

### 3 – Stratification

La formation des prix des logements dépend fortement de la localisation du bien. Pour prendre en compte cet aspect dans le calcul des indices, le territoire a été divisé en strates pour chaque type de bien (appartements et maisons), l'idée étant de distinguer différents marchés immobiliers : dans chaque strate sera ensuite calculé un modèle hédonique propre.

L'objectif recherché à travers ce découpage géographique est d'obtenir des strates homogènes en matière de valorisation des logements. Un équilibre a cependant dû être trouvé. Il était d'une part recherché une stratification très fine séparant au mieux la variabilité des prix sur le territoire, mais il fallait d'autre part que chaque strate possède un nombre suffisant d'observations afin d'effectuer des traitements statistiques de bonne qualité. Pour cela, il a été imposé que chaque strate contienne un nombre minimum de transactions, arbitré à 110 par trimestre en moyenne sur la période d'étude<sup>15</sup>.

Étant donné cette taille minimale des strates et l'importance de la localisation sur les prix, il reste des effets locaux non identifiés par la stratification. Afin de les capter au mieux, un zonage détaillé est parfois défini au sein des strates, qui rentre dans l'étape de modélisation (cf. partie 4-2).

La stratification a été définie différemment en Île-de-France et en province, compte tenu des contextes différents des marchés immobiliers :

- en Île-de-France, où les prix de l'immobilier s'établissent en bonne partie selon le positionnement par rapport à Paris, la stratification a été construite par une approche statistique. Celle-ci a été déclinée sur trois zones ayant été distinguées au préalable : deux pour le marché des appartements, Paris composant une zone et le reste de l'Île-de-France l'autre zone, et une seule pour les maisons, car distinguer Paris du reste de l'Île-de-France n'aurait pas d'intérêt du fait du trop faible effectif des maisons à Paris. La démarche statistique consiste, à partir du zonage géographique le plus fin disponible<sup>16</sup>, d'effectuer des regroupements successifs de ces zones selon leurs similarités en fonction de différentes variables d'intérêt<sup>17</sup>. Ces dernières, relatives à la zone (il peut s'agir alors de valeurs moyennes sur la commune ou de la répartition des modalités d'une variable), sont listées dans la figure 3-1.

---

<sup>15</sup> Soit les années 2003 à 2013 en Île-de-France et 1998 à 2007 pour la province.

<sup>16</sup> Il s'agit des quartiers administratifs à Paris, qui sont au nombre de 80, et des communes pour le reste de l'Île-de-France.

<sup>17</sup> Outre les variables issues de la base BIEN, des variables communales produites par la statistique publique ont également été utilisées.

**Figure 3-1 : Zonage et variables de classification en Île-de-France**

Appartements		Maisons
Paris	Reste de l'Île-de-France	
Nombre de pièces	Nombre de pièces	Nombre de pièces
Surface habitable	Surface habitable	Surface habitable
Époque de construction	Époque de construction	Surface du terrain
Étage et présence d'un ascenseur (au-delà du 4 <sup>e</sup> étage)	Étage et présence d'un ascenseur (au-delà du 4 <sup>e</sup> étage)	Époque de construction
Nombre de garages ou parkings	Nombre de garages ou parkings	Présence d'une dépendance
Nombre de salles de bains	Nombre de salles de bains	Nombre de bâtiments
Présence d'une cave	Présence d'une cave	Nombre de garages ou parkings
	Distance à Paris (calculée à partir des coordonnées des centres des communes)	Nombre de salles de bains
	Part des résidences principales occupées par des locataires	Nombre de niveaux
	Nombre d'habitants	Distance à Paris (calculée à partir des coordonnées des centres des communes)
	Revenu médian	Revenu médian

Dans le détail, une analyse en composantes principales (ACP) a premièrement été effectuée sur les communes ou quartiers parisiens avec ces variables. Cette méthode permet de réduire la dimension de l'espace de données étudié, tout en préservant le maximum d'informations possible : de nouvelles variables sont obtenues, appelées « composantes principales ». Ce sont des combinaisons linéaires des variables initiales, non corrélées deux à deux, et de variance maximale<sup>18</sup>.

