



HAL
open science

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers

Ghislain Geniaux

► **To cite this version:**

Ghislain Geniaux. Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers. 1. Journée d'échange Praticiens-Chercheur " Prix et changements d'usage des sols ", May 2017, Avignon, France. 8 p. hal-01606778

HAL Id: hal-01606778

<https://hal.science/hal-01606778>

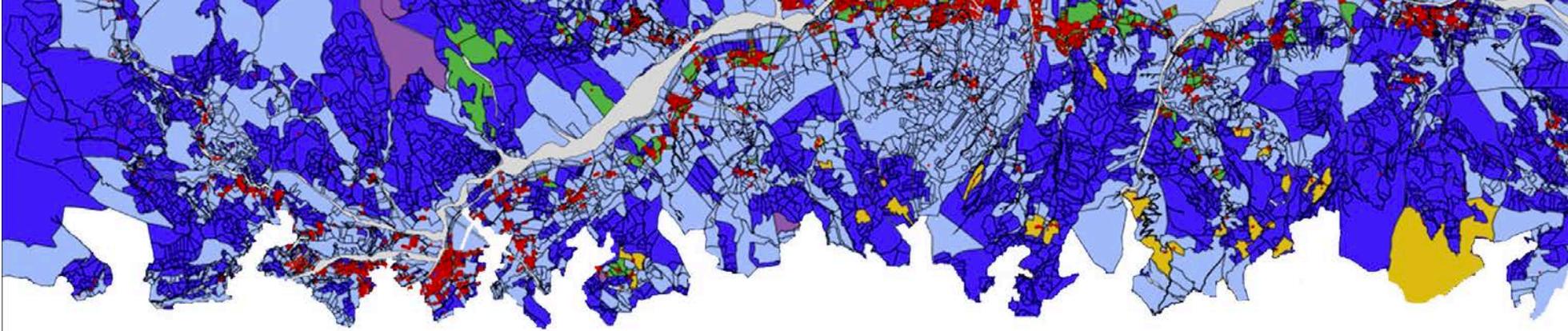
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

Ghislain Geniaux
INRA UR 767 Ecodéveloppement

17 Mai 2017

1^{ère} journée d'échange Praticiens-Chercheur
« Prix et changements d'usage des sols »
INRA/Cerema/UMR ESPACE

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

- **Objectif :**
 - Faire le point sur les avancés méthodologiques en matière de prédiction spatiale des prix de vente des terrains et des logements à partir de données de ventes géolocalisées.
- **Enjeux :**
 - disposer d'outils permettant de générer des cartes de valeur sur différents segments de marchés qui soient le plus fiable possible et avec des niveaux de résolution spatiale les plus fins.
- **Constat :**
 - L'utilisation d'aggrégats (moyenne, médiane) ou d'indices calculés sur des zones administratives préalablement définies reste la méthode la plus courante.

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

Une revue des performances prédictives de 8 méthodes statistiques

Un premier groupe de méthodes s'appuie sur des statistiques par zones péallablement définies :

1. Statistiques simple par zone (moyenne, mediane, quantile)
2. Régression multivariée (modèle hédonique)
3. Régression multivariée avec autocorrélation spatiale.
4. Machine learning (XGB).

Un deuxième groupe s'appuie sur des estimations locales s'affranchissant de tout découpage spatial préallable :

5. Statistiques simple locale (moyenne pondérée géographique)
6. Régression multivariée locale (GWR)
7. Régression multivariée locale avec autocorrélation spatiale (GWRSAR)
8. Régression multivariée locale avec autocorrélation spatiale (GWRSAR) et segmentation

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

Echantillon de travail

Segment des terrains à bâtir, soit environ 30 000 observations sur la période 2007-2016 en PACA.

Sélection de 10000 observations effectivement bâties après leur vente, ou avec permis de construire (Sitadel).

