



**HAL**  
open science

## Incertitudes liées aux échelles d'estimation des prix immobiliers

Didier Josselin, Delphine Blanke, Mathieu Coulon, Guilhem Boulay, Laure Casanova Enault, Antoine Peris, Pierre Le Brun, Thibault Lecourt

► **To cite this version:**

Didier Josselin, Delphine Blanke, Mathieu Coulon, Guilhem Boulay, Laure Casanova Enault, et al.. Incertitudes liées aux échelles d'estimation des prix immobiliers. Mireille Batton-Hubert; Eric Desjardin; François Pinet. Imperfection et information géographique 2 – Cas d'utilisation, ISTE Editions, 2024, Systèmes d'information, web et société, 9781784056254. hal-04665618

**HAL Id: hal-04665618**

**<https://hal.science/hal-04665618v1>**

Submitted on 31 Jul 2024

**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.

Copyright

DRAFT

July 31, 2024

---

# Incertitudes liées aux échelles d'estimation des prix immobiliers

---

Didier Josselin\*, Delphine Blanke\*\*, Mathieu Coulon\*, Guilhem Boulay\*, Laure Casanova Enault\*, Antoine Peris\*, Pierre Le Brun\*, Thibault Lecourt\*

\*UMR ESPACE 7300 \*\*Laboratoire de Mathématique d'Avignon

### 6.1. Introduction

Après être longtemps restées inaccessibles [BOU 12], les données sur les transactions foncières et immobilières sont désormais très facilement disponibles [CAS 19, JUI 19]. Conjugué au développement du secteur immobilier en ligne [LEF 15, SIT 17], cet état de fait a conduit à une diffusion massive des estimations de prix immobiliers, de nombreux sites prenant désormais le relais des marronniers de la presse écrite sur “les prix dans votre ville” [BOU 21]. Ainsi, que l'on soit professionnel de l'immobilier, particulier en recherche de bien ou acteur public en procédure d'expropriation, on sera amené à utiliser ces données pour s'informer sur la valeur d'un bien immobilier particulier, dans un secteur géographique donné.

Cet accès immédiat à l'information foncière et immobilière, s'il présente de nombreux aspects positifs, pose malgré tout des questions de fond : sur quelles méthodes les estimations des prix se fondent-elles ? A quelles conditions sont-elles valables ? Peu de sources donnent de telles informations, à tel point qu'on peut paradoxalement voir dans l'accessibilité accrue des estimations de prix immobiliers une transition entre l'ancienne “opacité de marché” (liée à l'absence d'information facilement accessible sur les marchés immobiliers) et l'actuelle “opacité méthodologique” (liée à l'absence de métadonnées ou de précautions d'usage des références de prix en ligne [BOU 21]).

Un enjeu méthodologique clé qui doit être éclairci est celui de la dimension géographique des marchés immobiliers : tout bien est localisé et sa valeur est fortement dépendante de son environnement – et ce à différentes échelles. Or, très souvent, l'estimation de prix locaux, ciblée dans le temps et dans l'espace géographique, aboutit à leur généralisation à d'autres échelles, plus vastes et parfois floues : le *quartier*, la *commune* ou la *région* par exemple. Cette utilisation non réfléchie des échelles, tout comme la généralisation des estimations d'une échelle à une autre, pose pourtant un grave problème au géographe, tant du point de vue méthodologique que de celui de l'étude des mutations des territoires : rien ne dit en effet que lesdites échelles de représentation des marchés immobiliers soient celles auxquelles ces derniers se structurent réellement. Ces représentations peuvent être cartographiques et/ou mentales, avec une convergence forte des deux acceptions, tant l'entité géographique délimitée renvoie à une perception partagée du territoire, et, en conséquence, à une valeur qu'on attribue aux biens échangés dans cette entité.

Les transactions immobilières, plutôt rares dans l'espace géographique, car en faible densité localement sur de courtes périodes en France métropolitaine (en dépit d'un nombre moyen supérieur au million par an ces dernières années), sont un point d'entrée privilégié pour étudier nombre de phénomènes de fond : ségrégation socio-résidentielle, inégalités socio-économiques dans les territoires, impact des politiques publiques sur les tissus urbains, etc. [LEG 21]. À partir de ces sources, il est possible d'évaluer l'effet potentiel de l'échelle géographique de restitution d'une estimation statistique de valeur immobilière et d'en tirer des recommandations quant aux échelles idoines de représentation de ces mouvements de marché. Est-ce que les références de prix dépendent des découpages administratifs ? Lorsqu'on agrège des transactions immobilières en fonction de différents découpages administratifs, les références de prix immobiliers varient-elles ? Le cas échéant, de quelle façon ? C'est ce que nous proposons d'étudier dans ce chapitre, en tirant parti de la possibilité d'estimer les prix à différentes d'échelles et rechercher, parmi celles-ci, celle qui semble la plus pertinente pour rendre compte des valeurs immobilières dans leur voisinage, indépendamment d'une maille territoriale qui nous serait imposée par les producteurs de données ou par une demande sociale particulière.

Ces questions renvoient finalement au problème de l'*incertitude*. Dans le cas particulier des estimations de prix immobiliers, cette dernière résulte de différentes sources : erreurs dans la saisie des données elles-mêmes, représentativité des échantillons de transactions pour estimer la valeur du stock et méconnaissance de la distribution statistique et spatiale des prix. Les deux premières sources d'incertitude ont été identifiées de longue date et sont relativement faciles à maîtriser. Les erreurs de saisie peuvent être corrigées par la détection de valeurs aberrantes [KON 06]. La correction de la représentativité de l'échantillon de transactions peut se faire par le recours à des méthodes hédoniques, qui prennent en compte les caractéristiques des biens et fournissent des estimations pour un "panier de biens" représentatif de la zone étudiée

[CAI 19]. En revanche, la dernière source d'incertitude, la distribution "statistico-spatiale" des prix, est encore largement méconnue, car trop peu souvent identifiée. Traiter de la question des échelles d'agrégation des prix immobiliers constitue donc un enjeu prioritaire d'investigation, à l'heure où la diffusion massive des représentations cartographiques et statistiques des prix immobiliers interfère largement avec le poids macro-économique croissant de l'immobilier [AAL 14].

Ainsi, nous soulignerons dans un premier temps l'insuffisante prise en compte du choix de l'échelle d'agrégation des prix dans les représentations des marchés immobiliers. Dans un deuxième temps, nous présenterons la méthodologie d'estimation de l'effet du support spatial sur les estimations des prix immobiliers, à travers deux panels de maillages territoriaux, l'un administratif et l'autre fonctionnel. Dans un dernier temps, enfin, nous présenterons et discuterons les résultats obtenus à travers plusieurs échelles de représentation des prix immobiliers.

## 6.2. L'effet du support spatial dans l'estimation des prix immobiliers

### 6.2.1. La génération d'incertitude dans les choix d'échelles d'agrégation

Le *Change of Support Problem* ou COSP [KIN 97, KIN 04], défini par Stan Openshaw comme le *Modifiable Areal Unit Problem* (MAUP) en géographie [OPE 84], découle directement de la partition des données spatiales en fonction d'un certain nombre d'éléments caractérisant l'espace géographique concerné : nombre d'échantillons, nombre d'entités par échantillon, limites des unités spatiales de ces échantillons. Potentiellement, il existe une infinité de découpages spatiaux dont la structure impacte directement les données statistiques manipulées. Cet effet de zonage est exacerbé lors de l'agrégation des valeurs mesurées à différents niveaux (les entités du niveau supérieur contiennent, dans la plupart des cas, les entités du niveau inférieur). Très concrètement, ces effets peuvent être observés sur les données immobilières que nous manipulons dans deux cas de figure : l'*upscaling* [JOS 19] et le *resampling* [LOU 16].

Les données de transaction immobilière ont pour particularité d'être attachées à des entités surfaciques – les parcelles – dont l'ensemble forme une *partition*. *Stricto sensu*, une partition correspond à une configuration dans laquelle toute donnée localisée appartient à une seule et unique entité. D'un point de vue spatial, cela signifie qu'il n'existe pas de superposition entre entités (sous-échantillons spatiaux), ni de *trous* dans la partition spatiale (exception faite du domaine public non cadastré, mais sur lequel aucune transaction immobilière à titre onéreux n'est légalement possible). Dans le premier cas, l'*upscaling* consiste à agréger les informations individuelles dans les différentes entités, de plus en plus généralisantes, qui composent la partition à une échelle donnée. Le second cas renvoie aux situations où les différents niveaux d'entités peuvent se superposer partiellement, sans parfaitement s'imbriquer hiérarchiquement. On parle alors de *resampling* puisque chaque partition correspond à un

nouveau découpage de l'espace. Dans les deux cas, à chaque échelle correspond une partition.

Les données immobilières sont susceptibles de relever tant du cas de l'*upscaling* que de celui du *resampling*. En effet, les transactions immobilières sont des individus statistiques enregistrés à l'échelle la plus fine qui soit – la parcelle cadastrale – et sont en général ensuite agrégées et restituées au public à des échelles plus larges : du quartier à la région, voire au pays, en passant par la commune ou le bassin de vie. Ce processus de production de l'information statistique sur les marchés immobiliers, globalement toujours ascendant, mobilise deux types de jeux de mailles. Le premier, fondé sur des périmètres administratifs, correspond à une structure hiérarchisée : les parcelles cadastrales s'imbriquent dans les sections cadastrales, qui s'imbriquent elles-mêmes dans les communes, lesquelles s'imbriquent dans les cantons, et ainsi de suite jusqu'aux régions. Le second jeu de mailles, qui renvoie plutôt à des périmètres fonctionnels et/ou vécus, ne correspond pas à ce cas de figure. En effet, si certains périmètres s'imbriquent (par exemple, les communes dans les structures intercommunales), ce n'est pas le cas de tous. Par exemple, les EPCI – Etablissements Publics de Coopération Intercommunale – ne s'agrègent pas hiérarchiquement (c'est à dire entièrement, en bloc) dans les aires d'attraction des villes, ni dans les départements. Même si les limites géographiques pourraient parfaitement coïncider, les communes les composant peuvent appartenir à différents départements. Ce cas particulier existe, mais demeure rarissime dans les partitions administratives (exemple de l'enclave des papes du Vaucluse). Les différentes mailles que nous utilisons ici mêlent ces deux types de partitionnement spatial. Cela nous permettra d'évaluer l'impact potentiel du découpage, selon s'il a été construit avec un objectif administratif, fonctionnel ou mixte (administratif et fonctionnel).

### **6.2.2. Des échelles de représentation des prix immobiliers peu questionnées**

Un rapport utilitaire à une donnée prétendument neutre [DES 01, GIT 13] se généralise avec le développement des études sur l'immobilier et de la *propotech* (contraction de “property” et “technology”, regroupant les nouveaux acteurs et start-up entendant révolutionner le secteur de la pierre et de l'immobilier), qui peut parfois réduire la production de référence de prix à un rapide calcul, occultant alors la robustesse des indicateurs, sensible à leur échelle d'agrégation [LOU 16, LEG 21]. Ainsi, bon nombre des représentations cartographiques des prix immobiliers ne questionnent pas véritablement le choix de telle ou telle échelle. On trouve principalement trois modalités de représentation des marchés immobiliers :

- 1) Des représentations du marché par points, comme sur les applications de visualisation de la base Demande de Valeurs Foncières (DVF), récemment mise à disposition de tous [CAS 19]. Dans ce cas-là, la question de l'échelle de l'agrégation ne se pose

pas directement, ce qui ne signifie pas pour autant que ce choix de représentation soit bon dans l'absolu, puisqu'il échoue à simplifier efficacement l'information statistique.

