MINISTÈRE
DE L'ÉDUCATION
NATIONALE,
DE LA JEUNESSE
ET DES SPORTS
Liberté
Égalité
Fraternité

depp | SSA

Le zonage en aires d'éducation

Gwenaëlle THOMAS (SSA d'Aix-Marseille), Christine CHEVALIER (SSA de Grenoble), Élisabeth VILAIN (SSA de Lille), Alice DEROSIER (SSA de Nantes), Paul-Olivier GASQ (SSA de Reims), Olivier MONSO (DEPP), Pascal ANDRÉ (DEPP)

Les publications du réseau DEPP-SSA
n° 2022-R01
Février 2022

Le zonage en aires d'éducation

Cet ouvrage est édité par le ministère de l'Éducation nationale,
de la Jeunesse et des Sports

Direction de l'évaluation, de la prospective et de la performance

61-65, rue Dutot
75732 Paris Cedex 15

Directrice de la publication :
Fabienne Rosenwald

Auteurs :
Gwenaëlle THOMAS (SSA d'Aix-Marseille), Christine CHEVALIER
(SSA de Grenoble), Élisabeth VILAIN (SSA de Lille), Alice
DEROSIER (SSA de Nantes), Paul-Olivier GASQ (SSA de Reims),
Olivier MONSO (DEPP), Pascal ANDRÉ (DEPP)

SOMMAIRE

↘ Introduction 6

↘ I. Les besoins d'un zonage pour l'étude et le pilotage du système éducatif 7

I.1 Les besoins au niveau national 7

I.2 Les besoins au niveau local 8

↘ II. La démarche et les choix méthodologiques 9

II.1 Le choix de l'outil Anabel 9

II.2 La description de la méthodologie et les choix de paramètres 10

↘ III. La mise en œuvre de l'algorithme et la consolidation des zonages 13

III.1 Les zonages Parcours et Collèges en sortie d'Anabel 13

III.2 La consolidation des zonages à l'aide de l'expertise académique 14

↘ IV. Le zonage final 16

IV.1 Cartographie des aires d'éducation 16

IV.2 Description des tailles des aires 17

↘ V. Vérifications et tests de robustesse 19

V.1 Présence d'établissements scolaires dans les aires d'éducation 19

V.2 Étude de la cohérence des aires à la rentrée 2020 21

V.3 Étude de la stabilité des aires au cours du temps 24

↘ VI. Déploiement et perspectives 27

VI.1 Livraison des zonages aux rectorats et à la DEPP 27

VI.2 Présentations des zonages 27

VI.3 Constitution d'un groupe de travail de valorisation des zonages 28

VI.4 Perspectives de mise à jour des zonages 28

↘ Remerciements 28

↘ Références bibliographiques 29

↘ Introduction

Le zonage en aires d'éducation, développé dans le cadre d'un groupe de travail « Aires d'éducation », entre la DEPP et les services statistiques académiques d'Aix-Marseille, Grenoble, Nantes, Lille et Reims, s'inscrit dans la logique d'élaboration de zonages d'études de la statistique publique. Les zonages sont construits principalement par l'Insee. Plusieurs d'entre eux s'appuient sur les flux entre domicile et lieu de travail (aires d'attraction des villes, zones d'emploi), et sont donc tout particulièrement adaptés aux analyses portant sur le marché du travail. Les bassins de vie sont construits sur une logique d'accessibilité aux services de la vie courante, permettant une variété d'analyses autour de ces thématiques[1].

Même si les zonages existants permettent une large gamme d'utilisations, certains besoins suscitent le développement de zonages spécifiques à un domaine particulier. Ainsi, le service statistique ministériel en charge du logement a développé un zonage adapté aux études sur le logement (Colin et Roussez, 2018).

Développé par des services statistiques du ministère de l'Éducation nationale, de la Jeunesse et des Sports (MENJS), le zonage en aires d'éducation s'inscrit dans cette lignée. Il s'appuie sur les flux d'élèves scolarisés dans le second degré des établissements de l'Éducation nationale. Ainsi, ce zonage est indépendant de tout autre zonage préexistant ; en particulier, il n'est pas impacté par la mise à jour de ces zonages. La méthodologie est définie de façon homogène sur le territoire national, ce qui facilite les comparaisons de territoires.

La méthodologie de construction respecte la logique principale des zonages statistiques, au sens où elle constitue une partition du territoire constitué par les départements de France métropolitaine et d'outre-mer[2]. Le zonage s'appuie sur des regroupements de communes. Ainsi, il permet une utilisation simple et la mise en relation avec la plupart des bases de données disponibles, incluant l'information sur la commune de scolarisation ou de résidence de l'élève.

Deux niveaux de zonage ont été produits : des aires « Parcours » construites sur la base des flux de l'ensemble des élèves du second degré ; des aires « Collèges » sur la base des flux des seuls collégiens.

Ce zonage s'applique à un large ensemble de thématiques portant sur les parcours et la réussite des élèves, et les facteurs pouvant les influencer (contexte sociodémographique, offre de formation). Il offre une maille territoriale pertinente pour le suivi fin des politiques éducatives. Un vaste ensemble de sujets peut donc être traité et il offre la possibilité de construire de nombreux indicateurs.

Le présent document décrit la démarche d'élaboration du zonage en aires d'éducation, qui a débuté en septembre 2020 et s'est achevée en septembre 2021 avec la livraison du zonage, pour les premières utilisations dans les services statistiques académiques et à la DEPP.

[1] Pour une revue de littérature sur les zonages de l'Insee, voir Aliaga (coord., 2015). Depuis ce document de référence paru en 2015, les zonages en aires urbaines et les zones d'emploi ont fait l'objet d'une refonte, et on peut se référer aux publications correspondantes (De Bellefon et al., 2020 ; Levy et al., 2020).

[2] Pour le zonage en aires d'éducation, les collectivités de Saint-Barthélemy et Saint-Martin sont également dans le champ. Celles-ci ont le statut de collectivité d'outre-mer depuis 2007, mais restent rattachées à la Guadeloupe pour l'éducation. Elles sont régulièrement incluses dans le champ des données publiées par la DEPP pour la France métropolitaine, les départements et régions d'outre-mer (DROM).

↘ I. Les besoins d'un zonage pour l'étude et le pilotage du système éducatif

I.1 Les besoins au niveau national

Les besoins qui ont motivé les travaux du groupe se déclinent au niveau national et au niveau local.

Au niveau national, pour ses travaux d'étude et de pilotage en lien avec les problématiques territoriales, la DEPP mobilise les zonages à des niveaux divers. Certains zonages sont importants pour caractériser les territoires en matière de degré de ruralité et d'urbanisation (Duquet-Métayer & Monso, 2019). Pour d'autres, leur utilité provient avant tout du fait qu'ils constituent une partition complète du territoire. Ils permettent notamment la représentation cartographique de données de contexte socioéconomique, d'offre de formation, de résultats et de parcours scolaire sur l'ensemble du territoire français, à une échelle plus fine que le département et l'académie. Les zonages utilisés dans ces travaux sont hétérogènes : zonages administratifs (comme l'arrondissement), statistiques (bassin de vie) ou ayant un statut mixte (canton-ville, qui correspond au canton administratif retravaillé par l'Insee pour faciliter les exploitations statistiques).

Par exemple, les cartes proposées par la DEPP et le Céreq dans l'Atlas des risques sociaux d'échec scolaire (Boudesseul et al., 2014 ; 2016), à l'échelle du canton, sont utilisées dans les dialogues de gestion entre l'administration centrale et les académies. Elles permettent d'appuyer les échanges sur une vision partagée du territoire académique en matière de contexte socioéconomique[3]. La publication Géographie de l'École (DEPP, 2021) a également mobilisé l'échelle du canton pour des représentations d'indicateurs de contexte socioéconomique. Ce choix du canton est cohérent avec le fait qu'il s'agit d'une échelle pertinente pour restituer l'information sur le contexte socioéconomique et les résultats scolaires (Murat, 2021). Il offre en effet un compromis entre la finesse de l'information d'un côté, la lisibilité et la robustesse des résultats de l'autre (environ 2 000 cantons actuellement en France).

Pour d'autres sujets, comme la mise en relation des parcours et de l'offre de formation, des zones plus étendues sont pertinentes. Par exemple, des travaux ayant trait aux parcours des élèves en lien avec l'offre de formation et la structure d'emploi, utilisant les zones d'emploi (Grelet, 2004) ou les arrondissements administratifs (Miconnet, 2016), ont mobilisé environ 350 zones en métropole.

Si, pour une grande partie des sujets, ces zonages préexistants offrent une solution satisfaisante, ils ne sont pas liés spécifiquement au système éducatif, ce qui peut rendre certains résultats difficiles à lire. Par exemple, des zonages statistiques tels que les zones d'emploi et les bassins de vie conduisent parfois à des zones très larges, qui reflètent la polarisation locale de l'emploi et des services mais sont moins pertinents pour l'éducation.

Ces zonages préexistants connaissent également des évolutions de périmètre d'une année sur l'autre, et plus occasionnellement, des refontes qui peuvent remettre en question certains travaux déjà réalisés. Pour les zonages administratifs, le redécoupage cantonal de 2013 en donne une illustration : le nombre de cantons a été réduit de moitié, et une partie des territoires français désormais plus découpée en cantons. Par ailleurs, les révisions de ces zonages statistiques de référence par l'Insee peuvent entrainer des ruptures de séries dommageables pour certains travaux engagés. Deux zonages de l'Insee ont ainsi fait l'objet de révision ou refonte en 2020 (aires urbaines devenues « aires d'attraction des villes », zones d'emploi).

À cette échelle, l'objectif du groupe de travail « Aires d'éducation » était ainsi de disposer d'un outil uniquement basé sur des données de l'Éducation nationale. Il s'agissait de construire un zonage pour les besoins d'éducation à des fins d'études et d'aide au pilotage qui ne dépende d'aucun autre zonage administratif ou d'aucun autre partenaire.

[3] Plusieurs académies les ont également utilisées pour le pilotage local (par exemple pour donner des éléments de contexte permettant d'appuyer l'auto-évaluation des établissements).

I.2 Les besoins au niveau local

La demande s'inscrit également dans un contexte thématique riche au niveau local.

À ce niveau, le besoin est de mettre à disposition des outils grâce auxquels les différents acteurs pourront répondre aux besoins d'études et de pilotage suivants :

- retracer, accompagner et rendre lisible le parcours des élèves de la maternelle à la fin de leur scolarité dans l'objectif de mieux les faire réussir ;
- faciliter la transition entre les niveaux, les degrés et les cycles d'apprentissage mais également accompagner l'entrée dans l'enseignement supérieur.

Par ailleurs, le besoin de développer des niveaux d'action et de décision pédagogiques en lien avec les territoires s'exprime de plus en plus (Azéma & Mathiot, 2019). Ainsi il s'agit d'avoir une approche plus large que l'établissement afin de donner aux élèves et aux familles une meilleure lisibilité sur les parcours possibles. L'enjeu est de faciliter la transition des élèves entre les différents niveaux et degrés. Le zonage doit aussi permettre des comparaisons entre territoires au niveau national.

Afin de répondre à ces besoins de pilotage territorialisé, les académies disposent d'un existant qui repose sur le découpage en circonscriptions pour le premier degré d'une part et d'un maillage en bassins d'éducation ou de formation (Bulletin Officiel n°26 du 28 juin 2001) qui répond à un besoin de travail en réseau collèges-lycées d'autre part. Ces deux zonages sont définis par l'autorité académique et répondent à des logiques administratives ou de gestion des moyens.

