

---

# ÉLABORATION D'UNE MAILLE GEOGRAPHIQUE POUR L'HABITAT

*Solène COLIN(\*), Vivien ROUSSEZ(\*)*

*(\*) CGDD, Service de la Donnée et des Études Statistiques*

`solene-c.colin@developpement-durable.gouv.fr`

`vivien.roussez@developpement-durable.gouv.fr`

**Mots-clés.** Analyse spatiale, graph mining, clustering, analyse multidimensionnelle, cartographie.

---

## Résumé

Le SDES a lancé en 2017 un projet visant à construire un maillage du territoire à même de rendre compte des disparités territoriales sur les enjeux propres au logement. En effet, ni les échelles administratives, ni les zonages d'études de l'Insee (zones d'emploi, bassins de vie) ne sont adaptées pour l'analyse localisée du logement, car ils mêlent dans les mêmes mailles des types d'habitat différents (urbain et périurbains notamment). Pour cela, neuf indicateurs représentatifs de l'équilibre des marchés du logement ont été sélectionnés pour alimenter la méthode de **régionalisation**, qui regroupe les communes en bassins homogènes. La méthode retenue est l'algorithme **SKATER** (Spatial Klustering Analysis by Tree Edge Removal), qui s'appuie sur l'exploration de graphe et notamment la notion d'arbre portant minimal.

Différentes simulations ont été menées pour déterminer la taille de ces mailles, et les résultats ont fait l'objet de tests au niveau régional. Ces mailles mettent au premier plan les disparités propres au logement (pouvoir d'achat immobilier, taille des ménages), alors que la maille communale fait principalement ressortir les disparités liées au degré d'urbanité. Elle permet également de distinguer les villes-centre de leur périphérie, et donc d'isoler les enjeux propres à ces espaces très différents sur le plan du logement.

L'analyse de ces mailles permet de mettre en exergue six types de marchés locaux du logement, principalement distingués par le niveau des prix, la composition des ménages (leur taille) et du parc de logement. Au fil du temps, les spécificités de ces marchés se renforcent et ceux déjà en tension voient leurs déséquilibres se renforcer. À l'inverse, le ralentissement démographique et le vieillissement à l'œuvre dans les espaces faiblement peuplés contribuent à la faible dynamique des marchés où la demande est déjà modérée. Par ailleurs, ces marchés présentent un degré inégal d'homogénéité, et on trouve d'avantage de diversité entre les communes de marchés de l'Est de la France.

# Abstract

Partant du constat que les mailles géographiques usuelles ne permettent pas une analyse fine des marchés du logement, le SDES a élaboré une maille ad hoc. Pour cela, 9 indicateurs emblématiques de la demande et de l'offre des marchés locaux du logement ont été sélectionnés et ont alimenté une méthode de **régionalisation** utilisant la théorie des graphes et les arbres portants minimaux. Ce maillage permet une lecture facilitée des dynamiques territoriales sur les plans du logement et de la démographie.

## Introduction

Les mailles habitat constituées par le SDES ont pour objectif de constituer une échelle pertinente pour l'observation et l'analyse des enjeux territoriaux liés à l'habitat. Elle permet de répondre à trois objectifs :

- Lisser visuellement l'information pour donner des cartes lisibles
- Conserver au maximum les disparités territoriales, tout en faisant ressortir les enjeux propres à l'habitat
- Alimenter la connaissance et les diagnostics locaux

Pour cela, le SDES a construit des ensembles de communes qui *se ressemblent* sur le domaine de l'habitat. Cette approche diffère de celle de l'Insee qui, pour ses zonages d'études, regroupe les communes qui sont fortement connectées les unes aux autres. Ce degré de connexion s'apprécie à l'aune du nombre de navettes domicile-travail entre les communes (pour les zones d'emploi) et les flux (théoriques) d'habitant se déplaçant de leur commune de résidence vers le pôle de services le plus proche (pour les bassins de vie). Pour plus d'information sur les maillages produits par l'Insee, consultez cette page.

## 1 Sélectionner les indicateurs pertinents

Choisis dans le cadre de deux comités regroupant des experts des territoires et du logement, les 9 indicateurs mobilisés sont représentatifs des marchés du logement, à la fois dans l'offre, la demande, et les équilibres de marché.