Une classification ascendante hiérarchique (CAH) est ensuite appliquée sur ces composantes principales. Cette deuxième méthode, algorithmique, permet de créer un arbre de classification<sup>19</sup> en partant des « individus » (dans ce cas les communes, ou les quartiers parisiens) qui sont regroupés pas à pas. Différents critères existent pour effectuer

<sup>18</sup> La méthode crée initialement autant de nouvelles variables que d'anciennes, mais non corrélées et triées par leur part d'inertie (généralisation de la variance dans le cas multidimensionnel) expliquée par rapport à l'inertie totale du nuage de points (correspondant à la somme des variances de toutes les variables). Cela permet alors, en appliquant un seuil, de retenir moins de variables (les n premières, selon le seuil choisi) tout en conservant le maximum de variabilité entre les différents individus.

<sup>19</sup> La classification est effectuée en partant des individus jusqu'à ce qu'ils soient tous regroupés en une seule classe. Le résultat peut être présenté sous forme d'arbre, permettant de visualiser tous les regroupements effectués successivement et de choisir à quelle étape on souhaite s'arrêter pour déterminer les classes définitives.

les regroupements ; c'est la méthode de Ward<sup>20</sup>, optimale du point de vue de l'homogénéité des agrégats construits, qui a été utilisée ici. Dans l'arbre ainsi calculé, la classification retenue est la plus fine respectant le critère des 110 transactions minimales par trimestre pour chaque agrégat.

Afin d'affiner une dernière fois le résultat obtenu, on vérifie s'il y a de « grands agrégats » (plus de 450 transactions par trimestre) qui seraient le fruit d'un regroupement ayant eu lieu avant l'étape finale, que l'on scinde alors à nouveau sous réserve que toutes les subdivisions respectent le critère des 110 transactions trimestrielles.

La constitution des strates est alors terminée ; il est à noter que les communes (ou quartiers pour les appartements parisiens) composant une strate ne sont pas forcément contiguës, l'important étant qu'elles possèdent des caractéristiques proches concernant leurs marchés immobiliers.

Au total, on obtient 18 strates pour les appartements parisiens (figure 3-2), 29 strates pour les appartements d'Île-de-France hors Paris (figure 3-3) et 22 strates pour les maisons (figure 3-4).

