



HAL
open science

Urbansimul, de l'observation à l'action foncière : présentation de l'outil

Ghislain Geniaux

► To cite this version:

Ghislain Geniaux. Urbansimul, de l'observation à l'action foncière : présentation de l'outil. Journée d'études Numérique et Immobilier (NUMIMMO), Sep 2019, Avignon, France. hal-02952536

HAL Id: hal-02952536

<https://hal.inrae.fr/hal-02952536>

Submitted on 29 Sep 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.

Urbansimul, de l'observation à l'action foncière : présentation de l'outil

Ghislain Geniaux,

UR Ecodéveloppement 767, INRA Avignon

1. Panorama du projet UrbanSIMUL
 1. Video
2. Le module marché foncier
3. Le module Dynamique des marchés
4. Modélisation statistique
 1. Les modèles du module PRONOSTIC
 2. Les prix et la simulation de l'étalement urbain
 3. L'effet prix de l'offre foncière



VIDEO :

<https://www.urbansimul.fr/support/presentation>

Panorama du projet UrbanSIMUL

4

- ▶ Un SaaS (software as a service) en ligne, avec plus de 500 utilisateurs.
- ▶ il inclut des données sur 6 millions de parcelles suivies annuellement et couvre 2 régions avec 1 500 municipalités.
- ▶ Il s'agit d'un projet gagnant-gagnant entre les chercheurs et les planificateurs/gouvernements locaux.

The screenshot displays the UrbanSIMUL web application interface. On the left, a 'Fiche d'information' panel provides details for a specific parcel:

UF	UF mobilisable
Surface de l'UF 2869 m ²	
Surface disponible 2852 m ²	
Surface bâti 0 m ²	
Nombre de logements 0	
Date de mutation 19/05/2009	
Nombre de propriétaires 1	
Type de propriétaire PERSONNE PHYSIQUE	

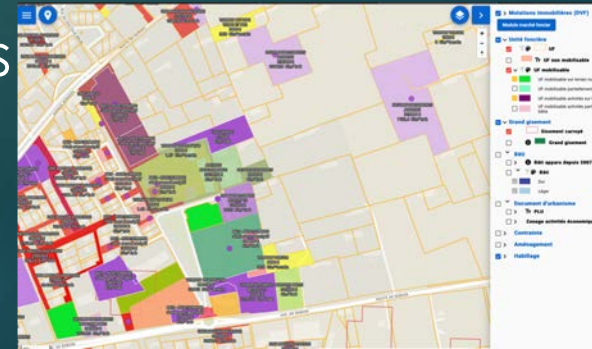
At the bottom of the panel is a 'Fermer' button. The central part of the interface shows a 3D aerial map of a rural area with several parcels highlighted in green and blue. On the right, a legend titled 'UrbanSIMUL : Carto Web' lists various data layers:

- > Mutations immobilières (DVF)
- Module marché foncier
- v Unité foncière
 - UF
 - Tr UF non mobilisable
 - > UF mobilisable
- v Grand gisement
 - Gisement carroyé
 - Grand gisement
- v Bâti
 - Bâti apparu depuis 2007
 - Bâti
 - Dur
 - Léger
- v Document d'urbanisme
 - > Tr PLU
 - > Zonage activités économiques
- > Contrainte
- > Aménagement
- > Habillage

At the bottom of the map, the text 'ON (84007) PLU : 2015 Cadastre : 2018' is visible.

Module marché foncier

- ▶ Améliorer la donnée :
 - ▶ Coupler la données DVF à MAJIC, et autres informations sur les unités foncières d'UrbanSIMUL (zonage, équipement, aménités)
 - ▶ Mieux qualifier les biens et leur segment de marché notamment pour les terrains à bâtir
- ▶ Faciliter la prise en main pour les opérationnels du foncier
 - ▶ En facilitant la sélection et filtrage des biens
 - ▶ Par des statistiques simples
 - ▶ Utile pour le montage de dossier à destination des élus de la justice administrative



Module Dynamique des marchés fonciers



DYNMARK : un outil d'exploration des dynamiques des marchés fonciers et immobiliers à partir des données DVF

Choisissez un territoire, un segment de marché et une période

Territoire :

département

Nom du territoire

06 Alpes-Maritimes

Segment de marché

- Maisons
- Maisons anciennes
- Maisons récentes, neuves ou VEFA
- Appartements
- Appartements anciens
- Appartements récents, neufs, VEFA
- Terrains à Batir

Année de début



Année de fin



Année de référence



Nombre minimal de transaction par an (courbes prix/volume) :