2) Des représentations à une seule échelle (souvent le quartier ou la commune, plus rarement l'EPCI), qui ne sont en général pas explicitement justifiées. On peut faire l'hypothèse que ces choix résultent de deux attitudes :

- une attitude "naïve", fondée sur le recours à un périmètre connu de tout le monde et qui semble doté d'une certaine "évidence" ; certains professionnels créent également leurs propres partitions par quartier en associant à chacun un nom supposé signifiant pour leurs clients ;

- une attitude plus pragmatique dictée par la disponibilité d'autres sources de données, permettant d'enrichir l'analyse des transactions ; l'IRIS (ou Îlots Re-groupés pour l'Information Statistique, découpage territorial infra-communal proposé par l'INSEE, en vigueur depuis 1999) est, par exemple, très fréquemment utilisé dans cette optique, notamment par des sites commerciaux spécialisés.

3) Enfin, certains supports proposent de visualiser les prix à différentes échelles, qui correspondent souvent à des découpages non explicitement justifiés et/ou imposés par le recours à d'autres sources statistiques. Sur de nombreux sites, commerciaux ou professionnels, il est ainsi possible de visualiser les prix à l'échelle des régions, des départements, des communes ou des IRIS, par exemple. C'est dans cette approche que cet article se positionne, avec une mesure quantitative systématique et comparative à travers les échelles, qui va au-delà des représentations non explicitement justifiées.

Il convient toutefois d'assortir ce constat de plusieurs remarques. Tout d'abord, les échelles de représentation considérées par les professionnels varient selon les marchés et sous-marchés sur lesquels ils exercent leur activité.

Certains d'entre eux peuvent ainsi privilégier la représentation par points dans des territoires (communes rurales, petites villes) ou de sous-marchés (l'immobilier haut de gamme) à faible nombre de transactions, pour lesquels l'agrégation présenterait des effets trop déformants. Autre exemple, les découpages à caractère administratif (communes et, dans une moindre mesure, EPCI) sont préférés dans la représentation des marchés du neuf, en raison de la place importante qu'occupent les décideurs publics locaux dans leur fonctionnement (élaboration des Plans Locaux d'Urbanisme (PLU) ou PLUI (PLU Intercommunaux), validation des permis de construire, etc.). À l'image du balancement entre attitudes naïve et pragmatique, les professionnels semblent alterner entre critères commerciaux et rigueur statistique.

Ensuite, la mobilisation arbitraire d'échelles de représentation des marchés immobiliers n'est pas propre à la presse ou aux professionnels du secteur. Elle est également largement répandue dans les publications scientifiques. Les chercheurs partagent d'ailleurs certaines contraintes avec les professionnels du secteur, susceptibles de guider leurs choix de mailles. Ces choix découlent en général des objectifs en partie contradictoires de disposer, d'un côté, de la maille la plus fine et, de l'autre, de la

possibilité de croiser les données immobilières avec d'autres bases agrégées et/ou de respecter des seuils de quantité minimale de ventes par unité spatiale, notamment pour des raisons de significativité statistique ou de confidentialité. Toutefois, l'absence de préoccupation commerciale amène les universitaires à mobiliser certains découpages ignorés par les professionnels, à l'image du carroyage de l'INSEE [BOU 11, LEC 19], que ces derniers jugent souvent illisibles pour leurs clients. Pourtant, si ce carroyage ne résout pas les questions de variation de taille d'échantillonnage impactant les statistiques agrégées, ils a le mérite de fournir un découpage spatialement homogène.

Enfin, la mobilisation non questionnée d'échelles de représentation du marché n'exclut pas pour autant l'existence d'une multitude d'autres méthodes qui prennent en compte le contexte local des prix immobiliers. On peut par exemple mentionner les méthodes de lissage spatial qui, en intégrant des effets de voisinage, permettent d'atténuer l'incertitude liée à l'impact ponctuel de valeurs extrêmes ou du manque de données [LEG 21]. Ces méthodes ne permettent cependant pas de résoudre les effets du MAUP, car si le lissage donne l'impression d'une certaine continuité, celui-ci reste calculé à partir de données agrégées à des échelles discutables. Une autre manière de limiter l'incertitude est de recourir à des périmètres ad-hoc qui visent à maximiser la ressemblance des entités agrégées à l'intérieur de mailles supérieures [FEU 21]. Ces méthodes, qui visent à produire des maillages pertinents pour l'analyse des marchés, ont notamment été mobilisées dans des études portant sur les prix du logement, du foncier ou des loyers du parc locatif [DUR 19, BRE 20, CAS 22]. Enfin, la famille des méthodes de régressions locales permet, elle aussi, de tenir compte de la détermination multi-niveaux des prix. Certaines applications mobilisent par exemple des maillages ad-hoc supra-communaux afin d'assurer un échantillon de transactions suffisant, mais appliquent des effets fixes communaux afin de tenir compte des deux échelles dans la modélisation des prix [DUR 19, BRE 20]. Dans cette même veine, les développements méthodologiques récents portant sur les régressions multiples géographiquement pondérées multi-échelles (*MGWR*) permettent eux aussi de mesurer et analyser la variabilité spatiale des processus influençant les prix immobiliers à différents échelles géographiques [SAC 22].

### **6.2.3. Différentes échelles de structuration des prix immobiliers**

Les marchés se structurent et se segmentent à différentes échelles du fait de la complexité des contextes spatiaux qui influencent la dynamique des prix. À l'échelle nationale, on observe par exemple des différences marquées entre les régions, en grande partie en fonction de leur attractivité. La productivité [LAF 03] et la fonction touristique [BUH 06] semblent notamment jouer un rôle crucial dans la structuration des prix à cette échelle large. Ces grands facteurs expliquent par exemple les très forts différentiels de prix entre les très chères régions Île-de-France ou PACA et la diagonale des faibles densités, en déprise démographique et industrielle depuis plusieurs décennies [OLI 16], caractérisée par des niveaux de prix particulièrement bas. Le ciblage



géographique de certaines politiques publiques alimente également une différenciation locale des prix. Les aides fiscales à l'investissement locatif des ménages, inflationnistes [BON 19], ne sont accessibles que dans un nombre de plus en plus restreint de communes [LEB 22]. À des échelles plus fines, infra-régionales ou départementales, d'autres effets de contexte interviennent et déterminent souvent des gradients de prix : les prix décroissent ainsi à mesure que l'on s'éloigne de la frontière suisse ou d'un littoral, ou enfin quand on s'éloigne des grandes agglomérations. La segmentation des prix s'observe également à l'échelle intra-urbaine. Des gradients de prix, mais également des discontinuités très nettes, peuvent découler de la ségrégation socio-résidentielle. L'appartenance à une commune générera d'importants effets-prix, qui se différencieront encore à des échelles plus fines comme le montre par exemple le traditionnel différentiel de prix entre les deux rives de la Garonne à Bordeaux. À l'échelle intra-urbaine également, les zones du renouvellement urbain – Zones d'Aménagement Concerté (ZAC), zones conventionnées avec l'Agence Nationale pour la Rénovation Urbaine (ANRU), Quartiers Prioritaires de la Politique de la ville (QPV), opérations de revitalisation de territoire – ont des effets nets sur les structures locales du marché (prix maîtrisés ou abaissés par rapport voisinage, surtout dans les cas de quartiers stigmatisés). Enfin, la proximité d'infrastructures et d'aménités environnementales peut aussi jouer de manière positive ou négative sur la valorisation immobilière. La littérature sur la modélisation hédonique des prix s'est par exemple intéressée à l'effet de la proximité de transports publics [MAL 09], de parcs urbains [TRA 13], de grands axes routiers [BR18] ou de la carte scolaire [FAC 09] sur les prix de l'immobilier. Des effets de discontinuité dans les prix de l'immobilier peuvent ainsi parfois s'observer très localement, à l'échelle de quartiers, voire parfois de rues [GUÉ 09].

Dans un contexte applicatif, plusieurs échelles peuvent être intéressantes, en fonction de la problématique, selon le facteur influençant la structuration des prix. Que l'on s'intéresse à des niveaux *macro*, *méso* ou *micro*, il faut pouvoir déterminer quelle maille est la plus adéquate pour étudier le niveau en question. C'est notamment la conclusion de plusieurs travaux de recherche qui, depuis les articles fondateurs de M. Straszheim et R. Palm sur San Francisco [STR 75, PAL 78], ont insisté sur l'existence de sous-marchés au sein des espaces urbains, régulièrement prouvés depuis [WAT 01, BOU 11]. Cette mise en évidence de la coexistence de différents sous-marchés à différentes échelles met en avant la dimension théorique et problématique du choix de la maille, puisqu'elle sous-entend que derrière le *continuum* apparent des prix dans l'espace se cache en fait "une structure d'exclusion mutuelle de sous-marchés hiérarchisés" [TOP 84]. En fonction d'une maille choisie, que l'on peut considérer comme "adéquate" dans un contexte et pour un objectif donnés, on discrétise les gradients de prix, révélant différentes structures spatiales. Nous verrons que ces discrétisations spatiales induisent une variabilité non négligeable des estimations des prix immobiliers.

### 6.3. Données et indicateurs pour estimer la sensibilité des prix immobiliers à l'échelle d'agrégation

#### 6.3.1. Différentes mailles territoriales pour tester les effets de l'agrégation

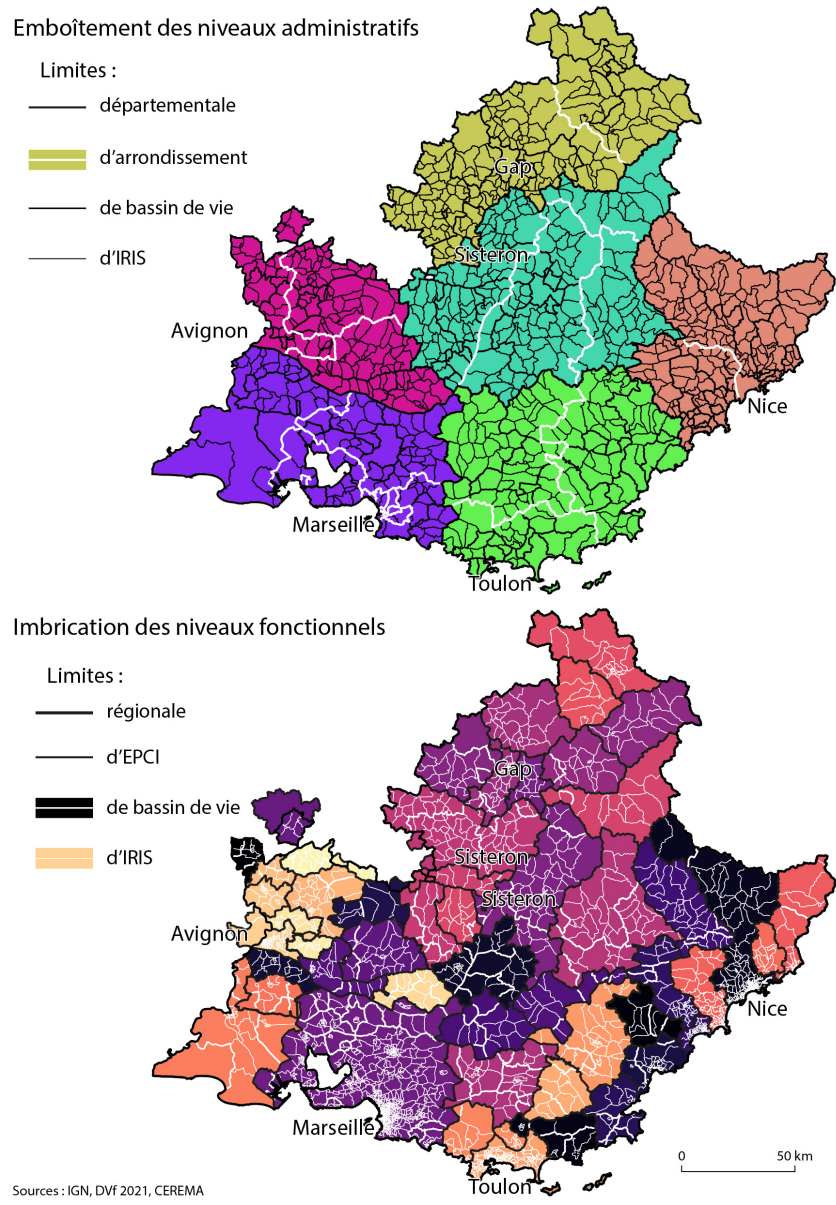
Dans l'optique d'évaluer au mieux l'effet du MAUP sous l'angle de l'*upscaling* et du *resampling*, on utilise ici un panel de 10 mailles s'imbriquant (ou pas, dans le cas du *resampling*), les unes dans les autres. L'expérimentation est réalisée sur le périmètre de la région Provence Alpes Côte d'Azur de 2014 à 2020, afin de disposer d'un corpus suffisamment important et compatible avec des temps de calcul supportables, soit 165 816 transactions immobilières concernant les maisons et 362 249 pour les appartements. La Figure 6.1 illustre l'imbrication des découpages administratifs, de la parcelle à la région et situe les EPCI, les bassins de vie et les IRIS de l'INSEE. La Figure 6.2 place les carreaux de l'INSEE de 1km par rapport aux entités administratives.