### 1.1 Quels indicateurs

Ces indicateurs sont calculés à partir de trois sources : recensements de la population, Filocom (Fichier des Logements localisés à la Commune), bases notariales (PERVAL et BIEN). Certaines communes ne présentant pas de donnée pour un ou plusieurs indicateurs, une **imputation** a été réalisée en prenant (de façon spatialement récursive) la moyenne de cet indicateur sur les communes voisines. Ces indicateurs sont présentés en détail dans le rapport méthodologique disponible sur le site internet du SDES.

1. L'indicateur de jeunesse du parc : part des logements récents (construits après 1975) rapportés à la part des logements anciens (construits avant 1949).
2. Durée d'occupation médiane des logements
3. Taux de transactions dans le marché de l'ancien (nombre de transactions rapporté au nombre de logements)
4. Part de logements sociaux
5. Part de logements en situation de suroccupation
6. Prix au  $m^2$  dans l'ancien rapportés au revenu médian communal
7. Nombre de personnes par ménage

8. Part des résidences secondaires
9. Part de logements vacants

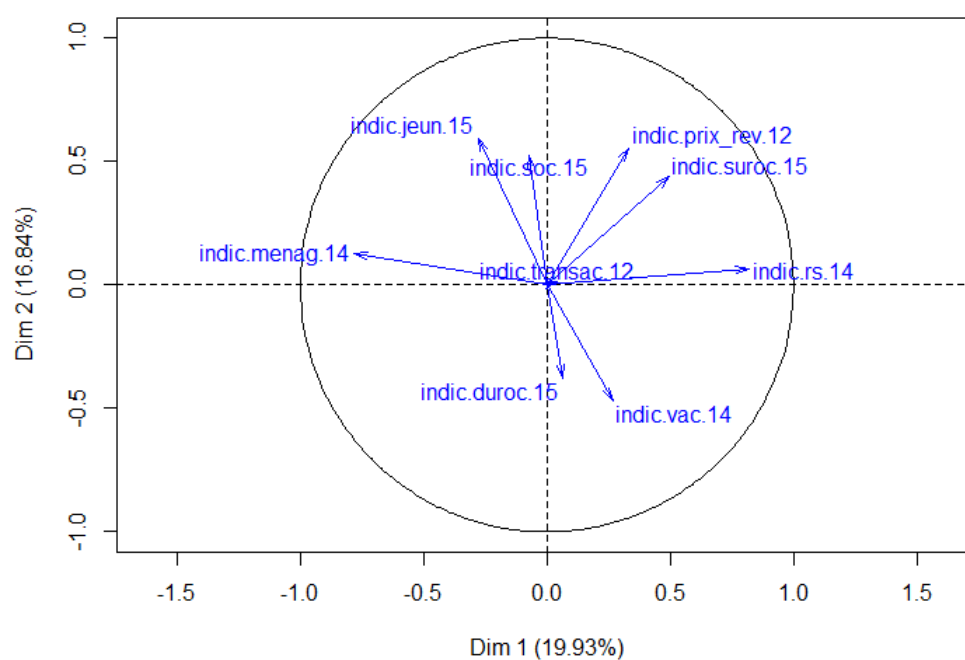
## 1.2 Analyse de quelques indicateurs

## 1.3 Validation des indicateurs

### 1.3.1 Analyse en composantes principales

Pour valider le choix des indicateurs sélectionnés (plusieurs itérations ont été nécessaires), une analyse en composantes principales (ACP) a été réalisée sur la base des indicateurs retenus. L'objectif consistait à retenir un minimum d'indicateurs pertinents et surtout **non redondants**, mais en conservant également quelques mesures "incontournables" de l'état du marché du logement (la part de logements sociaux, notamment).