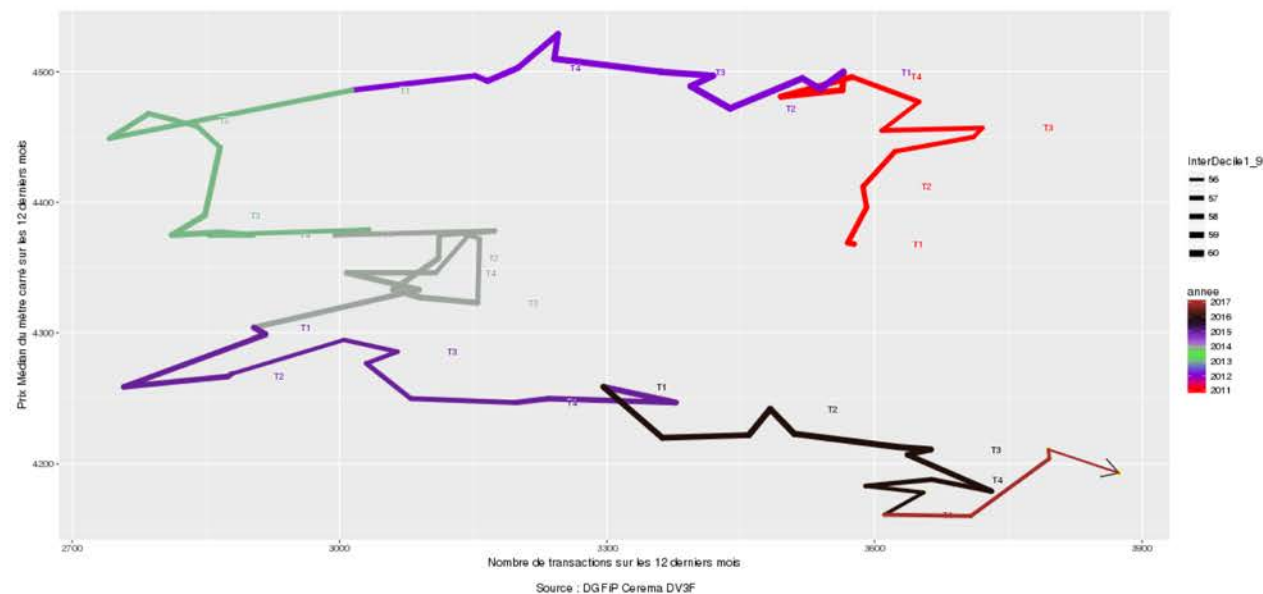


Conception : Ghislain Geniaux (INRA) et Bertrand Leroux (Cerema)

Source : DV3F DGFIP traitement Cerema

Evolution Cartographie Courbes Prix / Volume Dispersion en Euros/m² Dispersion en Euros

Relation Volume/Prix médian Segments des Maisons



Les courbes prix/volume renseignent sur la dynamique des marchés. Chaque point de la courbe correspond aux prix médian en Euros courant et au nombre de biens vendus durant les 12 derniers mois. Un label 'T' permet d'indiquer le trimestre tandis que la couleur permet d'identifier l'année. L'épaisseur du trait correspond à la dispersion des prix ; plus il est épais plus la dispersion interdécile (10-90) des prix est importante.

Quand la ligne va vers le bas le prix baisse.

Quand la ligne va vers le haut le prix augmente.

