obtenir une représentation graphique améliorée de l'information initiale, et rien n'interdit d'avoir recours à une technique *a priori* inadaptée, si le

résultat correspond à nos attentes, à la fois du point de vue de la qualité visuelle du rendu et de son interprétabilité. On pressent néanmoins que cet argument libéral possède une limite logique, liée à l'échelle d'analyse. Au niveau de la C.U.D.L., une représentation continue d'événements discrets est acceptable. Par contre, à une échelle plus fine, des contradictions émergent : peut-on réaliser la même manipulation à l'échelle d'un quartier urbain ? La proportion d'espaces non concernés par l'événement "accident", donc illégitimes du point de vue de la méthode, ne devient-elle pas visuellement trop prégnante ?

Nous laisserons cette question épineuse en suspens, afin d'axer notre propos sur la démonstration de l'intérêt de deux méthodes appliquées à la distribution spatiale des accidents de la route

## 2. Présentation des méthodes

### 2.1. Méthode par carroyage

Une première méthode, largement diffusée, permet d'obtenir une représentation surfacique d'un semis de points. Il s'agit de superposer une grille sur le semis de points et d'estimer un nombre d'événements pour chaque entité de cette grille (LAJOIE, 1992). L'utilisation d'un pavage permet de ne retenir que le caractère spatial de l'événement. En effet, la neutralité du découpage géométrique place cet événement en dehors de tout contexte spatial ne le concernant pas directement (cf. fig. 2).

La construction de ce découpage géométrique régulier nécessite la prise en compte de certains paramètres clés. La forme et la taille des entités conditionnent ainsi la pertinence du pavage. Parmi les différentes formes utilisables (carré, triangle, hexagone), l'hexagone présente des avantages indéniables quant à l'indice de forme, mais ces particularités ne sont guère utiles ici. Nous retiendrons donc le carré comme forme de base. Par ailleurs, la taille des entités dépend de la précision de la mesure que l'on souhaite obtenir.

Les événements (accidents) contenus dans chaque entité spatiale sont ensuite dénombrés, et la valeur ainsi obtenue est pondérée par la surface de l'entité correspondante. Une procédure de discrétisation est alors appliquée, permettant d'affecter une couleur à chacune des unités du quadrillage.

La technique dite du quadtree [12] permet d'adapter la représentation graphique surfacique de l'information à sa richesse locale : les carreaux couvrant des zones à forte densité d'événements sont ainsi subdivisés en carreaux de taille inférieure équivalente.

La représentation finale est donc composée de carreaux de différentes tailles. Dans les zones où les événements sont rares et dispersés, des carreaux de taille importante suffisent à résumer l'information, en évitant toute redondance éventuellement produite par un nombre trop important de petits carreaux. Inversement, dans les zones à forte concentration d'événements, les structures spatiales locales sont mieux dégagées par des petits carreaux. "*La structure arborescente de l'information permet de ne travailler, à petite échelle, que sur les données du premier niveau, et d'effectuer des zooms sur certaines parties de la carte, pour lesquelles une information plus détaillée est utile*" [6].

Le quadtree complexifie légèrement la technique du carroyage puisqu'il fait intervenir des entités de tailles variables, mais il apporte un réel avantage par rapport au pavage uniforme<sup>2</sup>. Une seconde méthode, basée sur des fenêtres mobiles, permet de produire de l'information surfacique à partir d'une information ponctuelle discrète.

### 2.2. Fonction de densité de Kernel

D'un point de vue formel, il s'agit d'estimer, en tout point de l'espace, l'intensité du phénomène observé. Pour ce faire, la zone d'étude est balayée systématiquement par une fonction tri-dimensionnelle mobile. Chaque événement rencontré est alors pondéré, en fonction de sa distance au point d'où est estimée l'intensité [1], [5].