FIGURE 1 – ACP sur les indicateurs retenus



Les résultats de l'ACP montrent que les indicateurs ne présentent pas de redondance d'information, ce qui aurait pu conduire à sur-pondérer une dimension (le clivage urbain / rural par exemple). Les 6 axes retenus permettent de mettre en lumière un certain nombre de caractéristiques territoriales, résumées dans le tableau suivant :

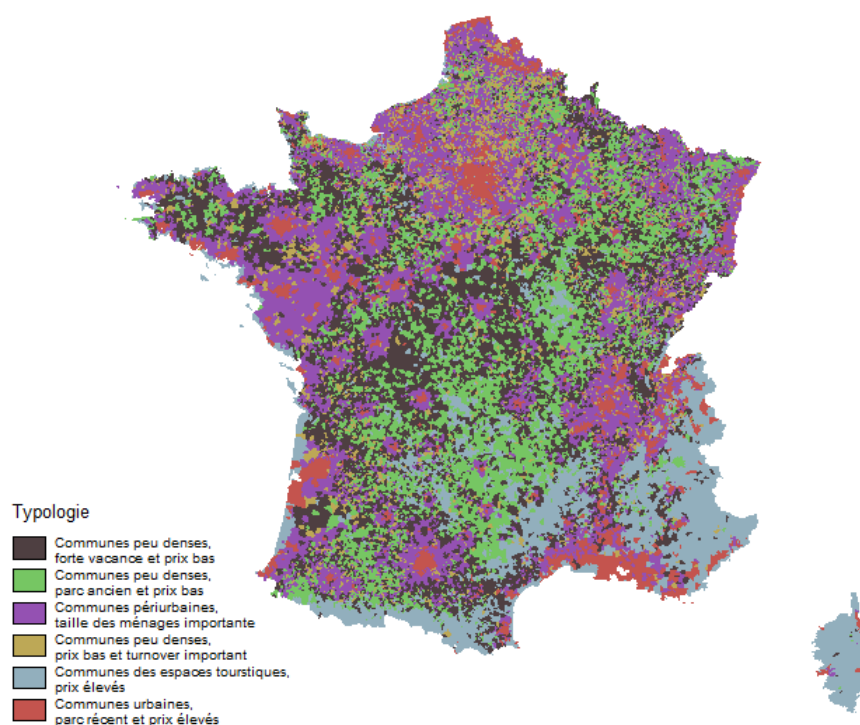
FIGURE 2 – Principaux axes factoriels

Axe factoriel	Variables constitutives	Signe	Disparités territoriales révélées
Axe 1	Jeunesse du parc Taille des ménages	+ +	Gradient urbain
Axe 2	Prix Suroccupation Taille des ménages	+ + -	Gradient de tension
Axe 3	Vacance Logements sociaux Durée d'occupation	- - +	Zones touristiques
Axe 4	Taux de transactions	+	Coeurs des AU Marchés dynamiques de l'Ouest et du Nord
Axe 5	Durée d'occupation Social	+ +	Zones Nord-Est (propriétaires) versus Sud et Sud-Est
Axe 6	Suroccupation	+	Couronnes périurbaines et zones touristiques du Sud-Est

### 1.3.2 Classification communale

Par la suite, on effectue une classification des communes par centres mobiles, opérée sur les coordonnées factorielles des 6 premiers axes. Cette classification permet de mettre en exergues six groupes de communes qui permettent de rendre compte des principales disparités territoriales attendues.

FIGURE 3 – Classification des communes



Ces types sont cohérents avec d'autres études déjà réalisées, soit sur la dimension territoriale, soit sur le logement. Ils confirment le choix des indicateurs qui a été fait. C'est donc sur cette base qu'est faite la régionalisation.

## 2 La régionalisation

La **régionalisation** désigne l'opération qui consiste à regrouper des unités géographiques élémentaires en un ensemble **contigu**, selon des critères statistiques. Il existe un grand nombre de méthodes de régionalisation, dont une partie est comparée dans cet article. Le SDES a retenu l'algorithme SKATER (Spatial Klustering Analysis by Tree Edge Removal), implémenté et mis à disposition dans le package SPDEP du logiciel statistique libreR.

