

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Modèles espace-temps du risque de cambriolage

Application à la Belgique

Augustin Ador, Florian Héraud, Guillaume Salha

CONFIDENTIEL

20 mai 2015

Introduction

Statistiques générales

Carte

Durbin Spatial

Gradient Boosting

Comparaison

Conclusion

Risque de cambriolage :

- Plus de 75 000 cambriolages recensés en Belgique en 2012
- En hausse depuis plusieurs années
- Problème majeur pour AXA en tant qu'assureur

Objectif : étudier les caractéristiques spatiales et temporelles du risque de cambriolage en Belgique

Introduction

Statistiques générales

Carte

Durbin Spatial

Gradient Boosting

Comparaison

Conclusion

D'après Gary S. Becker, un individu commet un crime si :

$$E(U) = qB - Cp > 0$$

$B \geq 0$ l'utilité apportée par le crime s'il réussit

$C \geq 0$ son coût en cas de condamnation

p la probabilité que l'agent se fasse attraper

q la probabilité de succès du crime

Facteurs pouvant influencer cette espérance d'utilité :

- La richesse du quartier
- L'équipement de la victime potentielle en termes de sécurité
- La proximité avec certains établissements publics ainsi que l'isolement
- La météo

Introduction

Statistiques générales

Carte

Durbin Spatial

Gradient Boosting

Comparaison

Conclusion

Variables à tester dans notre analyse spatio-temporelle :

- **la frontière zone riche/zone pauvre**, Rengert et al. (1999), Rossmo (1999) ;
- **la richesse du quartier**, Kennedy et Forde (1990), Bursik et Grasmick (1993), Pratt (2001) ;
- **le taux de chômage**, Roundtree et Land (2000), Hartnagel (2004),
- **le système de protection dans le quartier**, Paternoster et Bushway (2001) ;
- **l'évolution du nombre d'habitants dans la ville et la part de minorités**, Ceccato, Haining, et Signoretta (2002).

Introduction

Statistiques générales

Carte

Durbin Spatial

Gradient Boosting

Comparaison

Conclusion

Données AXA

- 374 397 assurés belges
- 2010-2013
- 232 variables

Données extérieures

- OpenStreetMap
- Données météorologiques (Institut Royal Météorologique de Belgique)

Influence des caractéristiques de l'habitation

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

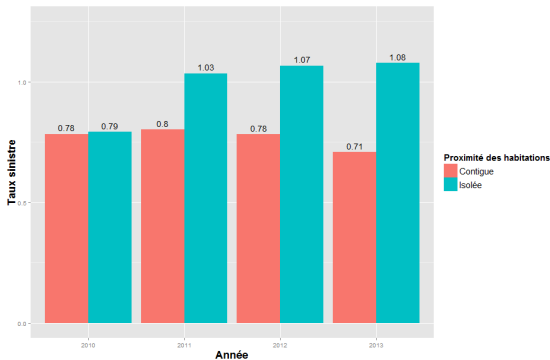


Figure: Taux de sinistres et isolement de l'habitation

- Maison/appartement
- Propriétaire/locataire

Influence de la météorologie

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

- 37% de nos cambriolages ont eu lieu lors d'un jour de pluie
- Il pleut plus d'un jour sur deux en Belgique (55% en moyenne)