Toutefois, plusieurs académies ont fait le constat selon lequel le maillage en bassins d'éducation et/ou de formation n'était parfois plus pertinent et le découpage en circonscriptions ne répondait pas à tous les besoins de par sa trop grande finesse.

De plus, le besoin de comparaisons entre zones d'académies différentes est croissant, notamment dans le contexte de la création des régions académiques. Or, la méthode de construction des bassins de formation diffère d'une académie à l'autre et ne rend pas possible ces comparaisons.

Pour répondre à ces limites, la conception d'un nouveau zonage doit compléter les zonages déjà existants en retenant une approche plus adaptée à l'étude des parcours des élèves, et selon une méthodologie homogène sur l'ensemble du territoire. Ce faisant, ce zonage doit pouvoir mettre à disposition des pilotes en local des zones établies sur un maillage territorial qui se base sur la « réalité du terrain » et les comportements des familles. Ce nouveau découpage doit permettre de réunir tous les acteurs qui jalonnent le parcours de l'élève. On ne cherche pas directement un zonage de pilotage mais un zonage qui appuiera les décisions au niveau local.

Enfin, les besoins remontés de la part des académies ont suggéré que ce nouveau zonage, pour cette version-ci, devrait s'inscrire dans le cadre des frontières académiques, qui est structurant pour la prise de décision dans le système éducatif. En, revanche, il devait s'affranchir des limites interdépartementales, afin de favoriser le pilotage interdépartemental.

↘ II. La démarche et les choix méthodologiques

II.1 Le choix de l'outil Anabel

La méthodologie devait s'appuyer sur des indicateurs de flux, compte tenu de leur caractère structurant pour l'éducation, en particulier pour appuyer l'étude et le suivi du parcours des élèves. C'est pourquoi l'agrégation des communes se fait à partir de l'analyse des flux entre commune de résidence et de scolarisation. La méthodologie recourt à l'application Anabel (Analyse Bilocalisée pour les Études Locales) créée par l'Insee. Elle permet de construire des zones de façon homogène, ayant des taux de stabilité élevés, au sens où les élèves qui y résident, pour leur grande majorité, y sont également scolarisés.

Plus précisément, l'outil Anabel repose sur un algorithme qui agrège une à une les communes fortement liées entre elles, au sens des déplacements entre lieux de résidence et de scolarisation. Cet outil a déjà été utilisé par l'Insee pour la construction de zonages nationaux (aires urbaines, bassins de vie, zones d'emploi), appuyés sur les trajets entre lieu de résidence et de travail (Aliaga, coord., 2015).

Dans le cadre du zonage en aires d'éducation, le choix d'Anabel s'est également imposé par l'analogie de la problématique avec celle qui a guidé ces zonages nationaux, et plus particulièrement la construction des zones d'emploi (avant la refonte de 2020). En effet, le zonage recherché devait répondre à un impératif de stabilité élevée des mailles construites sur l'observation des déplacements des élèves de leur domicile à leur lieu d'étude, tout comme les zones d'emploi le font en ce qui concerne les déplacements des actifs vers leur lieu de travail.

L'outil Anabel avait déjà été utilisé pour l'analyse des flux entre commune de résidence et de scolarisation. Ces travaux ont été menés à l'échelle régionale, par exemple, dans le cadre des travaux sur les projections de lycéens (horizon 2025 / 2050). Ces projections ont été calculées avec l'outil méthodologique de projection d'habitants, d'actifs, de logements et d'élèves (Omphale) sur des zones d'au moins 50 000 habitants. Celles-ci étant attendues au lieu de scolarisation afin d'être mobilisées de façon opérationnelle par les académies et les collectivités locales compétentes et les projections démographiques étant réalisées au lieu de résidence, ces travaux devaient être réalisés à l'échelle d'une maille géographique offrant une stabilité optimale. C'est pourquoi l'outil Anabel a été utilisé dans plusieurs rectorats, en collaboration avec les directions régionales de l'Insee.

Par ailleurs, dans plusieurs académies, le zonage a été utilisé au-delà des projections, également pour une description complète du territoire en matière de contexte socioéconomique et d'offre de formation. Ce fut le cas notamment pour l'académie de Nantes (région Pays de la Loire) dans le cadre de la réalisation de l'ouvrage « Lycées et lycéens des Pays de la Loire : état des lieux et perspectives à l'horizon 2025 » (Ropers et al., 2013).

En fin de compte, l'adéquation entre la méthodologie d'Anabel et les objectifs du zonage en aires d'éducation, ainsi que l'expertise des services statistiques académiques sur cet outil (à l'occasion des travaux réalisés avec les directions régionales de l'Insee) ont contribué à justifier le choix d'Anabel. Si d'autres travaux menant à la construction de systèmes territoriaux étaient connus des membres du groupe de travail (par exemple à l'occasion de travaux réalisés en direction régionale de l'Insee mobilisant la théorie des graphes[4]), il a semblé que l'outil Anabel pouvait répondre au besoin d'un zonage reflétant la réalité du terrain et mobilisable de façon opérationnelle. Ce constat a été confirmé par le fait que dans certaines académies (par exemple à Lille ou encore à Aix-Marseille et Nice), la confrontation de zonages issus d'Anabel avec les zonages opérationnels a indiqué de fortes similitudes.

La pertinence de l'utilisation s'est également trouvée confirmée a posteriori par l'obtention au niveau national de zones stables, pertinentes et qui parlaient aux acteurs de terrain, d'après les retours de plusieurs académies qui ont participé au groupe de travail ou ont présenté le zonage en interne dans leur académie.

[4] Pour une application à l'ancienne région Nord-Pas-de-Calais, voir par exemple Laval et Vilain (2014).

II.2 La description de la méthodologie et les choix de paramètres

Pour le cas des aires d'éducation, l'objectif est d'étudier le lien entre la commune de résidence de l'élève et la commune d'implantation de l'établissement d'inscription. Les données utilisées pour décrire ces liens sont issues de l'enquête dite de Constat de Rentrée des effectifs d'élèves (système d'information Scolarité) à la rentrée 2020 et sont hébergées dans la Base Centrale de Pilotage de la Depp. Le champ correspond aux élèves, sous statut scolaire, du second degré public et privé sous contrat inscrits dans les établissements sous tutelle du MENJS (quel que soit le type d'établissement). Il exclut les élèves des classes préparatoires aux grandes écoles, des sections de techniciens supérieurs et autres formations post-baccalauréat. Les flux sont calculés en croisant la commune de résidence et la commune de scolarisation des élèves.

L'agrégation des communes est réalisée en utilisant l'algorithme Anabel, dont on trouvera une description détaillée dans le document de référence de l'Insee sur les zonages supracommunaux (Aliaga, coord., 2015). Nous présentons ci-dessous le principe de la méthode et les choix de paramètres réalisés pour constituer les aires d'éducation.

Afin de permettre leur utilisation pour le pilotage au niveau de l'académie, une première contrainte géographique a été ajoutée en amont du processus de calcul : seuls les flux « intra-académiques » ont été pris en compte. À l'intérieur d'une même académie, les premières communes agrégées sont celles qui envoient tous leurs élèves résidents (100 %) dans une même commune d'études. S'ensuivent les deux communes ou zones les plus liées entre elles, jusqu'à atteindre un critère d'arrêt.

Sans critère d'arrêt, Anabel construit une zone unique. L'un des enjeux principaux de la méthode d'agrégation est donc de définir des critères d'arrêt du processus d'agrégation en fixant un seuil de lien minimal entre deux zones : par exemple en fixant un seuil à 0,5 (50 %), Anabel n'agrège deux zones que si au moins 50 % des élèves de l'une des deux zones étudient dans l'autre.

Une deuxième contrainte géographique est imposée via l'application Anabel : le critère de contiguïté. Selon ce critère, une commune ne peut être agrégée à une zone que si elle partage une frontière commune avec cette zone. Ce critère est implémenté via un fichier fourni par l'IGN sur la géographie 2020.

Au final, un paramètre externe (restriction aux flux intra-académiques) et deux paramètres internes à l'application Anabel (le critère du seuil et la prise en compte de la contiguïté) servent de contraintes dans l'élaboration du zonage. Ces critères sont fixés de manière uniforme au niveau national, dans un souci de permettre des comparaisons entre zones d'académies différentes.

L'application Anabel permet de mobiliser d'autres critères comme l'isolation partielle : il s'agit de la taille maximale d'un « satellite » pouvant être agrégé à une zone déjà existante dans l'exécution de l'algorithme. Il a été décidé de rendre ce critère non impactant : en effet, l'utilisation du seul critère de seuil a permis d'arriver à des résultats satisfaisants, sans avoir besoin de complexifier la démarche par l'ajout d'autres critères. Situé par défaut dans l'application à 15 000, le critère d'isolation partielle est valorisé à 60 millions, et n'intervient donc jamais dans la constitution de zones.

Pour des raisons similaires, le critère d'isolation totale a également été neutralisé. Il s'agit de la taille maximale au-delà de laquelle une zone ne peut plus évoluer, c'est-à-dire ni être rattachée à une zone existante ni en absorber d'autres. Le paramètre correspondant dans Anabel a été laissé à sa valeur par défaut de 60 millions, donc inopérante pour la construction du zonage.

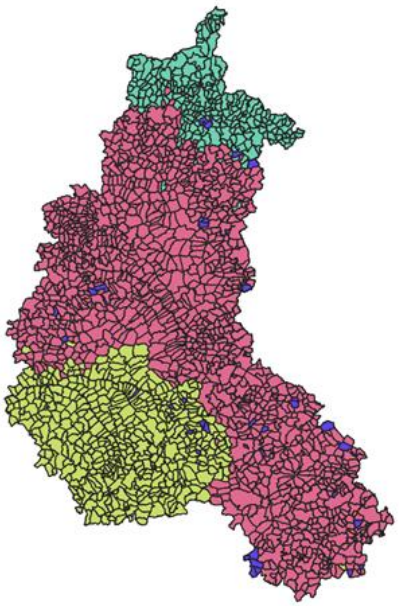
Le seuil a été identifié empiriquement, validé a posteriori par le groupe de travail sur la base de l'exécution de l'algorithme pour cinq académies en prenant en compte les critères suivants : la stabilité dans la zone (mesurée par la part d'élèves scolarisés dans la zone parmi ceux qui y résident), la taille de la zone (en nombre d'élèves et de communes) et le nombre de zones.

L'algorithme a ainsi été exécuté successivement en utilisant les seuils suivants.

Aucun seuil (zonage « naïf »)

On voit ici dans l'académie de Reims que sans critère d'arrêt, une zone couvrant le département de Marne, de la Haute-Marne et une partie de l'Aube et des Ardennes se forme (figure 1). Une seule zone couvre presque l'ensemble du territoire à partitionner, de sorte qu'on peine à voir la valeur ajoutée d'un tel zonage.

FIGURE 1 • Exemple sans seuil dans l'académie de Reims



Test de différents seuils

Par conséquent, compte tenu de la nécessité de fixer un seuil dans Anabel, les seuils suivants ont été testés :

- seuil à 0,5 (50 % des élèves de l'une des deux zones étudient dans l'autre) ;
- seuil à 0,2 (20 % des élèves de l'une des deux zones étudient dans l'autre) ;
- seuil à 0,15 (15 % des élèves de l'une des deux zones étudient dans l'autre) ;
- seuil à 0,1 (10 % des élèves de l'une des deux zones étudient dans l'autre).