Voici la liste des découpages utilisés, de la plus petite (échantillons à effectif élevé) à la plus grande maille (échantillons à faible effectif) :

1) La parcelle cadastrale est l'unité de base du cadastre. Elle appartient au même (groupe de) propriétaire(s) qui bénéficie de droits réels sur cette portion de territoire. Elle peut supporter (ou pas) un (ou plusieurs) logements, mais n'équivaut pas nécessairement à la "maison", telle qu'on l'entend dans le langage courant (puisque la propriété de quelqu'un peut porter sur plusieurs parcelles, y compris attenantes). Pour ce niveau particulier où, dans les déclarations des notaires, une même valeur de transaction immobilière peut être dupliquée sur plusieurs parcelles, nous avons effectué un tri éliminant tous les doublons. Nous disposons à ce niveau de 145 144 parcelles de maisons ayant muté (un peu plus de 7% des biens ont changé de main plusieurs fois sur la période) et 76 275 parcelles concernant les appartements (les ventes peuvent regrouper plusieurs appartements, identifiés par transaction et associés à la même parcelle).

2) La section cadastrale – qui n'existe que dans les communes les plus grandes – regroupe plusieurs parcelles cadastrales au sein d'une même commune (parcelles cadastrales, sections cadastrales et communes s'imbriquant donc). Il s'agit d'un découpage proprement administratif, qui reflète l'évolution de l'occupation du sol (de nouvelles sections sont régulièrement créées quand une commune s'urbanise, par exemple). Les parcelles s'agrègent respectivement, pour les maisons et les appartements, dans 11 803 et 7 089 sections, les appartements étant plus concentrés en zone urbaine dense.

3) Le carreau d'1 km<sup>2</sup> (au nombre de 9 645 pour les transactions concernant les maisons et 4 470 pour les appartements) se démarque par sa "neutralité" et son caractère arbitraire, contrairement aux maillages administratifs par exemple" et protège ainsi des "biais dus aux inégalités de forme et dimension des mailles" [PUM 97], sans pour autant être une solution au MAUP [BOU 11], à cause des variations de taille des



**Figure 6.1.** Découpages administratifs imbriqués en région Provence Alpes Côte d'Azur : de la parcelle à la région, via des découpages administratifs et/ou fonctionnels

échantillons dans chaque maille. Le nombre relativement moins élevé de carreaux de l'INSEE que de sections s'explique, d'une part, par le fait que les maisons éparses isolées ne sont pas prises en compte dans les mailles, et d'autre part parce que les petites sections sont plus fréquentes en zones urbaines, là où l'essentiel des mutations de parcelles se produit.

4) L'IRIS constitue la brique de base en matière de diffusion de données infra-communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps<sup>1</sup>. Il est à noter que toutes les communes ne sont pas découpées en IRIS, mais que lorsque c'est le cas, ceux-ci s'imbriquent dans les communes. On compte des quantités assez proches d'IRIS agrégeant les transactions immobilières pour les deux types de logements : 2 270 pour les maisons et 2 250 pour les appartements.

5) La commune est la plus petite des collectivités territoriales. Maillage avant tout administratif, elle correspond également en partie à un espace fonctionnel : malgré l'étalement des villes et la généralisation des navettes domicile-travail à l'échelle supra-communale, les communes fournissent des services essentiels à leurs habitants et sont identifiées comme des entités relativement homogènes. 951 communes sont concernées pour les maisons, 836 pour les appartements.

6) Le canton n'est désormais plus qu'une circonscription administrative destinée à l'organisation des élections. Les transactions touchent 107 cantons pour les maisons comme pour les appartements. L'agrégation à des mailles de plus en plus grosses fait converger les valeurs des effectifs des échantillons pour les deux catégories de logement.

7) L'Établissement Public de Coopération Intercommunale (EPCI) est un regroupement de communes administrant, par délégation, certaines compétences des communes. Celles-ci s'imbriquent dans les EPCI, mais ces derniers ne s'imbriquent dans aucune autre maille. Initialement imaginés par le législateur comme un moyen de faire correspondre les périmètres administratifs aux aires fonctionnelles des villes, ils se sont finalement fréquemment constitués selon des logiques locales de solidarités et d'opposition entre communes polarisées par un même centre. Depuis 2017, toutes les communes de France appartiennent à un EPCI. On en compte 52 dans nos données régionales.

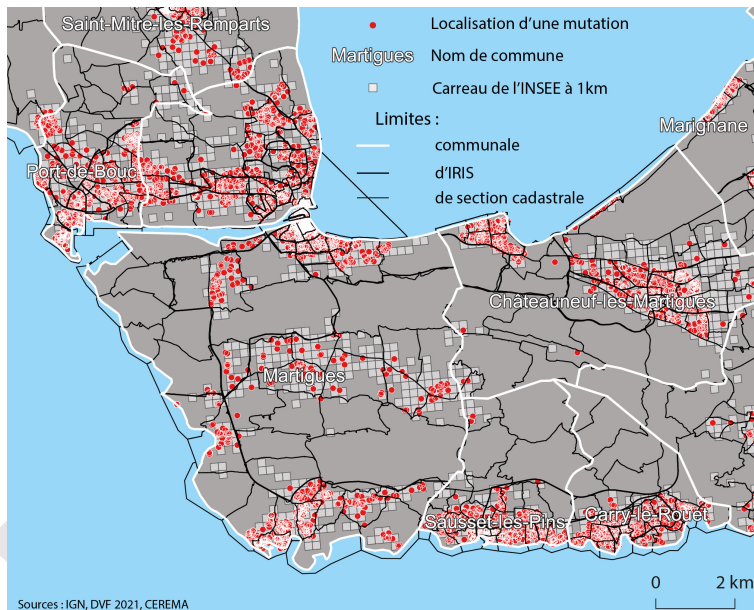
8) Les bassins de vie (22 au total dans nos données) correspondent au "plus petit territoire sur lequel les habitants ont accès aux équipements et services les plus courants"<sup>2</sup>. Ils ne s'imbriquent donc pas dans d'autres périmètres.

9) L'arrondissement (18 en tout) est l'échelon administratif intermédiaire entre le canton et le département, correspondant à l'aire d'action d'un sous-préfet.

---

1. <https://www.insee.fr/fr/metadonnees/definition/c1523/>

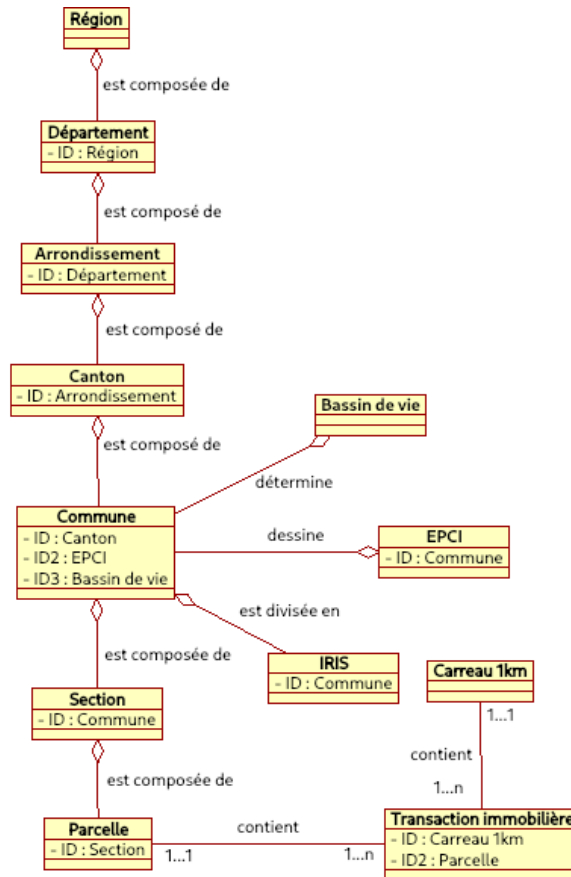
2. <https://www.insee.fr/fr/information/2115016/>



**Figure 6.2.** Zoom sur le secteur de Martigues : des mutations localisées, des carreaux de l'INSEE et des découpages (communes, IRIS et sections cadastrales)

10) Le département (6 en PACA) est une collectivité territoriale intermédiaire entre l'arrondissement et la région.

Ce panel de mailles exclut volontairement des maillages fréquemment utilisés comme les aires d'attraction des villes ou, plus généralement, tous les périmètres liés à des phénomènes de polarisation. En effet, ces derniers présentent l'inconvénient de ne pas recouvrir l'intégralité du territoire national, ni celui de la région PACA, et fausseraient donc le test de l'effet du changement d'échelle sur les références de prix obtenues puisque bon nombre de transactions échapperaient aux statistiques. Notons par ailleurs l'extrême amplitude des effectifs des échantillons d'entités (de 11 803 sections cadastrales à 6 départements). Cela induit en conséquence une très grande variation des effectifs des 528 065 transactions (au total pour les maisons et les appartements) réparties au sein des entités des découpages considérés. Le diagramme UML de la Figure 6.3 détaille les niveaux d'échelle et leurs relations.



