

---

## SECTORISATION DES ÉLÈVES AU COLLÈGE : UNE MÉTHODE D'IMPUTATION POUR RECONSTITUER LES CONTOURS INCONNUS DE LA CARTE SCOLAIRE

Sylvain MAUGIS (\*), Mustapha TOUAHIR (\*\*)

(\* ) MEN-DEPP

(\*\* ) MEN-DEPP

[sylvain.maugis@education.gouv.fr](mailto:sylvain.maugis@education.gouv.fr)

**Mots-clés** : carte scolaire, collèges, géolocalisation des élèves, évitement, imputation, ségrégation

### Résumé

En France, le collège de scolarisation d'un élève est déterminé par le conseil départemental à partir de son lieu de résidence. Dans ce contexte, l'information sur le collège de sectorisation de chaque élève n'est pas disponible au niveau national. Quelques initiatives locales ont néanmoins permis d'obtenir la carte scolaire pour les départements de la Gironde, du Puy-de-Dôme et de Paris. L'objectif de cet article est de proposer une méthode qui permette de déterminer les contours de la carte scolaire quand celle-ci est inconnue. Pour ce faire, nous utilisons le fichier géolocalisé 2015 qui recense tous les élèves inscrits à la rentrée 2015, avec leurs caractéristiques personnelles et scolaires et leurs coordonnées géographiques. Le fait de connaître les secteurs réels dans les trois départements cités est utilisé pour juger la qualité de la reconstitution.

Nous sectorisons les élèves grâce à ce que nous nommons la « méthode des voisins » : pour un élève donné, nous déterminons ses 10 voisins les plus proches ; si, parmi eux, un collège de scolarisation majoritaire se dégage, ce collège est attribué à l'élève. Cette opération est effectuée après avoir retiré des élèves voisins ceux dont on sait qu'ils sont plus souvent voire toujours scolarisés hors secteur, autrement dit les élèves scolarisés dans le privé, ceux qui le sont dans un autre département ou très loin de leur domicile, les élèves ayant changé de collège au cours de leur scolarité et ceux qui sont inscrits dans des formations atypiques. Cette première étape permet d'attribuer un collège de secteur à la quasi-totalité des élèves. Le taux d'erreur dépend cependant beaucoup de l'ampleur de la majorité qui se dégage : à plus de 9 voisins sur 10 scolarisés dans le même collège, on se trompe dans seulement 1 % des cas ; dans le cas d'une majorité relative, le taux d'erreur s'élève à 48 %. Les résultats de la première itération nous permettent (dans le cas d'une majorité forte) de corriger et d'enrichir la table des voisins initiale, en ajoutant notamment des élèves du privé pour lesquels un collège public a été attribué. À l'issue d'une deuxième itération du processus de détermination des voisins, un collège de secteur est attribué à 99 % des élèves à sectoriser, le taux d'erreur global s'élevant à 6 %. On attribue enfin aux 1% d'élèves non sectorisés le collège attribué à l'élève le plus proche d'eux.

Les élèves ainsi sectorisés peuvent se classer en deux groupes : ceux dont le collège attribué est quasi certain (groupe 1) et les autres (groupe 2). Dans le cadre d'une estimation statistique, nous supposons que seul le groupe 2 est susceptible de biaiser les résultats. Pour contrôler ce biais, nous développons une approche en « situations extrêmes ». Par exemple, pour la part d'élèves hors secteur par département, nous considérons que les estimations extrêmes sont obtenues en considérant que soit tous les élèves du groupe 2 sont hors secteur, soit aucun d'entre eux ne l'est.

Ces deux situations permettent d'obtenir des intervalles de confiance, dont nous vérifions alors qu'ils contiennent bien les valeurs réelles pour les trois départements étudiés. L'estimation est plus délicate au niveau établissement, le périmètre de recrutement de chaque collège étant lui-même soumis à incertitude. Nous montrons que l'estimation reste possible, en adaptant l'approche en situations extrêmes. Les résultats obtenus pour l'estimation de la part d'élèves recourant au privé laissent penser que la méthode est fiable même au niveau de l'établissement.

Dans le cadre de l'estimation d'une statistique plus complexe (par exemple, l'impact du recours au hors secteur sur la ségrégation entre établissements), une approche en « situations probables » semble plus judicieuse. L'idée consiste à lire dans le collège de leurs voisins l'établissement dont dépendent les élèves. Nous présentons le cadre théorique de cette approche et esquissons les exploitations qu'elle permet d'envisager.

## **Abstract**

In France, it is the department council which determines, from the location of each pupil, the high school (named "collège", usually between 11 and 15 years old) where he is supposed to be. In this context, the information on the sectoring school of each student is not available at the national level. Nevertheless, thanks to some local initiatives, the school map is known for the departments of Gironde, Puy-de-Dôme and Paris. The purpose of this article is to propose a method to determine the contours of the school map when it is unknown. To do this, we use the 2015 geolocated file that lists all students enrolled in the 2015 school year, with their personal and school characteristics and their geographic coordinates. Knowing the real sectors in the three departments mentioned is used to judge the quality of the reconstitution.

We sectorise pupils through what we call the "neighbors method": for a given pupil, we determine his 10 closest neighbors; if, among them, there is a majority school, this school is attributed to the pupil. This operation is carried out after having withdrawn from neighboring pupils those who are known to be more often or always schooling out of their sector, in other words the pupils schooled in the private sector, those who are in another department or very far from their home, pupils who have changed schools during their schooling and those who are enrolled in atypical courses. This first step makes it possible to assign a sector school to almost all pupils. The error rate, however, depends very much on the size of the majority that emerges: where more than 9 out of 10 neighbors in the same school, only 1% of cases are mistaken; in the case of a relative majority, the error rate is 48%. The results of the first iteration allow us (in the case of a strong majority) to correct and enrich the initial table of neighbors, including private students for whom a public school has been attributed. At the end of a second iteration of the process of determining neighbors, 99% of pupils are assigned to a sector school, with an overall error rate of 6%. Finally, we attribute to the 1% of non-sectoral students the school attributed to their closest neighbour.

The students thus sectorised can be classified in two groups: those whose assigned school is almost certain (group 1) and the others (group 2). In a statistical estimation, we assume that only group 2 is likely to skew the results. To control this bias, we develop an approach in "extreme situations". For example, for the proportion of out of sectors pupils by department, we consider that extreme estimates are obtained by considering that all students in group 2 are out of school, or none of them are. These two situations make it possible to build confidence intervals, which contain real values for the three studied departments. The estimate is more delicate at the establishment level, the recruitment perimeter of each school being itself subject to uncertainty. We show that estimation remains possible, adapting the approach in extreme situations. The results obtained for the estimation of the share of pupils using the private sector suggest that the method is reliable even at the school level.

In the estimation of a more complex statistic (for example, the impact of off-sector use on segregation between institutions), a "likely situations" approach seems more appropriate. The idea is

to read in the school of their neighbors the school on which the students depend. We present the theoretical framework of this approach and outline the operations it allows to consider.

## I. Contexte et objectifs

### Brève histoire de la carte scolaire

La carte scolaire désigne la procédure qui, à chaque élève, attribue une école ou un établissement de scolarisation en fonction de son lieu de résidence.

Dans un contexte où les enjeux de mixité à l'école sont jugés importants, les débats sur la sectorisation des élèves sont nombreux. Au cours des années passées, ils ont notamment porté sur l'assouplissement de la carte scolaire mis en place par la réforme de 2007. Cette réforme avait pour but de donner davantage de liberté aux familles dans le choix de l'établissement de leur enfant, et de corriger ainsi les contraintes résidentielles. Mais elle s'est traduite par des effets contrastés sur la mixité des établissements : la ségrégation aurait tendance à augmenter en fonction du nombre de dérogations accordées. Plus récemment, la question de l'assouplissement de la carte scolaire a laissé place au débat sur la définition des contours de secteurs. Depuis la rentrée 2016, des secteurs multi-collèges sont ainsi expérimentés dans plusieurs départements.

L'objectif de mixité assigné à la carte scolaire est relativement récent. Créée en 1963, la carte scolaire visait à l'origine à mieux gérer les effectifs d'élèves accueillis au sein des établissements, à éviter les déséquilibres entre eux. Pour cela, le principe de secteurs de collèges dépendant du domicile de l'élève a donc été retenu. La promotion de la mixité sociale est apparue plus tard. Elle est évoquée en particulier dans la circulaire n°98-263 du 29 décembre 1998. Celle-ci encourage les redécoupages, précise que les dérogations doivent faire l'objet d'un contrôle strict et que les options proposées dans chaque établissement peuvent servir à lutter contre la ségrégation scolaire.

### Le collège, lieu de cristallisation des enjeux

C'est au collège que la question de la carte scolaire se pose avec le plus d'acuité. En effet, on observe, à l'entrée en sixième, un afflux vers les établissements privés qui n'a pas d'équivalent aux autres niveaux de la scolarité. À la rentrée 2015<sup>1</sup>, 14,1 % des élèves de CM2 sont scolarisés dans le secteur privé ; à la rentrée 2015, c'est le cas de près de 22 % des élèves de 6<sup>ème</sup>. Près de 9 % des élèves de sixième sont scolarisés dans un collège privé alors qu'un an plus tôt, ils étaient scolarisés en CM2 dans une école publique.

La place particulière qu'occupent les « années collège » s'observe également en fin de troisième. En sortie de collège, un afflux, certes moins fort mais tout de même important, met en évidence le retour de certains élèves vers l'enseignement public : près de 7 % des élèves sont scolarisés en seconde générale et technologique dans un lycée public alors qu'ils fréquentaient un collège privé un an auparavant. Le flux inverse (passage du secteur public au privé) est deux fois moins important (cf. tableau 1).

---

<sup>1</sup> Les chiffres présentés ici portent sur le millésime 2015 par souci de cohérence avec les données qui seront exploitées dans la suite de l'article.

**Tableau 1**  
**Changements de secteur entre les rentrées 2014 et 2015 (en %)**

<b>Transition entre</b>	<b>Privé - public</b>	<b>Public - privé</b>	<b>Public - public</b>	<b>Privé - privé</b>	<b>Total</b>
la PS et la MS	0,4	0,7	88,5	10,3	100,0
la MS et la GS	0,4	0,8	88,5	10,3	100,0
la GS et le CP	0,4	1,9	87,0	10,7	100,0
le CP et le CE1	0,4	0,8	86,7	12,1	100,0
le CE1 et le CE2	0,4	0,9	86,5	12,3	100,0
le CE2 et le CM1	0,4	0,9	85,9	12,8	100,0
le CM1 et le CM2	0,3	0,8	85,5	13,3	100,0
le CM2 et la 6 <sup>e</sup>	<b>2,3</b>	<b>8,9</b>	76,1	12,7	100,0
la 6 <sup>e</sup> et la 5 <sup>e</sup>	0,7	0,8	77,8	20,7	100,0
la 5 <sup>e</sup> et la 4 <sup>e</sup>	0,8	0,9	77,9	20,3	100,0
la 4 <sup>e</sup> et la 3 <sup>e</sup>	0,8	1,0	77,9	20,3	100,0
la 3 <sup>e</sup> et la 2 <sup>nde</sup> GT	<b>7,0</b>	<b>3,5</b>	72,3	17,1	100,0
la 2 <sup>nde</sup> GT et la 1 <sup>ère</sup> GT	1,2	1,1	77,8	20,0	100,0
la 1 <sup>ère</sup> GT et la terminale GT	0,4	0,3	78,3	21,1	100,0

Sources : Fichier Scolarité 2015 – Diapre 2015

Champ : France métropolitaine + DOM

Note de lecture : parmi les entrants en 6<sup>ème</sup> à la rentrée 2015, 8,9 % sont scolarisés dans le privé alors qu'ils étaient dans une école publique l'année précédente.