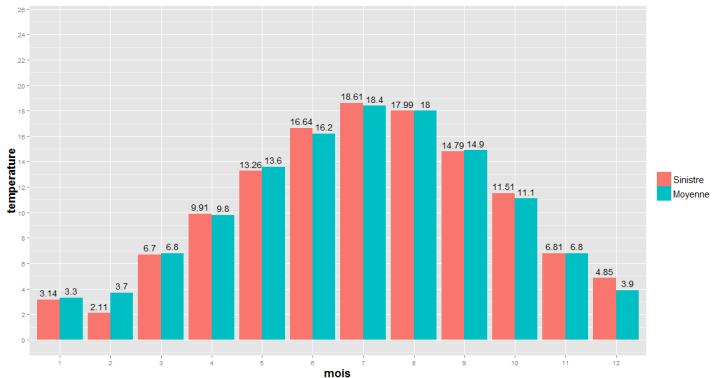


Figure: Taux de sinistres et température

Introduction

Statistiques générales

Carte

Durbin Spatial

Gradient Boosting

Comparaison

Conclusion

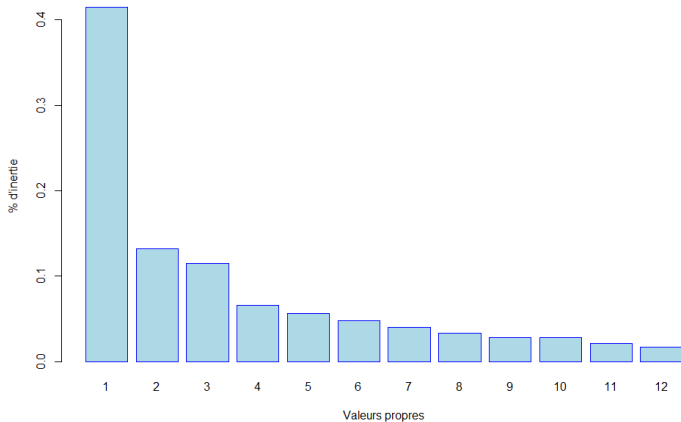


Figure: ACP : ébouli des valeurs propres

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

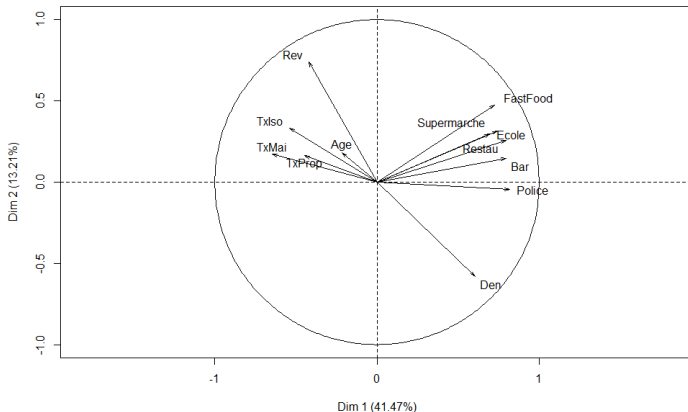


Figure: ACP : projection des variables dans le plan engendré par les C.P. 1 et 2

Importance de l'aspect temporel

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

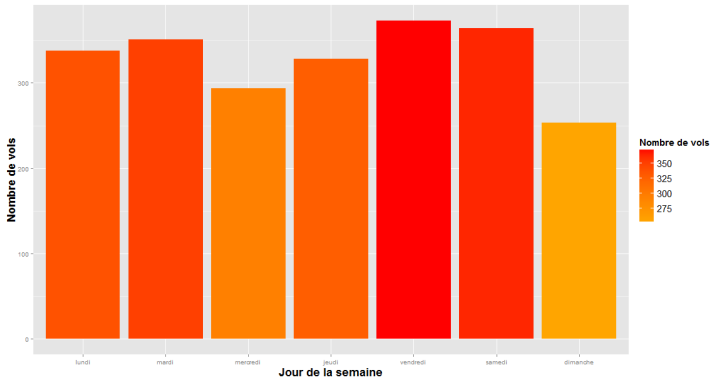


Figure: Taux de sinistres journaliers

Commune	Nb vols	Nb assurés	Ratio	Détails	Nb d'hab
Charleroi	716	33517	2.1%		204 670
Uccle	372	20745	1.8%	Banlieue Sud Brux	79 766
Anderlecht	220	13261	1.7%	Banlieue Ouest Brux	112 258
Ixelles	280	17335	1.6%	Banlieue Sud Est Brux	84 073
Molenbeek-Saint-Jean	162	10150	1.6%	Banlieue Ouest Brux	94 798
Liège	464	32886	1.4%		197 013
Seraing	142	10607	1.3%	Banlieue de Liège	63 968
Bruxelles	265	19801	1.3%		177 849
Woluwe-Saint-Lambert	161	12035	1.3%	Banlieue Est Brux	51 937
Mons	233	18133	1.3%	Ouest de Charleroi	93 366