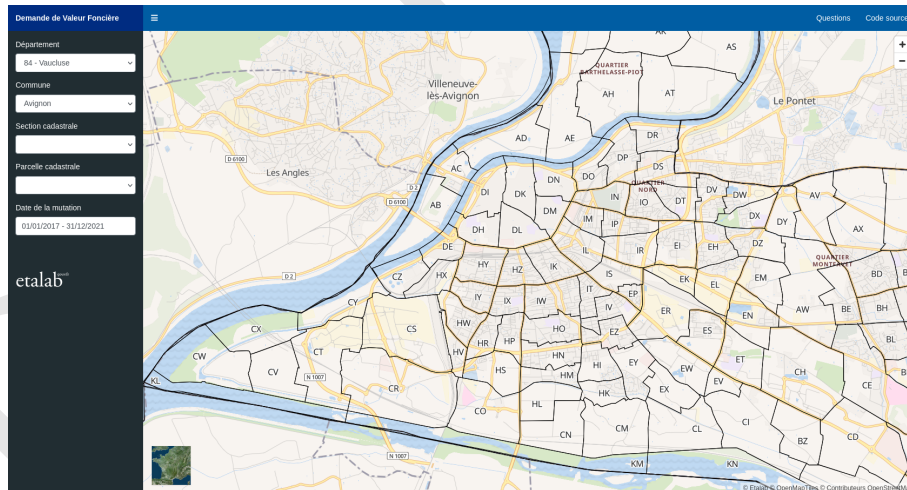
**Figure 6.3.** Diagramme UML présentant l'organisation de la base de données multi-échelles

### 6.3.2. Une base nationale de transactions immobilières géolocalisées : DVF

Les prix immobiliers ont tendance à être considérés comme une variable continue dans l'espace. Cette impression est alimentée par la diffusion de cartes agrégées à différents niveaux d'échelle ou encore par la diffusion de cartes lissées d'estimation des prix immobiliers à l'échelle de l'ensemble du territoire national. Or, ce sont les ventes immobilières, recensées en différents points du territoire, qui fondent la construction de cette variable, intrinsèquement discrète. La distribution spatiale des ventes se caractérise ainsi par son irrégularité à travers l'espace, sa concentration au sein des territoires urbains et également par sa variabilité au cours du temps. Plusieurs bases

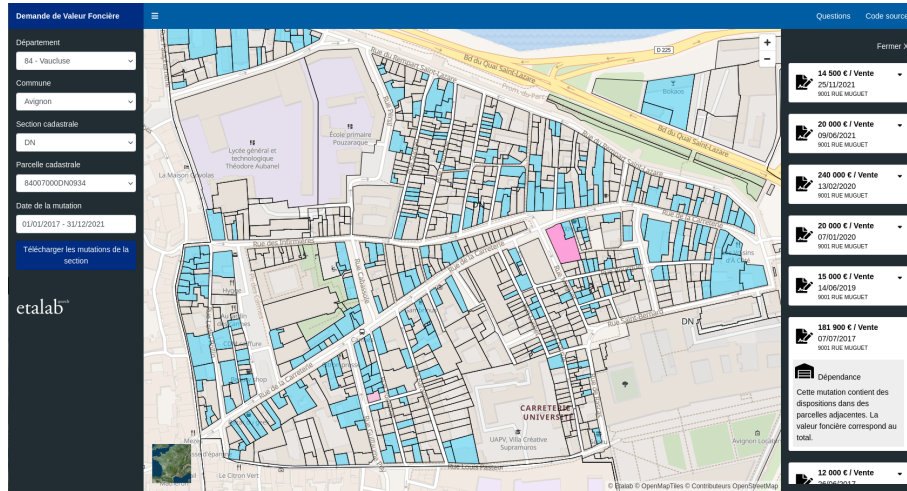


de données recensent aujourd'hui les ventes immobilières et leurs caractéristiques. Elles se fondent toutes sur une même et unique source, l'acte notarié [CAS 17] et divergent du point de vue du producteur de la donnée (notaires, Direction Générale des Finances Publiques ou DGFIP, entreprises privées), de la structuration de la base, ainsi que du canal de transmission et des conditions d'accès à l'information. En effet, pendant longtemps deux bases de données coexistaient : celle produite par les notaires (PERVAL-BIEN) et celle produite par la DGFIP (Demandes de Valeurs Foncières ou DVF, Figures 6.4 et 6.5).



**Figure 6.4.** Exploration de sections cadastrales de la ville d'Avignon sur le site DVF d'Etalab (CEREMA) : <https://app.dvf.etalab.gouv.fr/>

L'ouverture de DVF au grand public en 2019 a donné lieu à une multiplication de versions de cette base de données ainsi qu'au développement de nouvelles bases intégrant tout ou partie de ces données. L'intérêt pour les chercheurs est de pouvoir disposer désormais d'une base de données désagrégées (l'individu statistique étant la vente immobilière) et géolocalisées, qui est exhaustive et disponible à l'échelle du territoire national (sauf l'Alsace et la Moselle) depuis 2014. Ainsi, la base de données en open access DVF est exploitable à l'échelle de l'ensemble du territoire national pour les sept années disponibles (2014 à 2020). En sont extraites les ventes portant spécifiquement sur le marché du logement et par bien unique en région PACA. En effet, dans le cas de "multi-ventes", il n'est pas possible de connaître la part du montant total de la vente affectée à chaque bien, ce qui introduit de l'imperfection d'information (Follin dans [BAT 19]).



**Figure 6.5.** Exploration des parcelles mutées d'une section cadastrale de la ville d'Avignon sur le site DVF d'Etalab (CEREMA) : <https://app.dvf.etalab.gouv.fr/> ; à droite : la liste des transactions immobilières, dont le montant (en €) et la date de mutation sont mentionnés.

Afin de maximiser la significativité des indicateurs aux mailles d'agrégation les plus fines, c'est-à-dire de conserver le plus grand nombre de ventes dans notre échantillon, le filtrage des données porte sur le seul critère de type de biens sur une période de 7 ans. Les ventes immobilières sont ainsi segmentées en deux sous-marchés, celui des maisons et celui des appartements, qui se distinguent à la fois par des niveaux de prix différents et par une géographie différenciée (les appartements étant sur-représentés dans les pôles urbains et les maisons dans les zones périurbaines).

Le travail de préparation des données a consisté à extraire une version complète de la base de données DVF puis, en utilisant un Système d'Information Géographique (PostGIS), via les fichiers eux-mêmes pour les partitions administratives ou par des requêtes spatiales pour les carreaux INSEE, disposer des identifiants des entités des échelles supérieures. Par exemple, à la racine du processus, on doit connaître les identifiants respectifs du carreau et de la parcelle dans lesquels s'est produite la transaction immobilière, comme le montre la Figure 6.3. Cette information doit être connue jusqu'à la partition totalement agrégée, qui est la région dans notre cas. Nous avons ainsi dû supprimer certaines transactions mal géolocalisées ou dont le vecteur d'appartenance aux entités des différentes partitions était erroné ou incomplet (transaction hors carreaux de l'INSEE). Les données dédoublées (concernant les mêmes transactions avec des montants identiques) ont été réduites à un unique individu statistique, affecté à la première des parcelles mutées associées.



Dans la suite de l'article, nous parlerons alternativement de transaction immobilière, de mutation de parcelle, de valeur foncière ou de prix immobilier, l'ensemble de ces vocables faisant référence à un prix établi au moment de la transaction.

### 6.3.3. Représentation des statistiques agrégées sous forme de scalogrammes

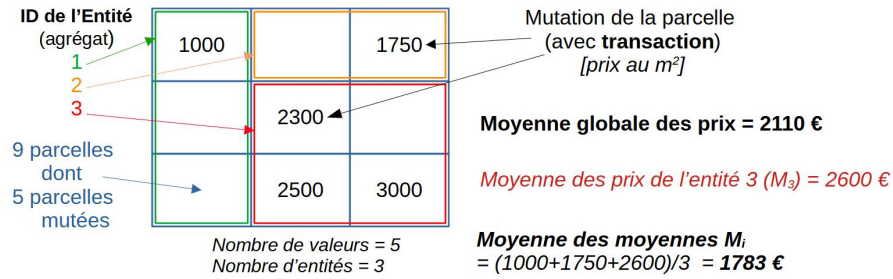
Le *scalogramme* est un graphique qui croise un estimateur statistique agrégé à travers les échelles (cf. Figure 6.8). Plus précisément, il s'agit d'un nuage de points où :

- en X, on positionne les niveaux d'agrégation dans un ordre décroissant de taille d'échantillon (du plus gros échantillon composé d'entités possédant peu de valeurs au plus petit échantillon dont les entités contiennent le plus de valeurs) ; dans notre cas, il s'agit de niveaux de découpage territorial ;
- en Y, on calcule un estimateur de valeur centrale ou de dispersion de la variable étudiée.

Le tracé du scalogramme dessine l'évolution de l'estimateur le long des échelles, permet de les comparer et d'appréhender leur impact sur l'estimateur statistique étudié. Pour réaliser un scalogramme, nous disposons, dans notre cas, d'un fichier complet de données fidèle au diagramme UML de la figure 6.3, issu de DVF en PACA entre 2014 et 2020, pour les transactions concernant les maisons et les appartements. Nous devons aussi choisir des estimateurs statistiques. Pour les valeurs centrales, nous utilisons la moyenne et la médiane, qui nous permettront d'estimer le prix au m<sup>2</sup>. Pour la dispersion, nous choisissons l'écart-type, qui nous fournira une information duale sur l'incertitude entachant l'estimation du prix. Pour l'analyse statistique des permutations aléatoires sous contraintes spatiales, nous ferons aussi appel aux boîtes à moustaches, représentation riche et synthétique de ces deux valeurs centrales et de la forme de la distribution.

La méthode de calcul consiste en une double agrégation statistique en fonction des échelles :

- 1) Au niveau de chaque entité, pour une partition spatiale à une échelle donnée, nous recensons les transactions immobilières et leur appliquons un indicateur (moyenne, médiane ou écart-type) ;
- 2) Pour une partition donnée, nous récupérons la liste des valeurs calculées pour toutes les entités de la partition lors de la première étape de calcul et leur appliquons un indicateur statistiquement consistant avec le premier. Selon la norme mathématique considérée, cela amène à calculer respectivement, pour  $L_2$ , les moyennes des moyennes et des écarts-types, pour  $L_1$ , les médianes des médianes des prix des entités du découpage manipulé. La Figure 6.6 illustre le mode de calcul appliqué dans le cas de la moyenne.



**Figure 6.6.** Exemple de calcul statistique agrégé pour 9 parcelles et 3 entités théoriques ; on constate que les valeurs diffèrent entre la moyenne globale des 5 prix (2110 €) et la moyenne des moyennes des prix des 3 entités (1783 €)

#### 6.4. Méthodologie d'étude de la variation des estimations des prix immobiliers en fonction des échelles

##### 6.4.1. Éléments méthodologiques préliminaires

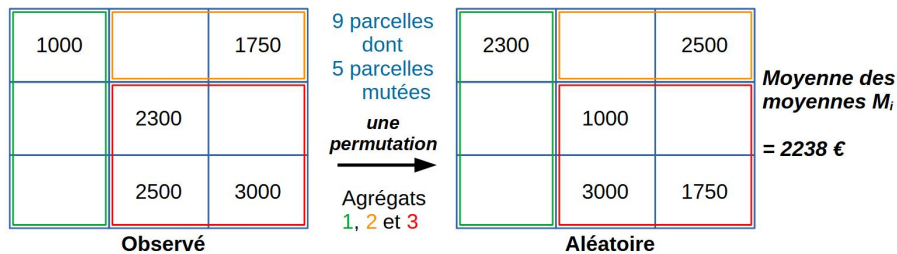
La première hypothèse sous-jacente à ces recherches est que nous supposons que les découpages ont une influence non négligeable sur l'estimation des prix agrégés. Nous souhaitons l'identifier et en connaître l'ampleur, et ce aux différentes partitions étudiées, qu'elles soient administratives (e.g. section, département), statistique (carreau de l'INSEE) ou fonctionnelles (EPCI, bassin de vie). Essentiellement, l'impact de ces découpages tient au lien fort qui existe entre les statistiques employées et les échantillons manipulés. En particulier, la taille de l'échantillon est un élément déterminant dans sa capacité à conserver une stabilité dans l'estimation. Considérant la loi des grands nombres et le théorème central limite, plus un échantillon est grand, plus l'impact probable de l'ajout d'un nouvel individu sur la variance est faible, l'estimateur convergeant vers l'espérance mathématique. Dans notre double processus d'agrégation, ce biais lié à la taille des échantillons peut s'observer au niveau :

- des prix recensés au sein de chaque entité spatiale,
- des prix agrégés dans les entités composant chaque partition.

Or, plus le nombre moyen de transactions augmente dans les entités d'une partition, plus le nombre d'entités de cette partition diminue. On retrouve cet "effet de balance" en analyse de variance (inter vs intra), mais on ignore son rôle effectif croisé en lien avec le COSP ou le MAUP, en fonction des sensibilités combinées des variances. Ce que l'on sait, c'est que produire des statistiques agrégées sur de petits échantillons comporte un risque et génère une incertitude. Dans notre cas, ce risque est omniprésent et agit aux deux niveaux d'agrégation mentionnés, sous l'effet de balance évoqué.