La quasi-totalité des travaux relatifs à la carte scolaire traitent des questions de mixité et de ségrégation au collège. Les chercheurs s'attachent notamment à étudier l'évitement des collèges publics. Oberti [2007] montre que les familles, notamment celles des classes moyennes et supérieures, intègrent la sectorisation scolaire dans leur projet de déménagement. Gilotte et Girard [2005] évaluent quant à eux la proportion d'élèves évitant leur collège public de secteur. D'autres auteurs évaluent la ségrégation scolaire, qui résulte à la fois de la ségrégation résidentielle (en lien avec le lieu d'habitation de l'élève) et du découpage des secteurs [Fack, Grenet et Benheda 2014]. D'autres travaux encore font ressortir des différences de valeur immobilière d'un secteur de collège à l'autre : plus un collège est attractif et plus les logements se vendent cher dans son périmètre de recrutement [Fack et Grenet 2009].

### **Le collège public de secteur : une information non disponible au niveau national**

Les études sur la carte scolaire seraient sans doute plus riches encore si l'on disposait d'une donnée supplémentaire dans les systèmes d'information nationaux : l'établissement de secteur dont dépend chaque élève. À ce jour, cette information n'est disponible ni pour le premier ni pour le second degré. Une telle absence s'explique notamment par le fait que la sectorisation est une compétence des collectivités territoriales (les communes s'occupant des écoles, les départements des collèges).

Les bases élèves permettent de décrire, plus ou moins précisément, à chaque rentrée scolaire, la formation suivie par les élèves, mais le seul établissement figurant dans ces fichiers correspond à celui dans lequel l'élève est scolarisé.

Si les données sur le collège de secteur font défaut au niveau national, elles sont en revanche parfois disponibles au niveau local. En effet, certaines collaborations entre conseils départementaux et Services Statistiques Académiques (SSA) ont permis de constituer des fichiers où figure, pour chaque élève, le collège de secteur. De telles initiatives locales ont notamment été menées dans les académies de Bordeaux, de Clermont-Ferrand, respectivement pour les départements de la Gironde et du Puy-de-Dôme. L'étude sur la ségrégation dans les collèges de Clermont-Ferrand par exemple

s'est appuyée sur les données ainsi récupérées (Cadoret, 2017). Du reste, les secteurs sont connus à Paris : la ville, qui en a la compétence, les a mis à disposition.

### **Reconstituer la carte scolaire à partir des coordonnées géographiques des élèves**

Le collège public de secteur est certes inconnu pour la quasi-totalité des élèves de France, mais d'autres données géographiques sont disponibles au niveau individuel, susceptibles d'aider à la reconstitution de la carte scolaire. En particulier, dans le second degré, les bases Scolarité contiennent les adresses des élèves, lesquelles peuvent être géocodées. La DEPP procède à la géolocalisation des élèves depuis la rentrée 2009. Compte tenu de la relative complexité de l'opération – toutes les adresses saisies dans les systèmes de gestion ne sont pas normalisées même si leur qualité s'améliore – l'opération n'est conduite qu'à une fréquence bisannuelle. Elle concerne tous les élèves de France métropolitaine et des DOM. Le dernier fichier disponible à ce jour est celui de la rentrée 2015. Il fournit, pour chaque élève, des coordonnées x et y exprimées dans une projection (Lambert 93) qui permettent de le situer sur une carte et de calculer les distances qui le séparent des autres élèves et des établissements alentour.

L'objectif de cet article est de proposer une méthode d'attribution d'un collège de secteur à chaque élève en s'appuyant sur les données de géolocalisation. La méthode présentée ici est appelée « méthode des voisins ». Elle repose sur l'idée qu'un élève est, en toute logique, sectorisé dans le même établissement que ses plus proches voisins. Bien sûr, nous ne connaissons pas le collège de secteur des élèves voisins. Mais dans la mesure où, a priori, une grande majorité des élèves sont scolarisés dans leur collège de secteur, nous pouvons émettre l'hypothèse qu'il en est de même pour les élèves voisins. Une autre manière de décrire l'intuition sous-jacente à cette méthode consiste à voir dans l'évitement scolaire la volonté de certaines familles d'échapper précisément à leur voisinage : les élèves qui ne sont pas scolarisés dans leur collège de secteur le sont souvent dans un établissement différent de celui que fréquentent leurs voisins.

Concrètement, on peut considérer que le collège de secteur d'un élève est, le cas échéant, celui dans lequel sont scolarisés une majorité de ses voisins. Cette définition pose plusieurs questions : quels voisins doit-on prendre en compte, et combien ? De quelle majorité parle-t-on ? Que faire lorsqu'aucune majorité ne se dégage ? Ces points sont discutés dans la suite de l'article. Nous traiterons aussi la question de la fiabilité de la méthode. Pour ce faire, nous nous appuyerons d'une part sur les vrais secteurs de collège des départements de la Gironde, de Paris et du Puy-de-Dôme ; grâce à cette information, nous serons en mesure de dire, pour chaque élève d'un des trois départements, si le collège fictif attribué est le bon établissement ou non<sup>2</sup>. D'autre part, la méthode « du collège le plus proche » pourra aussi servir d'élément de comparaison. Cette méthode consiste à sectoriser l'élève dans le collège le plus proche de son domicile.

Remarque : pour la Gironde, l'opération de géocodage a été effectuée par deux services différents : le SSA de Bordeaux pour le fichier contenant les vrais secteurs, l'Insee pour le fichier national Scolarité de la Depp. Pour un même élève ou une même adresse, des coordonnées différentes peuvent ainsi être observées, même si ces différences sont relativement faibles.

---

<sup>2</sup> Les fichiers géolocalisés enrichis de l'information sur le vrai secteur de collège seront par la suite appelés fichiers géolocalisés enrichis.

## II. La méthode du collège le plus proche

### Premiers résultats obtenus par la méthode « du collège le plus proche »

Les élèves que nous essayons de sectoriser sont donc les élèves résidant dans les départements du Puy-de-Dôme, de la Gironde et de Paris scolarisés dans un collège à la rentrée 2015. Tous les élèves sont pris en compte, qu'ils soient scolarisés dans un collège public ou privé. En effet, un des intérêts du travail présenté ici réside dans le fait qu'on cherche à attribuer un collège public de secteur y compris aux élèves qui ont recours à l'enseignement privé.

Intuitivement, la méthode la plus simple pour essayer de déterminer, à partir des données géographiques, le collège dont dépend un élève consiste à repérer l'établissement le plus proche de son domicile.

Le calcul de la distance entre les élèves et les établissements est rendu possible par le fait que les uns comme les autres sont géolocalisés. Toutefois, la qualité du géocodage des élèves n'est pas toujours parfaite. Dans le « fichier géolocalisé 2015 », une variable décrit, pour chaque élève, la qualité de la géolocalisation. Elle reflète la distance entre les coordonnées x et y qui ont été associées à l'adresse de l'élève et les coordonnées x et y présentes dans le référentiel national des adresses. Dans la plupart des cas, cette distance est nulle : les coordonnées x et y associées à l'adresse de l'élève existent dans le référentiel. Lorsque cette distance est strictement inférieure à 100 mètres, la géolocalisation est considérée comme « bonne ». Elle est « acceptable » lorsque la distance est supérieure ou égale à 100 mètres et strictement inférieure à 500 mètres. La qualité est considérée comme « mauvaise » pour les distances supérieures ou égales à 500 mètres.

Les secteurs de collèges se dessinent autour de contours précis. Placer le domicile d'un élève à plus de 500 mètres de son véritable emplacement – particulièrement en ville où la superficie des secteurs est plus petite – est susceptible de produire une image fautive du secteur de collège. C'est pourquoi nous retirons du champ les élèves des zones urbaines dont la qualité de la géolocalisation est « mauvaise » (cf. tableau 2). Enfin, toutes les distances sont calculées à vol d'oiseau.

**Tableau 2**  
**Sélection des collégiens à sectoriser**

	<b>Gironde</b>	<b>Puy-de-Dôme</b>	<b>Paris</b>	<b>Total</b>
Nombre d'élèves résidant dans le département et disposant de coordonnées x et y	72 001	28 772	79 145	179 918
Dont la géolocalisation est « bonne », « acceptable » ou éventuellement « mauvaise » si l'élève habite une commune rurale	68 588	27 735	78 215	174 538
Dont scolarisés dans le public	56 799	22 496	54 331	133 626
Dont scolarisés dans le privé*	11 789	5 239	23 884	40 912
% d'élèves scolarisés dans le privé*	17,2%	18,9%	30,5%	23,4%

\* Y compris le collège parisien Rognoni, à recrutement sélectif

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

La méthode de l'établissement le plus proche attribue le bon collège de secteur à 72,1 % des élèves. Le taux d'erreur est en effet de 27,9 % (cf. tableau 3). Il est nettement plus élevé à Paris (33,8 %) qu'en Gironde et dans le Puy-de-Dôme (respectivement 23,6 % et 22,2 %). Ces différences reflètent les écarts d'équipements entre les zones urbaines et rurales. Loin de ressembler à de simples cercles, les secteurs de collèges sont des créations hétérogènes reflétant la géographie, les routes, les possibilités offertes par les transports ou des décisions politiques.

**Tableau 3**  
**Taux d'erreur obtenu avec la méthode du collège le plus proche**

Nombre d'élèves pour lesquels le secteur fictif	Gironde	Puy-de-Dôme	Paris	Total
est le bon	52 414	21 573	51 810	125 797
est erroné	16 174	6 162	26 405	48 741
Taux d'erreur (en %)	23,6	22,2	33,8	27,9

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

### Le cas des élèves résidant à proximité immédiate d'un collège

L'information sur l'établissement le plus proche du domicile de l'élève n'est cependant pas inexploitable. Combinée à la distance à cet établissement, elle met en lumière le résultat suivant : lorsque l'on habite à proximité immédiate d'un établissement, on en dépend forcément.