Table: Top 10 des villes aux taux de sinistrés les plus importants (2010-2013)

Ratio moyen = 0.84%

Variance du ratio = 0.85%

La carte

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

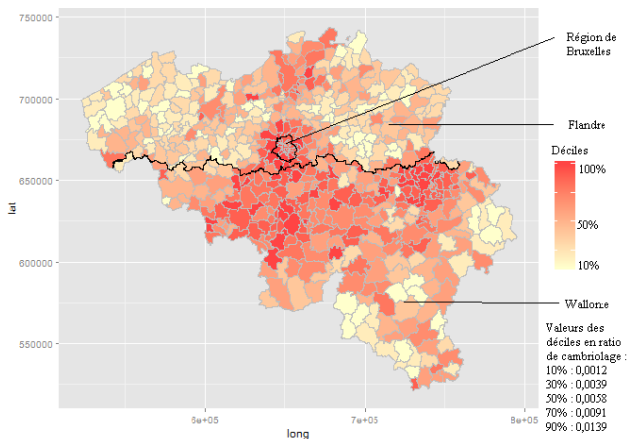


Figure: Carte des ratios de sinistralité sur les 4 ans : regroupement par déciles

La carte : principales informations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

- Les ratios de sinistralité par communes non répartis uniformément
- **Autocorrélation spatiale positive** de la sinistralité

La carte : principales informations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

- Les ratios de sinistralité par communes non répartis uniformément
- **Autocorrélation spatiale positive** de la sinistralité

Les zones les plus sinistrées :

- Nord de la Wallonie
- Région de Bruxelles
- Centre nord de la Flandre (région autour de la ville d'Anvers)

La carte : principales informations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Intuitions : les zones pauvres et denses sont les plus touchées.

Conclusion : L'autocorrélation spatiale positive semble apparaître nettement.

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Modèle de Durbin spatial :

$$Y = \rho WY + X_1\beta + WX_2\gamma + \epsilon$$

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Modèle de Durbin spatial :

$$Y = \rho WY + X_1\beta + WX_2\gamma + \epsilon$$

Y : Le nombre de cambriolages sur 4 ans

X_2 : Regroupant les variables explicatives pour lesquelles nous souhaitons récupérer les valeurs des plus proches voisins.

W : Matrice de poids

Durbin Spatial : le modèle

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Intérêts :

- Repérer l'effet spatial pur dans le ρ

Durbin Spatial : le modèle

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Intérêts :

- Repérer l'effet spatial pur dans le ρ
- Repérer l'effet des différences de richesse : frontières zones riches/zones pauvres

Durbin Spatial : les résultats

Introduction

Statistiques
générales

Carte

**Durbin
Spatial**

Gradient
Boosting

Comparaison

Conclusion

	Durbin Spatial
(Constante)	0.0011 (1.2789)
Densité	0.0034*** (0.0006)
Revenu	0.0005*** (0.0001)
Age	-0.0061 (0.0057)
Isolée	0.0043*** (0.0013)
Propriétaire	0.0078 (0.0053)
Maison	0.0091* (0.0037)
Police	-0.0062 (0.0082)
ρ	0.0315
P-value (two-sided)	0.0317
γ_{Rev}	-0.0402
P-value (two-sided)	< 0.001

*** $P(> |z|) < 0.001$, ** $P(> |z|) < 0.01$, * $P(> |z|) < 0.05$

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Facteurs aggravant le risque de cambriolage :

- Une augmentation de la densité

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Facteurs aggravant le risque de cambriolage :

- Une augmentation de la densité
- Une augmentation du revenu moyen

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Facteurs aggravant le risque de cambriolage :

- Une augmentation de la densité
- Une augmentation du revenu moyen
- Vivre dans une habitation isolée

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Facteurs aggravant le risque de cambriolage :

- Une augmentation de la densité
- Une augmentation du revenu moyen
- Vivre dans une habitation isolée
- Habiter dans une maison

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

**Durbin
Spatial**

Gradient
Boosting

Comparaison

Conclusion

Le coefficient de dépendance spatiale sur les cambriolages des voisins ρ significatif à 5% et positif.