Le seuil à 0,1 est apparu satisfaisant dans la plupart des cas, mais certaines zones étaient de grande taille, comme à Nantes où une zone était constituée par un seul département, la Mayenne (figure 2). Dans ce zonage, l'existence de zones de grande taille est à relier à la dispersion de l'offre de formation, notamment en lycée professionnel. Ces grandes zones forment une maille de décision stratégique, à une échelle large, par exemple celle de la région, et sur des sujets tels que les parcours des élèves du lycée vers l'enseignement supérieur.

Pour cette raison, la construction d'un second zonage est apparue pertinente, permettant de répondre aux besoins de pilotage plus fins, soit au niveau départemental, soit à l'infra-départemental. La construction de ce deuxième zonage utilise des données de flux plus inertes que celles décrivant l'ensemble des déplacements des élèves dans le second degré. Elle se focalise sur les flux des collégiens.

En effet, l'offre de formation en collèges est plus homogène sur l'ensemble du territoire, induisant des distances parcourues plus courtes qu'en lycée, où l'offre est plus diversifiée et les flux vers la voie professionnelle non soumis à sectorisation.

Le groupe de travail a donc arrêté la décision de la création de deux zonages : un zonage « Parcours » prenant en compte les flux domicile-lieu de scolarisation des élèves inscrits au collège ou au lycée, et un zonage « Collèges » prenant en compte les seuls flux des élèves inscrits dans un collège (ce qui exclut notamment les élèves en troisième prépa-métiers scolarisés dans un lycée).

Ainsi, pour reprendre dans le cas de Nantes, on vérifie que ce second zonage donne une version plus fragmentée du territoire, et en particulier que la Mayenne est désormais constituée de plusieurs zones distinctes (figure 3).

FIGURE 2 • Exemple de l'académie de Nantes, seuil à 0,1 pour les flux du second degré

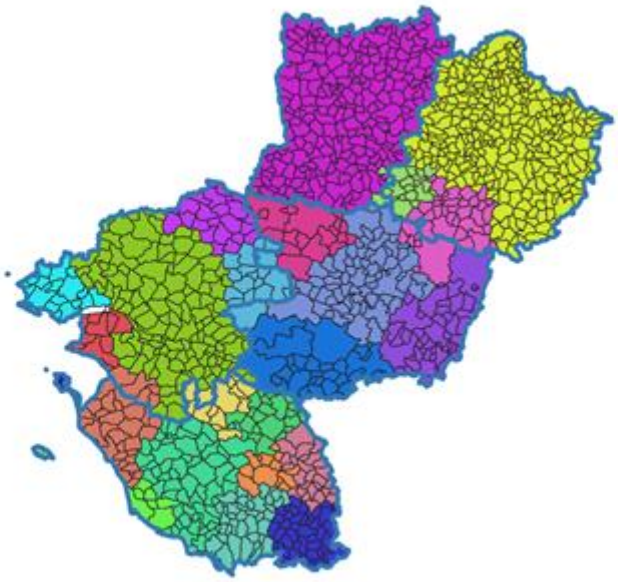
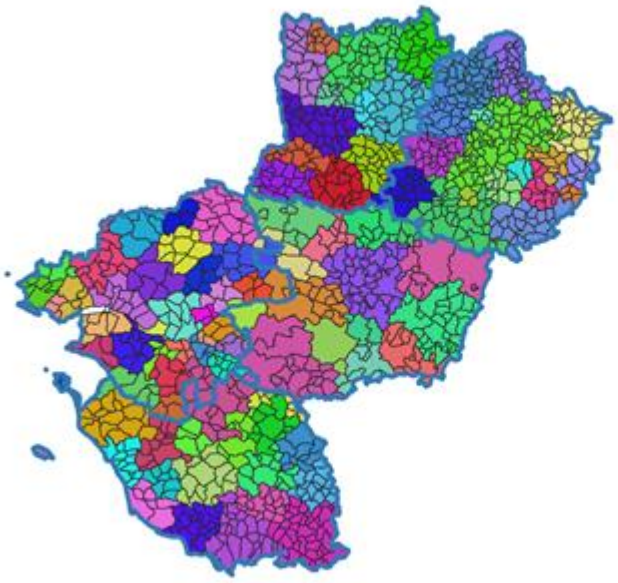


FIGURE 3 • Exemple de l'académie de Nantes, seuil à 0,1 pour les flux de collégiens

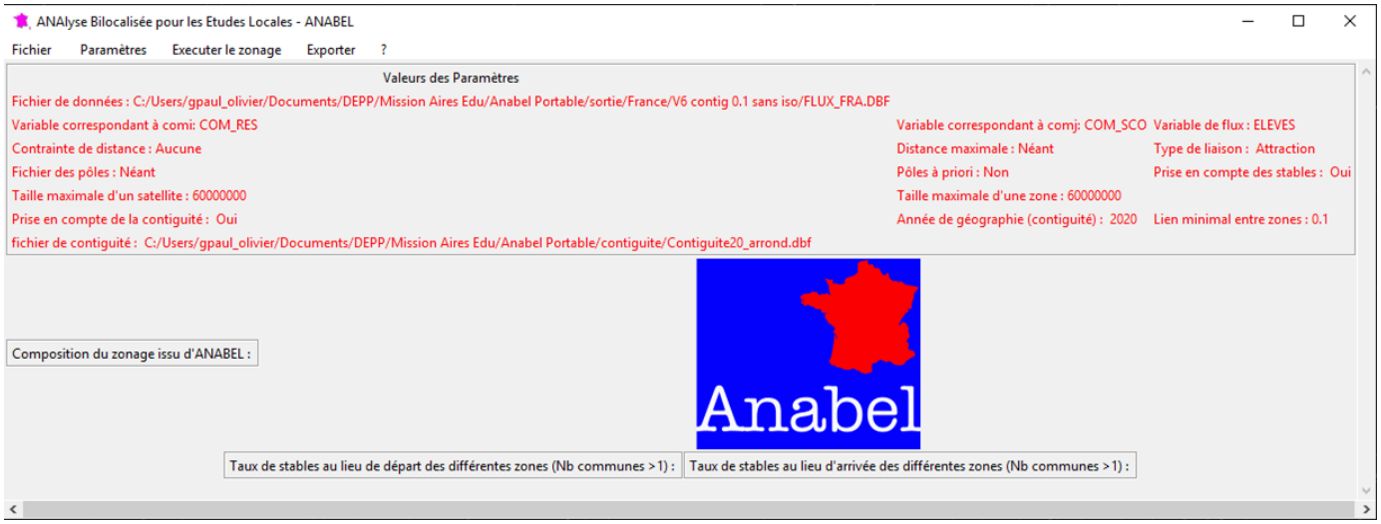


↘ III. La mise en œuvre de l'algorithme et la consolidation des zonages

III.1 Les zonages Parcours et Collèges en sortie d'Anabel

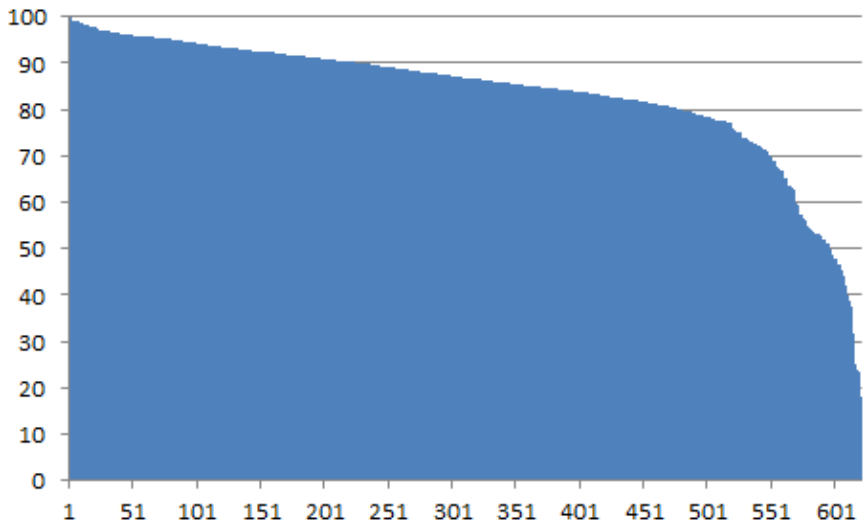
Deux jeux de données décrivant les flux d'élèves ont été constitués à partir de la Base Centrale de Pilotage de la Depp. L'application Anabel a ensuite été exécutée sur chacun de ces jeux de données avec les paramètres déjà décrits :

- seuil de lien minimal égal à 0,1 (10 %) ;
- prise en compte de la contiguïté ;
- taille maximale d'un satellite et d'une zone égales à 60 millions (non opérantes).



Résultat pour le zonage Parcours : l'algorithme, en 33 823 pas d'agrégation, a permis d'identifier 621 zones présentant pour la plupart une bonne stabilité, puisque 97,4 % d'entre elles affichent un taux de stabilité supérieur à 60 % (figure 4).

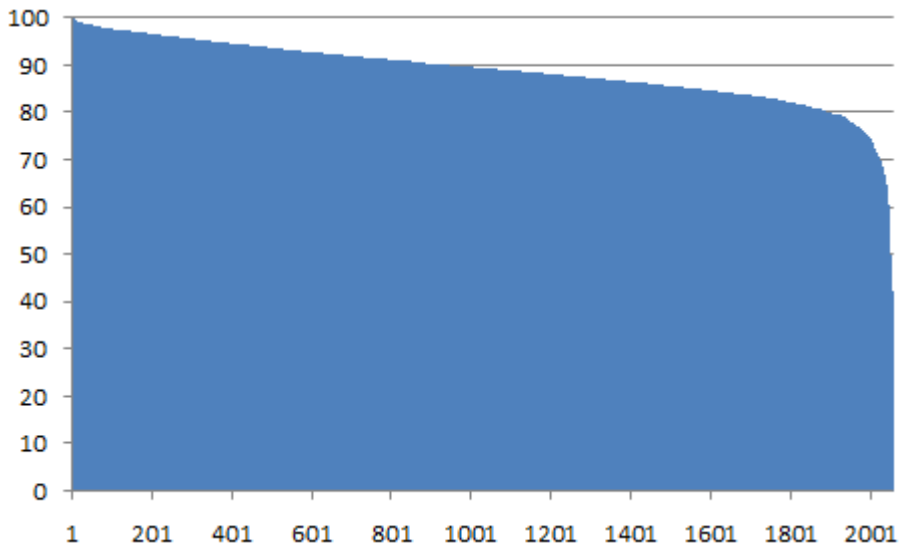
FIGURE 4 • Taux de stabilité du zonage Parcours en sortie d'Anabel, en %



Champ : France métropolitaine et DROM, second degré public et privé sous contrat, hors élèves qui ne sont pas scolarisés dans leur académie.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

Résultat pour le zonage Collèges : l'algorithme, en 32 026 pas d'agrégation, a permis d'identifier 2 058 zones, présentant une stabilité encore meilleure, puisque 99,4 % d'entre elles affichent un taux de stabilité supérieur à 60 %. Cette meilleure stabilité découle du fait, déjà évoqué, que les flux domicile-scolarisation des collégiens présentent par nature plus d'inertie que ceux des lycéens (figure 5).

FIGURE 5 • Taux de stabilité du zonage Collèges en sortie d'Anabel, en %



Champ : France métropolitaine et DROM, second degré public et privé sous contrat, hors élèves qui ne sont pas scolarisés dans leur académie.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

III.2 La consolidation des zonages à l'aide de l'expertise académique

Après la mise en œuvre d'Anabel sur les données, il est paru nécessaire de faire un « toilettage » notamment car certaines communes, n'ayant pas d'élèves dans les bases de l'année utilisée, se retrouvaient non zonées. Par ailleurs, d'autres zones pouvaient poser question, notamment pour les zones interdépartementales, ou encore présenter des problèmes de contiguïté non résolus par Anabel (communes isolées[5]), etc. La pertinence de corriger et donc de rattacher une commune ou une petite zone à une autre ne pouvait s'effectuer qu'à partir d'une connaissance locale du territoire.