Si on note  $X$  une localisation quelconque dans l'espace de la Communauté urbaine de Lille et  $X_1, X_2, \dots, X_n$  les localisations des  $n$  accidents de la circulation recensés en 1996, alors l'intensité  $\lambda(X)$  en  $X$  est estimée par la fonction :

$$\hat{\lambda}(X) = \frac{1}{r^2} \sum_{i=1}^n k\left(\frac{(X - X_i)}{r}\right)$$

avec  $r$  le rayon de la fenêtre mobile et  $k()$  une fonction symétrique, unimodale, bivariée et décroissante, qui affecte à chaque événement  $X_i$  un poids en fonction de sa distance au centre de la fenêtre mobile circulaire. Diverses expressions de  $k()$  peuvent être utilisées. On retiendra ici la formulation bi-carrée :

$$k(d) = \begin{cases} \frac{3}{\pi} (1 - d^2)^2 & \text{si } d \leq 1 \\ 0 & \text{autrement} \end{cases} \quad \text{avec } d = \left(\frac{(X - X_i)}{r}\right)$$

où  $d$  est la distance euclidienne séparant  $X$  de  $X_i$ . La formulation finale de la fonction de densité, en incorporant  $k()$  et en omettant de corriger d'éventuels effets de bord, devient alors :

$$\hat{\lambda}(X) = \frac{3}{\pi r^2} \sum_{d \leq r} \left(1 - \frac{d^2}{r^2}\right)^2$$

Le paramètre le plus important ici est  $r$ , le rayon de la fenêtre mobile : c'est lui qui détermine le degré de lissage de l'information. On pressent d'ailleurs qu'il serait intéressant de disposer de fenêtres mobiles de taille variable, s'adaptant à l'intensité locale du phénomène. Cet ajustement local peut être réalisé à l'aide de la technique dite des "fenêtres mobiles adaptatives", dans laquelle le paramètre  $r$  est remplacé par  $r(X_i)$ , une fonction de la présence d'événements à proximité de  $X_i$ . En tenant compte de ce "balloon estimator" [10] l'équation initiale devient alors :

$$\hat{\lambda}(X) = \frac{1}{r^2(X_i)} \sum_{i=1}^n k\left(\frac{(X - X_i)}{r(X_i)}\right)$$

où  $r(X_i)$  est obtenu en deux temps. Il s'agit tout d'abord d'estimer des densités initiales  $\tilde{\lambda}(X)$ , à partir de fenêtres de taille fixe  $r$ . Puis, pour chaque fenêtre, on détermine un nouveau rayon spécifique, en comparant la densité locale avec un indicateur de la densité initiale totale. Ainsi, la fonction  $r(X_i)$  peut être définie comme :

$$r(X_i) = r \left( \frac{\tilde{\lambda}_g}{\tilde{\lambda}(X_i)} \right)^a$$

où  $\tilde{\lambda}_g$  représente la moyenne géométrique de  $\tilde{\lambda}(X)$ . Le paramètre  $a$  est, quant à lui, un paramètre de sensibilité, fixé à 0,5 par défaut.

Concrètement, cette manipulation revient à réduire la taille des fenêtres mobiles dans les zones de forte densité, afin de mieux prendre en compte l'information locale, tandis que les zones de faible densité seront balayées par des fenêtres de taille plus grossière.

La représentation finale des densités ainsi obtenues suppose d'avoir recours à un algorithme d'interpolation quelconque [8], le principal critère étant en effet la qualité graphique des résultats finaux<sup>3</sup>.

### 3. Application aux accidents de la route dans la Communauté urbaine de Lille

L'application comparative des deux méthodes présentées aux accidents de la circulation dans la région lilloise, en 1996, permet d'illustrer leur intérêt.

L'utilisation d'un carroyage de 2000 m. de côté (Figure 2), et de fenêtres mobiles circulaires de 1000 m de rayon (Figure 3), donne des résultats intéressants mais néanmoins insuffisants<sup>4</sup>. Les représentations graphiques produites, trop lissées, apportent en effet peu d'information supplémentaire par rapport au semis de points initial et mériteraient d'être affinées.