Dans les départements de Paris, de la Gironde et du Puy-de-Dôme, tous les élèves résidant à moins de 25 mètres d'un collège font systématiquement partie du secteur de ce collège. Cette distance minimale est bien sûr différente selon le type de commune. Ainsi, les élèves des communes rurales résidant à moins de 153 mètres de leur établissement le plus proche sont toujours sectorisés sur celui-ci (cf. tableau 4). Dans une ville isolée, cette distance est de 550 mètres. Elle est plus faible pour les élèves habitant des communes de banlieue (100 mètres) ou le centre d'une agglomération (25,4 mètres). Au total, on peut inférer avec certitude le collège de secteur de tous les élèves se trouvant dans l'une de ces quatre situations, soit 1 650 élèves au total pour les trois départements étudiés.

**Tableau 4**  
**Proximité avec un collège et sectorisation certaine sur ce collège**  
**Par type de commune de résidence de l'élève**

Type commune	Distance en dessous de laquelle le collège le plus proche est toujours le collège de secteur	Nombre d'élèves concernés	Nombre d'élèves concernés scolarisés hors secteur (y compris dans le privé)
Commune rurale	153 mètres	111	15
Commune ville isolée	550 mètres	1 014	169
Commune centre d'une agglomération	25,4 mètres	291	126
Commune de banlieue	100 mètres	234	52
Total	Sans objet	1 650	362

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Bien sûr, les seuils déterminés ici l'ont été à partir de 3 départements seulement, mais l'hétérogénéité des territoires qu'ils couvrent laisse penser que les valeurs seraient tout aussi valables dans d'autres départements. En particulier, il paraît improbable d'obtenir, en zone urbaine, des seuils inférieurs à ceux de Paris. Dit autrement, dans l'optique d'une généralisation de ce résultat, tout élève qui résiderait en milieu urbain à moins de 25 mètres d'un collège pourra être considéré comme sectorisé sur celui-ci de manière certaine, que l'élève habite à Lyon, Marseille ou tout autre ville centre d'agglomération. Certes les effectifs d'élèves concernés sont faibles, mais ce critère de « la proximité immédiate » permet tout de même de repérer certains élèves hors secteur.

Dans les trois départements d'intérêt, parmi les 1 650 collégiens, 362 sont scolarisés dans un autre établissement que celui dont ils dépendent (cf. tableau 4).

### III. La méthode des voisins

#### Principe général : retrouver le collège de secteur de chaque élève à partir de celui que fréquentent ses voisins

Dans les départements de Paris, de la Gironde et du Puy-de-Dôme, les élèves habitant près d'un collège (« près » s'entend ici au sens des distances définies plus haut) représentent 0,9 % des effectifs totaux de collégiens. Afin d'attribuer un collège de secteur à tous les autres élèves, nous allons nous appuyer sur l'information apportée par leurs élèves voisins.

Comme évoqué précédemment, la méthode consiste donc à s'enquérir du collège d'inscription des plus proches voisins de chaque élève. Si une majorité de ces voisins est scolarisée dans le même établissement, on considère alors qu'il s'agit là du collège de secteur de l'élève considéré. Cette méthode suppose que, dans une zone donnée, les élèves scolarisés hors de leur secteur sont minoritaires. Elle peut par ailleurs aboutir à des cas d'incertitude, notamment en bordure de secteur. Nous verrons qu'en dépit de ces limites, la méthode des voisins permet d'obtenir de meilleurs résultats que celle du « collège le plus proche ».

Pour optimiser les chances de réussite de la méthode, il faut s'assurer que l'information apportée par les élèves voisins est de bonne qualité. Autrement dit, on aurait envie d'exclure ceux qui sont scolarisés hors de leur secteur. Pour cette raison, on exclut du champ des voisins les élèves scolarisés dans l'enseignement privé (ou ceux inscrits dans un établissement public à recrutement sélectif<sup>3</sup>).

À ce stade, les élèves retenus comme voisins potentiels sont tous scolarisés dans un collège public<sup>4</sup>, sans que l'on sache pour autant si ce collège est celui dont ils dépendent réellement ou non. Comme le montre le tableau 5, le fichier contenant les vrais secteurs des trois départements étudiés nous indique que plus de 20 % des collégiens publics sont scolarisés hors de leur secteur (20,5 % en moyenne sur les trois départements) : ils sont respectivement 15,7 % et 16,8 % à se trouver dans ce cas en Gironde et dans le Puy-de-Dôme ; à Paris, plus d'un élève sur quatre est scolarisé hors de son secteur.

**Tableau 5**  
**Répartition des élèves voisins scolarisés dans le public**

Département	Nombre d'élèves scolarisés dans leur secteur	Nombre d'élèves scolarisés hors de leur secteur	Part d'élèves scolarisés hors de leur secteur
Gironde	47 892	8 907	15,7%
Puy-de-Dôme	18 723	3 773	16,8%
Paris	39 591	14 737	27,1%
<b>Total</b>	106 206	27 417	20,5%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves inscrits dans un collège public

<sup>3</sup> Le collège public Rognoni à Paris est dans ce cas. Dans l'académie de Versailles par exemple, il ne faudrait pas prendre en compte les élèves du collège Franco Allemand de Buc.

<sup>4</sup> Les quelques élèves scolarisés dans le privé et résidant à proximité immédiate d'un collège public feront également partie du champ des voisins.

## Sélectionner les voisins fournissant une information fiable

Dans la mesure où l'on essaie de reconstituer des secteurs qui sont inconnus, il faut se placer dans une situation où les élèves hors secteur ne peuvent pas être identifiés individuellement en vue d'être retirés du champ des voisins. En revanche, à partir des fichiers contenant les vrais secteurs, il est possible d'identifier des « profils » d'élèves fréquemment scolarisés hors de leur secteur. L'idée consistera alors à exclure ces profils d'élèves, de sorte que l'information apportée par les voisins restants sera de meilleure qualité.

Les trois départements pour lesquels on dispose des vrais secteurs présentent un avantage : ils sont très différents. Paris est entièrement urbain et se caractérise par un poids du privé relativement important ; ses collèges se concentrent sur une petite superficie. A l'inverse, la Gironde et le Puy-de-Dôme comprennent à la fois des agglomérations et des communes rurales. Ces disparités laissent penser que les critères de sélection qui seront retenus pour les élèves voisins seront aussi valables pour d'autres territoires, départements ou académies dans lesquels l'information sur les vrais secteurs fait défaut.

Les élèves voisins sélectionnés répondent aux critères suivants :

- ils sont scolarisés dans leur département de résidence ;
- ils ne sont pas scolarisés trop loin de leur domicile ;
- ils n'ont pas changé d'établissement depuis leur entrée au collège ;
- ils ne sont pas inscrits dans une formation atypique.

Dans ce qui suit, nous expliquons pourquoi ces critères ont été retenus.

### **Les élèves scolarisés en dehors de leur département de résidence ou très loin de leur domicile sont systématiquement hors secteur**

La carte scolaire de chaque département est établie par le conseil départemental pour les élèves de son territoire. Mais des accords peuvent exister entre départements visant à sectoriser des élèves sur un département limitrophe. Ces cas sont, a priori, très rares. Ils ne concernent ni Paris, ni la Gironde, ni le Puy-de-Dôme. D'une part, le fichier contenant les vrais secteurs fait apparaître qu'aucun élève résidant à l'extérieur d'un de ces départements n'y est sectorisé. D'autre part, tous les élèves résidant à Paris, en Gironde ou dans le Puy-de-Dôme dépendent d'un collège de leur département de résidence. Autrement dit, les élèves parisiens scolarisés ailleurs qu'à Paris peuvent tous être considérés comme hors secteur. Il en est de même pour la Gironde et le Puy-de-Dôme (et a priori pour tous les autres départements). Les élèves dans cette situation sont donc exclus du champ des voisins potentiels car la contribution qu'ils apporteraient au repérage du vrai collège de secteur serait erronée. Seuls les élèves scolarisés dans un collège public de leur département sont ainsi conservés.

Environ 64,1 % des collégiens publics des 3 départements étudiés sont scolarisés dans l'établissement le plus proche de chez eux. Pour l'immense majorité d'entre eux (95,4%), il s'agit là du collège de secteur. La part d'élèves hors secteur est logiquement plus élevée pour les élèves inscrits dans le deuxième ou troisième collège le plus proche de leur domicile ; elle s'élève respectivement à 24,5 % et 40,7 %. De fait, plus un élève est scolarisé dans un établissement éloigné de son domicile (l'éloignement étant exprimé ici de manière ordinale<sup>5</sup>), plus la probabilité qu'il soit hors secteur est élevée. Au-delà du rang 11, le collège de scolarisation n'est jamais le collège de

---

<sup>5</sup> L'établissement le plus proche du domicile de l'élève est le collège de rang 1 ; le deuxième établissement le plus proche est le collège de rang n°2 ; etc.

secteur : 100,0% des élèves scolarisés dans un collège de rang 12 ou plus sont hors secteur, contre 16,4% pour les autres élèves (cf. tableau 6a). Compte tenu de ces résultats, on propose donc de ne conserver, parmi les voisins, que les élèves scolarisés dans un collège public dont le rang est au maximum égal à 11.

**Tableau 6a**  
**Répartition des élèves voisins**  
**Selon le rang du collège de scolarisation**

Département	Rang du collège de scolarisation	Nombre d'élèves scolarisés dans leur secteur	Nombre d'élèves scolarisés hors de leur secteur	Part d'élèves scolarisés hors de leur secteur
Gironde	<=11	47 892	7 530	13,6%
	>11	0	1 377	100,0%
	<b>Total</b>	47 892	8 907	15,7%
Puy-de-Dôme	<=11	18 723	3 172	14,5%
	>11	0	601	100,0%
	<b>Total</b>	18 723	3 773	16,8%
Paris	<=11	39 591	10 182	20,5%
	>11	0	4 555	100,0%
	<b>Total</b>	39 591	14 737	27,1%
Total	<=11	106 206	20 884	16,4%
	>11	0	6 533	100,0%
	<b>Total</b>	106 206	27 417	20,5%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves scolarisés dans un collège public

C'est à Paris et en Gironde que le « rang limite » est le plus élevé (11 contre 8 dans le Puy-de-Dôme).

### Les élèves qui ont changé de collège sont plus souvent hors secteur

Au cours de leur scolarité au collège, environ 20 % des élèves changent d'établissement<sup>6</sup>. Le déménagement ne constitue pas la seule cause de mobilité pour les élèves. D'autres motifs entrent en jeu. La mobilité touche en effet davantage les garçons, les élèves fragiles scolairement, ceux issus de milieu défavorisé. Autrement dit, un certain nombre d'élèves change de collège sans avoir toutefois changé de lieu de résidence. Dans ces conditions, il n'est pas étonnant de voir que les élèves mobiles sont plus fréquemment scolarisés hors secteur. Dans les trois départements étudiés, plus de 45 % des élèves mobiles sont dans ce cas, soit près de trois fois plus que les élèves n'ayant pas changé de collège (17,1 %, cf. tableau 6b). Comme précédemment, pour optimiser la qualité de l'information apportée par les élèves voisins, nous proposons donc d'exclure du champ les élèves qui ont changé d'établissement depuis leur entrée au collège.