L'ensemble des services statistiques académiques (SSA) ont donc été sollicités pour cette opération qui permettait aussi de faire une validation « pratique » des zonages construits. Cependant, seuls des arguments techniques pouvaient motiver ces corrections. Le toilettage s'est donc opéré sans consultation des corps d'inspection ou des secrétaires généraux, mais uniquement à partir de l'étude des flux d'élèves à un niveau très fin ainsi que de certains critères comme la carte des formations par exemple.

Cette opération ne devait pas être guidée par la recherche de cohérence avec les zonages préexistants en académie (sectorisation, bassins de formation). Il n'était pas autorisé de créer de nouvelles zones et il n'était pas demandé de chercher à homogénéiser la taille des zones : toute intervention dans ce sens risquait en effet d'introduire de l'arbitraire dans la construction du zonage et d'atténuer les possibilités de comparaisons entre les académies.

[5] Par « commune isolée », on fait référence à une commune qui constitue une zone à elle seule en sortie d'Anabel (à la différence des communes « non zonées », déjà décrites). Ce cas peut provenir d'un défaut de contiguïté : les élèves de la commune se rendent majoritairement dans une autre commune ou une zone existante, mais cette zone n'est pas contiguë. Cette situation peut se produire par exemple du fait de la sectorisation : la commune A est rattachée au secteur de collège de la commune B, mais il faut pour cela traverser une portion de territoire de la commune C, qui ne fait pas partie de ce secteur. Le travail de toilettage conduit à rattacher la commune A à la commune B.

Ce toilettage a été géré par le groupe de travail « Aires d'éducation » pour assurer l'homogénéisation de traitement au niveau national et le support technique auprès des SSA. Un guide du toilettage a été fourni aux services et des visio-conférences hebdomadaires ont été mises en place pour coordonner l'ensemble de ce travail. Tous les services statistiques académiques ont participé à ces visio-conférences et les échanges ont été riches et constructifs, permettant un partage de la connaissance et de la pratique.

Le toilettage des deux zonages a été effectué en deux temps : d'abord le zonage Parcours, ensuite le zonage Collèges. Toutefois, les zonages Parcours et Collèges sont toilettés de manière indépendante et ont chacun leur logique propre. Aucune contrainte quant à l'imbrication de l'un dans l'autre n'a été posée.

Dans les deux cas, il y avait une phase de visualisation cartographique du zonage pour repérer notamment toutes les communes non zonées ou isolées, et une phase d'analyse des fichiers de flux pour aboutir à une correction ou non. L'ensemble des modifications devaient, *in fine*, rester parcimonieuses, afin de préserver l'harmonisation de la méthode.

De nombreux cas particuliers ont été traités en ateliers ou en bilatéral avec les SSA concernés et partagés via une « Foire aux questions » afin que chaque SSA puisse appliquer les recommandations du groupe de travail. Des ateliers spécifiques ont également été organisés pour certains territoires présentant des caractéristiques particulières.

Atelier Île-de-France

L'atelier portant sur l'Île-de-France était notamment consacré à la question de savoir s'il était préférable ou non de faire des zones inter-académiques dans cette région, les flux entre académies franciliennes étant importants. Les besoins de pilotage en rectorat ont conduit à privilégier la construction de zones intra-académiques. L'Île-de-France n'a donc pas fait exception à la règle définie au plan national.

Atelier des communes de Paris, Lyon et Marseille

Cet atelier visait à définir si on appliquait Anabel à la commune ou si l'on tenait compte des arrondissements. Au final, les variantes menées en partant des communes entières ou des arrondissements de Lyon et Marseille ont abouti au même zonage (les arrondissements se regroupant au sein d'une même zone). Pour ce qui est de Paris, où une question similaire s'est posée, une variante du zonage Collèges a été élaborée à des fins techniques et d'analyse locale. Elle prend pour maille élémentaire les arrondissements de Paris et identifie 9 aires en complément du zonage Collèges.

Atelier des départements et régions d'outre-mer

L'atelier pour les départements et régions d'outre-mer (DROM) a notamment permis d'étudier le cas des flux entre îles. Celles-ci se retrouvaient isolées en sortie d'Anabel, faute de contiguïté, ce qui n'était pas toujours justifié. Les corrections ont été apportées en collaboration avec les SSA des DROM.

Tous les changements ont été documentés et justifiés par chaque service et certains ont été refusés par le groupe de travail pour garantir l'harmonisation nationale. Un bilan du toilettage a été présenté à l'ensemble des protagonistes, en même temps que le zonage final, en septembre 2021.

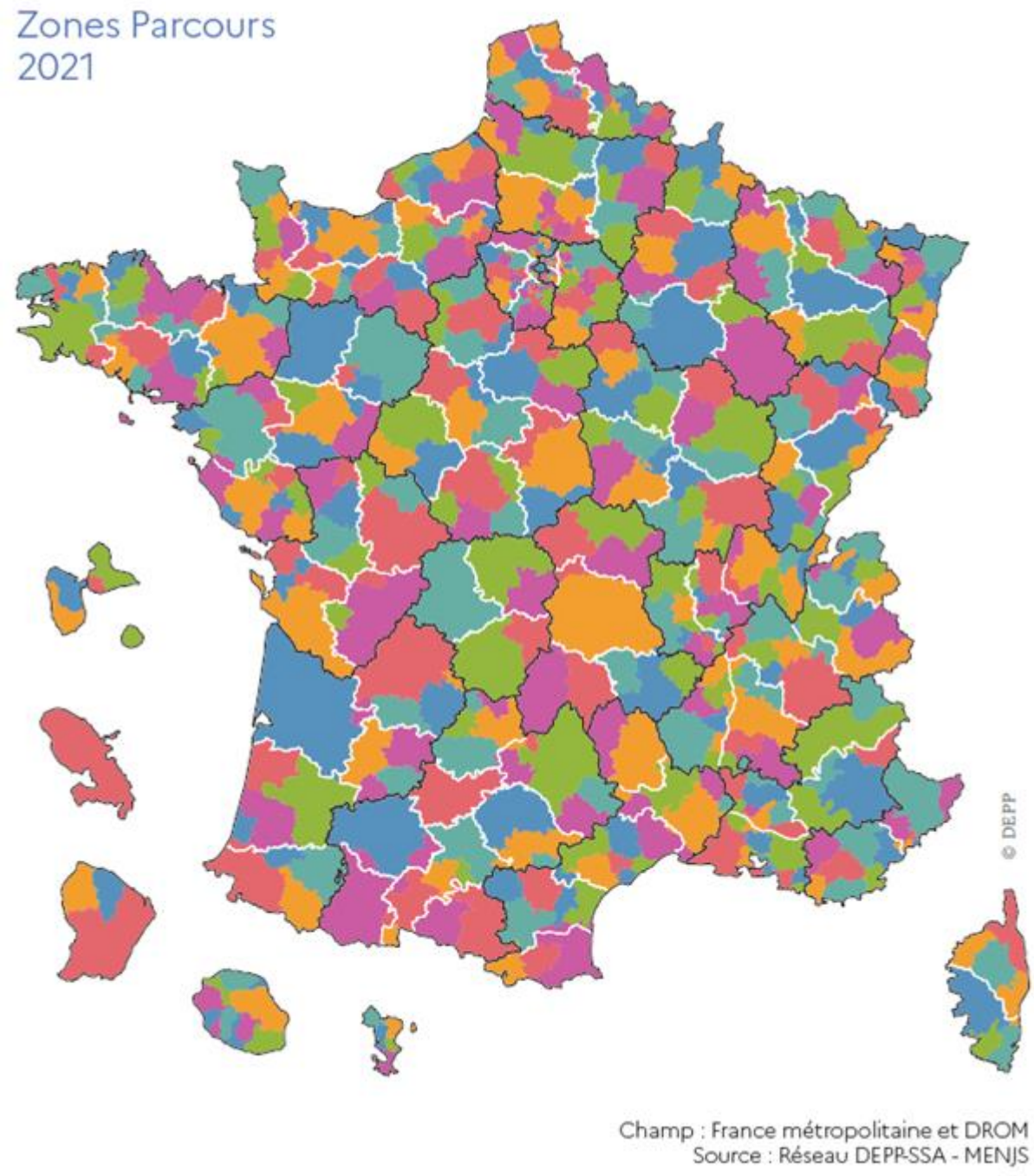
↘ IV. Le zonage final

IV.1 Cartographie des aires d'éducation

Deux niveaux de zonage ont donc été produits (figures 6 et 7) :

- **un zonage Parcours** construit sur la base des flux de l'ensemble des élèves du second degré (1er et 2nd cycle, voie générale, technologique et professionnelle) ;
- **un zonage Collèges** de granularité plus fine sur la base des flux des seuls collégiens (i.e. scolarisés dans un collège, excluant donc les flux des 3eprépa-métiers scolarisés en lycée).

FIGURE 6 • Les 548 aires du zonage Parcours



Champ : France métropolitaine et DROM
Source : Réseau DEPP-SSA - MENJS

FIGURE 7 • Les 1 986 aires du zonage Collèges



Champ : France métropolitaine et DROM
Source : Réseau DEPP-SSA - MENJS

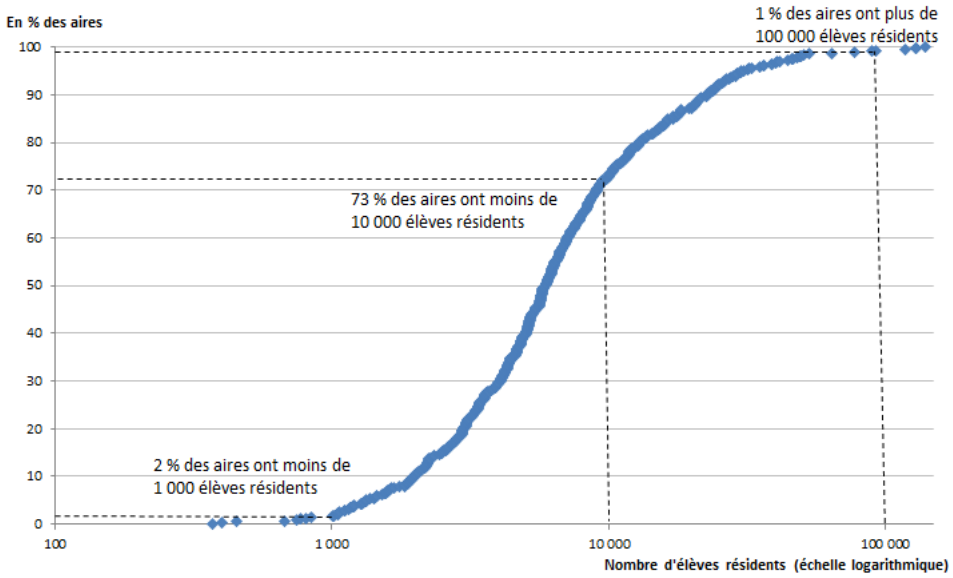
IV.2 Description des tailles des aires

La construction du zonage s'est faite sans poser de contrainte sur la taille des aires. Sa grande variabilité reflète la diversité de situations des aires (au voisinage d'une métropole ou plus éloignées ou encore insulaires) mais aussi la mobilité différente des jeunes d'un territoire à un autre (figures 8 et 9).

