

Mesure de l'artificialisation à l'aide des Fichiers fonciers Méthodologie



Crédit photo : © Arnaud Bouissou — Terra © Denis Gabbardo - Terra

Partenaires de l'étude

*Direction Générale de
l'Aménagement, du
Logement et de la Nature*

*Ministère de l'Agriculture
et de l'Alimentation*

*Commissariat général
au développement
durable*



MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE
ET SOLIDAIRE



MINISTÈRE
DE L'AGRICULTURE
ET DE
L'ALIMENTATION



MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE
ET SOLIDAIRE

Méthodologie de mesure de l'artificialisation à l'aide des Fichiers fonciers

Historique des versions du document

Version	Date	Commentaire
V1	21 juin 2019	Version de base

Affaire suivie par

Martin Bocquet - Département DATHa – Groupe Stratégies foncières et Expertises Géomatiques
<i>Tél. : 03 20 49 62 71</i>
<i>Courriel : martin.bocquet@cerema.fr</i>
Site de Lille : Cerema Nord-Picardie

Références

n° d'affaire :

maître d'ouvrage : DGALN

Rapport	Nom	Date	Visa
Établi par	Martin Bocquet	12 juin 2019	
Avec la participation de	Antoine Herman et Perrine Rutkowski		
Contrôlé par	Frédéric Lasseron	18 juin 2019	
Validé par			

Résumé de l'étude :

SOMMAIRE

Objet et contenu du rapport.....	6
Contexte.....	6
La méthode d'observation de l'artificialisation à l'aide des Fichiers fonciers.....	7
Méthode de calcul de la consommation d'espaces à partir des Fichiers fonciers.....	10
Données d'entrée.....	10
Définitions.....	11
Objectif de la méthode.....	11
Étape A : calcul du flux d'artificialisation.....	12
Problématique.....	12
Étape A-0 : traitement des parcelles problématiques.....	12
Étape A-1 : détermination des différences entre millésimes (algorithme Kirouge).....	13
Étape A-2 : transferts entre catégories : (algorithme Kaver).....	14
Données de sortie.....	14
Étape B : calcul du flux d'usage.....	15
Objectif de cette partie.....	15
Étape B-1 : élimination du « non affecté » (algorithme Kejaune).....	16
Étape B-2 : traitement des flux de changement d'usage (algorithme Korange).....	17
Étape C : Post-traitements.....	19
Point d'étape.....	19
Étape C-1 : Création d'une table multi-millésimes.....	19
Étape C-2 : Suppression des îlots « rebond » (algorithme Kanbleu).....	19
Étape D : Caractéristiques du produit final.....	21
Étape D-1 : Fusion à l'échelle de rendu.....	21
Description du produit final.....	21
Quelles variables utiliser ?.....	22

Annexe 1 : Logigramme de la méthode générale.....	23
Annexe 2 : établissement d'une table multi-millésime des Fichiers fonciers.....	24
Préalables.....	24
Objectif de l'étude.....	24
La première étude multi-millésime.....	24
Rappel : la parcelle.....	24
Rappel : les modifications de communes.....	24
Les remembrements.....	25
Les bases de données utilisées.....	26
Les Fichiers fonciers.....	26
Les DFI (documents de filiation informatisés).....	26
Notation dans le rapport.....	27
Traitement général.....	27
Résumé du traitement.....	27
Étape 0 : création des tables parcellaires.....	29
Traitement des bi-millésimes.....	30
Étape 1 : traitement des cas simples.....	30
Étape 2 : traitement des évolutions simples.....	30
Étape 3 : traitement des évolutions complexes.....	31
Étape 4 : élagage des arbres.....	33
Étape 5 : traitement des sauts de millésimes.....	33
Étape 6 : fusion des éléments restants.....	34
Contributions de chacune des étapes.....	35
Réalisation de la table multi-millésime.....	36
Réalisation de la table finale.....	36
Méthode pour créer la table finale.....	36
Signification thématique de l'îlot.....	37
Conclusion et intérêt de l'étude.....	38
Une table en cours de création.....	38
Ajout des champs thématiques.....	38
Annexe 2.1 : algorithme de recherche de composantes connexes d'un graphe. .	39
Lien avec le problème.....	39
Implémentation classique de l'algorithme.....	40
Implémentation actuelle de l'algorithme.....	42
Annexe 3 : Affectation de l'usage aux parcelles des Fichiers fonciers.....	44