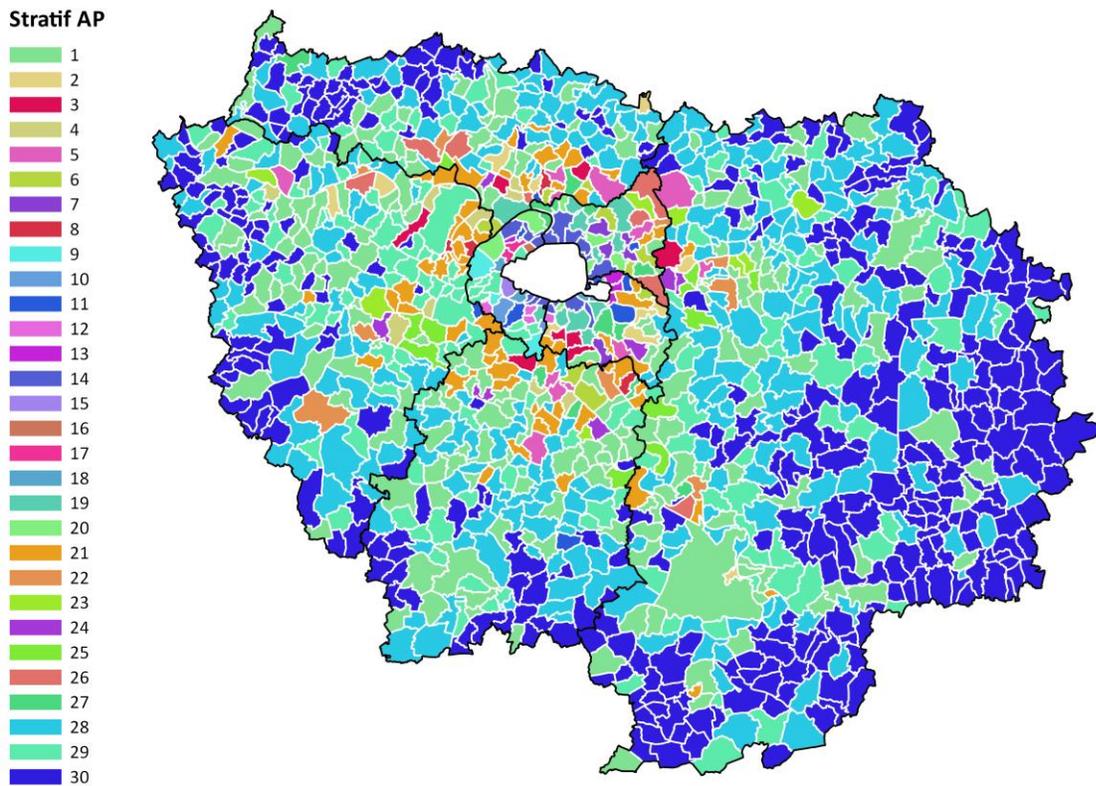
**Figure 3-2 : Stratification retenue pour les appartements parisiens**



Exemples de strates : la strate 1 regroupe les quartiers centraux de Paris, composée de 33 % d'appartements datant d'avant 1850 et de 38 % de la période haussmannienne (1850-1913). La strate 18, composée de quartiers du Nord parisien, est composée de seulement 2 % d'appartements datant d'avant 1850, mais de 54 % de la période haussmannienne.

<sup>20</sup> On peut décomposer, dans le cadre d'une classification, l'inertie totale en deux parties : l'inertie intra-classe (cumul des variabilités au sein de chaque classe) et l'inertie inter-classe (variabilité entre les classes). Avec la méthode de Ward, le regroupement effectué à chaque itération est celui qui conserve l'inertie inter-classe la plus grande possible, et inversement, qui minimise l'inertie intra-classe.