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

**Durbin
Spatial**

Gradient
Boosting

Comparaison

Conclusion

Le coefficient de dépendance spatiale sur les cambriolages des voisins ρ significatif à 5% et positif.

⇒ On a plus de risque d'être cambriolé lorsque son voisin est lui-même cambriolé.

Durbin Spatial : interprétations

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Le coefficient de dépendance spatiale sur les cambriolages des voisins ρ significatif à 5% et positif.

⇒ On a plus de risque d'être cambriolé lorsque son voisin est lui même cambriolé.

γ_{Rev} est négatif et aussi significatif à 5%

Introduction

Statistiques
générales

Carte

**Durbin
Spatial**

Gradient
Boosting

Comparaison

Conclusion

Le coefficient de dépendance spatiale sur les cambriolages des voisins ρ significatif à 5% et positif.

⇒ On a plus de risque d'être cambriolé lorsque son voisin est lui même cambriolé.

γ_{Rev} est négatif et aussi significatif à 5%

⇒ Lorsque nos voisins ont un revenu plus faible le risque d'être victime d'un cambriolage est plus important.

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Proposition d'une approche alternative par **gradient boosting**

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Proposition d'une approche alternative par **gradient boosting**

Objectifs :

- confirmer les précédentes interprétations

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Proposition d'une approche alternative par **gradient boosting**

Objectifs :

- confirmer les précédentes interprétations
- améliorer les prédictions

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Proposition d'une approche alternative par **gradient boosting**

Objectifs :

- confirmer les précédentes interprétations
- améliorer les prédictions

Différence avec le Durbin : l'aspect spatial n'est pas « forcé » *a priori*

Proposition d'une approche alternative par **gradient boosting**

Objectifs :

- confirmer les précédentes interprétations
- améliorer les prédictions

Différence avec le Durbin : l'aspect spatial n'est pas « forcé » *a priori*

Création d'un **indicateur de richesse relative** :

- rapport entre la richesse au niveau de l'habitation et la moyenne des richesses des zones voisines (rayon de 5 kms)

Proposition d'une approche alternative par **gradient boosting**

Objectifs :

- confirmer les précédentes interprétations
- améliorer les prédictions

Différence avec le Durbin : l'aspect spatial n'est pas « forcé » *a priori*

Création d'un **indicateur de richesse relative** :

- rapport entre la richesse au niveau de l'habitation et la moyenne des richesses des zones voisines (rayon de 5 kms)
- Indice > 100 : habitation dans une zone relativement plus riche que les zones voisines, et inversement

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Variable	Influence Relative (%)
Densité de population	35.75
Gradient de richesse	19.02
Richesse moyenne	16.88
Isolement	14.94
Age de l'assuré	7.47
Appartement ou maison	4.22
Propriétaire ou locataire	1.46
Police à proximité	0.18
École à proximité	0.09

Table: Influences relatives des variables du boosting

Gradient Boosting

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

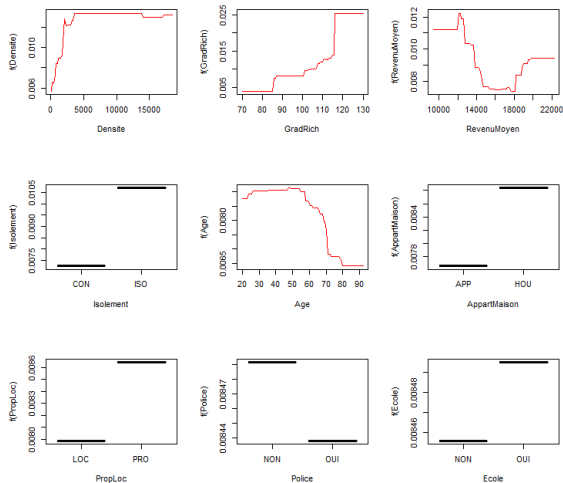


Figure: Dépendances partielles des variables du boosting

Comparaison des deux modèles

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Comparaison des deux modèles en termes de pouvoir prédictif,
selon trois critères :