<sup>6</sup> Pirus, 2015

**Tableau 6b**  
**Répartition des élèves voisins**  
**Selon que l'élève a changé de collège ou non**

Département	A changé de collège	Nombre d'élèves scolarisés dans leur secteur	Nombre d'élèves scolarisés hors de leur secteur	Part d'élèves scolarisés hors de leur secteur
Gironde	Non	43 673	6 631	13,2%
	Oui	4 219	2 276	35,0%
	Total	47 892	8 907	15,7%
Puy-de-Dôme	Non	17 411	2 859	14,1%
	Oui	1 312	914	41,1%
	Total	18 723	3 773	16,8%
Paris	Non	36 538	10 670	22,6%
	Oui	3 053	4 067	57,1%
	Total	39 591	14 737	27,1%
Total	Non	97 622	20 160	17,1%
	Oui	8 584	7 257	45,8%
	Total	106 206	27 417	20,5%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves scolarisés dans un collège public

Remarque : dans les fichiers utilisés, la mobilité des élèves peut être appréhendée à partir de l'établissement précédemment fréquenté. Un élève qui n'a pas changé de collège aura pour établissement précédent l'école où il était scolarisé dans le 1<sup>er</sup> degré. À l'inverse, s'il a changé de collège, l'établissement précédent sera son collège d'origine.

#### Le cas des collèges récemment ouverts

Pendant les quelques années qui suivent l'ouverture d'un collège, il est difficile de définir avec certitude les nouveaux contours des secteurs du collège. Durant cet intervalle de temps, la notion même d'élèves hors secteur doit être appréhendée avec précaution.

Pour traiter les cas d'ouverture de collège (et plus généralement les situations où un nouveau niveau de formation apparaît alors qu'il n'existait pas l'année précédente), la solution suivante est mise en œuvre. Dans la zone où le collège (ou le niveau) a ouvert, on ne retient, pour un élève en 6<sup>ème</sup>, que les élèves voisins qui sont aussi en 6<sup>ème</sup>. L'idée est de raisonner sur des élèves qui connaissent tous la même carte scolaire. De même, pour les élèves de 5<sup>ème</sup>, 4<sup>ème</sup> et 3<sup>ème</sup>, on prend en compte tous les élèves voisins à l'exception des sixièmes.

Ainsi, le collège La Rose Blanche (Paris 17<sup>e</sup>) a ouvert le 1<sup>er</sup> septembre 2015. À la rentrée 2015, cet établissement compte seulement 85 élèves de 6<sup>e</sup>. (Les 5<sup>e</sup> feront leur rentrée en 2016, les 4<sup>e</sup> en 2017). Les élèves de 5<sup>e</sup>, de 4<sup>e</sup> et de 3<sup>e</sup> qui résident alors dans le secteur de ce nouveau collège continuent à être scolarisés dans les établissements fréquentés avant cette ouverture : la plupart sont inscrits dans les collèges Boris Vian, Mallarmé, Ronsard, Balzac, Carnot et Chaptal. En tant que voisins, ces élèves fournissent forcément une information erronée car aucun d'entre eux n'est scolarisé au collège La Rose Blanche : ces élèves accroissent fortement le risque d'attribuer à un élève de 6<sup>e</sup> résidant dans le secteur de La Rose Blanche un autre collège alentour. Afin de neutraliser cet aléa, lorsqu'on cherche à déterminer le collège de secteur d'un élève de 6<sup>e</sup> qui réside aux alentours du collège La Rose Blanche, on ne tient compte que de ses voisins scolarisés en 6<sup>e</sup>.

## Le recours à la scolarisation hors secteur est plus fréquent pour les élèves qui suivent une formation particulière

En raison de leur situation particulière, certains élèves ont besoin d'une prise en charge éducative spécifique. Il en est ainsi des élèves en situation de handicap (scolarisés dans des Unités Locales pour l'Inclusion Scolaire), des élèves allophones nouvellement arrivés (scolarisés dans des Unités Pédagogiques pour Élèves Allophones Arrivants) ou encore des élèves scolarisés dans une Segpa (Section d'enseignement général et professionnel adapté). Pour ces derniers, la question de la sectorisation ne se pose pas dans les mêmes termes que pour les autres ; une part importante d'entre eux n'ayant pas vocation à être scolarisés dans le collège défini par la carte scolaire (à partir du lieu de résidence). Ces élèves sont retirés du champ des voisins.

D'autres formations atypiques peuvent être mises en avant, en particulier les classes à horaires aménagés (musique, danse et théâtre). Contrairement aux précédentes, ces formations correspondent davantage à un choix des familles. Mais elles se caractérisent, elles aussi, par une plus grande fréquence de la scolarisation hors secteur. Nous retirons également ces élèves du champ des voisins.

Enfin, tout en suivant une formation « classique », certains élèves choisissent des matières optionnelles qui ne sont pas proposées dans tous les établissements. Le nombre et l'intitulé des options suivies par les élèves informent sur le degré d'atypisme de leur parcours. S'agissant des langues vivantes étrangères (LVE) étudiées par exemple, il est courant, en 6<sup>e</sup> et en 5<sup>e</sup>, de n'étudier que l'anglais ou de l'allemand<sup>7</sup>. À partir de la 4<sup>e</sup>, l'étude de deux langues vivantes (anglais-allemand ou anglais-espagnol) devient la norme. Le latin, quant à lui, fait son apparition dès la 5<sup>e</sup> ; il n'est certes pas choisi par tous les élèves mais il est enseigné dans la quasi-totalité des collèges. Finalement, tous les élèves inscrits dans un parcours suivi par moins de 2% des élèves du département sont retirés du champ des élèves potentiellement voisins. Tous les autres élèves sont conservés, qu'ils soient latinistes ou non, qu'ils soient scolarisés en 3<sup>ème</sup> prépa-pro ou non.

La comparaison des élèves selon qu'ils sont inscrits dans une formation courante ou une formation particulière fait apparaître des écarts assez nets dans le recours à une scolarisation hors secteur : 34,5 % de ceux qui suivent une formation particulière sont scolarisés hors secteur contre seulement 17,0 % des autres (cf. tableau 8). Seuls ces derniers sont conservés dans le champ des voisins potentiels.

---

<sup>7</sup> À l'exception de Paris où la pratique de deux LVE (anglais-allemand ou anglais-espagnol) est très répandue dès la 6<sup>e</sup> et du Puy-de-Dôme où la réforme du collège de 2016 a été anticipée d'une année.

**Tableau 6c**  
**Répartition des élèves voisins**  
**Selon que la formation suivie est particulière ou non**

Département	Formation particulière	Nombre d'élèves scolarisés dans leur secteur	Nombre d'élèves scolarisés hors de leur secteur	Part d'élèves scolarisés hors de leur secteur
Gironde	Non	40 448	5 863	12,7%
	Oui	7 444	3 044	29,0%
	Total	47 892	8 907	15,7%
Puy-de-Dôme	Non	15 481	2 310	13,0%
	Oui	3 242	1 463	31,1%
	Total	18 723	3 773	16,8%
Paris	Non	32 748	10 023	23,4%
	Oui	6 843	4 714	40,8%
	Total	39 591	14 737	27,1%
Total	Non	88 677	18 196	17,0%
	Oui	17 529	9 221	34,5%
	Total	106 206	27 417	20,5%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves scolarisés dans un collège public

Le tableau 7 donne la part d'élèves hors secteur après application des trois critères de sélection. Cette part est de 12,0 % (bien inférieure aux 20,5 % observés avant sélection).

**Tableau 7**  
**Répartition des élèves voisins**  
**Après application des trois critères de sélection ci-dessus**

Département	Nombre d'élèves scolarisés dans leur secteur	Nombre d'élèves scolarisés hors de leur secteur	Part d'élèves scolarisés hors de leur secteur
<b>Gironde</b>	36 708	3 941	9,7%
<b>Puy-de-Dôme</b>	14 375	1 511	9,5%
<b>Paris</b>	30 211	5 624	15,7%
<b>Total</b>	81 294	11 076	12,0%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves scolarisés dans un collège public

### Mise en œuvre de l'algorithme d'attribution d'un collège fictif

#### Hypothèses formulées

Le champ des élèves voisins étant déterminé, on peut à présent essayer d'attribuer un collège de secteur à chacun des élèves habitant dans l'un des trois départements étudiés (hors élèves habitant à proximité immédiate d'un collège, déjà traités).

Nous formulons les hypothèses suivantes :

- Pour chaque élève (qu'on appellera alors élève de référence), on recherche les 10 voisins les plus proches. La distance calculée est une distance à vol d'oiseau. Une variante, plus complexe à mettre en œuvre, consisterait à recourir à un distancier.

- Si le onzième voisin est situé à la même distance de l'élève de référence que le dixième voisin, il est alors inclus dans le calcul (même raisonnement pour le douzième voisin, le treizième, etc.)
- Tout élève résidant dans la même commune que l'élève de référence est considéré comme plus proche que tout élève résidant dans une autre commune (et ce quelles que soient les distances en question). Ce critère permet d'exploiter le fait que la carte scolaire respecte généralement les limites communales.<sup>8</sup>
- Si une majorité d'élèves voisins est scolarisée dans un collège, ce collège est considéré comme le collège de secteur de l'élève de référence.
- Cette majorité peut être absolue ou relative : soit elle est strictement supérieure à 50% ; soit elle est inférieure ou égale à 50%, les proportions pour les autres collèges de scolarisation étant toutes inférieures.

Le choix du nombre 10 n'est pas arbitraire ; il a été fixé après que de nombreux tests ont été effectués. Il correspond à une sorte d'optimum, intermédiaire entre un nombre beaucoup plus petit (1 par exemple) et un nombre beaucoup plus grand (100 par exemple). Dans le premier cas (nombre 1), une majorité absolue de 100 % se dégage systématiquement, mais le taux d'erreur est élevé. Dans le second cas (nombre 100), le taux d'erreur est bas, mais les majorités qui se dégagent sont moins souvent absolues. Notons que les résultats obtenus pour des nombres situés autour de 10 (7 voisins, 8, 9, 11, 12, etc.) sont très proches. D'ailleurs, en raison des élèves qui habitent à la même adresse (ou à la même distance de l'élève de référence), le calcul ne s'effectue sur 10 voisins que dans 69,9 % des cas (cf. tableau 8). Cette dernière proportion est beaucoup plus faible à Paris (54,8 %) qu'en Gironde ou dans le Puy-de-Dôme (respectivement 82,0 % et 83,1 %). À Paris, le nombre maximum de voisins pris en compte est d'ailleurs égal à 67 (contre 43 en Gironde et 36 dans le Puy-de-Dôme).