Pour le zonage Collèges comme pour le zonage Parcours, les traitements ont permis d'arriver à des aires de taille suffisante pour pouvoir fournir des indicateurs robustes, y compris pour les plus petites aires.

Dans le zonage Parcours, environ la moitié des aires ont moins de 6 000 élèves (médiane à 5 917). Le minimum est à 371 élèves résidents (Bagnères-de-Luchon) et le maximum à 140 360 (Paris).

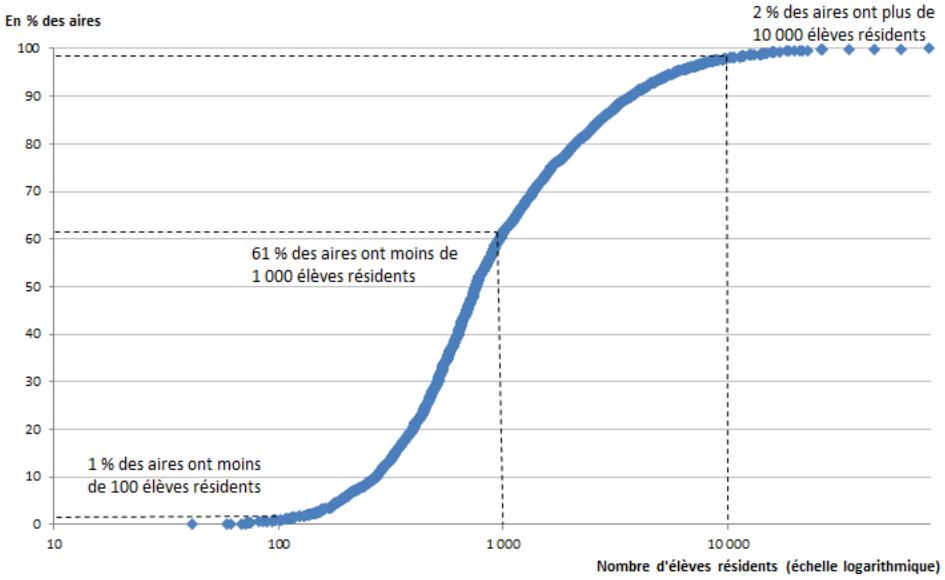
FIGURE 8 • Distribution de la taille des aires Parcours, rentrée 2020



Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

Dans le zonage Collèges, les aires sont, logiquement, de taille plus petite. Environ la moitié des aires ont moins de 800 élèves (médiane à 757). Le minimum est à 41 élèves résidents (La Désirade) et le maximum à 78 694 élèves (Paris - ville entière).

FIGURE 9 • Distribution de la taille des aires Collèges, rentrée 2020



Champ : France métropolitaine et DROM, élèves scolarisés en collège public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

↘ V. Vérifications et tests de robustesse

Suite à l'opération de toilettage, au terme de laquelle chaque commune française a été zonée, nous avons procédé à plusieurs vérifications permettant d'attester de la pertinence des aires constituées, au regard des objectifs assignés aux zonages.

V.1 Présence d'établissements scolaires dans les aires d'éducation

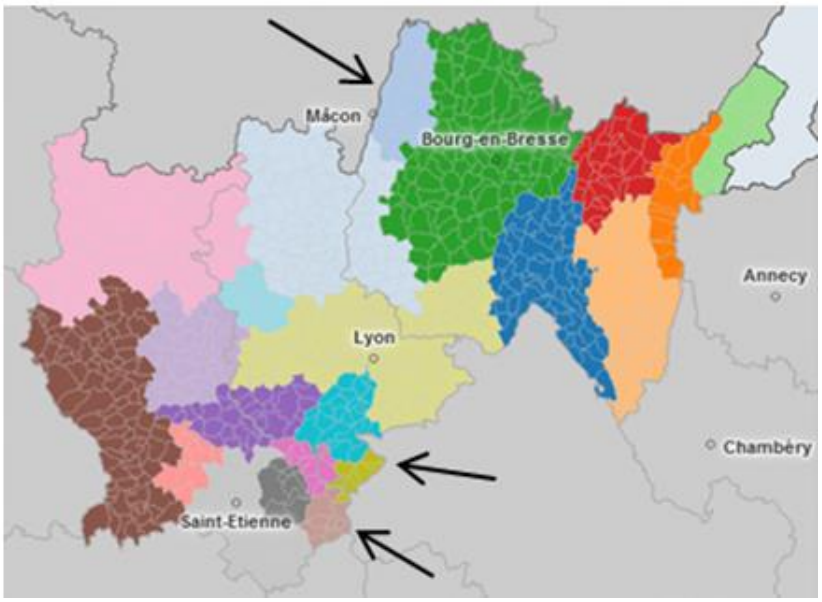
Tout d'abord, la logique de construction du zonage impliquait de créer des aires où les flux étaient polarisés par des communes dotées d'un ou plusieurs établissement(s) scolaire(s). La méthodologie mise en œuvre avec Anabel conduit à cette situation sauf dans des cas très spécifiques ou d'anomalies (communes isolées par défaut de contiguïté). Ces dernières ont été corrigées lors du toilettage.

Dans la version des zonages après toilettage, nous avons vérifié la présence d'au moins un collège sous tutelle de l'Éducation nationale dans toutes les aires Collèges. Une vérification similaire pour les lycées a été effectuée. Dans chaque aire Parcours, on trouve ainsi au moins un lycée sous tutelle de l'Éducation nationale, à 4 exceptions près.

Il s'agit tout d'abord de l'île de Saint-Barthélemy, pour l'académie de Guadeloupe, dont le cas a été arbitré en atelier DROM avec le SSA. Il n'était guère possible, et opportun, de l'affecter à une aire existante. Les flux sont faibles car d'une part, ils sont répartis entre plusieurs destinations et d'autre part, le passage au lycée peut être associé à un changement de lieu de résidence de l'élève quand il occasionne un départ de l'île. On note toutefois que, formellement, les élèves ont accès à une offre de formation en lycée dans l'île, même si elle est limitée (le collège Mireille Choisy a une seconde générale et technologique). Faute d'alternative satisfaisante, nous avons donc laissé l'île comme une aire Parcours spécifique.

Ensuite, dans le cas de l'académie de Lyon, trois aires sont concernées, autour de Bâgé-Dommartin, de Pélussin et de Condrieu. Au vu des flux au départ de ces trois aires, aucune autre aire ne dépasse le seuil de 10 % d'élèves justifiant un rattachement (figure 10).

FIGURE 10 • Localisation des trois aires de l'académie de Lyon concernées par la sectorisation inter-académique



©Géoclip-IGN GéoFla

Ceci s'interprète par leur caractère frontalier et l'existence de secteurs ou accords inter-académiques. Les élèves de Bâgé-Dommartin (au nord) sont sectorisés à Mâcon (Saône-et-Loire, académie de Dijon) en lycée GT. Les élèves de Condrieu et de Pélussin (au sud) bénéficient d'accords inter-académiques pour se rendre dans les lycées voisins de l'académie de Grenoble. Ils incluent notamment le lycée polyvalent de Saint-Romain-en-Gal, implanté physiquement dans le Rhône (commune de Saint-Romain-en-Gal), mais rattaché administrativement à l'académie de Grenoble (c'est pourquoi la commune d'implantation de l'établissement dans les bases scolaires est Vienne, Isère).

Pour statuer sur ces 3 cas repérés au toilettage, il est apparu nécessaire de s'assurer qu'ils étaient bien les seuls cas de figure au plan national. Cette vérification s'est appuyée sur la partition en sortie d'Anabel. Elle a permis d'identifier en tout 15 situations d'aires Parcours avec collège mais sans lycée (tableau 1).

TABLEAU 1 • Analyse des cas d'aires avec collège et sans lycée en sortie d'Anabel

Académie	Code zone Anabel	Libellé	Corrigé lors du toilettage	Île	Commentaire
10	01025	Bâgé-Dommartin			Cas de sectorisation inter-académique
10	42168	Pélussin			Cas de sectorisation inter-académique
10	69064	Condrieu			Cas de sectorisation inter-académique
13	17369	Saint-Martin-de-Ré	X	X	
13	17385	Saint-Pierre-d'Oléron	X	X	
14	29189	Plousgastel-Daoulas	X		Commune sectorisée à Brest, mais non contiguë. Correction faite par le SSA
14	59069	Groix	X		
14	56152	Le Palais	X	X	
17	85113	L'Île-d'Yeu	X	X	
17	85163	Noirmoutier-en-l'Île	X	X	
24	93047	Montfermeil	X		Commune sectorisée au Raincy, mais non contiguë. Correction proposée par le GT et validée par le SSA
32	97110	La Désirade	X	X	
32	97131	Terre-de-Haut	X	X	
32	97701	Saint-Barthélemy		X	
70	61484	Val-au-Perche	X		Commune frontalière de l'académie de Nantes mais qui est bien sectorisée en Normandie. Il y a plusieurs lycées de rattachement, ce qui peut expliquer cette particularité (aucun flux ne dépasse le seuil de 10 % avec chaque commune prise séparément). Dans tous les cas, le rattachement fait par le SSA à Montagne-au-Perche (l'une des communes de rattachement) est logique

Il s'est avéré qu'aucune des nouvelles situations ainsi repérées n'était liée à la sectorisation, ce qui les distinguait des cas lyonnais. Finalement, toutes ces situations (9 concernent les îles et 3 sont situées sur le continent) avaient été traitées lors du toilettage, en dehors de Saint-Barthélemy déjà mentionné.

En définitive, les 3 cas lyonnais sont apparus très spécifiques : il s'agissait d'aires entières reliées par la sectorisation du lycée à des aires limitrophes, situées dans d'autres académies, tout en ayant elles-mêmes un collège. Ces cas n'ont pas d'équivalent dans d'autres académies. Il était délicat de proposer des solutions de rattachement à d'autres aires de l'académie de Lyon. Il était tout aussi peu opportun de revenir sur le principe de ne pas construire d'aires inter-académiques. L'option retenue a donc été de laisser ces 3 aires telles quelles, et de mentionner leurs particularités dans la documentation. Cela permettra de les identifier et de les traiter de façon spécifique, si le besoin s'en fait sentir. Par exemple, pour certaines études, les indicateurs pourront être calculés en les associant avec les aires Parcours correspondantes des académies de Dijon et de Grenoble (Mâcon pour Bâgé-Dommartin et Vienne pour Pélussin et Condrieu).

V.2 Étude de la cohérence des aires à la rentrée 2020

Le fait qu'une aire soit dotée d'au moins un établissement scolaire sur son sol ne garantit toutefois pas sa cohérence. Comme indiqué initialement, le but de l'exercice est de parvenir à des aires dans lesquelles la majorité des élèves résidents sont également scolarisés. À défaut de parvenir à une telle cohérence, il sera difficile d'interpréter les indicateurs calculés sur ces aires, en matière d'études ou encore de pilotage (cf. partie I).

Cette cohérence du zonage peut se mesurer en calculant les taux de stabilité des aires. La formule la plus couramment adoptée pour définir ce taux est la suivante : rapport entre le nombre d'élèves scolarisés et résidents dans l'aire et le nombre total d'élèves résidents dans l'aire[6]. Outre de pouvoir être compris de façon intuitive, ce calcul est aussi cohérent avec le critère utilisé pour la construction du zonage.

Dans le zonage final, 91,8 % des élèves sont scolarisés dans leur aire de résidence pour le zonage Collèges et 90,5 % pour le zonage Parcours. La moitié des aires Collèges ont des taux de stabilité supérieurs à 88,3 % et la moitié des aires Parcours ont des taux de stabilité supérieurs à 86,7 % (tableau 2).