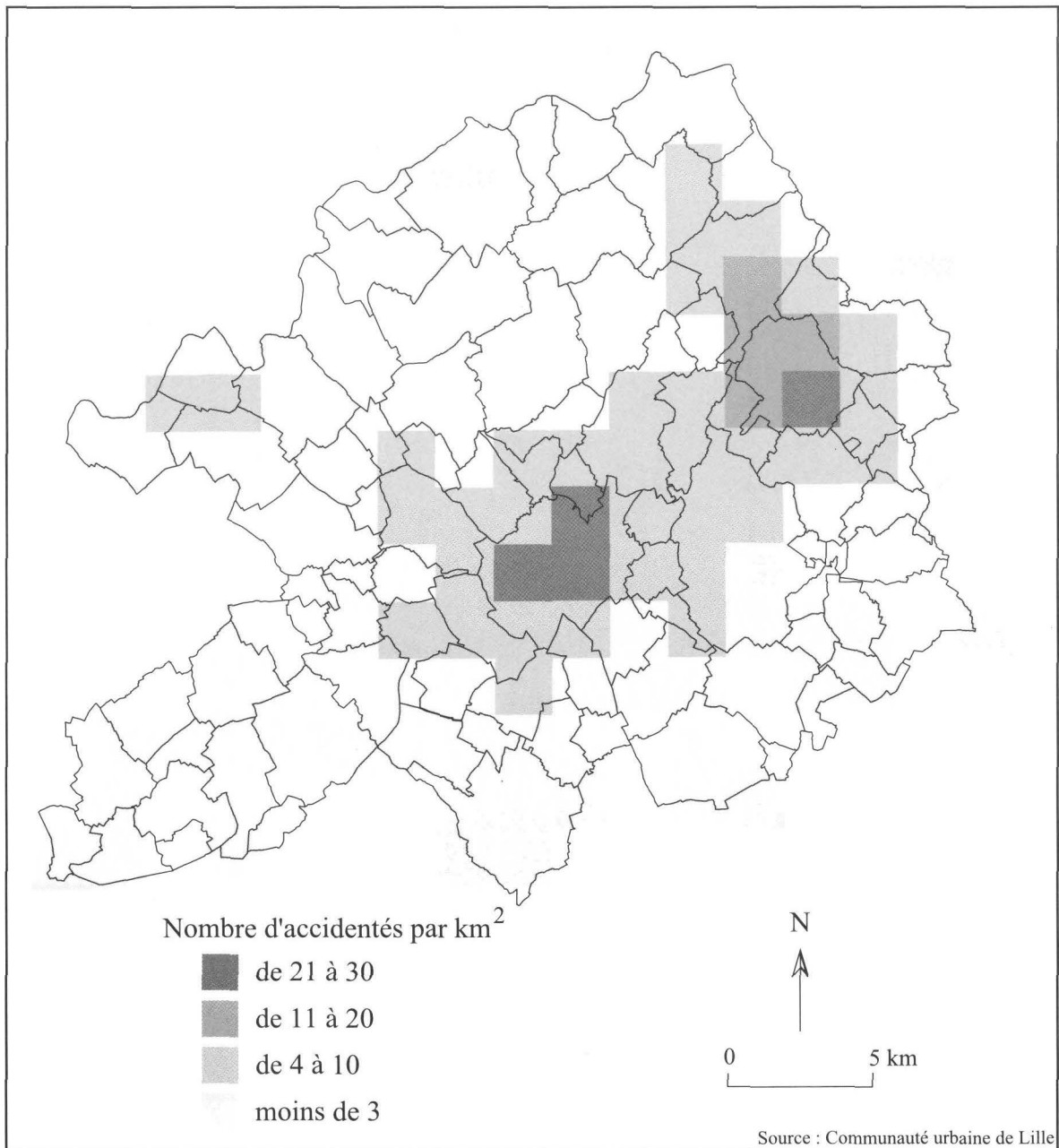


Figure 2 - Densités estimées d'accidents de la circulation dans la Communauté urbaine de Lille en 1996 (carroyage dont les carreaux mesurent 2000 mètres de côté)