**Tableau 8**  
**Répartition des élèves de référence selon le nombre de voisins sur lesquels le calcul est effectué**

Département		10 voisins	11 voisins	12 voisins	13 voisins	14 voisins	15 voisins ou plus	Total
Gironde	Nb élèves	55 527	8 116	1 838	762	424	1 068	67 735
	% élèves	82,0	12,0	2,7	1,1	0,6	1,6	100,0
Puy-de-Dôme	Nb élèves	22 611	3 048	618	258	161	510	27 206
	% élèves	83,1	11,2	2,3	0,9	0,6	1,9	100,0
Paris	Nb élèves	42 683	15 637	6 819	3 621	2 368	6 819	77 947
	% élèves	54,8	20,1	8,7	4,6	3,0	8,7	100,0
Total	Nb élèves	120 821	26 801	9 275	4 641	2 953	8 397	172 888
	% élèves	69,9	15,5	5,4	2,7	1,7	4,9	100,0

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Note de lecture : en Gironde, pour 12,0 % des élèves à sectoriser, le calcul s'effectue sur les 11 voisins les plus proches.

### Mise en œuvre en situation optimale théorique

Avant de mettre en œuvre l'algorithme avec les informations relatives au collège de scolarisation des voisins, nous pouvons observer ce que l'on aurait obtenu si :

<sup>8</sup> Le même raisonnement est étendu à l'IRIS : tout élève résidant dans le même IRIS que l'élève de référence est considéré comme plus proche que tout élève résidant dans un autre IRIS. Divers tests effectués ont montré que cette hypothèse permettrait de réduire le taux d'erreur.

- aucun élève n'avait été écarté du champ des voisins  
et

- tous ces élèves avaient été scolarisés dans leur collège de secteur.

Cette situation peut être décrite car nous disposons des vrais secteurs dans les trois départements étudiés. Elle correspond à une estimation des meilleurs résultats que l'on pourrait obtenir avec la méthode des voisins. Dans cette situation, les erreurs constatées peuvent être assimilées aux effets de bords, puisqu'il s'agit là de la seule source d'erreur.

**Tableau 9**  
**Taux d'erreur de la méthode des voisins dans le cas théorique où tous les voisins sont scolarisés dans leur collège de secteur**

	Pas de majorité	de ]0%;50[	[50%;100%]	dont 100%	Proximité immédiate	Total majorité
Erreur		26	2 292	0	0	2 318
Pas d'erreur	Sans objet	5	169 917	162 561	1 650	171 572
Taux d'erreur		83,9%	1,3%	0,0%	0,0%	1,3%
Pas de majorité	648			Sans objet		

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Elèves scolarisés dans un collège public

Note de lecture : pour 169 917 élèves, le bon collège ressort avec une majorité absolue parmi les voisins. Pour 2 292 élèves, une majorité absolue ressort, mais conduit à attribuer un collège erroné.

Dans cette situation optimale théorique, un collège majoritaire ressort pour 99,6 % des collégiens. Le taux d'erreur pour ces derniers est 1,3 % (tableau 9). Il est extrêmement bas et laisse penser qu'en s'approchant de la situation optimale, la méthode des voisins proposera des résultats suffisamment fiables.

### Mise en œuvre à partir des données de scolarisation des élèves : première itération

À partir du collège de scolarisation des élèves voisins (ceux sélectionnés précédemment), un collège majoritaire se dégage dans plus de 98,9% des cas (cf. tableau 10a). Pour plus de quatre élèves sur dix à sectoriser, cette majorité est égale à 100 %. Elle est supérieure à 90 % pour près de deux tiers des élèves. S'agissant des cas dans lesquels une majorité se dégage, le collège majoritaire est le bon dans 94,3 % des cas, avec d'importantes disparités selon l'ampleur de la majorité : lorsque tous les voisins sont scolarisés dans le même établissement, on se trompe dans 0,6 % des cas ; à l'inverse, dans le cas d'une majorité relative (i.e. avec moins de 5 voisins sur 10 dans le même établissement), le taux d'erreur s'élève à 48,1%.

**Tableau 10a**  
**Taux d'erreur de la méthode des voisins après une itération**

	Pas de majorité	de ]0%;50[	[50%;100%]	Dont [90%;100%]	Dont 100%	Proximité immédiate	Total majorité
Erreur		957	8 924	1 417	393	0	9 881
Pas d'erreur	Sans objet	1 033	159 989	110 503	69 789	1 650	162 672
Taux d'erreur		48,1%	5,3%	1,3%	0,6%	0,0%	5,7%
Pas de majorité	1 985			Sans objet			

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Note de lecture : pour 159 989 élèves, le bon collège ressort avec une majorité absolue parmi les voisins. Pour 8 924 élèves, une majorité absolue ressort, mais conduit à attribuer un collège erroné.

Paris se caractérise par un taux d'erreur plus élevé que la Gironde ou le Puy-de-Dôme (7,9 % contre respectivement 3,9% et 4,0%). D'une part, les majorités élevées sont moins fréquentes dans la capitale, d'autre part le taux d'erreur est plus élevé lorsque la majorité est de même ampleur. Les nombreux effets de bords, en lien avec la forte densité de collèges à Paris, expliquent ces disparités.

## Deuxième itération

Intuitivement, on aurait envie d'utiliser les résultats de la première étape pour améliorer encore la qualité des informations apportées par les élèves voisins. En particulier, on aimerait que les élèves dont le collège a été déterminé avec quasi-certitude apportent non plus l'information relative à leur collège de scolarisation, mais plutôt celle relative au collège qui leur a été attribué à l'issue de la première étape. Bien sûr, les deux collèges sont les mêmes dans la plupart des cas. Mais cette opération permet tout de même certaines corrections. Elle est réalisée pour la population suivante : les élèves dont les voisins sont scolarisés dans un même collège à 90 % ou plus (pour ces élèves, le collège attribué est le bon dans 98,7 % des cas).

Notons que l'attribution quasi-certaine d'un collège de secteur concerne aussi des élèves qui avaient été exclus du champ des voisins en raison de leur profil particulier. C'est le cas par exemple des collégiens du secteur privé : ceux d'entre eux auxquels un collège public a été attribué de manière certaine peuvent de nouveau être inclus dans le champ des voisins grâce à cette nouvelle information. Au total, ce sont 48 684 élèves qui sont ainsi ajoutés (cf. tableau 10b).

**Tableau 10b**  
**Amélioration apportée à la table des voisins**

Département	Nombre d'élèves présents dans la table des voisins originelle	Dont correction du collège de scolarisation	Nombre d'élèves ajoutés à la table des voisins originelle	Nombre d'élèves dans la table des voisins finale
<b>Gironde</b>	40 957	938	19 046	60 003
<b>Puy-de-Dôme</b>	16 113	319	7 535	23 648
<b>Paris</b>	35 969	880	22 103	58 072
<b>Total</b>	93 039	2 137	48 684	141 723

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Le tableau 10c montre que les résultats obtenus dans l'attribution d'un collège de secteur à partir de la table des voisins corrigée sont meilleurs. Le gain porte essentiellement sur les élèves pour lesquels aucune majorité ne se dégageait (baisse de 5 %). Au niveau global, il est relativement faible. Du reste, les tests effectués en itérant plusieurs fois le procédé de correction ont montré qu'il devenait très vite marginal.

**Tableau 10c**  
**Taux d'erreur de la méthode des voisins après une deuxième itération**

	Pas de majorité	]0%;50[	[50%;100%]	Dont 100%	Proximité immédiate	Total majorité
Erreur		890	9 034	579	0	9 924
Pas d'erreur	Sans objet	979	160 101	73 385	1 650	162 730
Taux d'erreur		47,6%	5,3%	0,8%	0,0%	5,7%
Pas de majorité	1 884			Sans objet		

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Finalement, à l'issue de ces premières étapes, on peut distinguer trois cas :

- Les élèves pour lesquels un collège ressort avec une majorité absolue (groupe 1) – ce groupe comprend aussi les élèves situés à proximité immédiate d'un collège ;
- Les élèves pour lesquels un collège ressort avec une majorité relative (groupe 2) ;
- Les élèves pour lesquels aucun collège majoritaire ne ressort (groupe 2bis).

La comparaison avec la méthode du collège le plus proche met en évidence que les taux d'erreur sont extrêmement bas pour les élèves du groupe 1. Ils sont élevés pour ceux du groupe 2. La question ne se pose pas pour les élèves du groupe 2bis puisqu'aucun collège ne leur a été attribué.

### Traitement des élèves sans collège majoritaire

Si l'on souhaite imputer un collège de secteur à tous les élèves y compris à ceux du groupe 2bis, plusieurs pistes sont envisageables. On peut décider d'affecter ces élèves dans l'établissement le plus proche de chez eux. On peut aussi décider de les affecter dans le collège attribué à leur plus proche voisin. Les tests effectués ont montré que la deuxième option était légèrement meilleure que la première, même si elle conduit, elle aussi, à se tromper pour près de 4 élèves sur 10 (cf. tableau 11). Il n'empêche qu'en traitant de la sorte les élèves pour lesquels aucun collège majoritaire ne se dégageait, on peut alors calculer un taux d'erreur global pour la « méthode des voisins ». Celui-ci s'élève à 6,1 % ; il est bien inférieur à celui obtenu par la « méthode du collège le plus proche » (27,9 %).

**Tableau 11**

#### Taux d'erreur de la méthode des voisins après traitement des élèves sans collège majoritaire

	<b>Groupe 1 Majorité absolue</b>	<b>Groupe 2 Majorité relative</b>	<b>Groupe 2bis (Collège du voisin le plus proche)</b>	<b>Total</b>
Erreur	9 034	890	711	10 635
Pas d'erreur	161 751	979	1 173	163 903
Taux d'erreur	5,3%	47,6%	37,7%	6,1%

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

Les illustrations qui suivent représentent la carte scolaire dans le 12<sup>ème</sup> arrondissement de Paris.