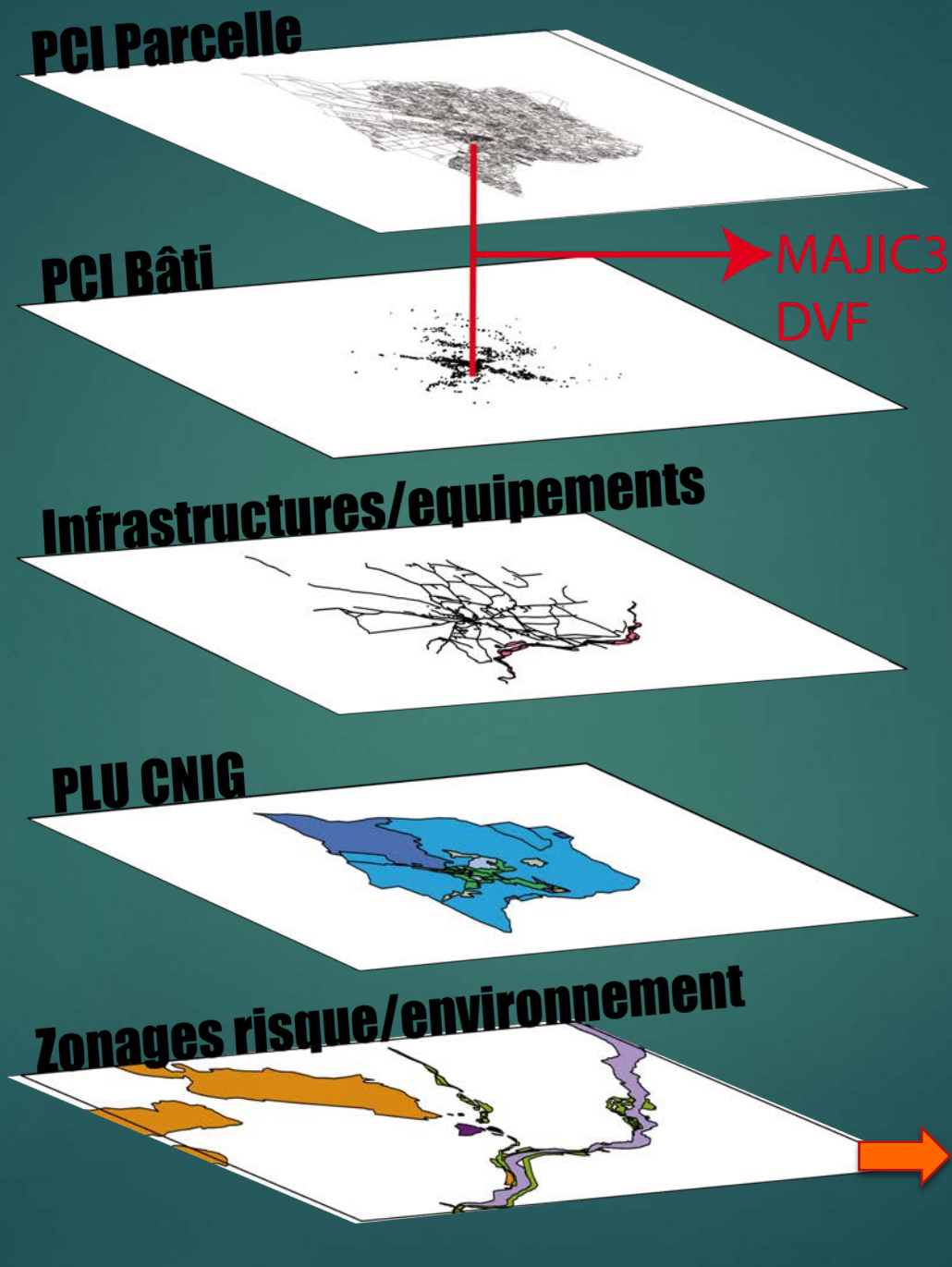
Quand la ligne va vers la gauche le nombre de biens vendus baisse.

Quand la ligne va vers la droite le nombre de biens vendus augmente.

Sur chaque segment, les observations avec les 1% de prix les plus élevés et les plus bas sont retirées.

DONNEES

Base de données
GéoHistorique
2007-2017



- 1 • Très forte contrainte
- 2 • Forte contrainte
- 3 • Contrainte moyenne
- 4 • Faible contrainte
- 0 • Aucune contrainte

DONNEES REMARQUABLES

8

▶ Données exhaustives sur :

▶ Les parcelles et les bâtis

▶ Les prix,

et surtout sur :