TABLEAU 2 • Taux de stabilité moyen des aires, rentrée 2020, en %

		Avant toilettage	Après toilettage
Collèges	Moyenne	91,5	91,8
	Médiane des zones	87,9	88,3
Parcours	Moyenne	89,8	90,5
	Médiane des zones	84,2	86,7

Lecture : dans le zonage en sortie d'Anabel, 91,5 % des élèves étaient scolarisés dans leur aire de résidence et 91,8 % après toilettage.
Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

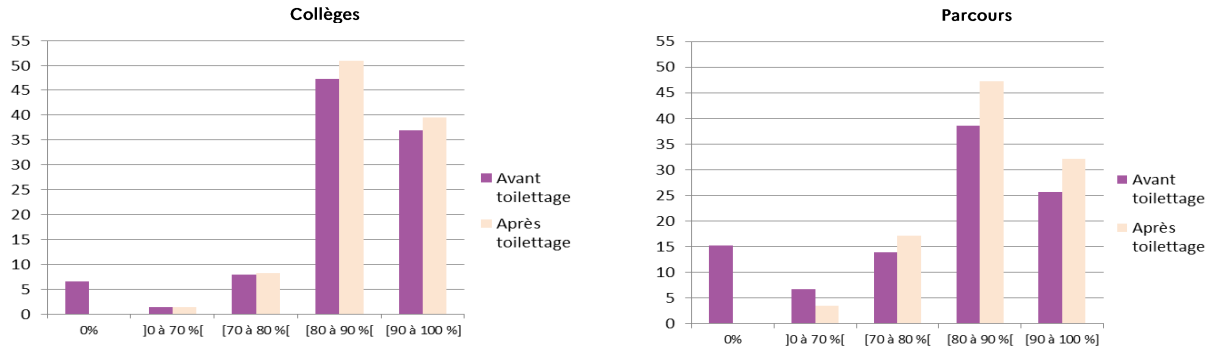
À titre de comparaison, pour les zones d'emploi 2020, déterminées à partir des flux domicile-travail, la moitié des zones d'emploi ont des taux de stabilité supérieurs[7] à 77,8 %. Du fait de la sectorisation en collège et en lycée GT, et du fait que les déplacements domicile-études sont plus courts que les déplacements domicile-travail, il a donc été possible, sur les données de l'Éducation nationale, de former des ensembles de communes présentant un fort degré de cohérence interne.

De plus, les travaux de toilettage effectués par les SSA, décrits ci-dessus, ont permis d'améliorer la cohérence des aires. Cela est dû au traitement des communes « isolées »[8] : celles-ci n'avaient pas été affectées par Anabel à leur aire de rattachement la plus logique (au vu des flux) faute de contiguïté avec l'aire en question. Le taux de stabilité moyen a légèrement progressé suite à l'opération de toilettage : de 91,5 % à 91,8 % pour le zonage Collèges et de 89,8 % à 90,5 % pour le zonage Parcours. Compte tenu du fait que les corrections portent sur des aires de petite taille, elles concernent peu d'élèves et comptent peu dans l'amélioration du taux de stabilité moyen. Les améliorations du zonage sont bien plus sensibles lorsqu'on considère les proportions d'aires à faibles taux de stabilité. En particulier, 6,5 % des aires Collèges et 15,2 % des aires Parcours avaient un taux de stabilité de 0 %, correspondant aux communes isolées par défaut de contiguïté (figure 11). Ces anomalies ont toutes été corrigées au moment du toilettage.

[6] C'est par exemple la façon dont l'Insee a communiqué sur la stabilité des zones d'emploi 2020. Voir la publication https://www.insee.fr/fr/statistiques/5015734?sommaire=4763049#graphique-figZE1_radio1. Ce taux peut aussi être calculé au lieu de scolarisation (rapport entre le nombre d'élèves scolarisés et résidents dans la zone et le nombre total d'élèves scolarisés dans la zone) ou sous forme composite, tenant compte à la fois des flux entrants et sortants de la zone (rapport entre le nombre d'élèves résidents ET scolarisés et le nombre d'élèves résidents OU scolarisés).
[7] Cf. la publication de l'Insee citée ci-dessus.
[8] Une part importante du travail des SSA a également consisté à affecter une aire à des communes non zonées en sortie d'Anabel, faute d'élèves résidents. Ces corrections n'ont, par construction, pas de conséquence sur les taux de stabilité calculés sur la rentrée 2020.Rappelons que, les SSA ont pu, exceptionnellement, procéder à d'autres types de modifications ayant un impact sur les taux de stabilité, notamment dans un souci d'utilisation opérationnelle du zonage.

On notera également que pour le zonage Parcours, 6,7 % des aires avaient un taux de stabilité inférieur à 70 % (sans être nul) en sortie d'Anabel. Après toilettage, il n'y en a plus que 3,5 %. Parmi elles, les taux de stabilité les plus faibles sont atteints dans l'académie de Lyon, pour les aires de Condrieu (38,8 %), Pélussin (52,9 %) et Bâgé-Dommartin (56,9 %). Ces situations s'expliquent par des spécificités déjà signalées liées à la sectorisation inter-académique : des communes de ces aires sont en effet reliées à des académies limitrophes (Besançon et Grenoble) par la sectorisation en lycée GT. Les autres aires Parcours ont des taux de stabilité supérieurs à 57 %.

FIGURE 11 • Taux de stabilité des aires avant et après toilettage, rentrée 2020, en %

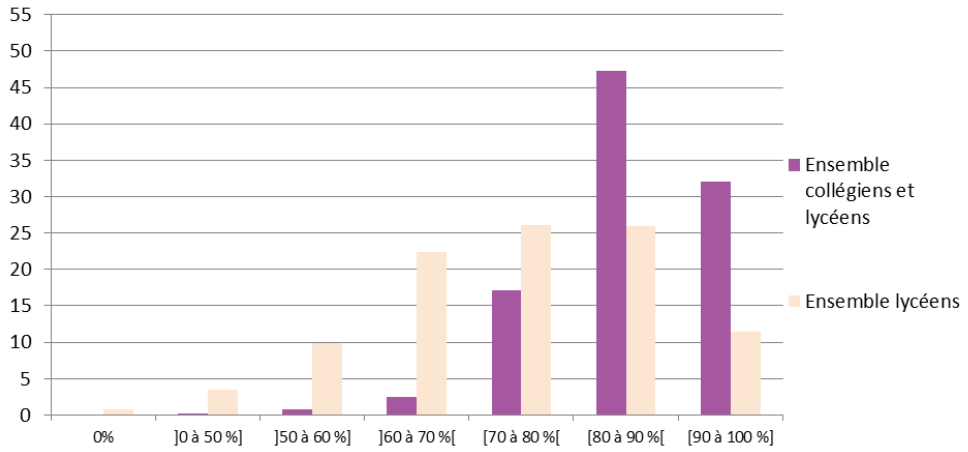


Lecture : dans le zonage en sortie d'Anabel, 6,5 % des aires Collèges ont un taux de stabilité nul, c'est-à-dire qu'aucun élève résident n'y est scolarisé. Ces cas n'existent plus après toilettage.
Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

Pour ce qui concerne les aires Collèges, 1,5 % des aires ont un taux de stabilité inférieur à 70 %. Parmi elles, les valeurs minimales sont atteintes pour les Îles-du-Ponant dans le Morbihan (22,9 %), et pour l'aire du Donjon dans l'Allier (46,9 %). Cette dernière est également concernée partiellement par la sectorisation inter-académique (Saône-et-Loire, académie de Dijon). Les autres aires Collèges ont des taux de stabilité supérieurs à 54 %.

Enfin, dans le cas du zonage Parcours, le taux de stabilité est plus faible lorsqu'on se restreint aux seuls élèves scolarisés en lycée[9], puisque les lycées professionnels ne sont pas sectorisés, et aussi parce que le zonage n'a pas été calculé spécifiquement sur cette population (figure 12).

FIGURE 12 • Taux de stabilité des aires Parcours sur l'ensemble du champ et sur les lycéens, rentrée 2020, en %



Lecture : dans le zonage Parcours, 0,7 % des aires ont un taux de stabilité nul pour les lycéens, c'est-à-dire qu'aucun lycéen résident n'y est scolarisé.
Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

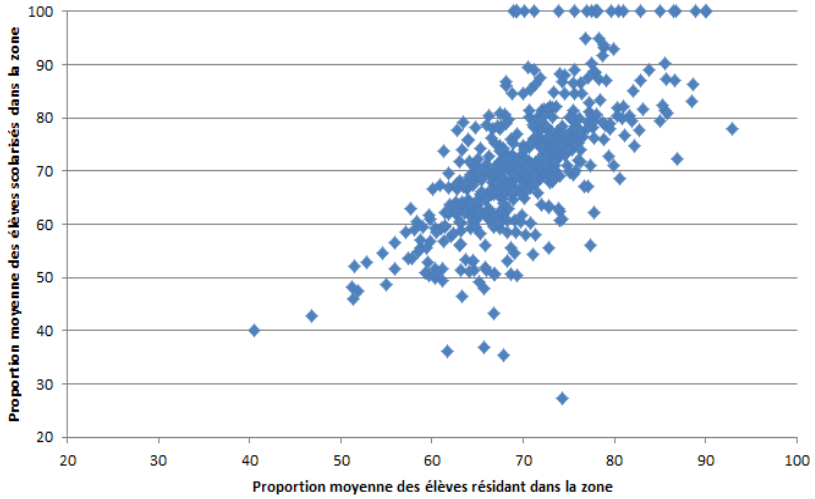
[9] Il s'agit des élèves scolarisés dans un établissement « lycée » et non en formation de lycée, de façon cohérente avec la logique de construction du zonage.

Comme attendu puisqu'elles ne comportent pas de lycée, les 4 aires de l'île de Saint-Barthélemy, Bâgé-Dommartin, Pélussin et Condrieu déjà évoquées affichent un taux de stabilité à 0%. Pour la quasi-totalité des aires Parcours (96 %), la majorité des lycéens qui y résident y sont également scolarisés (les taux de stabilité sont supérieurs à 50 %).

En définitive, si les aires forment donc des ensembles en général cohérents, le fait que les taux de stabilité ne sont pas égaux à 100 % implique que des décalages existent entre les indicateurs calculés au lieu de résidence et de scolarisation. Ce décalage peut être marqué si on travaille sur des populations à plus faible taux de stabilité.

À titre d'exemple, la proportion de lycéens scolarisés en voie générale et technologique a été calculée à la maille Parcours, en considérant alternativement l'aire où réside l'élève et son aire de scolarisation (figure 13). Les lycéens constituent, comme on l'a vu, une population à moindre stabilité. Pour cette raison, les taux calculés au lieu de résidence et de scolarisation peuvent s'éloigner sensiblement (coefficient de corrélation = 0,688). Dans certains cas, il n'y a pas de formation professionnelle implantée dans l'aire, ce qui induit un taux de scolarisation en voie générale et technologique de 100 % parmi les élèves scolarisés dans l'aire. Pour autant, une part non négligeable des élèves peut se rendre dans l'aire voisine pour suivre une formation professionnelle.