La seconde hypothèse que nous émettons est l'existence d'échelle(s) "pertinente(s)", revêtant une certaine optimalité. Si l'on parvient à comparer les variations d'estimation des prix entre les échelles, l'on peut être à même de déterminer l'échelle la plus fiable de restitution des prix moyens et médians en fonction de la problématique posée ou de la structure à observer. Cette échelle pourrait correspondre à celle qui s'éloigne le plus du cas aléatoire, c'est à dire celle qui maximise le poids de l'explication géographique dans l'organisation spatiale des entités de la partition considérée. En effet, puisqu'on s'éloigne d'une répartition aléatoire homogène, on peut penser que chaque agrégat de valeurs (à l'intérieur des entités spatiales) donne une idée assez juste de la spécificité locale fixant un niveau moyen de prix. Cette approche a été proposée et discutée dans des travaux précédents d'estimation de la biodiversité [JOS 08]. Elle nécessite la mise en place d'une procédure de ré-échantillonnage aléatoire dirigée.

#### 6.4.2. Un générateur aléatoire d'échantillons pour éliminer l'incertitude liée aux échelles



**Figure 6.7.** Exemple d'une permutation des prix sous contrainte de localisation des parcelles mutées ; on constate que la moyenne des moyennes des prix des 3 entités (2 238 € pour cette permutation-ci) diffère de la moyenne globale des 5 prix (2 110 €) et de la moyenne des moyennes observées (1 783 €); ce processus est répété 10 000 fois pour obtenir une bonne représentation de l'aléa.

L'approche que nous proposons dans cet article se décompose en trois étapes :

- Dans un premier temps, nous calculons les valeurs statistiques sur l'ensemble des partitions à toutes les échelles, pour construire le scalogramme. Cela revient à répéter pour tous les niveaux d'échelle géographique, l'opération décrite en Figure 6.6. Cela constitue le cas d'observation.

- Dans un second temps, nous réalisons le même calcul, mais cette fois-ci avec le ré-échantillonnage aléatoire, appliquant 10 000 permutations des transactions immobilières localisées. Il s'agit d'un calcul 10 000 fois plus long, qui a nécessité, pour notre corpus de données immobilières, le recours à un serveur relativement puissant,

mais qui, après différents tests empiriques, nous assure une convergence vers le cas aléatoire.

– Enfin, pour un indicateur statistique donné, nous calculons l'*écart relatif* entre l'observation et l'aléa, rapporté à l'aléa, qui permet de voir à quel point les données observées (les transactions immobilières) s'éloignent d'une répartition aléatoire.

L'étape 2 du calcul est illustrée par la Figure 6.7 : on redistribue spatialement et aléatoirement les prix des transactions dans l'ensemble des parcelles effectivement mutées et ce un nombre important de fois. Cela revient à "casser" les poches éventuelles de prix autocorrélés. Par exemple, les écarts de prix moyens, pouvant être très élevés en passant d'un quartier à un autre, sont ainsi très réduits, l'ensemble convergeant vers une valeur moyenne lissée de valeur des transactions, indépendante de la localisation des entités considérées (sections de cadastre, commune, EPCI, bassin de vie, selon les cas).

La troisième et dernière étape, qui consiste à calculer l'écart relatif (normalisé)  $ER$  entre l'observation ( $obs$ ) et le cas aléatoire ( $aléa$ ) est décrite dans l'équation suivante,  $S$  étant la statistique du prix moyen calculé pour une partition donnée :

$$ER(\%) = 100 * \frac{|S_{obs} - S_{aléa}|}{S_{aléa}}.$$

En soustrayant aux estimations statistiques  $S_{obs}$  des prix des transactions observées celles obtenues par ré-échantillonnage aléatoire  $S_{aléa}$  de ces mêmes données, sous contrainte de localisation, on estime l'écart (en proportion) de la statistique des transactions immobilières  $S$  observées au cas aléatoire. La valeur d' $ER$  est, la plupart du temps, négative pour les valeurs centrales, car l'estimation des valeurs "randomisées" lors des permutations est supérieure à celle de l'observation. Dans nos traitements, c'est systématiquement le cas pour les valeurs centrales, mais pas pour la variabilité, selon la variable et l'échelle considérées. C'est la raison pour laquelle nous ajoutons une valeur absolue, afin de quantifier l'écart relatif (non algébrique) au cas d'une répartition hypothétique, sans structure, ni autocorrélation, spatiales, des transactions immobilières.

#### **6.4.3. Présentation des éléments d'analyse sous forme d'un graphique composite**

L'analyse des résultats, présentés en détails dans la section suivante, s'appuie sur un *graphique composite* permettant une analyse croisée (cf. Figure 6.8). La variable étudiée est le prix au m<sup>2</sup> du bien ayant été échangé lors de la transaction immobilière sur la parcelle ayant muté. Le graphique associé, pour un type de bien immobilier

donné (maison vs appartement) de la région PACA entre 2014 et 2020, un ensemble de 4 graphiques et pour un type d'indicateur donné (valeur centrale vs estimateur de dispersion) :

– La partie inférieure de chaque graphique concerne les données de **l'observation** (réelles), où apparaissent les scalogrammes des prix observés des transactions immobilières, avec les médianes et les moyennes des prix pour chaque partition, auxquels s'ajoutent les *Écarts Relatifs ER* des observations au cas aléatoire, représentés par la surface du cercle ;

– La partie haute représente le **cas aléatoire** (théorique), fournissant des statistiques sur la variabilité des indicateurs après 10 000 redistributions aléatoires des prix au m<sup>2</sup>.

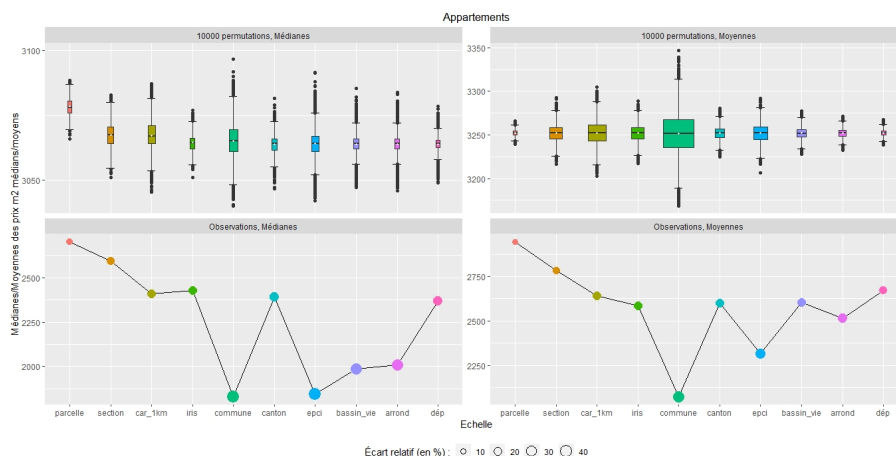
Les boîtes à moustaches s'interprètent de manière classique : médiane représentée par un segment horizontal, hauteur de la boîte correspondant à l'écart inter-quartile (longueur de l'intervalle regroupant 50% des observations, 25% étant au-dessus de la boîte, 25% en-dessous). Les segments ont une longueur égale à 1,5 fois l'écart inter-quartile, les points situés au-delà pouvant alors être interprétés comme des valeurs extrêmes (outliers) pour des données qui proviendraient d'une loi normale. Selon le graphique considéré, un point correspond ainsi à un des trois types de calcul (moyenne de moyennes, médiane de médianes ou moyenne d'écarts-types) par entité pour une redistribution aléatoire donnée.

En complément, chaque *point blanc* dans le diagramme des boîtes à moustaches représente la moyenne calculée pour l'indicateur et le niveau d'échelle considérés. La largeur des boîtes est proportionnelle à l'écart-type des données : on a ainsi deux mesures de dispersion distinctes représentée sur chaque axe (en largeur, l'écart-type et en hauteur, l'écart inter-quartile).

Pour les scalogrammes, la grosseur des points est proportionnelle à l'Écart Relatif (multiplié par 100) de l'observation avec la moyenne des moyennes des 10 000 redistributions aléatoires par entité. Cet Écart Relatif représente le pourcentage de baisse (si négatif), de hausse (si positif) observé par rapport à une distribution complètement aléatoire des prix sur le territoire. S'il est nul, la répartition observée des transactions est purement aléatoire.

## 6.5. Résultats : mise en évidence d'effets de structure liés aux mailles territoriales et saillance d'échelles

### 6.5.1. Analyse des variations des estimations des prix moyens et médians des appartements et des maisons de la région PACA de 2014 à 2020



**Figure 6.8.** Variations (en haut) des médianes de médianes (à gauche) et des moyennes de moyennes (à droite) par entité après redistribution aléatoire des prix au m<sup>2</sup> pour les appartements. Valeurs observées (en bas) pour ces indicateurs par échelle et cercles proportionnels à l'Écart Relatif (%)

#### 6.5.1.1. Le cas des appartements

La Figure 6.8 représente les variations des estimations des prix moyens et médians au m<sup>2</sup> des appartements. On constate que, d'une part, pour les calculs des moyennes avec les 10000 redistributions aléatoires des prix (colonne de droite), chaque moyenne (représentée par un point blanc) de moyennes et chaque médiane de moyennes (segment central des boîtes) sont pratiquement égales et constantes d'une échelle à l'autre. Cette valeur commune est proche de la moyenne globale calculée à environ 3 252 € (la médiane globale étant plus faible et égale à 3 064 €). Il est connu qu'une moyenne de moyennes sur des sous-groupes n'est en général pas égale à la moyenne globale (l'égalité étant vraie pour des effectifs égaux par sous-groupe). La variabilité observée est ici visualisée par les diagrammes en boîte à moustaches. Cependant, le nombre important de redistributions effectuées implique une stabilisation de la moyenne et de la médiane de ces 10 000 moyennes vers la moyenne globale. Le cas aléatoire reste ainsi "centré" sur la moyenne globale, quelle que soit l'échelle considérée.

Si l'on compare maintenant les échelles entre elles, on voit une symétrie apparaître sur la dispersion des moyennes d'une échelle à l'autre. En mettant de côté les IRIS et les cantons, la dispersion des moyennes augmente de la parcelle à la commune, pour décroître après jusqu'au département (phénomène observé aussi bien sur les écarts-types avec la largeur des boîtes que pour l'écart inter-quartile avec leur hauteur). Les prix ayant été redistribués de manière complètement aléatoire, cela met en avant le rôle de l'échelle, avec un effet particulièrement important pour la commune où les

dispersions entre les différentes moyennes simulées sont maximales, qui se vérifie également pour la médiane. Cette échelle semble être celle qui génère le maximum de variabilité des prix, à la fois à l'intérieur de chaque entité et dans l'échantillon des entités de la partition.