**Figure 3-3 : Stratification retenue pour les appartements en Île-de-France hors Paris**

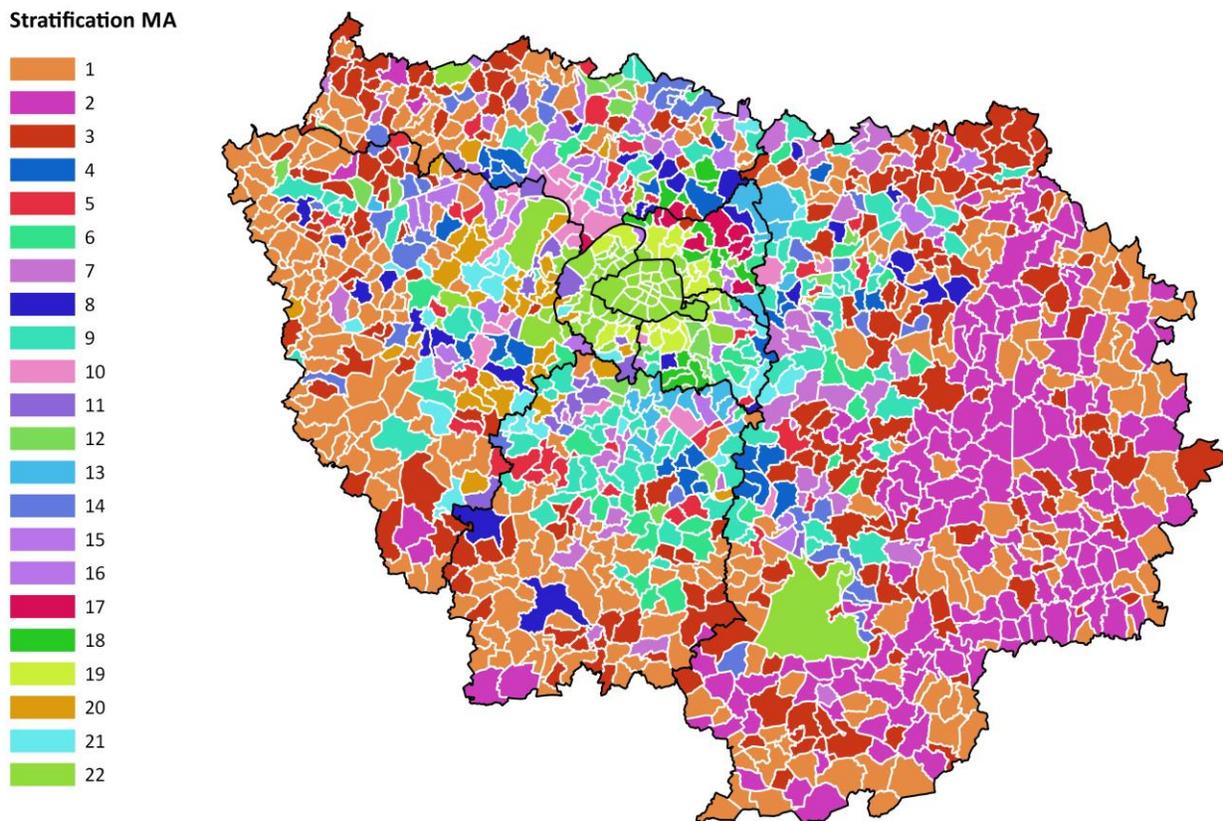


Note : la strate 30 représente les communes « orphelines », c'est-à-dire les communes (au nombre de 382) n'enregistrant aucune vente d'appartements sur la période 2003-2013 dans la base BIEN. Elles sont toutes rattachées à la strate 28, qui est la strate la plus proche en matière de typologie de commune basée sur les variables d'aménités du SDES et du recensement Insee (distance à Paris, population, part des résidences principales occupées par des locataires, revenu par ménage médian).

Exemples de strates : la strate 3 regroupe des communes à la périphérie de la petite couronne (Massy, Chelles, Poissy, Chevilly-Larue, Orly, Thiais, Franconville, Villiers-Le-Bel), situées en moyenne à 16 km du centre de Paris. Ce sont plutôt des grands appartements (8 % sont des studios, 34 % ont 4 pièces ou plus), construits sur la période après-guerre (74 % ont été construits entre 1948 et 1980, 24 % après 1980).

La strate 28 est surtout composée de communes de la grande couronne, situées en moyenne à 35 km du centre de Paris. Ce sont plutôt des petits appartements (22 % sont des studios, 13 % ont 4 pièces ou plus), construits récemment (17 % ont été construits entre 1948 et 1980, 55 % après 1980).

**Figure 3-4 : Stratification retenue pour les maisons en Île-de-France**



Note : 3 communes sont des communes « orphelines » pour les maisons d'Île-de-France. Elles ont été rattachées à la strate 1, qui est la strate la plus proche en matière de typologie de commune basée sur les variables d'aménités du SDES et du recensement Insee (distance à Paris, population, part des résidences principales occupées par des locataires, revenu par ménage médian).

Exemples de strates : la strate 1 regroupe 302 communes « rurales » de 2 000 habitants en moyenne, se situant en périphérie de la grande couronne, en moyenne à 50 km du centre de Paris. Cette strate est caractérisée par des maisons avec de grands terrains (62 % des maisons de cette strate ayant une surface de terrain supérieure au 3e quartile de l'Île-de-France), et relativement récentes (plus de 50 % des maisons construites après 1970). La part des résidences principales occupées par des locataires dans cette strate est de 16 %.