Figure 1a - Secteurs réels – Paris 12<sup>ème</sup>

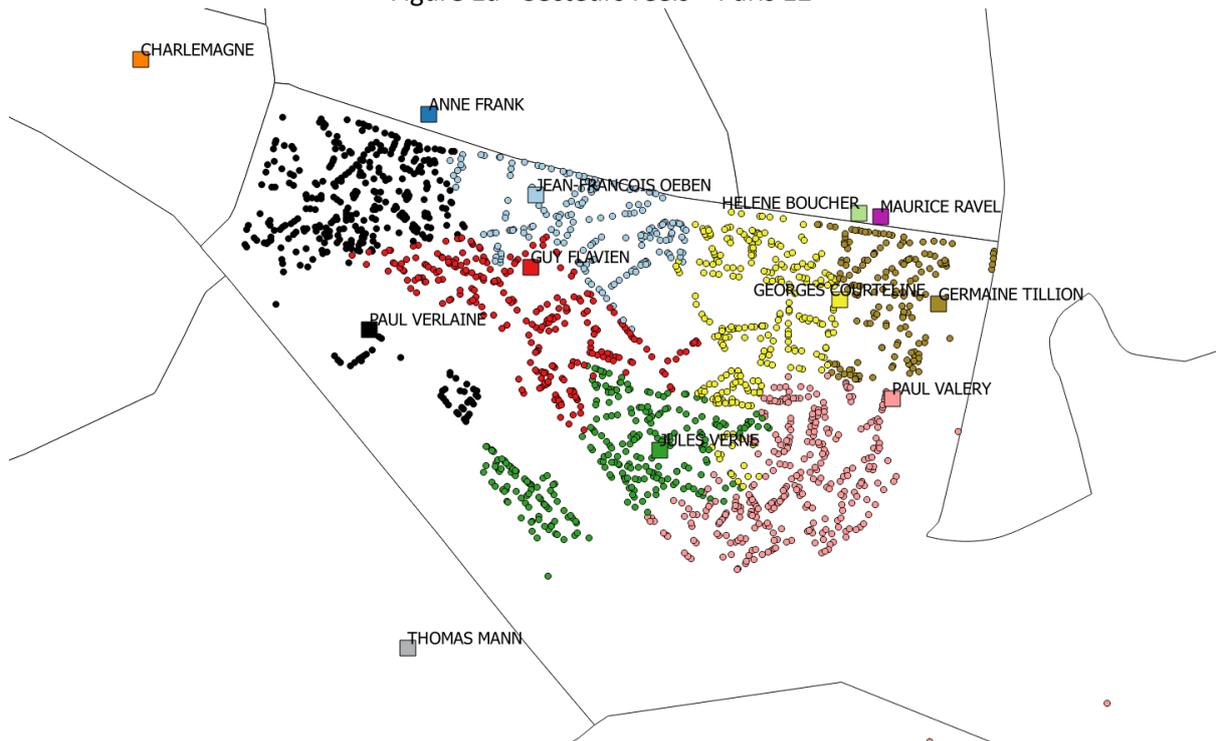


Figure 1b - Secteurs reconstitués (méthode des voisins) – Paris 12<sup>ème</sup>

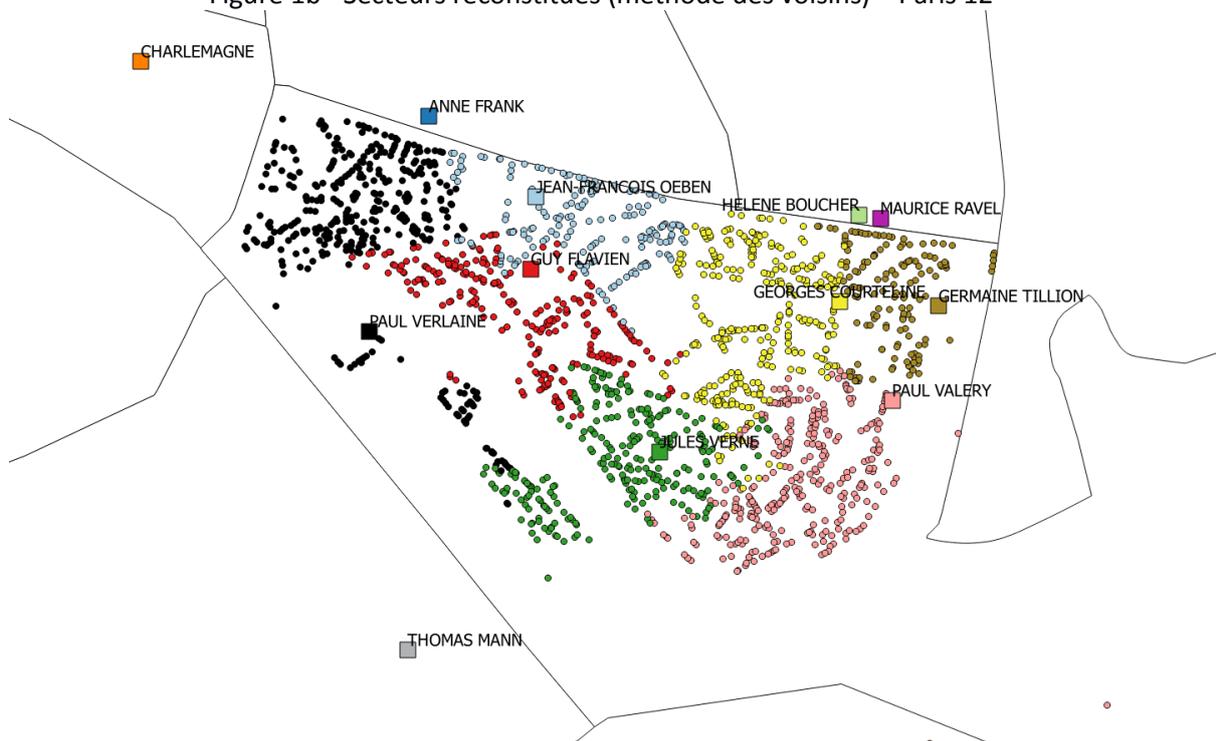
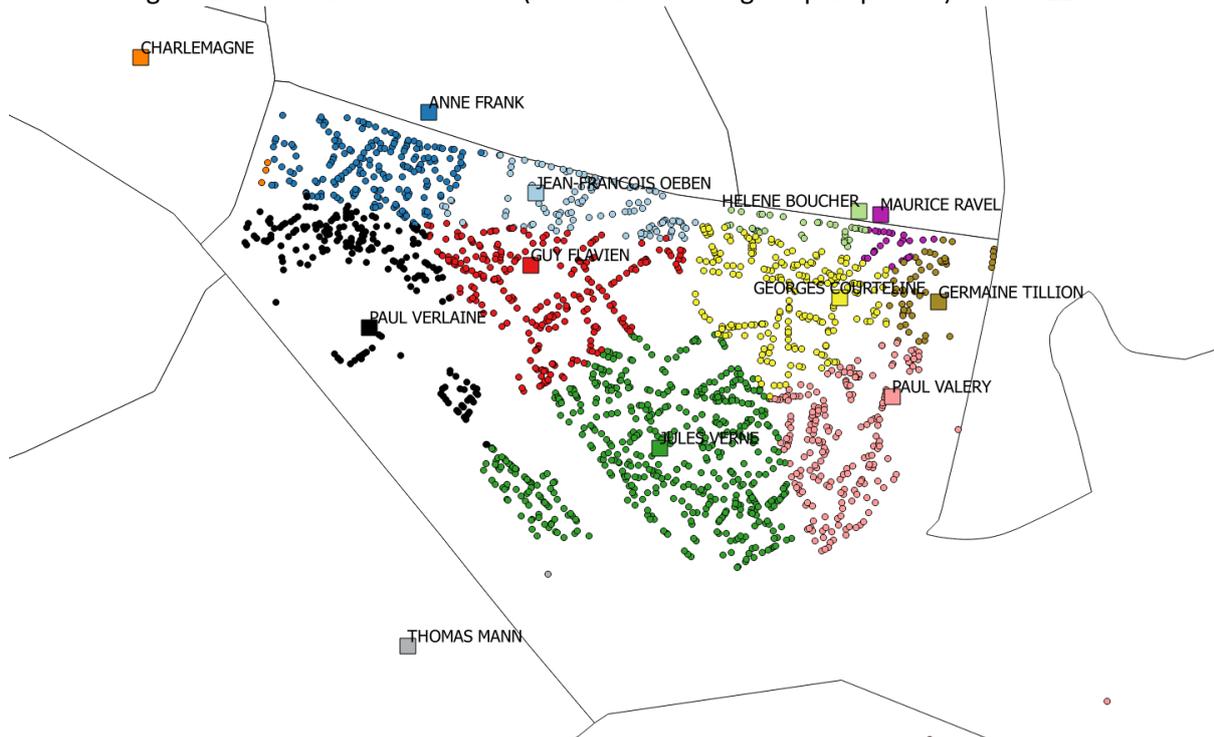


Figure 1c - Secteurs reconstitués (méthode du collège le plus proche) – Paris 12<sup>ème</sup>



#### IV. Fiabilité de la méthode

##### Existence éventuelle d'un biais

Pour juger de la qualité de la méthode proposée ici, il faudrait aller plus loin et s'interroger sur la gravité des erreurs commises. Les cas pour lesquels le collège attribué n'est pas le bon sont-ils sources de biais ?

Pour répondre à cette question, on peut émettre l'hypothèse que le biais est nul pour les élèves du groupe 1. Pour eux, l'absence de biais est justifiée par le faible taux d'erreur dans l'attribution du collège de sectorisation. Notons que, dans ce qui suit, on pourra étayer cette hypothèse avec d'autres arguments en fonction de la statistique étudiée.

Si les résultats obtenus pour le groupe 1 peuvent donc être considérés comme fiables, il n'en est pas de même pour les groupes 2 et 2bis. Le traitement des élèves pour lesquels aucun collège ne ressort avec une majorité absolue nécessite une vigilance particulière.

Une manière de traduire l'incertitude qui entoure ces élèves consiste à adopter une attitude prudente et à ne pas leur attribuer à tout prix un collège de secteur. Dans un objectif d'estimation statistique, on peut par exemple envisager les cas extrêmes liés à cette incertitude et s'appuyer sur ces cas pour calculer un intervalle de confiance.

##### Estimation de la part d'élèves hors secteur dans un territoire donné : approche en situations extrêmes

Supposons que, dans un territoire donné, nous souhaitons estimer la proportion notée  $p$  d'élèves scolarisés hors secteur. On note :

$N_1$  le nombre d'élèves du groupe 1

$N_2$  le nombre d'élèves des groupes 2 et 2bis

$p_1$  la proportion d'élèves hors secteur dans le groupe 1. Cette proportion, connue, est supposée non biaisée. En plus du faible taux d'erreur, on peut ajouter ici un argument :  $p_1$  correspond, par construction, à la proportion exacte d'élèves qui ne fréquentent pas le collège majoritaire de leurs voisins. Dans la mesure où, en estimant la part d'élèves hors secteur, on cherche peu ou prou à repérer les élèves qui évitent le collège de leur voisin, l'hypothèse d'une absence de biais semble tout à fait raisonnable.

$p_2$  la proportion d'élèves hors secteur dans les groupes 2 et 2bis. Cette proportion est supposée biaisée puisque le collège attribué à ces élèves est potentiellement erroné.

On a donc :

$$p = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$$

Les situations extrêmes sont obtenues dans les deux cas suivants :

- tous les élèves des groupes 2 et 2bis sont scolarisés dans leur collège de secteur ( $p_2 = 0$ )
- tous les élèves des groupes 2 et 2bis sont scolarisés en dehors de leur secteur ( $p_2 = 1$ )

On a ainsi :

$$p_{min} = \frac{N_1 p_1}{N_1 + N_2} \text{ et } p_{max} = \frac{N_1 p_1 + N_2}{N_1 + N_2}$$

L'intervalle de confiance s'écrit alors :  $IC = [p_{min}, p_{max}]$ .