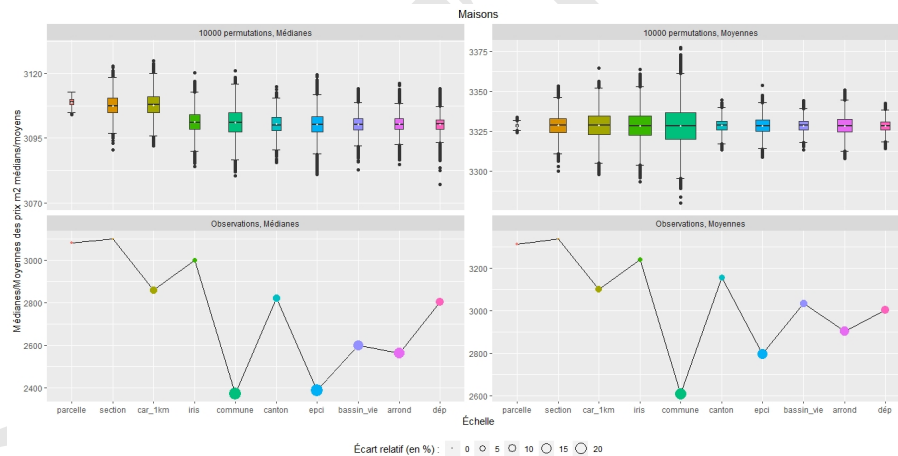
Toujours pour les moyennes, si l'on compare les résultats aléatoires à ce qui a été réellement observé (graphique en bas à droite), on observe tout d'abord une nette différence, avec des moyennes observées beaucoup plus basses pour chaque partition et un écart pouvant dépasser 30% pour les communes. On a tout d'abord une illustration du paradoxe de Simpson [SIM 51], puisque les moyennes sur chaque entité sont inférieures à la moyenne globale calculée à 3 252,11 €. De plus, leur éloignement par rapport aux valeurs simulées dans les boîtes à moustaches (représentant l'indépendance de distribution des prix) met en lumière l'effet de grappes (*cluster*). Cet effet est logiquement moins marqué pour les petites entités (parcelles) comprenant peu de transactions que pour les plus grandes. Encore une fois, le comportement des communes est singulier, ainsi que celui des EPCI en second lieu. Ces deux découpages sont ceux qui montrent le plus d'homogénéité dans les valeurs réelles des transactions par entité et, en parallèle, le plus d'hétérogénéité des prix immobiliers redistribués aléatoirement. Ce constat est corroboré par l'Écart Relatif qui est maximal pour ces deux niveaux d'échelle.

La colonne de gauche de la Figure 6.8 reprend les mêmes calculs, mais pour les médianes des prix médians. Par comparaison aux graphiques obtenus pour la moyenne, le caractère plus robuste de la médiane est clairement marqué, avec des dispersions beaucoup moins accentuées. De plus, on constate un phénomène de stabilisation vers la médiane globale (3 064 €), à partir des partitions d'effectifs d'entités plus faibles (IRIS). Il est également assez remarquable de noter, pour cet indicateur robuste, le rôle particulier des communes et des EPCI, avec l'effet de grappe nettement marqué par les valeurs minimales des prix médians au m<sup>2</sup>, les valeurs maximales des Écart Relatifs et l'étalement prononcé des diagrammes en boîte. L'allure relativement similaire des courbes des observations obtenues avec les médianes et les moyennes peut s'interpréter également comme la manifestation de l'effet géographique, puisque les résultats ne semblent dépendre que des échelles et pas des métriques employées.

Quant à l'Écart Relatif, il peut s'approcher de 40 % (médianes des communes, des EPCI). Cependant, pour la plupart des découpages, il reste assez faible, ce qui signifie que la répartition des valeurs des transactions est relativement proche d'une répartition aléatoire, indépendante des découpages. Il existe par ailleurs un décrochage plus prononcé pour les médianes des prix aux échelles des bassins de vie et des arrondissements que pour leurs moyennes respectives (valeurs médianes plus faibles que les valeurs moyennes). Cela révèle une forme asymétrique de la distribution des prix observés à ces deux échelles (il y a un plus grand nombre de faibles valeurs de prix), qui d'ailleurs ne se retrouve pas dans les deux distributions aléatoires correspondantes.

(cela étant dû à la “symétrisation” de la distribution induite par les permutations aléatoires).

### 6.5.1.2. Le cas des maisons



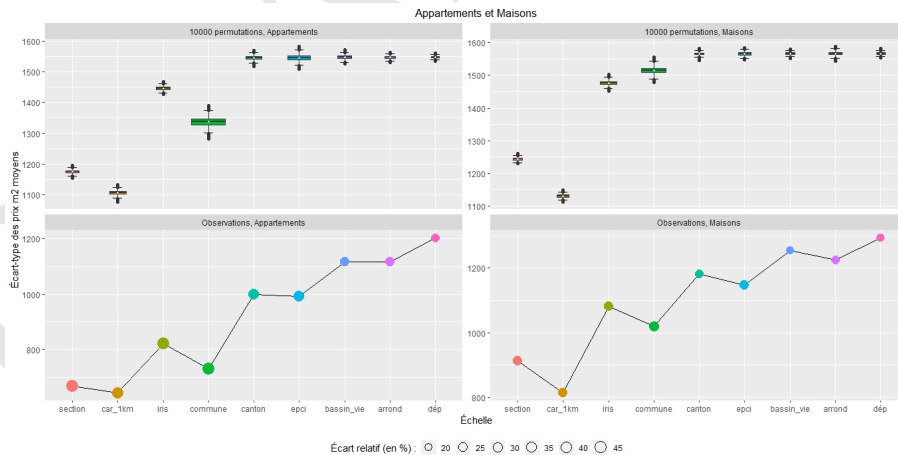
**Figure 6.9.** Variations (en haut) des médianes de médianes (à gauche) et des moyennes de moyennes (à droite) par entité après redistribution aléatoire des prix m<sup>2</sup> pour les maisons. Valeurs observées (en bas) pour ces indicateurs par échelle et cercles proportionnels à l'Écart Relatif (%).

Le même graphique composite est présenté pour les transactions relatives aux maisons dans la Figure 6.9. La stabilisation, respectivement, vers la médiane globale (3 100 €) et la moyenne globale (3 328 €) est apparente à partir de l'IRIS pour les médianes et pour toutes les entités pour la moyenne. Par ailleurs, la symétrie des formes des boîtes à moustaches centrée sur les communes ne semble plus d'actualité : du canton au département, les distributions aléatoires se ressemblent (leur variabilité ne décroît pas au fil de l'agrégation). On peut également remarquer que l'Écart Relatif est quasi-nul pour les petites entités (parcelles et sections où il devient même positif pour ces dernières) et très faible pour les IRIS : pour ces petites divisions administratives, on ne met plus en évidence d'effet de grappe, la répartition des prix au m<sup>2</sup> étant quasiment aléatoire. Globalement, pour les maisons, il y a moins d'autocorrélation spatiale que pour les appartements, quel que soit le niveau d'échelle considéré, avec un écart relatif divisé par deux. On note cependant toujours le rôle particulier joué par les communes (ainsi que les EPCI pour les médianes).



**6.5.2. Analyse des variations des estimations des écarts-types des prix moyens des appartements et des maisons de la région PACA de 2014 à 2020**

Comme nous l’avons évoqué précédemment, il est aussi intéressant de considérer la variabilité des prix des transactions immobilières étudiés, afin de qualifier l’incertitude associée aux estimations de prix en fonction d’une échelle donnée. Nous l’appréhendons en calculant les moyennes des écarts-types des prix pour chaque partition. La Figure 6.10 met en regard les appartements (à gauche) et les maisons (à droite), avec les mêmes types de représentations statistiques que précédemment : les boîtes à moustaches des 10 000 écart-types des prix moyens au m<sup>2</sup> et les scalogrammes correspondants, avec les écarts relatifs par niveau d’échelle.



**Figure 6.10.** Variations (en haut) des moyennes d’écarts-types pour les appartements (à gauche) et les maisons à droites (à droite) par entité après redistribution aléatoire des prix m<sup>2</sup>. Valeurs observées (en bas) pour cet indicateur, par échelle et type de bien.

Dans cette figure, les parcelles ont été éliminées du graphique car la variabilité interne n’était pas suffisante pour être calculée. Pour cet indicateur, la variabilité est relativement faible pour les permutations aléatoires (écrasement généralisé des boîtes à moustaches, sauf, à nouveau, pour les communes). En mettant de côté les carreaux de l’INSEE de 1km et les communes pour les appartements et, dans une moindre mesure les EPCI – qui, en quelque sorte, “décrochent” de la courbe –, on observe que la dispersion augmente au fur et à mesure que les entités croissent en taille. On peut noter que la stabilisation des écart-types moyens se produit à partir des cantons, aussi bien pour les appartements et les maisons (l’écart-type global pour ces biens étant respectivement égal à 1 547 € et 1 567 €). S’agissant du moment

d'ordre 2, la stabilisation par la loi des grands nombres demande plus d'observations, car la variabilité devient d'ordre 4. Cette contrainte rend plus délicate l'estimation d'indicateurs de variabilité, car nécessitant plus d'effectifs.

Il est également remarquable d'observer une courbe relativement similaire entre les 10 000 permutations aléatoires et les valeurs observées, ce qui n'était pas le cas pour les valeurs centrales (notamment les moyennes qui restaient particulièrement stables à travers les échelles). Cela montre l'impact déterminant du support spatial sur la variance des prix, phénomène connu dès lors que l'on manipule des indicateurs de variabilité (écarts-types, indices de diversité comme l'entropie, par exemple) [MAH 07, JOS 09]. Il est également satisfaisant de voir que les écarts-types sont nettement inférieurs pour les observations que dans le cas des répartitions aléatoires. Ce dernier graphique met ainsi en avant une variance *intra* moindre au sein des entités et ainsi leur effet de grappe marqué. À noter que les carreaux de l'INSEE dévoilent les plus faibles écart-types des échantillons aléatoires et observés. Leur régularité géométrique a un effet net de réduction de l'incertitude. Pour ce qui concerne les Écarts Relatifs des écarts-types, ils s'étendent cette fois-ci sur une plage moins large (de 20 à 45 % environ), mais se situent globalement à des valeurs plus élevées.

### **6.5.3. Analyse transversale aux graphiques composites**

#### **6.5.3.1. L'effet de grappe du support spatial**

Il est intéressant de remarquer que le changement de type de biens n'a que peu d'influence sur l'allure générale des courbes obtenues, que ce soit pour les séries de boîtes à moustaches ou pour les scalogrammes, pour les valeurs centrales (moyennes et médianes des prix), comme pour leur écart-type. Il règne donc une certaine similarité dans les données des transactions immobilières de maisons et d'appartements à travers les échelles, qui s'explique essentiellement par la relation qu'entretient la variable étudiée (les prix au m<sup>2</sup>) avec la structure topologique des entités composant la partition à une échelle donnée (localisation des transactions et forme des partitions). On notera toujours les cas particuliers des communes et des EPCI, qui présentent davantage d'autocorrélation spatiale des prix immobiliers que les autres découpages.

L'on peut également séparer assez distinctement l'analyse des deux types d'indicateurs de centralité ou de dispersion des prix immobiliers, appliqués indifféremment aux appartements ou aux maisons. Pour le cas de l'échantillonnage aléatoire, ce sont les valeurs centrales qui connaissent à la fois le plus de stabilité des valeurs et le plus de variations des distributions des prix estimées à travers les échelles (cela est un peu moins vrai pour le cas des médianes des prix). On constate le phénomène exactement inverse pour la variation des écarts-types, dont la valeur estimée se stabilise à partir des cantons, mais qui est très variable pour les découpages plus fins. Par ailleurs, on note une grande fluctuation des valeurs centrales à travers les échelles (sauf pour les

carreaux de l'INSEE de 1km et, dans une moindre mesure, les communes et les EPCI), pendant que la variance ne cesse de croître des échelles fines (section) aux partitions plus généralisantes (département), montrant l'impact régulier du support spatial sur l'incertitude.

Enfin, l'Écart Relatif (ER) se comporte aussi de façon homogène, que l'on considère les appartements comme les maisons. Ses valeurs sont élevées pour les moyennes et les médianes des prix immobiliers (bien que les valeurs moyennes et médianes soient assez faibles), plus faibles pour la dispersion (écarts-types) des prix (alors que les valeurs des écart-types sont plus élevées). Enfin, on notera l'évidente relation entre l'Écart Relatif et la dispersion des échantillons aléatoires : plus les échantillons aléatoires génèrent une instabilité de la distribution des prix simulés, plus les valeurs observées moyennes et médianes sont faibles, et plus l'ER est élevé (en valeur absolue), montrant que l'observation s'écarte davantage de l'aléa. C'est le signe d'un effet de grappe plus marqué pour les partitions spatiales dont les permutations aléatoires ont généré davantage de variabilité et donc d'une plus forte autocorrélation spatiale interne aux entités de ces partitions.