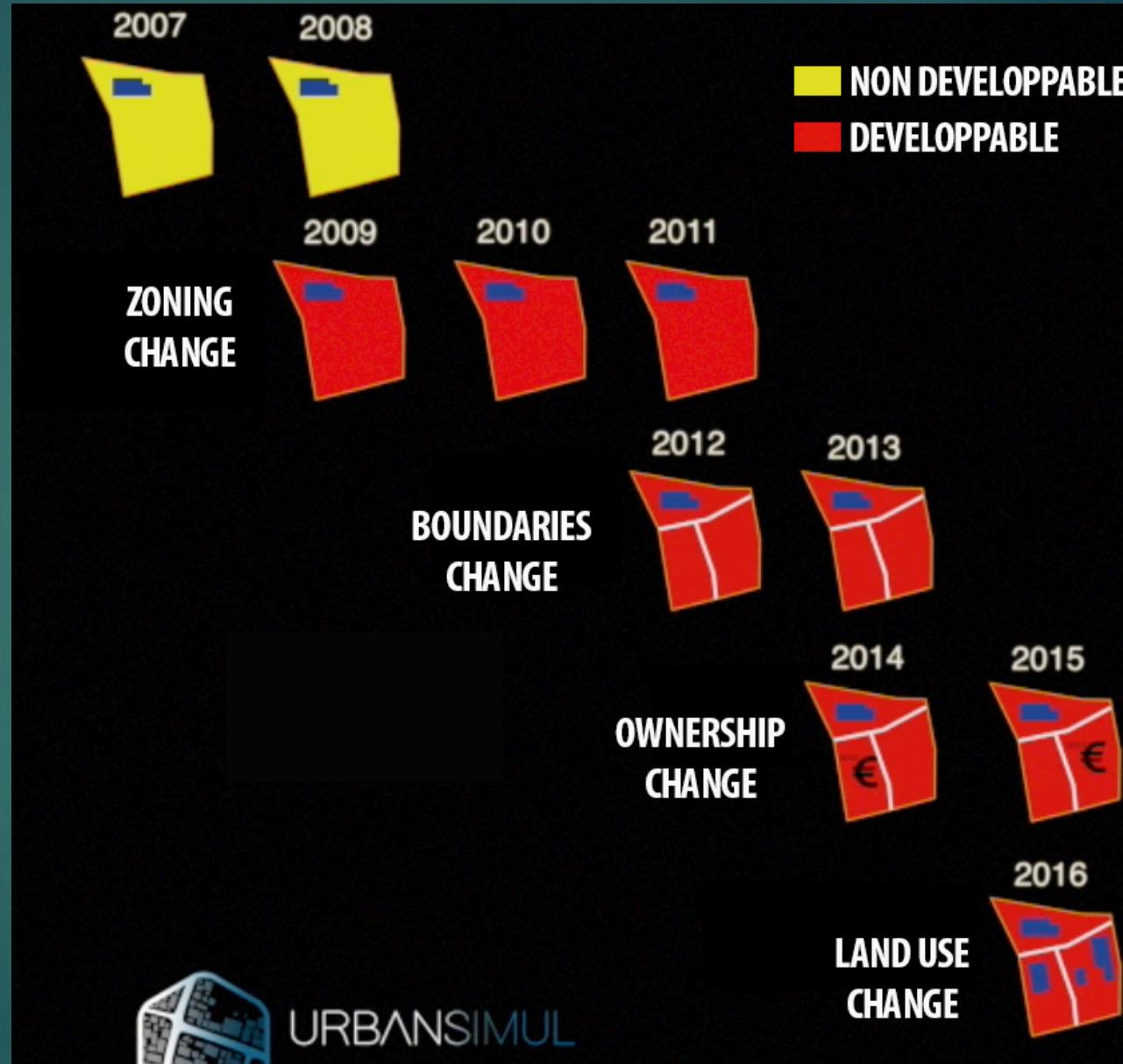
▶ Les propriétaires

▶ La réglementation en matière d'urbanisme



Offre foncière communale et
portefeuille foncier/immobilier

DES DECISIONS DANS L'ESPACE ET LE TEMPS



URBANSIMUL

GHISLAIN GENIAUX & BERTRAND LEROUX

Développements de méthodes statistiques sur données de hautes dimension spatiales

- ▶ **Modèle de choix discret avec autocorrélation spatiale**
 - ▶ Package R *ProbitSpatial* 2016 (7135 téléchargements en 30 mois), Martinetti and Geniaux 2017 RSUE
- ▶ **Hétérogénéité spatiale + autocorrélation spatiale**
 - ▶ Package R *mgwrsar* 2017 (697 téléchargements en 5 mois), Geniaux and Martinetti 2017 RSUE
 - ▶ **+ non linéarité** Package R *mgwrsar* 2018
- ▶ En cours de développement :
 - ▶ **Fast-version of *mgwrsar*, CRAN 12/2019**
 - ▶ **algorithmes de *boosting* pour données spatiales/spatio-temporelles 2020**