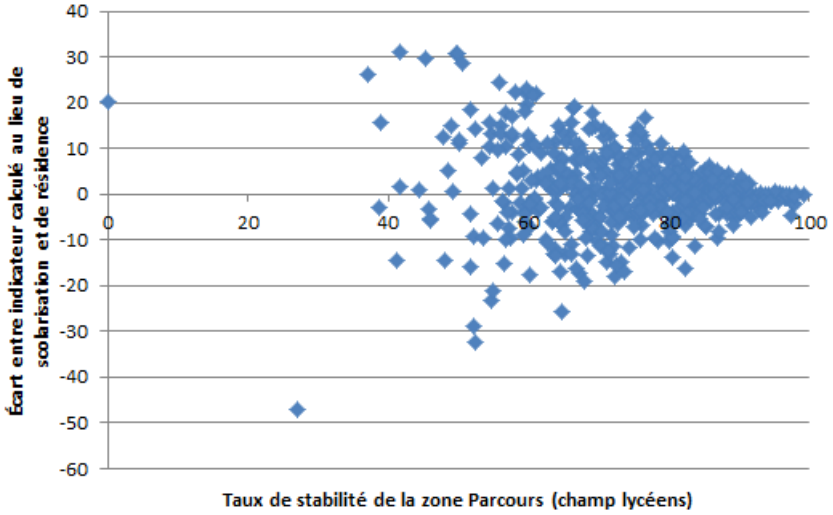
FIGURE 13 • Proportion de lycéens scolarisés en voie générale et technologique, rentrée 2020, en %



Lecture : le point ayant pour coordonnées (80,85) se lit ainsi : la proportion de lycéens, résidant dans cette aire, qui sont scolarisés en voie générale et technologique est de 80 % ; la proportion de lycéens, scolarisés dans cette aire, qui sont scolarisés en voie générale et technologique est de 85 %.
Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

De façon attendue, l'écart entre les indicateurs est corrélé au taux de stabilité précédemment calculé sur les lycéens (figure 14). Un faible taux de stabilité dans l'aire doit donc alerter sur le risque que les indicateurs calculés au lieu de résidence prennent des valeurs éloignées de celles calculées au lieu de scolarisation.

FIGURE 14 • Taux de stabilité des aires Parcours et écart entre indicateur (proportion de lycéens en voie générale et technologique) calculé aux lieux de résidence et de scolarisation, rentrée 2020, en % et points de %



Lecture : le point ayant pour coordonnées (80,10) se lit ainsi : pour cette aire, la proportion d'élèves en lycée, qui résident dans leur aire de scolarisation (taux de stabilité) est de 80 %. L'écart entre le taux de scolarisation en voie GT, parmi les lycéens, calculé au lieu de scolarisation et de résidence est de 10 points.
Champ : France métropolitaine et DROM, second degré public et privé.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

V.3 Étude de la stabilité des aires au cours du temps

Enfin, le fait que le zonage est calculé une année donnée (pour la rentrée 2020) pose la question de la stabilité au cours du temps. D'une année sur l'autre, les flux d'élèves entre communes peuvent être modifiés, du fait de variations dans l'offre scolaire (fermeture ou ouverture d'établissements ou de formations), de la répartition des familles entre les communes et de leurs choix d'établissement. Ces modifications peuvent également être simplement le fruit des évolutions démographiques dans les communes, notamment pour des communes de petite taille, où les flux, calculés sur un nombre réduit d'élèves, peuvent faire passer la commune « juste en dessous » ou « juste au-dessus » du seuil de 10 % constituant le critère d'agrégation à une autre aire. Or, l'algorithme est construit de telle sorte que ces conséquences de ces franchissements de seuil, à une étape de l'arbre de classification, ont des conséquences sur les regroupements entre les autres communes, et in fine sur la constitution des aires.

Afin de donner quelques éléments d'analyse sur la stabilité au cours du temps, on propose ci-dessous un exercice de variante sur le zonage « Parcours » : il s'agit de comparer le zonage obtenu en 2020 avec celui obtenu avec des données similaires, correspondant aux années 2010 et 2019. Refaire l'exercice complet de construction du zonage, c'est-à-dire en incluant le toilettage par les académies pour chaque année, serait trop coûteux. C'est pourquoi cette comparaison est menée uniquement d'après les aires en sortie d'Anabel.

L'indice de Rand[10] (1971) est régulièrement utilisé pour comparer la cohérence de classifications entre elles. Il correspond à la proportion de couples de communes classées de façon cohérente aux deux dates dans les zonages. Pour que l'indicateur ait du sens et ne donne pas des résultats exagérément optimistes (deux communes à un bout et l'autre de la France ne seront jamais regroupées), on se restreint, dans ce qui suit, aux communes d'un même département[11]. Pour chaque couple de communes, nous repérons ceux qui sont classés de façon cohérente par les différents zonages, d'une année sur l'autre. Ainsi, lorsqu'on compare deux années, on considère que deux communes sont classées de façon cohérente si :

- elles appartiennent à une même aire à chacune des deux dates ;
- elles n'appartiennent pas à la même aire à chacune des deux dates.

[10] William M. Rand (1971), « Objective Criteria for the Evaluation of Clustering Methods », *Journal of the American Statistical Association*, vol. 66, n°366, p. 846-850.
[11] On a exclu de l'analyse Paris, les Bouches-du-Rhône et le Rhône, dont les résultats sont sensibles à la prise en compte des arrondissements de Paris, Marseille et Lyon, comme des entités séparées ou regroupés en communes entières.

L'appartenance des communes aux aires est définie à partir des fichiers en sortie d'Anabel. On ne calcule l'indice que sur les communes présentes aux trois dates, en excluant donc les communes qui ont fusionné avec d'autres sur la période.

88 % des communes sont classées de façon cohérente dans les zonages en 2010 et en 2020 (colonne « Ensemble des communes », tableau 3). Sur un intervalle d'une année (entre 2019 et 2020), et de façon logique, ce taux de cohérence est plus élevé encore, atteignant 95 %. La plupart des départements ont des taux de cohérence très élevés, puisque la moitié ont un taux de cohérence supérieur à 94 %.

TABLEAU 3 • Cohérence entre les zonages Parcours construits à partir d'Anabel en 2010, 2019 et 2020 (indices de Rand), en %

	Ensemble des communes			Aires > 2 communes			Pondération par les flux		
	2010-2019	2019-2020	2010-2020	2010-2019	2019-2020	2010-2020	2010-2019	2019-2020	2010-2020
Indice moyen	90	95	**88**	90	95	**88**	95	98	**95**
Médiane départements	94	98	**94**	95	98	**95**	96	99	**96**
Département avec la valeur minimale	33	38	**30**	34	38	**30**	81	84	**83**

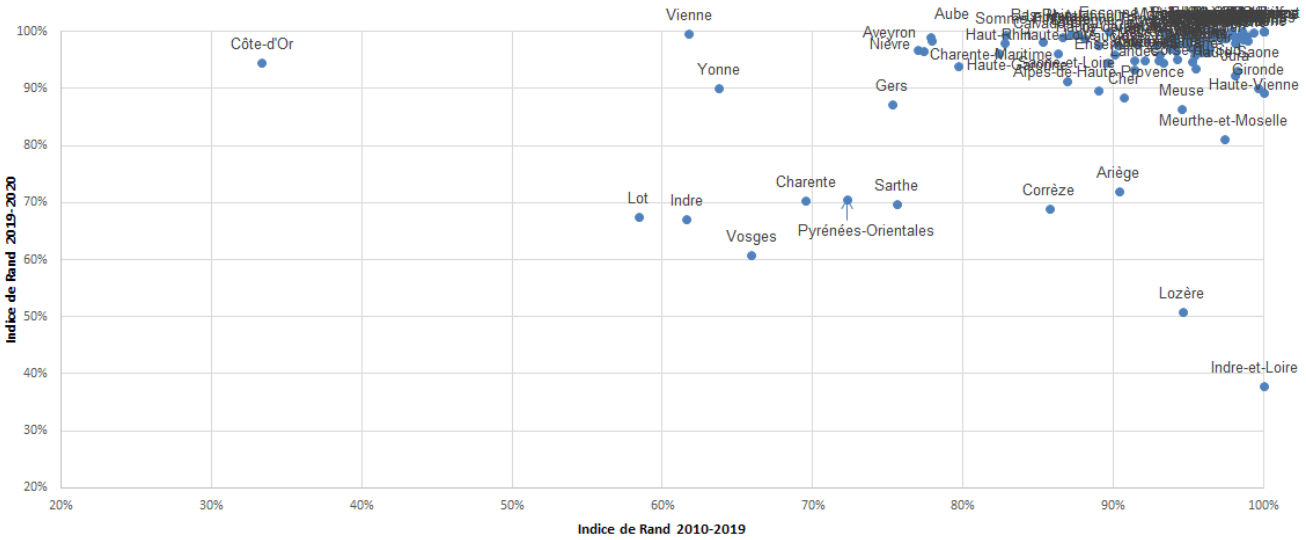
Champ : France métropolitaine + DROM hors Bouches-du-Rhône, Paris, Rhône, second degré public et privé sous contrat.
Lecture : pour 88 % des couples de communes d'un même département, le zonage de 2020 est cohérent avec le zonage de 2010. Dans la moitié des départements, ce taux de cohérence est supérieur à 94 %. Le département avec le taux de cohérence le plus faible a un taux de cohérence de 30 %.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

Comme on l'a constaté plus haut, l'un des défauts du zonage en sortie d'Anabel est d'inclure des aires « factices » composées d'une seule commune, isolée par défaut de contiguïté. Afin de limiter l'impact de ce biais, nous nous restreignons également aux aires ayant au moins deux communes (colonnes « Aires > 2 communes »). Les résultats sont quasiment inchangés.

Enfin, pour des raisons similaires, on peut aussi pondérer les indicateurs par les flux moyens (moyenne des flux d'élèves sur les dates 2010, 2019 et 2020). On s'intéresse alors aux élèves qui sont classés de façon cohérente dans la même aire, dans les deux zonages. Fort logiquement, le taux de cohérence ainsi défini augmente. En effet, les facteurs d'instabilité sont d'abord liés à des petits flux, dont le poids est plus faible dans cette variante de l'analyse. L'indice de Rand global passe à 95 % sur la période 2010-2020. De plus, tous les départements ont désormais un indice de Rand au moins égal à 83 %.

Dans certains départements, toutefois, des petites aires créent des instabilités. Sur la figure 15, on a croisé l'indice de Rand sur longue période (2010-2019) et sur courte période (2019-2020). Certains départements se distinguent par un faible indice, notamment inférieur à 80 % sur la période 2019-2020 et à 70 % sur 2010-2020. Dans certains cas, comme en Côte-d'Or, ces constats peuvent être mis en relation avec une forte variation du nombre d'aires (5 créées par Anabel en 2010 et 10 en 2019 et 2020) mais ce n'est pas systématique.

FIGURE 15 • Cohérence entre les zonages Parcours construits à partir d'Anabel en 2010, 2019 et 2020 par département (indices de Rand), en %



Champ : France métropolitaine + DROM hors Bouches-du-Rhône, Paris, Rhône, second degré public et privé sous contrat.
Lecture : pour 33 % des couples de communes de la Côte-d'Or, le zonage de 2019 est cohérent avec le zonage de 2010. Cette proportion s'élève à 95 % si on compare le zonage de 2019 et celui de 2020.
Source : DEPP, système d'information Scolarité ; Insee, outil Anabel, traitement GT DEPP-SSA Aires Éducation.