Performance Prédictive :

MAPE, Le pourcentage d'erreur absolue moyen

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\text{Prix}_i - \text{Prediction}(\text{Prix}_i)}{\text{Prix}_i} \right|$$

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

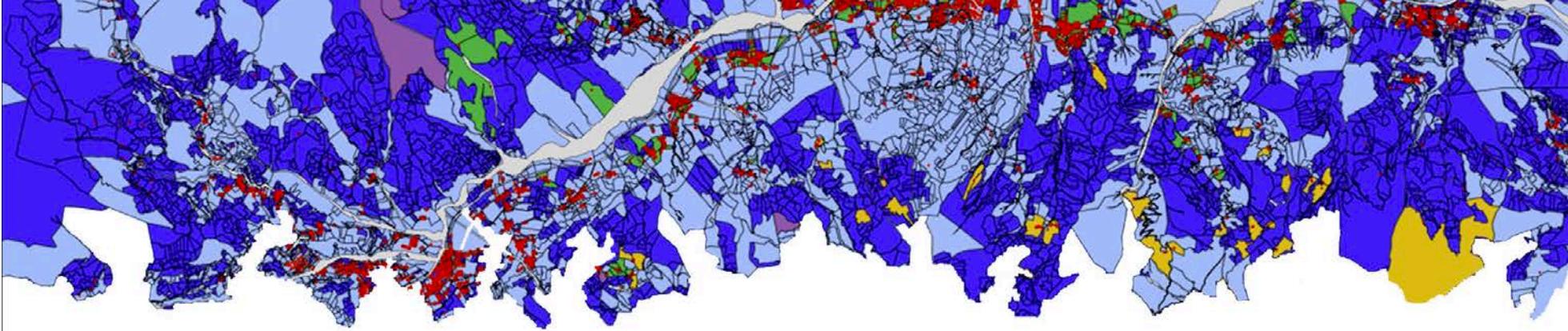
METHODE	MAPE	Remarques	Difficulté
1. statistiques simple par zone (moyenne, mediane, quantile)	55.35 %	On limite à la période 2000-2009 et aux communes avec plus de 10 obs	1 à 2
2. Régression multivariée linéaire (modèle hédonique)	47.49 %	Modèle simple à 6 variables : log(prix)=surface + date + Dpole_urbain + Dnoyau_villageois + Droute_secondaire+erreur (35.14 avec toutes les variables)	3
2. Régression multivariée non linéaire (modèle hédonique)	40.81 %	Modèle simple à 6 variables , Linéaire par morceaux (MARS) (29.44 avec toutes les variables)	4
3. Régression multivariée avec autocorrélation spatiale.	32.36 %	Modèle simple à 6 variables + autocorrélation spatiale de l'endogène	5

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

METHODE	MAPE	Remarques	Difficulté
1. statistiques simple par zone (moyenne, mediane, quantile)	55.35 %	On limite à la période 2000-2009 et aux communes avec plus de 10 obs	1 à 2
2. Régression multivariée linéaire (modèle hédonique)	47.49 %	Modèle simple à 6 variables : log(prix)=surface + date + Dpole_urbain + Dnoyau_villageois + Droute_secondaire+erreur (35.14 avec toutes les variables)	3
2. Régression multivariée non linéaire (modèle hédonique)	40.81 %	Modèle simple à 6 variables , Linéaire par morceaux (MARS) (29.44 avec toutes les variables)	4
3. Régression multivariée avec autocorrélation spatiale.	32.36 %	Modèle simple à 6 variables + autocorrélation spatiale de l'endogène	5
4. Machine Learning (XGB).	28 à 32 %	Modèle limité à 6 variables : Efficace pour les prédictions mais pas interprétable.	4

Les avancées méthodologiques en statistiques spatiales dans l'identification des prix fonciers ou immobiliers.

METHODE	MAPE	Remarques	Difficulté
5. moyenne pondérée géographique	46.52 %	La difficulté est de trouver le bon voisinage (ici noyau gaussien avec seuil à trois voisins)	3
6. Régression multivariée géographique (GWR)	30.75 %	Modèle simple à 6 variables	4
7. GWR + autocorrélation spatiale (GWRSAR)	29.54 %	Modèle simple à 6 variables <i>Geniaux et Martinetti 2017 RSUE</i>	5
8. GWRSAR + segmentation endogène	25.89 %	Modèle simple à 6 variables 3 segments 0-500 m ² ; 500-2500 m ² , plus de 2500 m ² <i>Geniaux et Martinetti 2017 RSUE + Li Racine 2010 Econometric Theory</i>	5



Merci de votre attention



Provence
Alpes
Côte d'Azur



Centre Régional
de l'Information
Géographique

Provence-Alpes
Côte d'Azur