

Interstats Méthode N° 11

Le lissage spatial de la délinquance enregistrée et la représentation sous forme de cartes de chaleur

Gâce au géocodage des enregistrements de la police et de la gendarmerie, des techniques de lissage spatial permettent de représenter la répartition de la délinquance sous forme de « cartes de chaleur ».

Sans rapporter à ce stade cette délinquance au nombre d'habitants (ou de logements), les cartes produites permettent seulement de répondre à la question « où sont perpétrés les cambriolages ? ». En revanche, elles ne permettent nullement de répondre à la question « où risque-t-on le plus d'être victime de cambriolage ? » : ainsi, ces cartes ne doivent en aucun cas être interprétées comme des cartes de risque individuel dans les zones concernées.

Afin de pouvoir réaliser des analyses territoriales fines, le SSMSI procède, en collaboration avec l'Institut national de la statistique et des études économiques (Insee)¹, au géocodage exhaustif de certaines formes de délinquance constatées par les forces de sécurité sur l'ensemble du territoire. Cette opération de géocodage a jusqu'à présent porté sur les infractions de niveau criminel et délictuel enregistrées par la police et la gendarmerie nationales au cours de l'année 2016 et 2017. Ainsi, pour toutes les formes de délinquance retenues *a priori*, le SSMSI dispose de la base exhaustive des coordonnées géographiques des lieux de commission des crimes et délits.

Ces données géolocalisées rendent possible la création de représentations géographiques détaillées, permettant de distinguer les densités territoriales de délinquance, indépendamment des frontières administratives habituelles (départements, zones de compétence des services, communes, etc.). Pour parvenir à remplir le double objectif de proposer des représentations structurelles de la délinquance (sans bruit micro-local ou conjoncturel), tout en veillant au respect du secret statistique, les techniques de lissage spatial ont été utilisées et représentées sous forme de « cartes de chaleur ».

Ce document a pour objectif de présenter la technique de lissage spatial utilisée par le SSMSI dans le cadre de la confection des cartes de chaleur associées à la délinquance enregistrée.

Le géocodage des adresses renseignées dans les logiciels de rédaction des procédures

L'Insee a développé, pour les besoins de la statistique publique, un outil de géolocalisation permettant la reconnaissance textuelle des adresses saisies et l'affectation de coordonnées géographiques (XY) adéquates. En pratique, les adresses de commission renseignées dans les systèmes d'information du ministère de l'intérieur sont de qualité variable : certaines sont rédigées de façon approximative ou incomplète (voire tout simplement manquante), et ce dans des proportions non négligeables. En revanche, la commune de commission des faits est systématiquement saisie. Ainsi, l'outil de géolocalisation de l'Insee ne parvient pas toujours à reconnaître l'adresse précise et doit procéder à des imputations, c'est-à-dire, sélectionner aléatoirement une adresse en limitant au maximum le risque d'écart au réel. Par exemple, lorsque le numéro de voie est manquant mais que l'algorithme reconnaît le nom de la voie (exemple : « rue de Rivoli » à Paris), une adresse est tirée au sort dans la voie. En pratique, le tirage est pondéré par la population résidente : plus l'adresse est peuplée, plus elle a de chance d'être tirée au sort. Dans ce cas, les coordonnées géographiques associées à l'infraction sont celles de l'adresse tirée au sort.

Enjeux sur la qualité et la représentativité des données

La qualité des adresses renseignées dans les logiciels de rédaction des procédures dépend du type de délinquance considéré. En effet, les adresses saisies par les plaignants sont en moyenne bien plus fréquemment reconnues par l'algorithme de l'Insee dans les cas de cambriolages de logements, vols de véhicules et dégradations que dans les cas de vols et violences sur les personnes ou encore d'infractions à la législation sur les stupéfiants (**voir Annexe**). Ceci s'explique en partie par la plus ou moins grande difficulté pour la victime à connaître ou mémoriser l'adresse de commission des faits : elle est évidente pour une victime d'un cambriolage, mais nettement moins pour une victime de vol à l'arraché sur la voie publique.

Enfin, les enregistrements des forces de sécurité ne recensent qu'une partie de la délinquance effective : ils sous-estiment parfois lourdement le nombre d'actes de délinquance, notamment pour les types d'infractions qui ne font pas de victime directe (infractions à la législation sur les stupéfiants) ou celles pour lesquelles les taux de

¹ Ces cartes de chaleur ont pu être réalisées grâce au concours et à l'expertise de la division des statistiques et des analyses urbaines de l'Insee.

plainte sont faibles (notamment les violences aux personnes). De plus, ces remontées dépendent des pratiques de gestion et de saisie des forces de sécurité. Ainsi, les enregistrements géolocalisés sont de bons indicateurs de la délinquance effective dans le cas des atteintes pour lesquelles les victimes portent fréquemment plainte, et pour lesquelles les adresses de commission des faits sont révélées et enregistrées de façon fiable. Pour autant sous l'hypothèse que le processus de révélation et d'enregistrement des crimes et délits soit homogène sur le territoire, les enregistrements des forces de sécurité sont un indicateur satisfaisant en relatif : une zone où ont été constatées davantage d'infractions qu'ailleurs est généralement une zone plus touchée par des problèmes réels de délinquance.