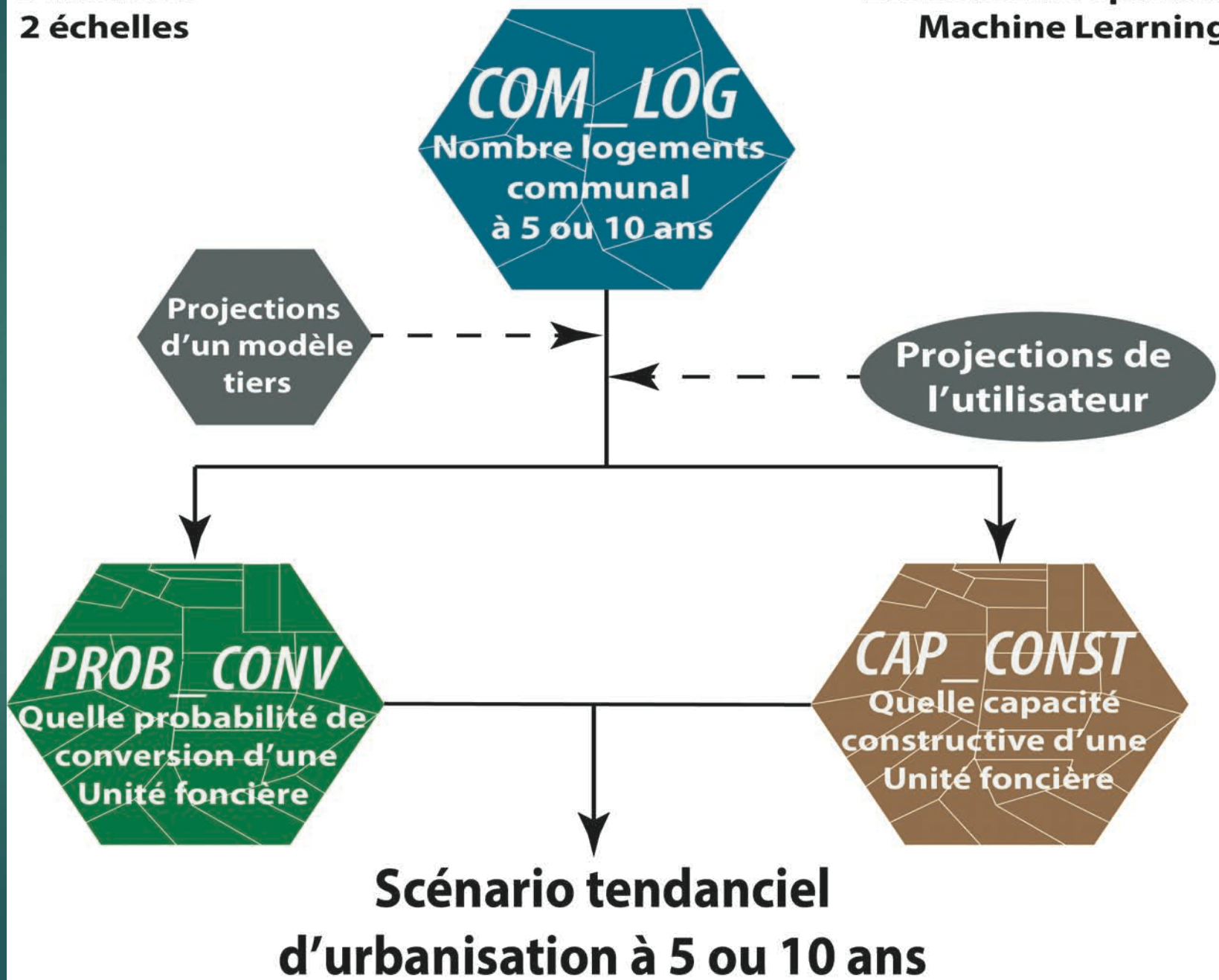
Simulation
de
l'urbanisation

MODULE PRONOSTIC

3 modèles
2 échelles

METHODES

Econométrie spatiale
Machine Learning



Les variables d'intérêt

12

CAP_CONST Les capacités constructives des Unités Foncières :

- ▶ $\text{nblog}(t+8, UF)$ sachant $\text{nblog} > 0$ (multinomial, count data)
- ▶ $\text{surflog}(t+8, UF)$ sachant $\text{nblog} > 0$

PROB_CONV Les probabilités de conversion des Unités Foncières :

- ▶ $\text{Proba}(\Delta \text{surflog}(t+8, UF) > 0)$

PROB_CONV Les probabilités de construction des Unités Foncières.

13

- ▶ Données 2009-2017 : 3 millions d'Ufs, **210 000 Ufs constructibles non bâties**
- ▶ Construction de l'endogène : projection des nouveaux bâtis 2017 sur les UF de 2009
- ▶ X 330 variables initiales sur les Ufs, leur propriétaire, leur environnement, les règles d'urbanisme, les distances minimales à ...
- ▶ 7000 variables de lissage spatio-temporel construites à partir de ces 330 variables initiales
 - ▶ K premiers voisins : 2,5,10,20,50
 - ▶ Noyau bisquare adaptatif
 - ▶ sans distinction, et/ou bâti ou pas et/ou de même type de zonage et/ou de taille comparable, en considérant l'évolution du bâti sur 3, 5, 7 ou 9 ans.

```
Library(mgwrsar)  
W=KNN(coords,k=10,kernel='gauss_adapt')  
WX=W %*% X
```

Méthodes

1. Estimation d'un modèle binomial
2. Utilisation de méthode de *gradient descent boosting* (*xgboost*)
3. Modèle *Probit Spatial* (*ProbitSpatial*)
4. Méthode de Model averaging

Cross validation : binomial prob_conv

15

	Reference	
Prediction	0	1
0	216334	8256
1	3584	17593

Forte proportion
de permis de
construire

Overall Statistics

Accuracy : 0.9583

95% CI : (0.9574, 0.9591)

No Information Rate : 0.934

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6042

Statistics by Class:

Sensitivity : 0.9887

Specificity : 0.5280

Pos Pred Value : 0.9674

Neg Pred Value : 0.7671

Prevalence : 0.9340

Detection Rate : 0.9234

Detection Prevalence : 0.9546

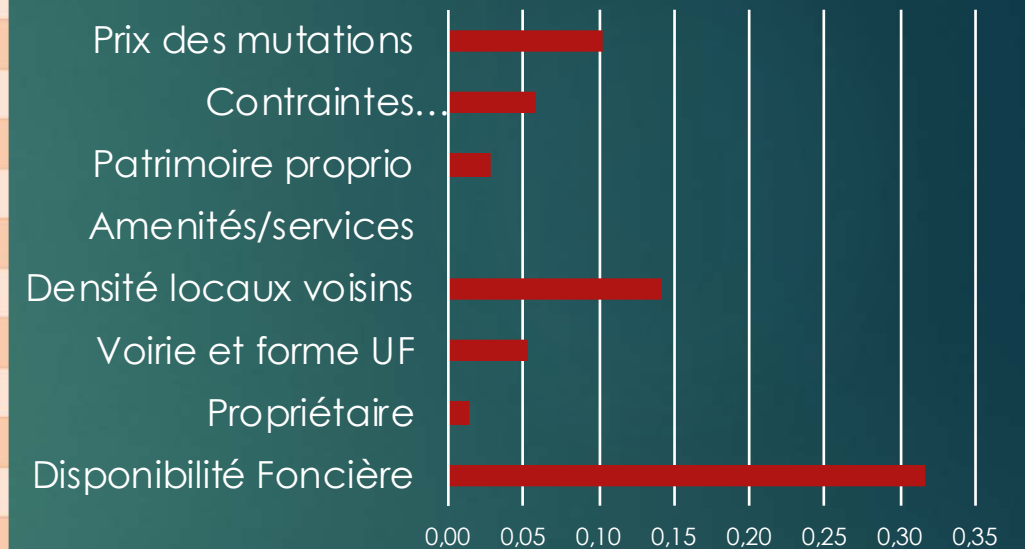
Balanced Accuracy : 0.7583

Les principaux *drivers* des probabilités de construction

16

Variables	Type de variables	Variable Importance
d_bati_pci	densité locaux voisins	0,14080
surfdispo	disponibilité foncière	0,08432
cgrnumdxtTERRAINS_A_BATIR	mutation	0,08106
surfdispo2	disponibilité foncière	0,08021
amutp	mutation	0,06921
surf	disponibilité foncière	0,05205
ze_min1	contrainte réglementaire	0,04381
frrue_rel	Voirie et forme UF	0,02706
prop_spar	patrimoine proprio	0,02101
is	Voirie et forme UF	0,01557
ncont_niv1	contrainte réglementaire	0,01371
mediane	mutation	0,01171
e_m2_terr	mutation	0,01096
ispomobilisable_terrain_nu	disponibilité foncière	0,01001
age	propriétaire	0,00925
prop_sloc	patrimoine proprio	0,00714
WK4_e_m2_terr	mutation	0,00677
long	Voirie et forme UF	0,00592
libniv1PERSONNE_PHYSIQUE	propriétaire	0,00530
surf_plane	disponibilité foncière	0,00480
long2	Voirie et forme UF	0,00468
ccogrmNA	propriétaire	0,00442
nb_pro	propriétaire	0,00439
WK10_r9_ces	disponibilité foncière	0,00414
WK4_amutp	mutation	0,00367

Variable Importance



CAP_CONST Les capacités constructives des Unités Foncières.

- ▶ Données 2009-2017 : 3 millions d'Ufs, 40 000 Ufs constructibles non bâties en 2009 devenues construites en 2017
- ▶ Même lot de variables que pour le modèle de probabilité de conversion.

Cross validation : multinomial 3 classes

Reference

Prediction	1	2-6	>6
1	23228	3720	194
2-6	3644	5122	764
>6	277	542	1810

Overall Statistics

Accuracy : 0.7674

95% CI : (0.7632, 0.7716)

No Information Rate : 0.6908

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.4947

Statistics by Class:

	Class1	Class2	Class3
Sensitivity	0.8556	0.5458	0.65390
Specificity	0.6779	0.8527	0.97758
Pos Pred Value	0.8558	0.5375	0.68847
Neg Pred Value	0.6775	0.8568	0.97388
Prevalence	0.6908	0.2388	0.07043
Detection Rate	0.5910	0.1303	0.04605
Detection Prevalence	0.6906	0.2425	0.06689
Balanced Accuracy	0.7667	0.6992	0.81574

Cross validation : comptage

19

Le modèle de comptage produit des erreurs moyennes à l'échelle de l'unité foncière qui ne sont pas encore pleinement satisfaisante avec une erreur moyenne de la prédiction out-sample à 40 % à l'échelle de l'UF. En revanche, on peut voir que lorsqu'on cumule ces prédictions à des échelles supérieures les prédictions deviennent plus satisfaisantes:

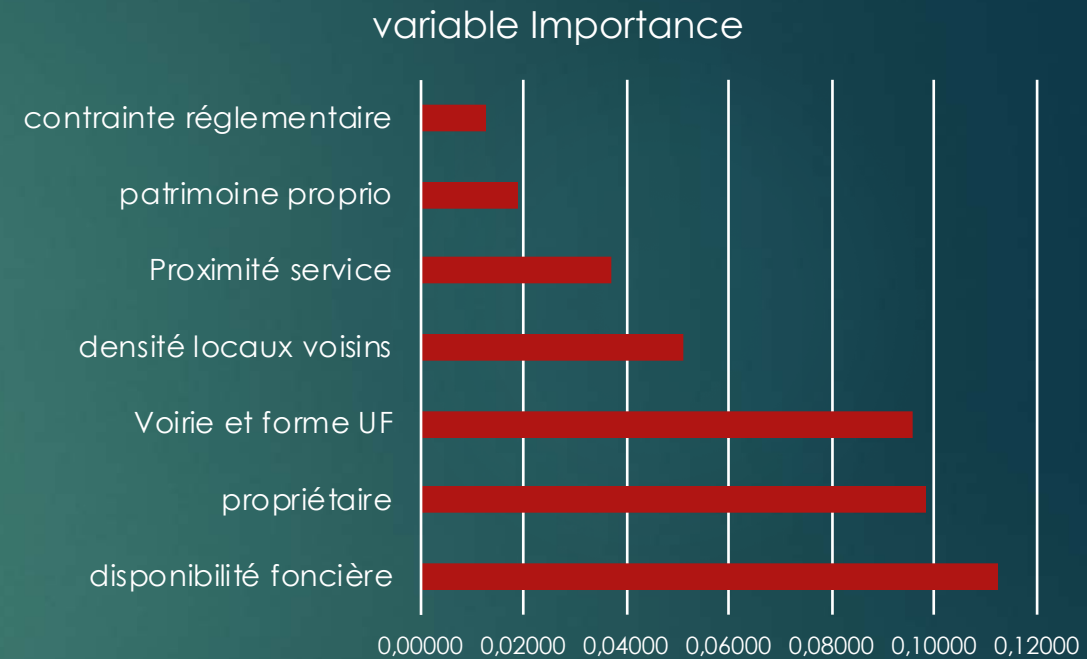
Distribution des erreurs cumulées

Échelle	Commune x zone_plu			Commune :		
Stat	median	mean	p75	median	mean	p75
Err %	0.1852	0.2321	0.3104	0.1005	0.1776	0.2229

Les principaux *drivers* de la capacité constructive

20

Variables	Type de variables	Variable Importance
frrue_abs	Voirie et forme UF	0,07011
ccogrmNA	propriétaire	0,05679
surf_plane	disponibilité foncière	0,04771
surfdispo2	disponibilité foncière	0,03377
d_bati_pci	densité locaux voisins	0,03255
libniv1PERSONNE_PHYSIQUE	propriétaire	0,02858
frrue_rel	Voirie et forme UF	0,01866
surf	disponibilité foncière	0,01820
d_pharmacie	Proximité service	0,01565
WK50_nlocappt	densité locaux voisins	0,01352
ncont_niv1	contrainte réglementaire	0,01272
surfdispo	disponibilité foncière	0,01270
prop_spar	patrimoine propriétaire	0,01231
WK50_nblocal	densité locaux voisins	0,01202
d_ecole_ma	Proximité service	0,01186
amutp	Date de mutation	0,01104
d_ecole_el	Proximité service	0,00975
dlogtuf_com.mediancom	densité locaux voisins	0,00726
WK4_frrue_rel	Voirie et forme UF	0,00684
age	propriétaire	0,00655
prop_sloc	patrimoine propriétaire	0,00646
dlogtuf_complu.median	densité locaux voisins	0,00624
WBR9S0.85_dlogtuf	densité locaux voisins	0,00616
nb_pro	propriétaire	0,00613
WBR5S0.85_dlogtuf	densité locaux voisins	0,00598



LOG_COM La démographie et le parc de logement

Le nombre de logements, les prix et la population à la commune (~ Jeanty Partridge & Irwin 2010):

$$\begin{cases} \Delta \text{pop}_c = f(X_c, \Delta \text{logt}_c, \Delta \text{Indiceprix}_c, \Delta \text{OffreFoncière}_c) \\ \Delta \text{logt}_c = g(X_c, \Delta \text{pop}_c, \Delta \text{Indiceprix}_c, \Delta \text{OffreFoncière}_c) \\ \Delta \text{IndicePrix}_c = h(X_c, \Delta \text{logt}_c, \Delta \text{pop}_c, \Delta \text{OffreFoncière}_c) \\ \Delta \text{OffreFoncière}_c = k(X_c, \Delta \text{logt}_c, \Delta \text{pop}_c, \Delta \text{IndicePrix}_c) \end{cases}$$

Avec $f()$, $g()$, $h()$, $k()$ calibrés à partir de méthodes de boosting.



**MODELE
COM_LOG**

$$\begin{cases} \Delta^M pop_c^{t+h} = Fpop(X_{c,t}, h) \\ \Delta^M log_c^{t+h} = Flog(X_{c,t}, \Delta pop_c^{\hat{t+h}}, h) \end{cases}$$