Sa longueur est donnée par  $p_{max} - p_{min}$ . Elle vaut  $\frac{N_2}{N_1 + N_2}$ .

Elle correspond au poids relatif des élèves des groupes 2 et 2bis par rapport à l'ensemble des élèves. Ce résultat est logique : l'estimation est d'autant plus précise que les élèves pour lesquels le collège est attribué avec quasi-certitude sont nombreux (groupe 1).

### Approche en situations extrêmes : résultats

**Tableau 12a**  
Estimation de la part d'élèves hors secteur par département

Département	% élèves scolarisés		% estimé d'élèves scolarisés dans un collège public hors secteur	
	Dans le privé	Dans un collège public hors secteur	Minimum	Maximum
<b>Gironde</b>	17,7	11,3	10,3	11,1
<b>Puy-de-Dôme</b>	19,4	12,2	11,0	12,7
<b>Paris</b>	31,0	17,7	16,6	18,4

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Hors Segpa et UPE2A

Le tableau 12a donne les résultats obtenus, par la méthode des voisins, pour l'estimation de la part d'élèves hors secteur dans les trois départements étudiés. On constate que l'intervalle de confiance contient la vraie valeur pour Paris et le Puy-de-Dôme. Pour la Gironde, la valeur issue du fichier contenant les vrais secteurs est à la marge de l'intervalle de confiance (11,3 % contre 11,1 % pour la borne supérieure) ; cela tient sans doute aux différences de géocodage dans les deux fichiers que l'on compare. Rappelons que l'estimation qui est faite ici est, par construction, une estimation non

biaisée d'une autre donnée : la part d'élèves qui ne fréquentent pas le même collège public que leurs voisins.

On peut légitimement considérer que la méthode proposée pour reconstituer les secteurs de collège est fiable et permet d'envisager la diffusion de résultats à l'échelle des départements dont la carte scolaire n'est pas disponible. Par ailleurs, les fichiers géolocalisés de la DEPP ayant été produits pour les années scolaires 2009-2010, 2011-2012 et 2013-2014, il est possible d'étudier l'évolution de la part d'élèves hors secteur entre ces différentes dates. Ces analyses pourront être produites dans le cadre d'une prochaine étude.

L'exploitation des résultats peut bien sûr s'effectuer à d'autres niveaux que le département, même si cet échelon semble incontournable dans la mesure où il correspond à l'instance qui définit la carte scolaire (le conseil départemental). En particulier, il est possible de se placer à l'échelle de territoires comme l'académie, la région académique, l'EPCI (établissement public à caractère intercommunal) ou l'unité urbaine. Le calcul ne pose pas de problème particulier : le fait de connaître le lieu de résidence pour chaque élève permet d'agréger facilement les données à l'échelon souhaité. De plus, comme l'information sur la sectorisation a été reconstituée au niveau individuel, il est possible de décrire le profil des élèves hors secteur à l'aide des caractéristiques disponibles dans les fichiers géolocalisés comme le sexe ou encore l'appartenance à un quartier relevant de la politique de la ville (tableau 12b).

**Tableau 12b**  
**Recours au privé ou au public hors secteur en fonction du sexe et de l'appartenance à un QPV**

	% élèves scolarisés		% estimé d'élèves scolarisés dans un collège public hors secteur	
	Dans le privé	Dans un collège public hors secteur	Minimum	Maximum
<b>Sexe</b>				
<b>Garçons</b>	24,2	14,9	13,7	15,2
<b>Filles</b>	23,7	13,8	12,7	14,2
<b>Appartenance à un QPV</b>				
<b>Oui</b>	14,0	21,8	20,3	21,8
<b>Non</b>	24,6	13,9	12,8	14,2

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris – Hors Segpa et UPE2A

Note de lecture : 21,8 % des collégiens résidant en QPV sont scolarisés dans un collège public hors secteur. La méthode des voisins estime cette part entre 20,3 % et 21,8 %.

Par ailleurs, le fichier géolocalisé pouvant être apparié avec les données des panels d'élèves (sur la base de l'adresse), d'autres exploitations sont envisageables au niveau individuel. En effet, dans les panels, les parents sont interrogés sur le recours éventuel à une scolarisation hors secteur pour leur enfant. Cette information est déclarative. En comparant la donnée fournie par les parents à celle obtenue par la méthode des voisins, il est possible de juger la qualité de la réponse des parents. Il est aussi possible de produire des analyses plus riches encore sur le recours au hors secteur en s'appuyant sur les nombreuses données socio-économiques recueillies dans le cadre des panels. L'exploitation pourrait porter sur le panel d'élèves entrés en sixième en 2007 et sur le fichier géolocalisé 2009 (à cette date, les élèves du panel sont encore tous scolarisés en collège).

## Estimation au niveau établissement

Dans le cas d'une diffusion au niveau de l'établissement, la question est moins triviale. En effet, si l'on veut estimer la part d'élèves hors secteur pour un collège donné (et mesurer ainsi la proportion d'élèves qui évitent ce collège), le champ des élèves à retenir ne peut pas être défini de manière certaine, contrairement aux cas évoqués plus haut. Il correspond en fait aux élèves sectorisés sur ce collège et est donc sujet, pour une partie des élèves, à une incertitude. L'exemple qui suit porte sur le recours au privé par secteur de collège : il propose une manière d'appréhender la méthode évoquée ci-dessus.

Considérons un secteur de collège public et notons :

$N_1$  le nombre d'élèves sectorisés dans ce collège de manière certaine (groupe 1)

$p_1$  la proportion d'élèves du groupe 1 qui sont scolarisés dans le secteur privé

$N_2$  le nombre d'élèves pour lesquels le collège  $j$  n'atteint pas la majorité absolue, mais fait partie des collèges de scolarisation probables en ce sens qu'au moins un des voisins  $y$  est scolarisé (groupe 2<sup>9</sup>)

$p_2$  la proportion d'élèves du groupe 2 qui sont scolarisés dans le privé

$N_j$  l'effectif d'élèves sectorisés dans le collège  $j$ . Ce nombre est inconnu ; il est compris entre  $N_{jmin} = N_1$  et  $N_{jmax} = N_1 + N_2$

Pour l'estimation de la part d'élèves ayant recours au privé, le repérage des situations extrêmes s'effectue de manière légèrement différente. La situation qui minimise cette proportion est celle correspondant :

- au numérateur le moins élevé : les seuls élèves scolarisés dans le privé sont ceux du groupe 1 et
- au dénominateur le plus élevé : les élèves sectorisés dans le collège  $j$  sont tous ceux du groupe 1 et tous ceux du groupe 2 qui sont scolarisés dans un collège public

$$\text{On a ainsi : } p_{min} = \frac{N_1 p_1}{N_1 + N_2 (1 - p_2)}$$

De même, la situation qui maximise la part d'élèves ayant recours au privé est celle correspondant :

- au numérateur le plus élevé : les élèves du privé sont ceux du groupe 1 et tous les élèves du groupe 2 qui sont scolarisés dans le privé
- et
- au dénominateur le moins élevé : les élèves sectorisés dans le collège  $j$  sont tous ceux du groupe 1 et tous ceux du groupe 2 qui sont scolarisés dans le privé

$$\text{On obtient ainsi : } p_{max} = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2 p_2}$$

Une autre manière de décrire ces deux situations consiste à dire que le groupe 2 apporte au collège  $j$  :

- aucun élève scolarisé dans le privé, la part d'élèves ayant recours au privé  $p$  est ainsi minimale.
- uniquement des élèves qui sont scolarisés dans le privé,  $p$  est ainsi maximale

$$\text{On a bien : } \frac{N_1 p_1}{N_1 + N_2 (1 - p_2)} \leq p \leq \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2 p_2}$$

---

<sup>9</sup> Ce groupe est l'équivalent des groupes 2 et 2bis définis précédemment.

La longueur de l'intervalle de confiance dépend, comme précédemment, des élèves pour lesquels le collège n'est pas déterminé avec certitude. Cette longueur est nulle si  $N_2 = 0$ . L'estimation est d'autant plus précise que l'effectif  $N_2$  est faible. Rappelons que nous raisonnons ici à collège donné et que le calcul de la précision est donc spécifique à chaque établissement. Dans ces conditions, il est vraisemblable que la précision soit jugée suffisante pour certains collèges et insuffisante pour d'autres.

Le tableau 13 donne les résultats obtenus pour l'estimation, par secteur de collège, de la part d'élèves ayant recours au secteur privé. Dans beaucoup de cas, les intervalles de confiance estimés contiennent la vraie valeur. Bien sûr, à l'échelle de l'établissement, l'inclusion dans l'intervalle de confiance peut être définie de manière un peu indulgente : si la vraie valeur est distante de l'intervalle de confiance d'au plus de 2 points, on peut considérer que le résultat est satisfaisant. On pourrait même étendre cette approche aux cas où la distance est comprise entre 2 et 5 points. Seuls cinq collèges font l'objet d'une distance comprise entre 5 et 10 : un collège situé en Gironde, département où la géolocalisation des élèves a été réalisée par le SSA (8,7 points de distance) ; un collège du Puy-de-Dôme à très faibles effectifs (7,3 points de distance et seulement 38 élèves à la rentrée 2015) ; les trois collèges restants sont situés à Paris (distances comprises entre 5,1 et 6,5 points).

**Tableau 13**

**Part d'élèves ayant recours au privé : inclusion de la vraie valeur dans les intervalles de confiance**

L'écart entre la vraie valeur et l'intervalle de confiance est	Nombre de collèges
<b>Nul</b>	84
<b>Compris entre 0 et 2 points</b>	172
<b>Compris entre 2 et 5 points</b>	14
<b>Compris entre 5 et 7 points</b>	3
<b>Compris entre 7 et 10 points</b>	2
<b>Supérieur à 10 points</b>	0

Source : Fichiers géolocalisés enrichis 2015

Champ : Départements de la Gironde, du Puy-de-Dôme et de Paris

En supposant que la vraie valeur est bien incluse à chaque fois dans l'intervalle de confiance, c'est la longueur de celui-ci qui doit alors être prise en compte pour la diffusion des résultats. Les secteurs pour lesquels la longueur de l'intervalle de confiance est faible (de l'ordre de quelques points) pourraient effectivement donner lieu à diffusion.

C'est là l'un des principaux avantages de la méthode des voisins : non seulement les taux d'attribution du bon collège sont élevés, mais surtout nous pouvons quantifier la fiabilité de l'imputation, et ce même à un niveau très fin : nous savons quels sont les secteurs pour lesquels l'information reconstituée est fiable ou non, ce que ne permet pas de savoir la méthode du collège le plus proche par exemple. Finalement, la situation peut être considérée comme analogue à celle d'une enquête dont les taux de réponse seraient suffisants pour certaines entités mais insuffisants pour d'autres.