↘ VI. Déploiement et perspectives

VI.1 Livraison des zonages aux rectorats et à la DEPP

Le nouveau zonage a été livré aux SSA en septembre 2021, sous la forme de deux fichiers Excel. Le premier liste les aires Collèges et Parcours, ainsi que l'appartenance des communes aux aires[12]. Pour chaque zonage, un onglet 2020 et un onglet 2021 ont été proposés. Les onglets 2020 correspondent à l'appartenance des communes aux aires directement issue de l'élaboration du zonage. Les onglets 2021 tiennent compte des fusions, rétablissements de communes, changements de code Insee et de noms de communes intervenus entre 2020 et 2021. Enfin, un onglet spécifique à Paris et au zonage Collèges définit la composition des neuf aires sur Paris[13], qui ont été obtenues en appliquant la même méthodologie aux arrondissements de Paris (au lieu de laisser la commune entière comme une seule aire). Cette variante est mise à disposition pour des besoins spécifiques, par exemple pour l'Île-de-France. Dans le cas général, toutefois, il est recommandé de s'appuyer sur la composition communale définie sur les communes entières, par souci d'homogénéité. Un second fichier présente les taux de stabilité pour chaque aire.

VI.2 Présentations des zonages

Le déploiement du nouveau zonage dans les académies s'est accompagné d'une communication de la DEPP à l'occasion des réunions nationales périodiques des recteurs et secrétaires généraux d'académies. Une fiche rédigée par le groupe de travail a ainsi été diffusée de manière à expliciter la démarche, présenter les zonages sous forme de cartes, donner de premières indications sur leur taille et leur stabilité et indiquer de premières expérimentations d'utilisation en cours. Cette communication a été conçue de manière à susciter l'intérêt des responsables académiques tout en laissant aux services statistiques académiques la main pour saisir localement toute occasion de promouvoir l'utilisation des nouveaux zonages, notamment auprès des échelons départementaux et des corps d'inspection.

Par ailleurs, en vue d'assurer une large diffusion de ce nouvel outil d'analyse territoriale dans les études, le groupe de travail a proposé qu'une présentation des aires d'éducation soit programmée dans le cadre des « ateliers de la DEPP », format destiné aux personnels de la DEPP, des SSA et du SIES. Ce temps d'échanges entre statisticiens est destiné tant à présenter la méthodologie de construction du zonage que de premières investigations et analyses menées à cette maille d'observation.

Enfin, le groupe de travail a soumis une proposition de présentation, et de contribution écrite aux Journées de Méthodologie Statistique (JMS), prévues en mars 2022. Les JMS sont organisées par l'Insee et ont pour objectif de « faire connaître des travaux novateurs et d'en assurer la diffusion, mais elles ont aussi une ambition pédagogique : un de leurs objectifs est que les auditeurs trouvent matière dans leurs propres travaux à mettre en œuvre les techniques exposées » (cf. site jms-insee.fr). Le document proposé aux JMS doit aider à mieux faire connaître notre démarche de construction d'un zonage national adapté au système éducatif, mobilisant l'expertise nationale et locale. Le public-cible inclut en particulier les services statistiques ministériels qui partagent les mêmes problématiques (le travail du service statistique en charge du logement, déjà cité, y a été présenté en 2018) et les chercheurs intéressés par l'analyse des disparités territoriales.

[12] Cette table a été construite sur le modèle des tables d'appartenance des communes aux zonages de l'Insee usuels.
[13] Dans le cas de Lyon et Marseille, la variante prenant en compte les arrondissements n'est pas proposée, car l'application de la méthodologie donne des résultats identiques à ceux issus du traitement des communes entières, comme indiqué précédemment. C'est aussi le cas du zonage Parcours pour Paris. Pour des raisons pratiques (facilité d'utilisation des zonages avec les outils d'information géographique usuels), c'est la commune entière qui a été retenue dans le fichier Excel d'appartenance aux aires.

VI.3 Constitution d'un groupe de travail de valorisation des zonages

Les aires d'éducation permettent au réseau DEPP-SSA de disposer à partir de septembre 2021 d'un zonage construit « ad hoc », à partir de données propres à l'Éducation nationale. Basé sur la mobilité des élèves entre leur domicile et leur lieu de scolarisation, il peut tout à fait constituer un niveau de pilotage fin pour l'Éducation nationale et en particulier pour les académies et départements. Pour cela, il est essentiel que chaque acteur s'en empare à son niveau. Outre une communication large relative à l'existence de ce zonage, s'engager dans un travail de valorisation semble une condition sine qua non à son utilisation. Ce zonage ayant été construit par les académies (SSA) et la DEPP, à un niveau national, sa valorisation ne peut se faire que sur le même modèle : c'est-à-dire de manière collaborative, avec le concours de SSA et de la DEPP. Pour ces raisons, les membres du groupe de travail « aires d'éducation » ont souhaité la constitution d'un nouveau groupe qui, dans le prolongement de leurs propres travaux, aura pour vocation d'explorer et mettre en œuvre les valorisations possibles.

La création de ce nouveau GT a été actée avec la DEPP en novembre 2021, et sa première réunion a eu lieu en décembre 2021.

Parmi les pistes mises en avant lors de la réunion préalable, plusieurs ont retenu l'attention :

- la création d'une ou plusieurs typologie(s) permettant de caractériser les aires d'éducation. Ces typologies pourraient s'appuyer sur des données exogènes (contexte socioéconomique, démographiques…) ou endogènes (l'offre de formation par exemple) ;
- l'intégration dans les outils de la DEPP du zonage en aires d'éducation comme niveau de représentation et référence ;
- l'utilisation du zonage comme périmètre de description / représentation d'indicateurs territoriaux, pour les besoins du pilotage national et local.

VI.4 Perspectives de mise à jour des zonages

Lors des échanges avec les académies, le délai de mise à jour des zonages a également été abordé. L'arbitrage sur ce délai doit prendre en compte l'intérêt d'avoir des données fraîches, compte tenu de la modification des flux scolaires (fermeture ou ouverture d'établissement, changements de la sectorisation, de l'offre de formation des établissements, etc.) mais aussi, de l'autre côté, le coût d'actualisation (y compris d'un nouveau toilettage) et enfin, l'intérêt d'avoir un zonage stable au cours du temps, pour calculer des évolutions.

À titre de comparaison, l'Insee est actuellement sur un rythme décennal pour l'actualisation de la plupart de ses zonages d'études (unités urbaines, aires urbaines devenues aires d'influence des villes, zones d'emploi).

Compte tenu des modifications rapides de l'offre scolaire, et en tenant compte du coût associé à l'opération de révision du zonage, un intervalle de cinq ans entre ces actualisations a été proposé. Ce délai pourra être modulé en fonction des utilisations et des besoins. D'ici-là, la composition du zonage devra être actualisée chaque année a minima en tenant compte des révisions de la géographie communale, se traduisant dans des modifications des codes communes, telles que des fusions ou scissions de communes.

Remerciements

Nous tenons à remercier les personnes qui nous ont offert leur collaboration sur ce projet, en particulier l'ensemble des collègues des services statistiques académiques, Cécile Duquet-Métayer, Daphné Beauregard-Côté (DEPP C2SIG) et Julien Jamme (ENSAI) pour l'appui cartographique, et Fabienne Rosenwald pour la confiance qu'elle nous a témoignée.

↘ Références bibliographiques

Aliaga C. (2015), « Les zonages d'étude de l'Insee – Une histoire des zonages supracommunaux définis à des fins statistiques », *Insee Méthodes*, n°129. https://www.insee.fr/fr/information/2571258

Azéma A., Mathiot P. (2019), *Mission Territoires et réussite*, Rapport remis au ministre de l'Éducation nationale et de la Jeunesse. https://www.education.gouv.fr/rapport-mission-territoires-et-reussite-7577

Boudesseul G., Caro P., Grelet Y., Vivent (2014), *Atlas des risques sociaux d'échec scolaire : l'exemple du décrochage*, Céreq, DEPP. https://www.education.gouv.fr/atlas-des-risques-sociaux-d-echec-scolaire-l-exemple-du-decrochage-france-metropolitaine-et-dom-3989

Boudesseul G., Caro P., Grelet Y., Minassian L., Monso O., Vivent C. (2016), *Atlas des risques sociaux d'échec scolaire. L'exemple du décrochage, France métropolitaine et DOM*, Céreq, DEPP. https://www.education.gouv.fr/atlas-des-risques-sociaux-d-echec-scolaire-l-exemple-du-decrochage-france-metropolitaine-et-dom-3989

Colin S., Roussez V. (2018), « Projet MATELO : une maille géographique pour l'analyse du logement », *Actes des XIIIes Journées de méthodologie statistique de l'Insee*. http://www.jms-insee.fr/2018/S08_5_RESUME_COLIN_JMS2018.pdf

De Bellefon M.-P., Eusebio P. Forest J., Pégaz-Banc, Warnod R. (2020), « En France, neuf personnes sur dix vivent dans l'aire d'attraction d'une ville », *Insee Focus*, n°211. https://www.insee.fr/fr/statistiques/4806694

DEPP (2021), *Géographie de l'école*, treizième édition, DEPP. https://www.education.gouv.fr/geographie-de-l-ecole-323657

Duquet-Métayer C., Monso O. (2019), Une typologie des communes pour décrire le système éducatif », *Note d'Information*, n° 19.35, DEPP. https://www.education.gouv.fr/une-typologie-des-communes-pour-decrire-le-systeme-educatif-6524

Laval N., Vilain É. (2014), « Les systèmes territoriaux en Nord-Pas-de-Calais : une région façonnée par les déplacements », *Pages de Profils*, Insee, n°155. https://www.insee.fr/fr/statistiques/1292116

Levy D., Dubois M., Lefebvre M. (2020), « L'orientation économique des zones d'emploi : entre spécialisation et diversification des économies locales », *Insee Première*, n°1814. https://www.insee.fr/fr/statistiques/4653582?pk_campaign=avis-parution

Grelet, Y. (2004), « La reproduction sociale s'inscrit dans le territoire », *Formation Emploi*, n° 87, 2004, p. 79-98, Céreq. https://www.persee.fr/doc/forem_0759-6340_2004_num_87_1_1672

Miconnet N. (2016), « Parcours des élèves en difficulté scolaire », *Éducation & formations*, p. 53-75, DEPP. https://hal.archives-ouvertes.fr/hal-01350120/document

Murat F. (2021), « Les inégalités territoriales en matière d'éducation. Les écarts entre communes en termes de milieu social et de réussite au diplôme national du brevet », *Éducation & formations*, p. 67-102, DEPP. https://halshs.archives-ouvertes.fr/halshs-03347659/

Rocher T. (2016), « Construction d'un indice de position sociale des élèves », *Éducation & formations*, n°90, p. 5-27, DEPP. https://hal.archives-ouvertes.fr/hal-01350095

Ropers C., Fouchard C., Kerdommarec L., Rodrigues A. (2013), « Lycéens et Lycéens en Pays de la Loire : état des lieux et perspectives à l'horizon 2025, *Dossier Insee Pays de la Loire*, n°48 https://www.insee.fr/fr/statistiques/1294920

Le réseau de la statistique sur l'éducation

Service statistique ministériel, la direction de l'évaluation, de la prospective et de la performance (DEPP) dispose de relais dans les échelons administratifs déconcentrés que sont les académies, les régions académiques et les vice-rectorats. Ces relais, au nombre de 33, sont les services statistiques en académies (SSA), dont certains sont organisés en structures régionales ou inter-académiques, afin de fournir une information consolidée à l'échelle des régions. Ce vaste réseau contribue à la qualité de la statistique publique de l'éducation à tous les niveaux : collecte, analyse et diffusion. Services à haute technicité, les SSA sont une ressource précieuse et reconnue pour fournir aux autorités académiques l'information destinée à l'aide à la décision et au pilotage.

Réseau DEPP – Services statistiques en académies
Ministère de l'Éducation nationale, de la Jeunesse et des Sports
Contact : Reseau.depp.ssa@education.gouv.fr

depp | SSA