$$\Delta^M \hat{log}_c^{t+h}$$

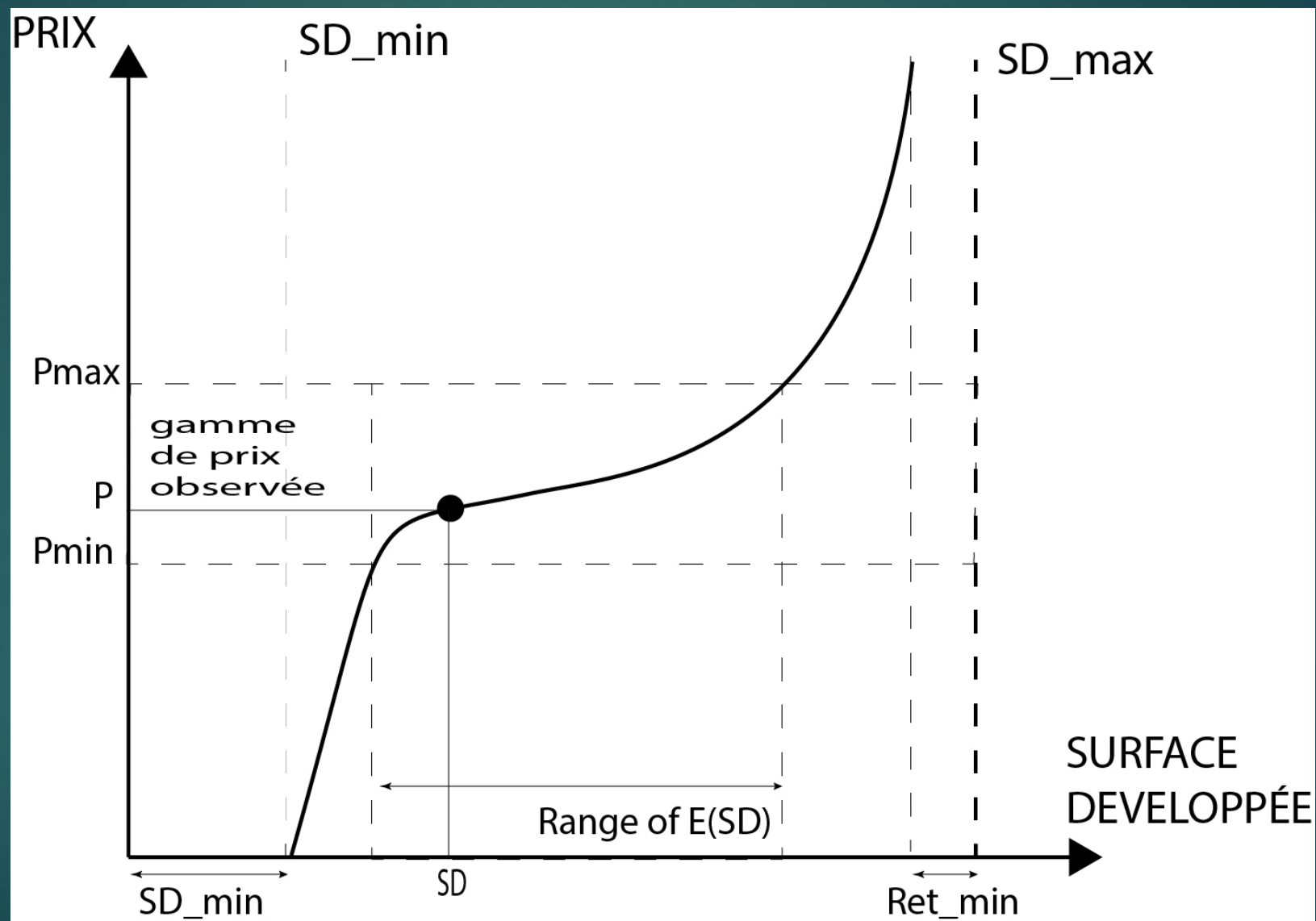
**Macro : Nombre de nouveaux
logements sur la commune c**

**Macro : Nombre de
nouveaux
logements sur la
commune c**

**micro : Nombre de
nouveaux
logements sur la
commune c**

?

DISCUSSION



DISCUSSION

$\exists \forall c, \bar{P}_c$ tel que

$$\sum_c \Delta^M \hat{\log}_c^{t+h} = \sum_c \Delta^m \log_c^{\hat{t}+h}(\bar{P}_c)$$



**MODELE Ajustement
micro/macro**



$$\tilde{P}_c = \arg \min_{\bar{P}_c} \sum_c (\bar{P}_c - Cost(\tilde{P}_c))$$

tel que $\forall c,$

$$\Delta^M \hat{\log}_c^{t+h} = \Delta^m \log_c^{\hat{t}+h}(\tilde{P}_c)$$

Avec $Cost(P_c) = P_c * OC$
Et $OC =$ Ordre de Contiguité

Coté évaluation des politiques publiques.

- ▶ Effet prix de l'offre foncière (2015),
 - ▶ actualisation du papier sur données 2019 avec un plus gros échantillon,
 - ▶ Utilisation de méthode contrefactuelle (contrôle synthétique)
- ▶ Effet du patrimoine et prix sur les décisions de zonage (suppression des zones NB).
 - ▶ Probit spatial : probabilité d'être du bon coté du trait selon son patrimoine et sa valeur.

Je vous remercie de votre attention

Ces travaux ont bénéficié de fonds issus d'un partenariat entre les institutions suivantes :

