

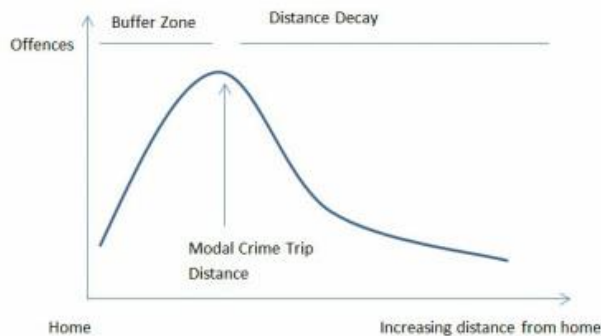
Sujet de « data sciences » appliquées à l'actuariat

2014-2015

Modèles espace-temps du risque de cambriolage

Application à Paris et Bruxelles à partir des données d'assurance

Le but de ce projet de « statistiques appliquées » est d'étudier les caractéristiques spatiales et temporelles de la criminalité vol afin de mettre en place un système d'alerte au vol et de mieux appréhender ce risque. En effet, l'influence spatiale de la criminalité est bien établie même si elle peut intervenir de façon complexe (par exemple, la sur-criminalité n'est pas directement liée à des caractéristiques socio-économiques mais à la proximité de certaines zones riches avec des zones plus défavorisées). Par exemple, selon Rengert et al 1999; Rossmo, 2000, la sinistralité est maximale à quelque distance du lieu d'habitation du voleur (Distance Decay Effect), avec une distance médiane de 2.6 km (distance euclidienne - Montréal, 2001):



Ce travail aura donc pour but d'établir un modèle spatiale basée sur la théorie de criminologie / cambriolage et sur l'application aux données assurantielles, puis d'étudier ce modèle de façon dynamique : comment en particulier détecter les phénomènes nouveaux de vol afin de mieux les prévenir ?

Pratiquement, le sujet pourra être découpé de la façon suivante :

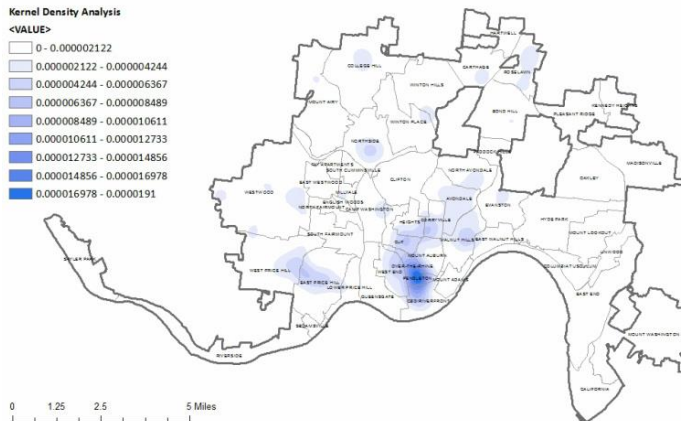
- Revue de la littérature théorique sur la criminalité / vol (théorie des opportunités criminelles,...)
- Proposition d'un ou plusieurs modèles spatiaux à tester sur les données d'AXA tenant compte de la littérature théorique ou d'approches purement statistiques (modèle d'Auto-régression conditionnelle (CAR)...).
- Application de ces modèles sur les données AXA (sinistre et police géocodé) France et Belge
- Recherche de données externes (socio-démographiques / criminelles) permettant d'améliorer le modèle.

Kernel Density Spatial Analysis of Robbery in Cincinnati, 2010

Data Source: Cincinnati Police Department/ Cincinnati Area Geographic Information System

Map Designed by Yihua Feng on Feb 13, 2014

Highest Score means highest concentrations of data points which also means the area is the most dangerous.



-
- Intégration de données supplémentaires sur les caractéristiques du sinistre (heure, type de vol...) pour affiner le modèle spatial

Étalement du noyau de densité des affaires de vol qualifié ayant eu lieu entre 7 h et 9 h, Winnipeg, 2001



- Discussion sur les différents modèles statistiques et théoriques
- Puis extension de ces modèles en intégrant l'approche dynamique. Cette partie pourra être l'occasion de tester l'efficacité des modèles alternatifs non paramétriques (gradient boosting),....
- Création d'un modèle prédictif de vol en fonction de cette approche dynamique afin de pouvoir prévenir en amont du risque de vol.

Littérature – quelques pistes d'introduction :

- http://www98.griffith.edu.au/dspace/bitstream/handle/10072/24437/49610_1.pdf?sequence=1 : a cross national assessment of space-time patterns of burglary risk, Shane D Johnson et al - 1998
 - Computational Actuarial Science with R, Arthur Charpentier (2014) : chapitre sur la gestion statistique des données spatiales sous R.
 - <http://www.statcan.gc.ca/pub/85-561-m/85-561-m2008015-fra.htm> : analyse spatiale de la criminologie au Canada
 - <http://evasdatavisualization.weebly.com/project-1.html> : Exemple sur la ville de Cincinnati
- exemple

- http://www.math.u-psud.fr/~stafav/IMG/pdf/Statistique_Spatiale.pdf : cours de statistique spatiale
- R : documentation des packages spatiaux geoR, gstat, RandomFields, spatial, spatstat, spdep

Pour plus de contacts : guillaume.gorge@axa.com et phileas.condemine@axa.com