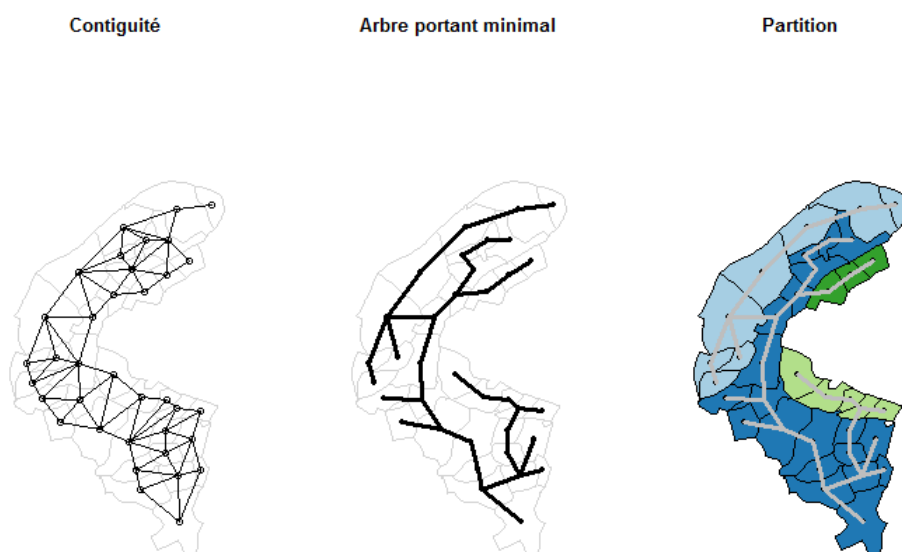
### 2.1 L'algorithme SKATER

Cet algorithme fonctionne en 4 étapes :

1. Construction de la matrice de contiguïté → obtention d'un *graphe* (un noeud = une commune et un lien = relation de contiguïté entre deux communes)
2. Pondération de ce graphe avec les dissimilarités calculées à partir des indicateurs (distance euclidienne)
3. Construction de l'**arbre** portant minimal, en retenant le lien avec le voisin le plus ressemblant pour chaque nœud du graphe
4. Suppression itérative des branches de l'arbre maximisant la variance inter-classes des sous-graphes obtenus après élagage.

Il peut donc être vu comme une forme de classification descendante hiérarchique opérée sur les liens de l'arbre portant minimal. Le fait de travailler à partir de cet arbre portant minimal plutôt que sur les nœuds du graphe garantit la contiguïté des mailles finales.