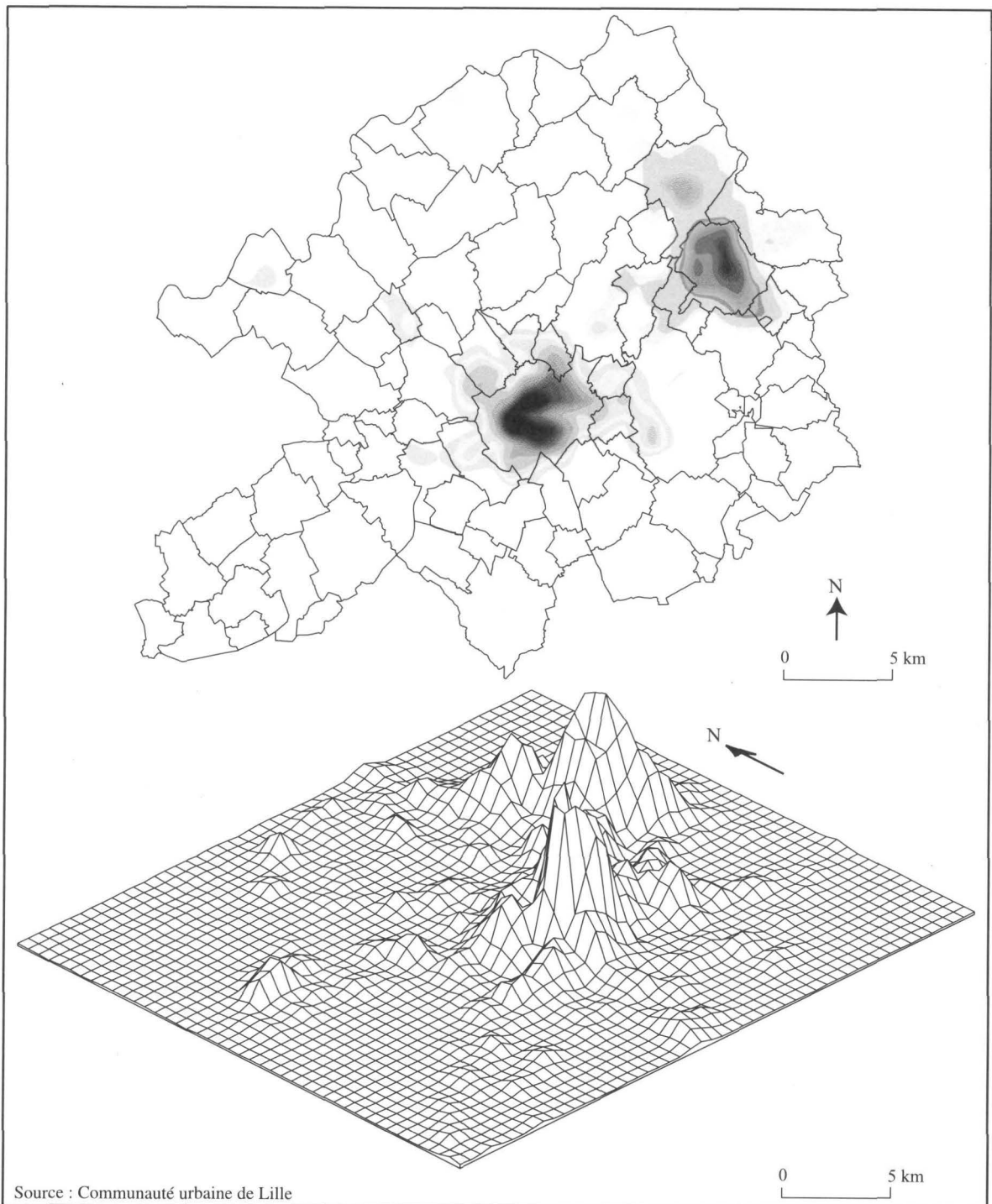


Figure 3 - Densités estimées d'accidents de la circulation dans la Communauté urbaine de Lille en 1996  
(à partir de 20000 fenêtres mobiles de rayon 1000 m)

D'un point de vue comparatif, la carte obtenue à l'aide des fonctions de densité de Kernel est indéniablement plus précise. La surface moindre des fenêtres mobiles n'est pas seule en cause : la fonction de pondération utilisée dans la méthode par noyaux joue ici un rôle déterminant. Par contre, la méthode par carroyage offre l'avantage d'être directement interprétable du point de vue de ses unités (nombre d'accidents par  $\text{Km}^2$ ).

Le recours aux techniques des quadtree (Figure 4) et des fenêtres mobiles adaptatives (Figure 5) permet alors d'améliorer significativement la connaissance des structures spatiales sous-jacentes. Les foyers accidentogènes sont bien mieux détaillés : des structures locales relativement complexes sont ainsi révélées, que l'on n'aurait pu mettre à jour autrement. On remarquera que la technique des quadtrees modifie uniquement la représentation des zones de forte densité, alors que celle des fenêtres mobiles adaptatives agit sur l'ensemble de l'information, selon un processus continu. Le résultat en est, une fois encore, plus précis. Les structures sont ainsi remaniées, le bruit gommé.