Si la diffusion de données à l'échelle d'un collège est possible (il faut bien sûr l'envisager avec prudence), celle à l'échelle d'un type de collège l'est a fortiori. Il est possible, par exemple, d'agréger des données de niveau établissement pour produire des résultats relatifs à l'ensemble des collèges de l'éducation prioritaire : quelle part d'élèves sectorisés en REP ou en REP+ évitent le collège de secteur ? Cette proportion a-t-elle évolué suite à la refonte de la carte de l'éducation prioritaire à la rentrée 2015 ? Les perspectives d'études qui sont esquissées ici renvoient à la question du « label éducation prioritaire » dont on interroge parfois l'effet qu'il peut produire sur le comportement des parents. Cette question pourra être traitée dans une prochaine étude.

## Limites de l'approche en situations extrêmes

Jusqu'ici, pour déterminer un intervalle de confiance, nous avons eu recours à l'approche en situations extrêmes. Cette approche était relativement simple à mettre en œuvre car la situation d'intérêt – élève scolarisé dans son secteur ou non – présentait seulement deux états. Dans ce contexte, les cas extrêmes pouvaient être décrits sans difficulté particulière : tantôt tous les élèves du groupe 2 étaient considérés comme scolarisés dans leur collège de secteur, tantôt ils étaient tous considérés comme étant hors secteur. Par ailleurs, nous avons vu que la diffusion de résultats à l'échelle d'un secteur de collège nécessitait une vigilance particulière. Dans ce cas-là, le champ même des élèves à prendre en compte pour envisager les cas extrêmes était incertain et le calcul de l'intervalle de confiance était alors légèrement complexifié. On peut penser que si la statistique étudiée était plus élaborée que « la part d'élèves hors secteur » ou la « part d'élèves ayant recours au privé », les cas extrêmes seraient encore plus difficiles à repérer.

Supposons qu'on s'intéresse à l'estimation des effets de la scolarisation hors secteur sur la ségrégation entre les établissements. Comme dans les exemples présentés ci-dessus, il convient certes, dans un premier temps, de distinguer les élèves dont le collège est supposé connu de ceux dont le collège attribué est incertain. Mais une fois cette sélection opérée, le traitement à effectuer auprès des élèves du groupe 2 pour appréhender les situations extrêmes est loin d'être évident. Pour minimiser la ségrégation entre collèges, on aurait envie d'affecter les élèves du groupe 2 en fonction de leur milieu social de sorte que parmi ces élèves, ceux issus d'un milieu favorisé soient affectés dans les collèges les moins favorisés et inversement. De la même manière, pour maximiser la ségrégation, on aurait envie d'affecter les élèves favorisés du groupe 2 dans les collèges les plus favorisés et les élèves défavorisés dans les collèges les moins favorisés. Cette description impressionniste de ce que seraient les situations extrêmes semble tout à fait délicate à mettre en œuvre concrètement.

## Approche en situations probables

Une manière de contourner la difficulté évoquée ci-dessus consisterait à envisager non plus les cas extrêmes mais plutôt les cas probables. Jusqu'ici, les élèves pour lesquels aucun collège majoritaire ne ressortait faisaient l'objet d'un traitement des plus prudents. Cette prudence peut sembler un peu radicale, car elle n'exploite aucune des informations relatives à ces élèves. Or, leurs voisins nous renseignent tout de même sur les collèges probables de sectorisation : on peut en effet émettre l'hypothèse que le vrai collège de secteur est l'un des collèges dans lesquels sont scolarisés les voisins.

Pour formaliser ces hypothèses, on peut raisonner comme suit :

Dans un département donné, on note  $N$  le nombre total de collégiens et  $J$  le nombre total de collèges.

Soient  $i \in \llbracket 1; N \rrbracket$  et  $j \in \llbracket 1; J \rrbracket$ . L'événement élémentaire « l'élève  $i$  est scolarisé dans le collège  $j$  » est noté  $\omega_{i,j} = \{c_{i,j}\}$ .

L'ensemble  $A$  des répartitions possibles des élèves dans les collèges s'écrit :

$$A = \left\{ \bigcup_{i=1}^N \omega_{i,j}, \quad j \in \llbracket 1; J \rrbracket \right\}.$$

On a :  $\text{Card}(A) = N^J$

On note :  $p_{ij} = P(\omega_{i,j})$ . L'hypothèse qui est faite est la suivante : la probabilité qu'un élève est scolarisé dans un collège  $i$  dépend du collège de sectorisation des élèves voisins. D'un point de vue empirique, elle est donnée par :

$$p_{ij} = \frac{1}{10} \sum_{k=1}^{10} \mathbb{1}_{i_k \in (\text{secteur\_college}_j)} \quad \text{où } i_1, i_2, \dots, i_{10} \text{ sont les 10 voisins de l'élève } i$$

Bien sûr, pour les élèves voisins  $i_1, i_2, \dots, i_{10}$ , on ne connaît que le collège de scolarisation. Mais puisque ces derniers ont été sélectionnés de sorte à optimiser la part d'élèves dans leur secteur, on peut considérer que, pour eux, collège de secteur et collège de scolarisation sont confondus. On obtient ainsi :

$$p_{ij} = \frac{1}{10} \sum_{k=1}^{10} \mathbb{1}_{i_k \in (\text{college}_j)}$$

Exemple : supposons qu'un élève  $i$  ait 3 voisins scolarisés dans un collège  $j_1$ , 3 dans un collège  $j_2$  et 4 dans un collège  $j_3$ . On a alors :

$$p_{ij_1} = \frac{3}{10}, p_{ij_2} = \frac{3}{10}, p_{ij_3} = \frac{4}{10} \quad \text{et} \quad p_{ij} = 0 \quad \forall j \neq \{j_1, j_2, j_3\}$$

Remarque : la formule proposée pour  $p_{ij}$  tient compte de la dépendance qu'il existe entre les 11 élèves  $i, i_1, i_2, \dots, i_{10}$ . Elle suppose par ailleurs l'indépendance avec tous les autres élèves.

En s'appuyant sur cette définition, il est possible d'effectuer un nombre important de tirages qui respecteraient ces lois de probabilités pour les élèves du groupe 2. En ajoutant les élèves du groupe 1, on obtiendrait ainsi  $L$  reconstitutions de secteurs, plus ou moins différentes. Le calcul d'une statistique  $X_l$  pour chaque reconstitution permettrait ensuite de calculer un intervalle de confiance. Pour cela, il convient de se placer dans le cadre de la loi des grands nombres, ce qui est vérifié dès lors que le nombre de tirages est suffisamment grand (compris entre 100 et 1000 par exemple). L'intervalle de confiance à 95 % s'écrit alors :

$$I_c = \left[ \frac{\bar{X} - 1,96\hat{\sigma}}{\sqrt{L}} ; \frac{\bar{X} + 1,96\hat{\sigma}}{\sqrt{L}} \right]$$

où  $\bar{X}$  est la moyenne de la statistique  $X$  calculée pour chaque échantillon  $l$  et  $\hat{\sigma}$  est la variance de  $X$  entre les différents échantillons.

## Conclusion et perspectives

Dans de prochains travaux, il sera possible de mettre en application le cadre théorique présenté ci-dessus. L'objectif pourrait être de caractériser les établissements dont la composition sociale est le plus affectée en raison de la scolarisation dans le privé ou dans un autre collège public. L'approche en situations probables pourrait aussi être mise en œuvre pour mesurer, à l'échelle d'un territoire, la ségrégation et ses différentes composantes : contribution résidentielle, contribution du secteur privé et du recours au public hors secteur.

De manière générale, la méthode des voisins, dont nous avons montré qu'elle permettait de dessiner des secteurs de collège extrêmement proches de la réalité, ouvre la voie à de nombreuses exploitations. Certaines seront plus complexes à mettre en œuvre que d'autres. S'agissant de l'approche en situations probables par exemple, un préalable consistera à optimiser les temps de calcul de l'algorithme, afin de pouvoir reconstituer des secteurs un nombre élevé de fois.

## Bibliographie

Agulhon C., Palma J., 2013, « Sectorisation et assouplissement de la carte scolaire. Des actions segmentées et contradictoires », *Éducation & formations*, n°83, p. 75-86.

Ben Ayed C., Broccolichi S., Monfroy B., 2013, « Quels impacts de l'assouplissement de la carte scolaire sur la ségrégation sociale au collège ? Tendances nationales et déclinaisons locales contrastées », *Éducation & formations*, n°83, p. 39-57.

Cadoret A., « La mixité sociale des collèges de Clermont-Ferrand », 2017, *Géographie de l'École 2017*.

Da Costa S., Van Zanten A., 2013, « La gestion de la carte scolaire dans la périphérie parisienne : enjeux, dynamiques et limites de la gouvernance éducative locale », *Éducation & formations*, n°83, p. 99-107.

Dutercq Y., Mons N., 2013, « Les principaux de collège face à l'assouplissement de la carte scolaire : un repositionnement stratégique », *Éducation & formations*, n°83, p. 87-98.

Fack G., Grenet J., 2009, « Sectorisation des collèges et prix des logements à Paris », *Actes de la recherche en sciences sociales*, 2009-5, n°108, p. 44-62.

Fack G., Grenet J., 2013, « Les effets de l'assouplissement de la carte scolaire dans l'éducation prioritaire », *Éducation & formations*, n°83, MEN-DEPP, p. 25-37.

Fack G., Grenet J., Benhenda A., 2014, *L'impact des procédures de sectorisation et d'affectation sur la mixité sociale et scolaire dans les lycées d'Île-de-France*, rapport IPP, n° 3.

Girard P., Gilotte O., 2005, « La sectorisation, l'affectation et l'évitement scolaire dans les classes de sixième à Paris en 2003 », *Éducation & formations*, n°71, MEN-DEPP, p. 137-149.

Lefresne F., 2013, « Que nous enseignent les comparaisons internationales en matière de carte scolaire ? », *Éducation & formations*, n°83, p. 109-114.

Murat, Thauvel-Richard, 2013, « Évolution des caractéristiques des collèges durant la mise en œuvre de l'assouplissement de la carte scolaire de 2007 », *Éducation & formations*, n°83, p. 11-23.

Oberti M., 2007, *L'école dans la ville : Ségrégation, mixité, carte scolaire*, Paris, Presses de Sciences Po.

Oberti M., Prêteceille E., 2013, « Dérogations et contextes scolaires : comparaison Hauts-de-Seine Seine-Saint-Denis », *Éducation & formations*, n°83, p. 59-72.

Pirus C., 2015, « Un élève sur cinq change d'établissement au cours de sa scolarité au collège », *Note d'information*, n° 15-32, MENESR-DEPP.

Van Zanten A., Obin J.-P., *La carte scolaire : faits, révélations, analyses*, PUF, coll. « Que sais-je ? », Paris, 2008.