La méthode de lissage spatial des enregistrements géolocalisés

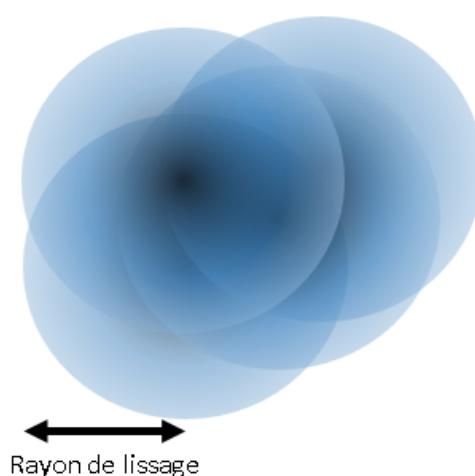
Le lissage de l'information géographique permet à la fois une visualisation structurale des actes de délinquance enregistrés, tout en préservant le secret statistique. En effet, la publication de « cartes à punaises » affichant un point pour chaque adresse de commission des infractions poserait des difficultés en termes de respect de la confidentialité des données, et ne permettrait que très difficilement d'apprécier des phénomènes structurels de forte ou faible concentration de la délinquance sur un territoire donné.

De façon simplifiée, le lissage consiste à passer d'un point (adresse d'une infraction) à une zone de chaleur circulaire définie par un rayon dit « de lissage ». Cette zone est plus chaude/concentrée en son centre (adresse de l'infraction) que sur ses bords. La superposition de zones de chaleur sur une carte permet une visualisation adaptée de la délinquance sur un territoire donné (**voir schéma**). Le choix d'un rayon de lissage par le statisticien/cartographe dépend avant tout de l'objectif final de la carte : plus le rayon est grand et plus la carte révélera des phénomènes de répartition structurels, mais ceci au détriment de particularités locales potentiellement intéressantes.

En pratique, le SSMSI a mis à profit un *package* R de lissage spatial (le package BTB, **voir bibliographie**), développé par des agents de l'Insee et partagé de manière libre.



La superposition de toutes les zones de chaleur sur un territoire forme une carte lissée, ou carte de chaleur



La représentation cartographique des données lissées

Pour effectuer l'opération de lissage, l'ensemble du territoire à cartographier est découpé en carreaux de petite taille et chaque carreau se voit affecter une valeur lissée (ou densité spatiale de délinquance en l'occurrence) correspondant à une moyenne pondérée de sa valeur initiale et des valeurs des carreaux voisins. Plus les carreaux sont éloignés du carreau central, moins ils vont contribuer à sa valeur lissée. Sur la base de cette valeur de lissage, une classification est réalisée sur l'ensemble des carreaux afin de faciliter la visualisation. Différents algorithmes peuvent être mobilisés en fonction des objectifs et des contraintes associés à la carte finale (temps de calcul, comparabilité dans le temps ou dans l'espace, maximisation de l'homogénéité des groupes formés, etc.).

Mise en garde dans l'interprétation des cartes de chaleur

Le lissage spatial permet de représenter des densités d'infractions, c'est-à-dire des nombres d'infractions par kilomètre-carré. Sans prise en compte de la densité de population, ou de la densité de logements, les niveaux de délinquance lissée ne doivent pas être confondus avec des taux de délinquance lissés (nombre de violences pour 1 000 habitants, nombre de cambriolages pour 1 000 logements, etc.). De surcroît, ces niveaux ne doivent pas être interprétés comme une représentation du risque individuel de victimation. Par exemple, à risque égal de cambriolage par logement, une zone dense en logements verra en moyenne davantage de cambriolages perpétrés qu'une zone faiblement dense en logements.

La possibilité de représenter des taux lissés d'infractions (pour 1 000 habitants, pour 1 000 logements, etc.) fait actuellement l'objet de travaux au sein du SSMSI. Elle consiste à rapporter, pour chaque carreau de l'espace, la valeur lissée de la densité de délinquance à la valeur lissée de la densité de population (ou de logements). Elle nécessite la mise au point d'une méthode permettant de limiter, à niveau géographique fin, l'influence des zones vides sur les zones habitées attenantes (par exemple certaines zones du centre-ouest de Paris : Esplanade des Invalides, la Seine, Jardin des Tuileries, Place de la Concorde, etc.). Ces travaux feront l'objet de prochaines publications.

Annexe : Part des adresses de commission des infractions enregistrées en 2017 selon la qualité du géocodage

Type de délinquance	Voie reconnue	Voie probable	Voie non reconnue
Cambriolages de logements	87 %	8 %	5 %
Vols de véhicules	86 %	6 %	7 %
Dégradations	86 %	5 %	9 %
Violences intrafamiliales	83 %	7 %	10 %
Vols dans et sur les véhicules	81 %	7 %	13 %
Violences	78 %	7 %	15 %
Vols violents	74 %	6 %	19 %
Violences sexuelles intrafamiliales	70 %	9 %	21 %
Usages de stupéfiants	72 %	6 %	22 %
Vols sans violence	67 %	7 %	25 %
Violences sexuelles	64 %	7 %	29 %
Trafics de stupéfiants	63 %	6 %	31 %

Note : tri croissant sur la part d'adresses dont la voie n'est pas reconnue par l'algorithme de géocodage.

Note de lecture : sur l'ensemble des cambriolages enregistrés en 2017, 87 % des voies de commission des faits sont reconnues par l'algorithme de l'Insee de manière sûre. Dans 8 % des cas, la voie est reconnue de manière probable, mais sans certitude, et, pour 5 % des adresses, l'algorithme ne peut reconnaître la voie.

Champ : France hors Mayotte.

Sources : SSMSI, Base des crimes et délits enregistrés par la police et la gendarmerie. Calculs SSMSI.

Bibliographie :

- Dos Santos A., Semecurbe F., Renaud A., Faivre C., Cornely T., Marouchi F, « Package BTB », mai 2018.
- Adaoust S., Rouaud P., « *Dans les quartiers de la politique de la ville, deux fois plus de jeunes ont quitté l'école et sont sans diplôme* », Insee Analyses n°60, juin 2018.