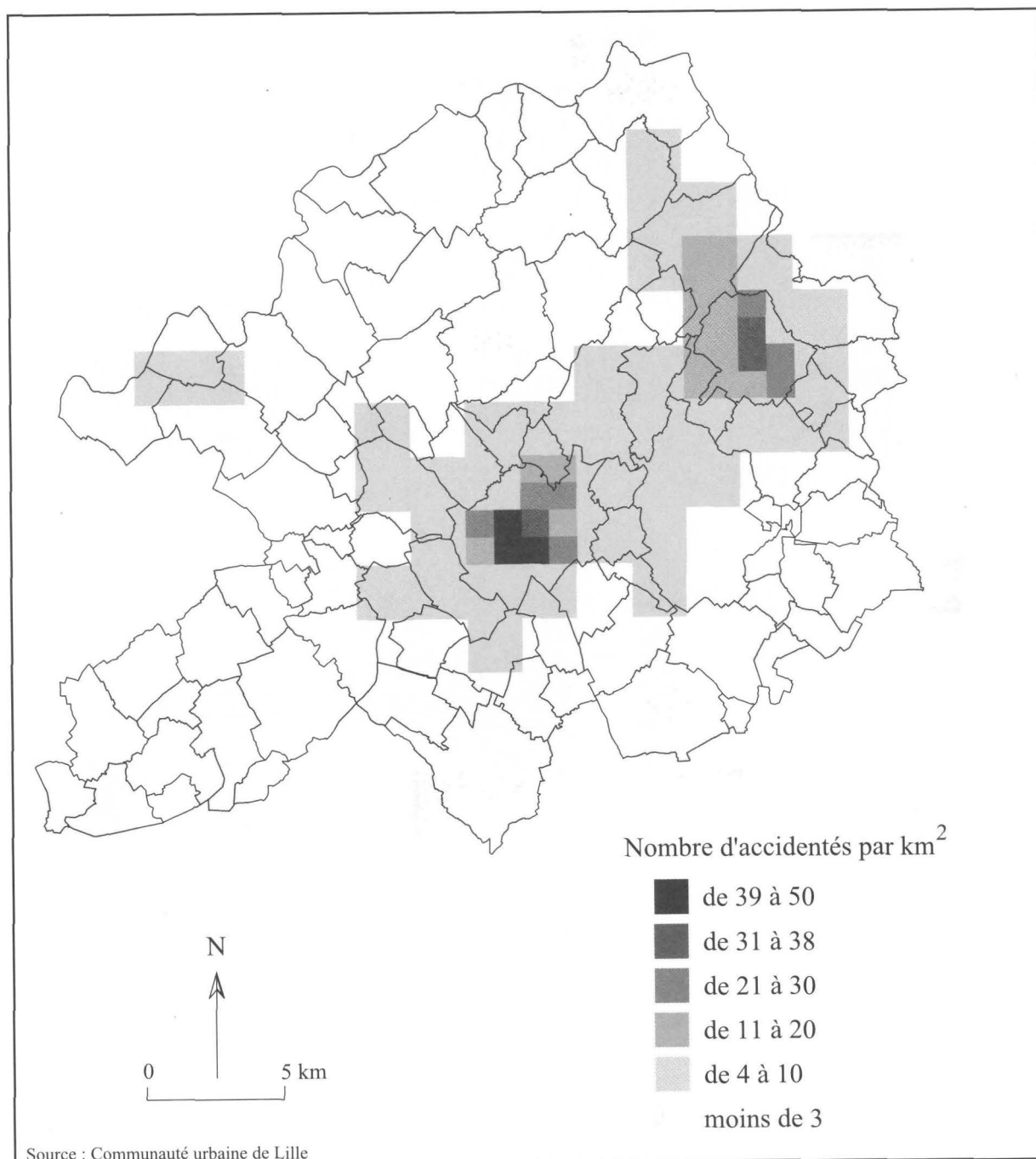


Figure 4 - Densités estimées d'accidents de la circulation dans la Communauté urbaine de Lille en 1996 (Carroyage-Quadtree 2000 et 1000 mètres de côté)