#### 6.5.3.2. Recherche d'échelle pertinente

La commune constitue à chaque fois une échelle pivot qui maximise l'écart au cas aléatoire. Suivent les EPCI et, dans une moindre mesure, d'autres découpages selon les métriques statistiques employées : bassin de vie et arrondissement pour les valeurs centrales, carreau de l'INSEE, section cadastrale et IRIS pour la dispersion. En appliquant les attendus de notre méthode d'élimination de l'effet du support spatial, on peut affirmer que, sur la série de données relatives aux prix immobiliers au m<sup>2</sup> en région PACA de 2014 à 2020, la commune est le découpage qui concentre le plus d'effet de grappe et d'autocorrélation spatiale, pour les appartements comme pour les maisons. Les EPCI, qui en sont constituées, reproduisent plus ou moins cette propriété. Pour l'estimation des prix immobiliers, la commune reste finalement l'échelle la moins biaisée, si l'on considère sa capacité locale de différenciation de la valeur des biens, d'autant plus qu'elle inscrit les plus grandes valeurs d'Écart Relatif, sur les valeurs centrales comme sur l'indicateur de variabilité – l'écart-type – des prix. Pour autant, l'échelon communal présente la plus grande incertitude parmi les partitions, avec une distribution aléatoire des écarts-types des échantillons simulés la moins concentrée, ainsi qu'une valeur moyenne élevée de variabilité, en particulier pour les maisons.

Nous pourrions penser que c'est la position intermédiaire de l'échelle communale, permettant un bon équilibre entre la taille des échantillons de prix au sein des entités et la taille de l'échantillon global des prix agrégés par entité pour toute la partition, qui pourrait induire de faibles variances *intra* et *inter*. Cette proposition est invalidée par la proximité des cantons, dont toutes les valeurs de prix (cas observé comme aléatoire) s'éloignent très rapidement des points ou des formes des boîtes à moustaches des communes. Cela ne résulte donc pas que d'un effet statistique, mais bien de

l'impact de l'organisation spatiale des transactions observées. Ce constat est renforcé par le revirement observé de la position des EPCI, qui se rapproche bien davantage des communes que des cantons, pourtant voisins directs dans le processus incrémental de changement d'échelle. Ce phénomène permet de mettre en avant la logique locale (ressemblance et homogénéité des prix des transactions au sein des entités) des EPCI, contrairement aux cantons qui ne sont que des circonscriptions d'élection.

Enfin, il convient d'évaluer la confiance que nous pouvons accorder à l'estimation des prix immobiliers moyens ou médians au m<sup>2</sup>. Pour ce faire, nous disposons, d'une part, des distributions (sous forme de boîtes à moustaches) et des valeurs centrales des prix des 10 000 échantillons, simulant le cas aléatoire à une échelle donnée, mais également de la série complète des écarts-types des prix moyens le long des échelles. Comme nous l'avons vu, cette valeur augmente quelque peu mécaniquement au fil du processus d'agrégation. On sait ainsi que plus la partition contient d'entités (de petite taille, c'est à dire contenant peu de transactions immobilières), plus la variabilité des prix estimée est faible. L'incertitude croît avec le niveau d'échelle, ce qui relativise l'existence d'une échelle pertinente par le risque d'entachement d'erreur, d'autant plus que les valeurs d'Écart Relatif des écarts-types des prix restent dans un mouchoir de poche, pour les appartements comme pour les maisons.

## 6.6. Conclusion, discussion

Appliquer la méthode d'évaluation des variations statistiques des prix immobiliers liées aux échelles et rechercher une éventuelle échelle pertinente présentent un double intérêt méthodologique et thématique. Du point de vue méthodologique, l'appréhension spatiale des transactions immobilières implique en effet de se confronter à des événements relativement rares et dont les semis de points sont très irréguliers, combinant zones de très forte concentration et vastes étendues marquées par une très faible densité de points. D'un point de vue plus thématique, l'identification d'échelles de représentation du marché immobilier qui fassent sens est d'autant plus importante que celui-ci occupe une place centrale [SCH 09, AAL 14, PIK 13] et croissante [PIK 13, WET 17, ADK 21] dans les inégalités patrimoniales. Les dynamiques qui traversent les marchés locaux peuvent renforcer l'inégalité de la distribution sociospatiale des richesses immobilières [LEG 21] et l'échelle de représentation cartographique choisie n'est pas neutre dans la compréhension de ces phénomènes [LEC 19].

Dans cet article, nous avons pu montrer que l'échelle d'agrégation avait un effet notable sur l'estimation des prix immobiliers moyens et médians, pour les appartements comme pour les maisons. D'une part, la variabilité des estimations de prix est directement impactée par le niveau d'agrégation et ce de façon quasi-linéaire. Elle est d'autant plus élevée que la partition contient peu de gros agrégats. D'autre part, l'écart entre les valeurs de prix observés et leur répartition aléatoire ne dépasse pas 40%. Dans certains cas (sections cadastrales, IRIS et, dans une moindre mesure, carreaux

de l'INSEE de 1km), la répartition des prix est très proche de l'aléa, ce qui signifie qu'il n'y a aucune autocorrélation spatiale ou effet de grappe (cluster) des prix immobiliers. Fournir une valeur moyenne sur la base de ces échelles géographiques, dont la répartition des prix de transactions est quasi-aléatoire, n'est pas utile : la valeur moyenne globale des prix suffirait. En revanche, les moyennes par niveau communal ou par EPCI sont très probablement plus justes, car elles tiennent compte des spécificités locales (en s'éloignant du cas de répartition aléatoire). En y adjoignant l'information duale sur l'incertitude des prix à ces échelles, on pourra conserver la commune qui, malgré tout, montre un écart-type des moyennes de prix assez faible, en dépit d'une variabilité relativement élevée (par exemple, par rapport au découpage par carreaux de l'INSEE). Elle constitue ainsi le meilleur compromis. Enfin, gardons en tête, d'un point de vue pratique, que plus l'échantillon considéré pour étalonner le prix est petit et plus la probabilité que la valeur diffère de la moyenne des valeurs de cet échantillon est élevée.

D'un point de vue géographique, la pertinence des communes, et, en second lieu des EPCI, est justifiée parmi les dix partitions proposées (administratives et/ou fonctionnelles). En effet, cela correspond aux niveaux d'échelles où, au sein des entités, les prix observés sont les plus homogènes et où leur redistribution aléatoire génère en conséquence la plus grande variabilité. Ces découpages résultent de la construction progressive de zonages fonctionnels (vécus et perçus), d'où leur propriété d'effet de grappe et d'autocorrélation spatiale. Ségrégation socio-spatiale, gentrification ou paupérisation localisée poussent ces territoires vers une spécification sociale, parfois en phase avec une volonté politique. Ce(s) profil(s), hérités de l'histoire locale, détermine(nt) en partie les prix de l'immobilier moyens d'une commune. Les professionnels du domaine (agences immobilières, notaires) ne sont d'ailleurs pas exempts de responsabilité dans cette spécification, dans la mesure où ils conseillent les futurs vendeurs ou acheteurs au niveau communal, en affectant à chaque municipalité quelques adjectifs descriptifs, en lien avec la valeur estimée des biens, de nature à orienter les choix des vendeurs ou des acheteurs : commune "de standing", "ayant mauvaise réputation", "dortoir", etc. S'ajoutent souvent des éléments à la granularité du "quartier", découpage dont nous ne disposons malheureusement pas dans cette étude mais qui, plus fin que les IRIS et moins technique que les sections, pourrait beaucoup apporter à la compréhension des répartitions spatiales des valeurs foncières.

Si la méthode appliquée dans cet article est réutilisable pour de futures études sur la géographie de l'immobilier en France, en matière de marché immobilier, comme dans d'autres domaines, il convient néanmoins de rester vigilant quant à la "pertinence" absolue d'un découpage :

– La pertinence d'une échelle résulte de processus géo-historiques susceptibles d'évoluer. Par exemple, l'instauration des plans locaux d'urbanisme intercommunaux (PLUI), en remplacement des PLU municipaux, risque de renforcer la pertinence du découpage par EPCI dans le sous-marché de l'immobilier neuf.

– Une échelle pertinente dans un contexte peut devenir une convention de représentation. À ce titre, l'échelon communal reste incontournable du simple fait qu'il correspond à un découpage de l'espace partagé par les individus. En outre, la pertinence d'une représentation s'établit en fonction des objectifs du cartographe : elle n'obéit pas aux mêmes critères dans la sphère académique et dans le domaine privé, où la finalité commerciale aboutit à privilégier les échelles perçues comme pertinentes par les professionnels (promoteurs, conseillers en gestion de patrimoine ou investisseurs), elles-mêmes fondées sur les préférences attribuées aux acquéreurs finaux des biens (les ménages). La pertinence d'une maille est donc également située socialement. La "convention" communale est ainsi plus souvent dépassée dans les milieux académiques (par exemple en la complétant d'un carroyage de l'INSEE), tandis que les experts privés continuent de privilégier des partitions infracommunales (quartiers, IRIS) ou des mailles qui en constituent des agrégations (EPCI).

– La pertinence d'une échelle, même si elle reste dépendante d'un contexte, peut s'auto-renforcer, via une "prophétie auto-réalisatrice" des estimations de prix par maille, puisque la majorité des intermédiaires du marché immobilier utilise des dispositifs d'estimation des prix fondés sur des découpages de l'espace. L'échelle de représentation des prix immobiliers joue ainsi le rôle d'un dispositif de marché [MUN 07], orientant la lecture des phénomènes par les acteurs et influençant leurs comportements dans le sens d'une validation *a posteriori* des découpages supposés pertinents [WEB 16]. De fait, la méthode que nous avons proposée possède la capacité de s'adapter à différentes variables à estimer le long de tout type de séries d'échelles, permettant une recherche localement et thématiquement adaptée. Qui plus est, elle produit une double estimation de la variable traitée : celle de la valeur et celle de la variabilité de cette valeur (information duale sur l'incertitude).

Enfin, ce travail de recherche sur la relation entre échelles et prix immobiliers amène à se poser la question de la place et du statut de l'incertitude dans la méthodologie utilisée [BAT 19] (agrégation de valeurs, changement d'échelle, comparaison de l'observation au cas aléatoire). Si l'on définit l'incertitude *aléatoire* comme la variation naturelle inhérente au phénomène aléatoire étudié et l'incertitude *épistémique* comme un défaut de connaissance dans une modélisation [SAL 13], on peut essayer de positionner notre approche par rapport à ces deux définitions. Dans notre cas, le générateur aléatoire de permutations modélise l'incertitude liée au support spatial, l'incertitude aléatoire des prix immobiliers eux-même étant estimée par les écarts-types des prix par niveau d'échelle. Quant à l'incertitude épistémique, elle résulte ici en grande partie de l'effet caché du support spatial. Toute personne qui souhaite estimer la valeur d'un bien immobilier ne peut pas connaître, par expertise, l'échantillon optimal à choisir pour réaliser cette estimation, encore moins l'effet d'un choix inapproprié d'échantillon sur la valeur estimée. C'est tout l'enjeu de ces recherches, qui visent à réduire les deux formes d'incertitude, par une analyse fouillée graphique et numérique, notamment via le calcul de l'Écart Relatif de l'observation à un cas

théorique aléatoire de répartition des transactions dépourvu de la géographie des prix immobiliers.