FIGURE 4 – Illustration des étapes de SKATER sur les Hauts de Seine

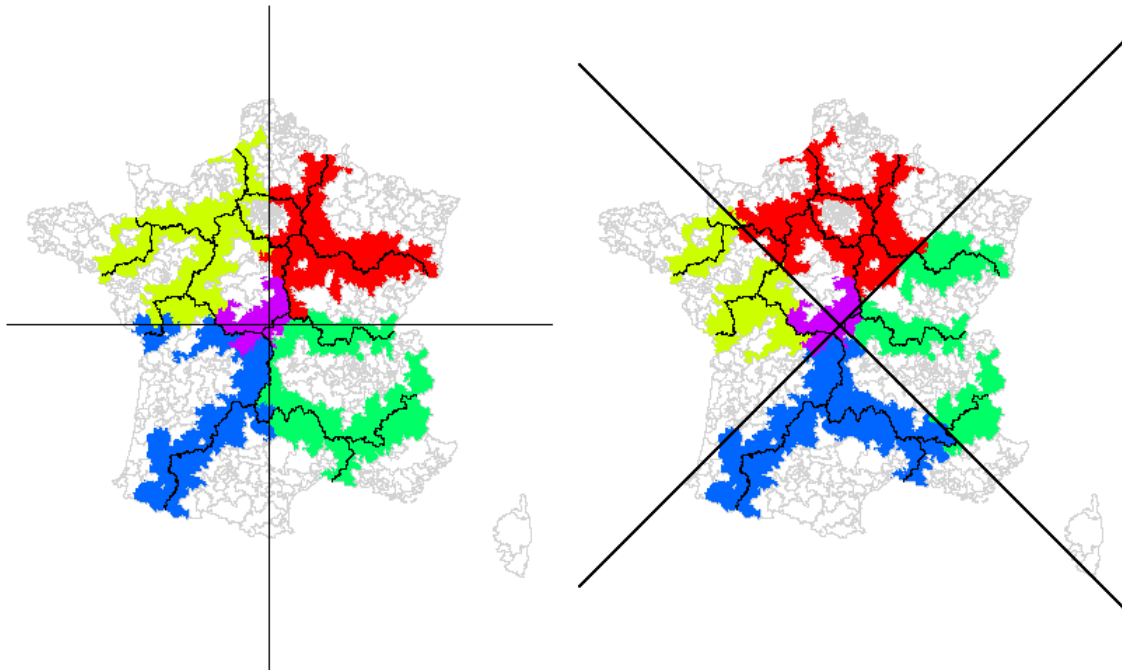


## 2.2 Découper le problème

Cet algorithme est très coûteux : sa complexité est en effet légèrement supérieure à  $o(n^2)$ , et il n'a pas été optimisé dans un langage plus efficace que R. En conséquence, il est impossible de traiter les près de 36000 communes françaises. En effet, le traitement de l'Île-de-France (moins de 1300 communes) prend environ 5 minutes, alors qu'il faut 3 heures pour le Grand-Est (environ 5000 communes). C'est pourquoi l'algorithme a été exécuté région par région. Toutefois, pour s'affranchir de ces limites administratives, on procède à un *lissage* des frontières régionales. Pour cela, on identifie les zones qui touchent une frontière régionale et on refait un zonage à partir de l'ensemble des communes qui en font partie.

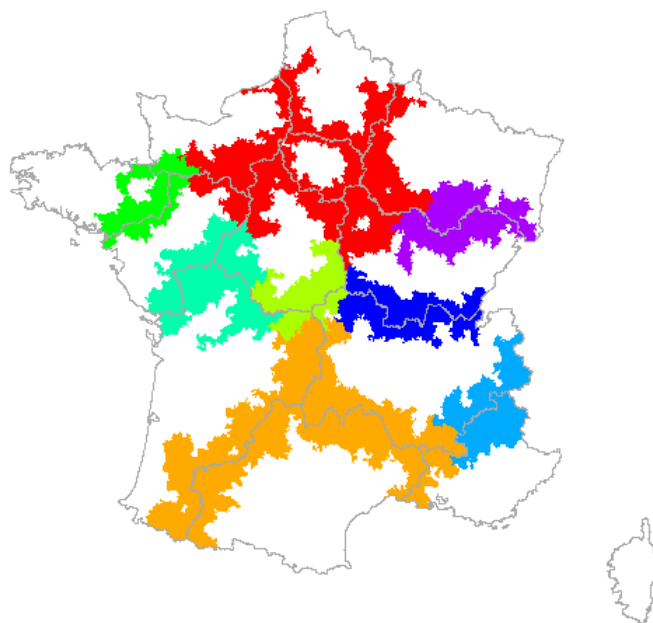
L'ordre de grandeur du nombre de communes qui se trouvent dans les zones concernées est de 16 000. La difficulté liée au temps de calcul apparaît donc de nouveau. Pour contourner ce problème, il faut couper cet ensemble en plusieurs sous-ensembles de taille plus petite. Pour réaliser ce découpage, on a repéré le barycentre du pays. La zone se trouvant sur ce barycentre, ainsi que les zones contiguës constituent une première *macro-zone*.

FIGURE 5 – Répartition des zones frontalières



On répartit ensuite le reste des communes dans les 4 cadrans du repère orthonormé ayant pour origine ce barycentre. Toutefois, les axes de ce repère passant le long de frontières, on conserve en partie l'effet de frontière. Pour éviter ce biais, on effectue une rotation du repère orthonormé de 45 degrés. Cette méthode fait néanmoins apparaître, selon les cas de figure, de très petites zones isolées, et des macro-zones qui ne sont pas contiguës. Une dernière étape consiste donc à rattacher les isolats à la macro-zone contiguë de taille suffisante ou à séparer les parties non contiguës des grandes macro-zones.

FIGURE 6 – Découpage final en macro-zones



## 2.3 Les différentes simulations

L'algorithme fonctionne à l'aide de deux paramètres qui permettent de faire varier le résultat final : le nombre de mailles finales souhaitées et une taille minimale (en termes de nombre de communes, de population, de nombre de logements...).

## 3 Caractérisation des mailles

### 3.1 L'habitat au premier plan

### 3.2 Typologie des mailles

### 3.3 Cohésion et dynamique de ces mailles