À l'inverse, la strate 22 regroupe 54 communes « urbaines » de 11 000 habitants en moyenne (dont Paris et 36 communes de la petite couronne), en moyenne à 13 km du centre de Paris. Cette strate concentre des maisons avec des petits terrains (43 % des maisons de cette strate ont une surface de terrain inférieure au 1er quartile de l'Île-de-France), et plus anciennes (seules 24 % ont été construites après 1970). La part des résidences principales occupées par des locataires dans cette strate est de 47 %. Ces maisons sont construites plus en hauteur, car elles possèdent dans un tiers des cas 2 étages ou plus, alors que seules 12 % des maisons d'Île-de-France ont plus d'un étage.

En province, on ne retrouve pas un effet central unique comme celui de Paris en Île-de-France. Il peut de plus exister des marchés locaux très cloisonnés telles les zones touristiques (stations balnéaires, stations de ski). Lors de la mise en place de la version 4 des indices, de nombreux tests ont été réalisés dans l'idée d'établir une stratification pour la province *via* des méthodes de classification, sur le même principe que pour l'Île-de-France, en ajoutant des variables distinguant les communes littorales et les stations de sport d'hiver et en utilisant non plus la distance à Paris mais la distance à la commune de plus de 50 000 habitants la plus proche. Cependant, la stratification obtenue par ce biais permettait une séparation des prix légèrement moins bonne que celle définie

précédemment et n'a donc pas été retenue. La stratification en province a ainsi été reconduite telle qu'elle était en version 3 des indices. La méthodologie est basée sur une démarche de regroupement selon la proximité géographique. Une première sélection a été effectuée à partir des unités de territoire suivantes :

- la commune : maille la plus fine. On isole dans un premier temps les communes qui dépassent le seuil minimum de 110 transactions par trimestre, qui définiront une strate à elles seules. Ceci n'est réalisé dans la pratique que pour les plus grandes communes (il y en a 33 pour les appartements et 9 pour les maisons) ;
- l'unité urbaine : on isole les unités urbaines qui dépassent le seuil minimum de 110 transactions par trimestre, en retirant au préalable les communes qui ont déjà été sélectionnées en tant que strate. Il peut donc y avoir des strates formées par une banlieue seule ou bien par l'agglomération tout entière quand une grande ville n'était pas suffisamment importante pour constituer une strate à elle seule ;
- les stations de sports d'hiver alpines : les communes ayant sur leur territoire une station de ski sont regroupées en strates par massif montagneux (massif du Mont Blanc, etc.) ;
- les communes littorales : les strates sont obtenues par regroupement géographique des communes littorales de même département ou région. Elles se distinguent des autres communes limitrophes par un prix moyen des logements plus élevé.

Toutes les autres communes, soit celles qui n'ont ni un nombre de ventes suffisant par trimestre pour être distinguées à ce titre, ni de caractéristiques particulières, ont été réparties en strates d'après les quantiles de revenu moyen par habitant de l'année 2006. Ce travail a été réalisé au cas par cas à partir d'un processus descendant en commençant au niveau des zones économiques et d'aménagement du territoire (ZEAT).

Si le seuil limite de transactions par trimestre pour une ZEAT était largement dépassé, on descendait au niveau régional. Il en va de même pour une région donnée, pour laquelle on a pu être amené à descendre à un niveau encore plus fin, le département. Dans d'autres cas, deux départements d'une même région ont été regroupés pour former une strate.

Depuis la mise en place de la V4, les 4 Drom et la Corse ont été introduits dans le champ des indices : la Corse forme à elle seule une strate et les 4 Drom ont été réunis dans la même strate faute d'activité suffisante pour pouvoir les étudier séparément.

Au total, on obtient en province 99 strates pour les appartements et 176 strates en maisons. Les cartes de la stratification « province » pour la métropole hors Corse sont fournies dans l'annexe 2 de l'*Insee Méthodes* V3 de juillet 2014 (<https://insee.fr/fr/information/2569926>).