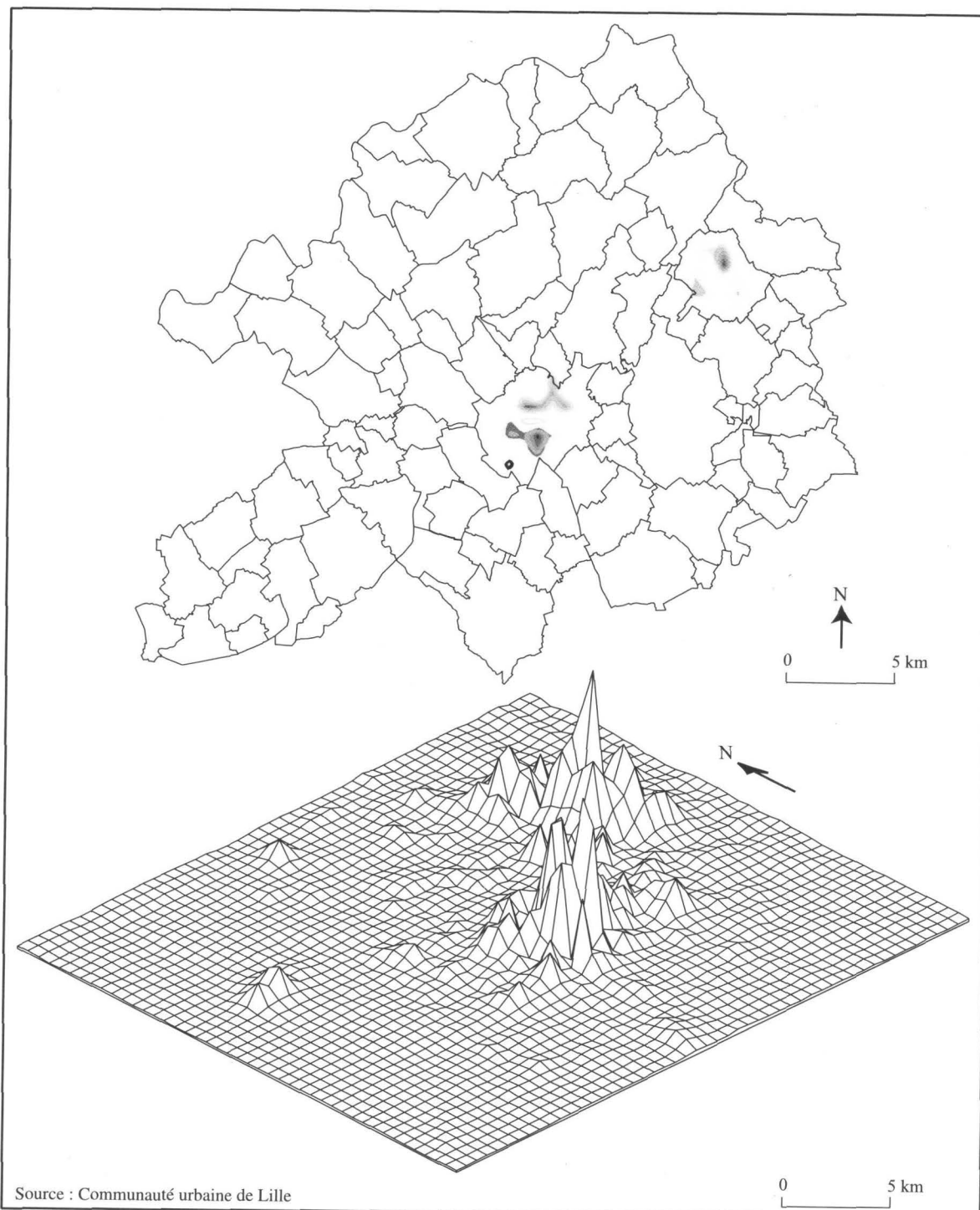


Figure 5 - Densités estimées d'accidents de la circulation dans la Communauté urbaine de Lille en 1996  
 (à partir de 20000 fenêtres mobiles adaptatives de rayon initial 1000 m)



La comparaison des deux méthodes doit également s'opérer, au niveau de leur facilité d'utilisation et d'interprétation. La méthode par carroyage est incontestablement plus simple à mettre en œuvre et à interpréter, même si l'utilisateur doit prendre quelques décisions-clés. La surface des carreaux, la méthode de discrétisation (mise en classes) ou la détermination du seuil à partir duquel décomposer chaque entité du pavage (quadtree), sont ainsi des contraintes impossibles à automatiser complètement.