DRAFT

---

## Bibliographie

---

- [AAL 14] AALBERS M. B., CHRISTOPHERS B., “The housing question under capitalist political economies”, *Housing, Theory and Society*, vol. 31, n° 4, p. 422–428, Taylor & Francis, 2014.
- [ADK 21] ADKINS L., COOPER M., KONINGS M., “Class in the 21st century: Asset inflation and the new logic of inequality”, *Environment and planning A: economy and space*, vol. 53, n° 3, p. 548–572, SAGE Publications Sage UK: London, England, 2021.
- [BAT 19] BATTON-HUBERT M., DESJARDIN E., PINET F., *L'imperfection des données géographiques 1 Bases théoriques*, ISTE, septembre 2019.
- [BON 19] BONO P.-H., TRANNOY A., “The Impact of the 'Scellier' Income Tax Relief on Building Land Prices in France”, *Economie et Statistique / Economics and Statistics*, , n° 507-508, p. 91–114, INSEE, 2019.
- [BOU 11] BOULAY G., *Le prix de la ville. Le marché immobilier à usage résidentiel dans l'aire urbaine de Marseille-Aix-en-Provence (1990-2010)*, PhD thesis, Université de Provence Aix-Marseille 1, 2011.
- [BOU 12] BOULAY G., BUHOT C., FOURNIER J.-L., “Les chercheurs exclus de l'open data? Appel à signatures lancé à la communauté de la recherche”, *Cybergeo: European Journal of Geography*, CNRS-UMR Géographie-cités 8504, 2012.
- [BOU 21] BOULAY G., BLANKE D., CASANOVA ENAULT L., GRANIÉ A., “Moving from Market Opacity to Methodological Opacity: Are Web Data Good Enough for French Property Market Monitoring?”, *The Professional Geographer*, vol. 73, n° 1, p. 115–130, Taylor & Francis, 2021.
- [BRE 20] BREUILLÉ M., GRIVAULT C., LE GALLO J., *Les indicateurs de loyers dans le parc locatif privé. Note méthodologique*, Rapport, Ministère chargé du logement, décembre 2020.
- [BR18] BRÉCARD D., LE BOENNEC R., SALLADARRÉ F., “Accessibilité, pollution locale et prix du logement : le cas de Nantes Métropole, France”, *Economie et statistique*, vol. 500-501-502, p. 97–115, octobre 2018.



- [BUH 06] BUHOT C., *Marché du logement et division sociale de l'espace dans les îles du Ponant*, PhD thesis, Université de Bretagne occidentale-Brest, 2006.
- [CAI 19] CAILLY C., CÔTE J.-F., DAVID A., FRIGGIT J., GREGOIR S., NOBRE A., PROOST F., SCHOFFIT S., TAUZIN N., THÉLOT H., "Les indices Notaires-Insee des prix des logements anciens Méthodologie v4", 2019, Publisher: Insee.
- [CAS 17] CASANOVA ENAULT L., BOULAY G., GÉRARD Y., YAHY L., "Deux bases de données, aucune référence de prix", *Revue d'Economie Regionale Urbaine*, , n° 4, p. 711–732, Armand Colin, 2017.
- [CAS 19] CASANOVA ENAULT L., BOULAY G., COULON M., "Une aubaine pour les géographes ? Intérêts des fichiers open DVF sur les transactions foncières et immobilières et précautions d'usage", *Cybergeo: European Journal of Geography*, CNRS-UMR Géographie-cités 8504, 2019.
- [CAS 22] CASANOVA ENAULT L., PERIS A., "L'articulation des prix fonciers et immobiliers en France : une géographie des marchés locaux", *Foncier en débat*, 2022.
- [DES 01] DESROSIÈRES A., "Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative", *Genèses*, , n° 2, p. 112–127, Belin, 2001.
- [DUR 19] GÉNÉRAL AU DÉVELOPPEMENT DURABLE C., *Une nouvelle grille de lecture des territoires pour le logement, la maille habitat*, Rapport, juillet 2019.
- [FAC 09] FACK G., GRENET J., "Sectorisation des collèges et prix des logements à Paris", *Actes de la recherche en sciences sociales*, vol. 180, n° 5, p. 44–62, 2009, Paris, Le Seuil.
- [FEU 21] FEUILLET T., *Les effets de contexte en géographie. Des fondements théoriques à la modélisation statistique*, Habilitation à diriger des recherches, Université Lyon 3 Jean Moulin, septembre 2021.
- [GIT 13] GITELMAN L., *Raw data is an oxymoron*, MIT press, 2013.
- [GUÉ 09] GUÉROIS M., LE GOIX R., "La dynamique spatio-temporelle des prix immobiliers à différentes échelles: le cas des appartements anciens à Paris (1990-2003)", *Cybergeo: European Journal of Geography*, CNRS-UMR Géographie-cités 8504, 2009.
- [JOS 08] JOSSELIN D., MAHFOUD I., FADY B., "Impact of a Change of Support on the Assessment of Biodiversity with Shannon Entropy", *Spatial Data Handling, SDH'2008*, Montpellier, June, 23-25, p. 109-131, 2008.
- [JOS 09] JOSSELIN D., MAHFOUD I., FADY B., "Analyse exploratoire des effets de support spatial et de robustesse statistique sur la fiabilité de la mesure de la (bio)diversité", *Photo-interprétation / European Journal of Applied Remote Sensing*, vol. 45, p. 3-11;35-41, 2009.
- [JOS 19] JOSSELIN D., MAHFOUD I., FADY B., "Impact of the Scale on Several Metrics Used in Geographical Object-Based Image Analysis: Does GEOBIA Mitigate the Modifiable Areal Unit Problem (MAUP)?", *ISPRS International Journal of Geo-Information*, vol. 8, n° 3, Page156, MDPI, mars 2019.
- [JUI 19] JUILLARD C., "Le tournant numérique des données immobilières : permanences et recompositions", 2019.
- [KIN 97] KING G., *A solution to the ecological inference problem. Reconstructing individual behaviour from aggregate data*, Princeton University Press, 1997.

- [KIN 04] KING G., ROSEN O., TANNER A. M., (dir.), *Ecological Inference. New Methodological Strategies*, Cambridge University Press, 2004.
- [KON 06] KONTRIMAS V., VERIKAS A., “Tracking of doubtful real estate transactions by outlier detection methods: a comparative study”, *Information Technology and Control*, vol. 35, n° 2, juin 2006, Number: 2.
- [LAF 03] LAFERRÈRE A., DUBUJET F., “Niveau et évolution des prix des logements anciens en province”, *Données Urbaines*, vol. 4, p. 293–306, 2003.
- [LEB 22] LE BRUN P., “L’Etat dans la régulation locale : le cas des aides fiscales à l’investissement locatif des ménages en France”, *Revue d’Economie Régionale & Urbaine*, vol. [Prépublication], Armand Colin, 2022.
- [LEC 19] LE CORRE T., Paris à tous prix. Analyse des inégalités par une géographie de l’investissement sur le marché immobilier résidentiel en Île-de-France., PhD thesis, Université Paris 1-Panthéon Sorbonne, 2019.
- [LEF 15] LEFEBVRE T., “Une nouvelle ère pour l’intermédiation en immobilier résidentiel : fondements, digitalisation et limites”, 2015, thèse de doctorat en Sciences de Gestion, Université Paris Dauphine.
- [LEG 21] LE GOIX R., CASANOVA ENAULT L., BONNEVAL L., LE CORRE T., BENITES-GAMBIRAZIO E., BOULAY G., KUTZ W., AVELINE-DUBACH N., MIGOZZI J., YSEBAERT R., “Housing (In) Equity and the Spatial Dynamics of Homeownership in France: A Research Agenda”, *Tijdschrift voor economische en sociale geografie*, vol. 112, n° 1, p. 62–80, Wiley Online Library, 2021.
- [LOU 16] LOUVET R., JOSSELIN D., GENRE-GRANDPIERRE C., ARYAL J., “Impact des niveaux d’échelle sur l’étude des feux de forêts du sud-est de la France”, *Revue Internationale de Géomatique*, Lavoisier, décembre 2016.
- [MAH 07] MAHFOUD I., JOSSELIN D., FADY B., “Sensibilité des indices de diversité à l’agrégation”, *Revue Internationale de Géomatique*, vol. 3-4, p. 293-308, 2007.
- [MAL 09] MALLE R., “Méthode hédonique et loyers des bureaux en Île-de-France”, *Revue d’Economie Régionale & Urbaine*, vol. décembre, n° 5, p. 905–933, 2009, Place: Paris Publisher: Armand Colin.
- [MUN 07] MUNIESA F., MILLO Y., CALLON M., “An Introduction to Market Devices”, *The Sociological Review*, vol. 55, n° 2, p. 1–12, SAGE Publications Sage UK: London, England, 2007.
- [OLI 16] OLIVEAU S., DOIGNON Y., “La diagonale se vide? Analyse spatiale exploratoire des décroissances démographiques en France métropolitaine depuis 50 ans”, *Cybergeo: European Journal of Geography*, CNRS-UMR Géographie-cités 8504, 2016.
- [OPE 84] OPENSHAW S., *The modifiable areal unit problem*, Norwich: Geo Books, CATMOG 38, 1984.
- [PAL 78] PALM R., “Spatial segmentation of the urban housing market”, *Economic Geography*, vol. 54, n° 3, p. 210–221, Taylor & Francis, 1978.
- [PIK 13] PIKETTY T., *Le capital au XXIe siècle*, Le Seuil, 2013.

- [PUM 97] PUMAIN D., SAINT-JULIEN T., *L'analyse spatiale. Localisations dans l'espace*, Armand Colin (Coll. CURSUS « Géographie »), 1997.
- [SAC 22] SACHDEVA M., FOTHERINGHAM S., LI Z., "Do Places Have Value?: Quantifying the Intrinsic Value of Housing Neighborhoods Using MGWR", *Journal of Housing Research*, vol. 31, n° 1, p. 24–52, janvier 2022.
- [SAL 13] SALLAK M., AGUIRRE F., SCHON W., "Incertitudes aléatoires et épistémiques, comment les distinguer et les manipuler dans les études de fiabilité ?", *QUALITA2013*, Compiègne, France, mars 2013.
- [SCH 09] SCHWARTZ H. M., SEABROOKE L., "Varieties of residential capitalism in the international political economy: Old welfare states and the new politics of housing", *The politics of housing booms and busts*, p. 1–27, Springer, 2009.
- [SIM 51] SIMPSON E., "The Interpretation of Interaction in Contingency Tables", *Journal of the Royal Statistical Society - Series B (Methodological)*, vol. 13-2, p. 238-241, 1951.
- [SIT 17] SITTLER P. et al., "Digitalization in real estate", *24th Annual European Real Estate Society Conference; European Real Estate Society: Delft, The Netherlands*, 2017.
- [STR 75] STRASZHEIM M. R., "Front matter," An Econometric Analysis of the Urban Housing Market", *An Econometric Analysis of the Urban Housing Market*, p. 16–0, NBER, 1975.
- [TOP 84] TOPALOV C., *Le Profit, la rente et la ville: éléments de théorie*, Economica, 1984.
- [TRA 13] TRAVERS M., APPERE G., LARUE S., "Évaluation des aménités urbaines par la méthode des prix hédoniques : une application au cas de la ville d'Angers", *Économie et Statistique*, vol. 460, n° 1, p. 145–163, 2013.
- [WAT 01] WATKINS C. A., "The definition and identification of housing submarkets", *Environment and Planning A*, vol. 33, n° 12, p. 2235–2253, SAGE Publications Sage UK: London, England, 2001.
- [WEB 16] WEBER R., "Performing property cycles", *Journal of Cultural Economy*, vol. 9, n° 6, p. 587–603, Taylor & Francis, 2016.
- [WET 17] WETZSTEIN S., "The global urban housing affordability crisis", *Urban Studies*, vol. 54, n° 14, p. 3159–3177, SAGE Publications Sage UK: London, England, 2017.