Comparaison des deux modèles

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Comparaison des deux modèles en termes de pouvoir prédictif, selon trois critères :

- visualisation des courbes ROC

Comparaison des deux modèles

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Comparaison des deux modèles en termes de pouvoir prédictif, selon trois critères :

- visualisation des courbes ROC
- analyse des fonctions de perte

Comparaison des deux modèles

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Comparaison des deux modèles en termes de pouvoir prédictif, selon trois critères :

- visualisation des courbes ROC
- analyse des fonctions de perte
- comparaison du nombre de zéros modélisés et observés
(*non évoquée ici*)

Comparaison des deux modèles

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

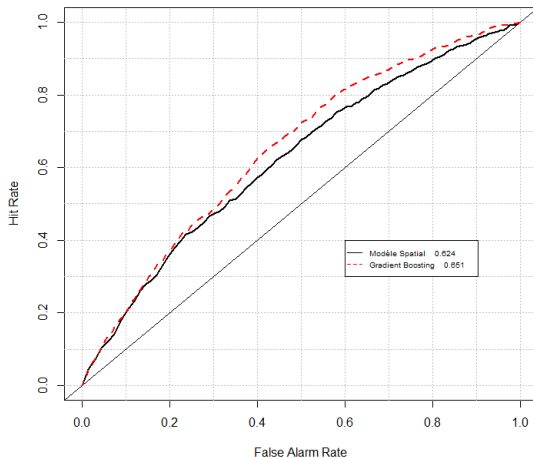


Figure: Courbes ROC

Un autre critère : les **fonctions de perte**

Modèle	Perte Poisson	Perte Quadratique
Modèle spatial	103482.6	840.26
Gradient boosting	103254.4	839.43

Table: Comparaison des modèles à partir des fonctions de perte

Un autre critère : les **fonctions de perte**

Modèle	Perte Poisson	Perte Quadratique
Modèle spatial	103482.6	840.26
Gradient boosting	103254.4	839.43

Table: Comparaison des modèles à partir des fonctions de perte

Conclusion : le gradient boosting a un meilleur pouvoir prédictif que le modèle spatial

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Nous pouvons désormais à l'aide de ces modèles identifier la présence :

- d'une zone à risque à l'aide de caractéristiques spatiales

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Nous pouvons désormais à l'aide de ces modèles identifier la présence :

- d'une zone à risque à l'aide de caractéristiques spatiales
- d'un individu ou d'une habitation à risque

Introduction

Statistiques
générales

Carte

Durbin
Spatial

Gradient
Boosting

Comparaison

Conclusion

Nous pouvons désormais à l'aide de ces modèles identifier la présence :

- d'une zone à risque à l'aide de caractéristiques spatiales
- d'un individu ou d'une habitation à risque

Nous proposons à AXA d'utiliser plutôt un **gradient boosting** pour réaliser des prédictions sur de nouveaux assurés

Nous pouvons désormais à l'aide de ces modèles identifier la présence :

- d'une zone à risque à l'aide de caractéristiques spatiales
- d'un individu ou d'une habitation à risque

Nous proposons à AXA d'utiliser plutôt un **gradient boosting** pour réaliser des prédictions sur de nouveaux assurés

Quelques pistes d'amélioration :

- mieux lier les aspects spatiaux et temporels (limite des données)

Nous pouvons désormais à l'aide de ces modèles identifier la présence :

- d'une zone à risque à l'aide de caractéristiques spatiales
- d'un individu ou d'une habitation à risque

Nous proposons à AXA d'utiliser plutôt un **gradient boosting** pour réaliser des prédictions sur de nouveaux assurés

Quelques pistes d'amélioration :

- mieux lier les aspects spatiaux et temporels (limite des données)
- utiliser l'API professionnelle Google Map