La méthode des noyaux mobiles est, quant à elle, plus complexe. Elle fait appel à des notions statistiques plus élaborées et repose sur des algorithmes de calcul intensif, parfois délicats à mettre en œuvre, en dépit de la rapidité des ordinateurs modernes. Il ne s'agit pas non plus d'une méthode complètement automatisée : la détermination du rayon des fenêtres mobiles est à la charge de l'utilisateur, et des diagnostics graphiques réellement efficaces doivent encore être imaginés. L'estimation de densités par cette méthode et leur représentation par krigeage est donc un défi dans les SIG actuellement disponibles. Pourtant, les représentations graphiques obtenues sont indéniablement parlantes et utiles, même en l'absence d'unités interprétables.

La représentation surfacique d'une information ponctuelle discrète offre de multiples avantages. Les structures spatiales sous-jacentes ainsi mises à jour guident l'analyste dans sa quête de sens. La comparaison visuelle rapide de cartes nombreuses (comparaisons chronologiques ou attributaires) en est grandement facilitée, tandis que des comparaisons statistiques bivariées, aisément interprétables, entre deux semis de points sont rendues possibles [2]. Enfin, les méthodes présentées permettent de respecter d'éventuelles contraintes d'anonymat de l'information [3], tout en conservant au mieux sa richesse.

On rappellera néanmoins la limite scalaire des méthodes décrites. Peut-on déterminer des échelles spatiales de pertinence ? Existe-t-il des seuils objectivables, indépendants de la seule intuition de l'utilisateur ?

## Références bibliographiques

- [1] BAILEY T., GATRELL A., 1995 : *Interactive Spatial Data Analysis*, Longman Scientific and Technical, London, 413 p.
- [2] BITHELL J., 1990 : An application of density estimation to geographical epidemiology, *Statistics in Medicine*, Vol. 9, pp. 691-701
- [3] BIZET F., 1997 : Carroyage et SIG urbain : les chômeurs à Rouen, *Mappemonde*, n° 2, pp. 32-34
- [4] BOOTS B., GETIS A., 1988 : *Point pattern Analysis*, Sage Publications, Newbury Park, 93 p.
- [5] BRUNSDON C., 1991 : Estimating probability surfaces in GIS : an adaptive technique, *EGIS '91, Proceeding, Second European Conference in GIS*, Brussels, Belgium, April 2-5, vol. 1, EGIS Foundation, pp. 155-164
- [6] GUERMOND Y., 1995 : Information, informatique et système d'information géographique, in *Encyclopédie de Géographie*, Economica, Paris, pp. 295-310
- [7] HUGUENIN-RICHARD F., 1999, Recherche en analyse spatiale en vue de proposer des modes opératoires de diagnostics et d'évaluation du risque routier en milieu urbain et périurbain. Premières considérations, *Actes des Quatrièmes Rencontres de Théo Quant*, Besançon, pp. 223-234
- [8] ISAAKS E., SRIVASTAVA R., 1989 : *An introduction to applied geostatistics*, Oxford University Press, Oxford, 561 p.
- [9] LAJOIE G., 1992 : *Le carroyage des informations urbaines*, Rouen, Presses Universitaires de Rouen, 238 p.
- [10] SAIN S., 1994 : *Adaptive Kernel density estimation*, Thesis, Rice University, Houston, Texas, 128 p.

[11] TIERNEY L., 1990 : *Lisp-Stat, an object-oriented environment for statistical computing and dynamic graphics*, John Wiley and Sons, New York, 397 p.

[12] VOIRON C., 1995 : *Analyse spatiale et analyse d'images par la morphologie mathématique*, Montpellier, GIP Reclus, 1995, 190 p.

## Notes

- 1 - Par commodité, nous utiliserons le sigle officiel (Communauté urbaine de Lille) dans la suite du texte.
- 2 - Ces différentes techniques ont été développées par J. BOLOT au sein de l'environnement de programmation Map Basic du SIG commercial Map Info.
- 3 - Les différentes techniques présentées ont été développées par A. BANOS au sein de l'environnement de programmation statistique Xlisp-Stat (TIERNEY, 1991).
- 4 - La représentation tri-dimensionnelle permet de mieux visualiser les apports de la technique dite des fenêtres mobiles adaptatives, lorsque l'on compare la figure 3 avec la figure 5.