Contexte et présentation de l'étude.....	44
Présentation de l'étude.....	44
Contexte de l'étude.....	44
Principe de base et présentation du rapport.....	44
Étape 1 : Affectation initiale des parcelles.....	46
Bases de données utilisées.....	46
Liste des parcelles à traiter.....	46
Affectation des parcelles.....	46
Étape 1.0 : préparation de la table.....	46
Étape 1.1 : détermination des cas.....	47
Étape 1.2 : calcul des scores.....	47
Étape 1.3-cas1 : gestion des parcelles avec des locaux industriels (cas 1).....	48
Étape 1.3-cas2 : gestion des parcelles avec un commerce avec boutique (cas 2).....	50
Étape 1.3-cas3 : gestion des autres cas ambigus (cas 3).....	50
Correction des affectations pour le public.....	52
Étape 1.4 :Croisement entre la parcelle et la BD Topo.....	52
Affectation des parcelles artificialisées non bâties.....	54
Point d'étape.....	54
Problématique : les parcelles artificialisées sans local.....	55
Affectation des polygones « non affecté » par voisinage.....	57
Affectation plus lointaine.....	60
Résultat et conclusion.....	61
Résultat final.....	61
Contribution de chaque étape.....	62
Résultats en termes d'affectation.....	62
Conclusion.....	62
Annexe 4 : description des algorithmes Kaver et Korange.....	63
Annexe 4.1 : description de l'algorithme Kaver.....	63
Données de base et hypothèses.....	63
Conséquences et mise en place de l'algorithme.....	65
Annexe 4.2 : description de l'algorithme Korange.....	66
Principe de l'algorithme Korange.....	66
Gestion des cas simples.....	66
La gestion des cas complexes : algorithme Kaunoir.....	67
Annexe 5 : structure et contenu des données.....	70
Contenu des données.....	70
Deux produits en diffusion.....	70
Données et nom des champs.....	71

Objet et contenu du rapport

Contexte

« *Tous les 10 ans, l'équivalent d'un département français disparaît sous le coup de l'urbanisation* ». Cette formule-choc marque depuis 2010 les communications autour de la nécessité de réduire l'artificialisation des terres.

Cependant, les mesures du phénomène opérées au niveau national sont ne permettent pas de corroborer cette affirmation¹, et présentent des résultats différents. En effet, s'il existe un consensus politique pour réduire l'artificialisation, le phénomène est fondamentalement mal connu, faute d'une définition précise et d'une mesure associée cohérente.

Fort de ce constat, le plan biodiversité adopté le 4 juillet 2018 contient, dans son action 7, l'engagement de « *[publier], tous les ans, un état des lieux de la consommation d'espaces et [mettre] à la disposition des territoires et des citoyens des données transparentes et comparables à toutes les échelles territoriales.* ». Dans ce contexte, le ministère de la transition écologique et solidaire a missionné le Cerema, l'IGN et l'IRSTEA pour produire ces données.

Des données annuelles issues des Fichiers fonciers

Une partie de cette mission consiste à produire des données annuelles de flux de consommation d'espaces, à une maille communale. Cela comporte ainsi :

- Un recensement des sources de données potentiellement utilisables.
- Un rapport sur les questions de définition et sur les apports et limites des Fichiers fonciers pour l'observation de l'artificialisation.
- Ce présent rapport méthodologique.
- Les données elles-mêmes, accompagnées d'outils de visualisation en ligne.

Ces éléments sont présents sur le portail internet <https://artificialisation.biodiversitetousvivants.fr/>.

Objet du présent rapport

Ce rapport explicite le processus de création d'indicateurs de consommation d'espaces à partir des Fichiers fonciers restituée à une maille communale et calculée de manière annuelle en flux, en en détaillant les limites d'utilisation.

Il est admis, en préalable, que le lecteur est au fait de la base de données « Fichiers fonciers », de sa structure, ainsi que des principaux avantages et limites de la base. Si ce n'est pas le cas, il est conseillé de lire le premier rapport traitant de la source de données² et/ou de consulter le site internet consacré aux données foncières³.

Ce rapport est ainsi destiné à un public de techniciens désireux de connaître la manière dont a été constituée la donnée. Il a donc été fait ici le choix de présenter de la manière la plus exhaustive possible les traitements réalisés.

1 Panorama de la quantification de l'évolution nationale des surfaces agricoles https://agriculture.gouv.fr/sites/minagri/files/documents/pdf/140514-ONCEA_rapport_cle0f3a94.pdf

2 Mesure de l'artificialisation à l'aide des Fichiers fonciers : définition, limites et comparaison avec d'autres sources, présent sur <https://artificialisation.biodiversitetousvivants.fr/>

3 <https://datafoncier.cerema.fr>

Calculer le « combien » pour mieux évaluer le « Pourquoi »

L'exercice mené ici permet d'obtenir des données comparables à toutes les échelles de territoire. Il s'agit de données solides permettant aux territoires de se saisir de ces chiffres. L'apport de cette étude est donc de répondre à la question du « combien ».

Cependant, il s'agit ici d'une méthode nationale, qui se positionne en complément de méthodes d'observation locales parfois mieux adaptées aux spécificités des territoires.

En outre, il faut rappeler que le nombre d'hectares artificialisés n'est ici qu'une partie du problème. S'il s'agit d'atteindre le zéro artificialisation, il est nécessaire de se pencher sur les causes de cette artificialisation, et sur les leviers permettant de la réduire. En cela, des analyses nationales et locales plus poussées, notamment dans le cadre d'observatoires fonciers déjà constitués seront nécessaires pour comprendre le « pourquoi ».

La méthode d'observation de l'artificialisation à l'aide des Fichiers fonciers

Ce rapport retrace le processus de traitement des données Fichiers fonciers. Étape par étape, il sera fait le suivi de la transformation de la donnée de base en des chiffres communaux annuels sur la période 2009-2017. De manière condensée, la méthode est la suivante.

- **1** – Dans un premier temps, pour tous les millésimes, on classe chaque parcelle des Fichiers fonciers, selon son caractère artificialisé ou non. Ensuite, si elle est artificialisée, il sera précisé son usage (habitat, activité ou mixte).
- **2** – Une fois cette action réalisée, l'objectif est d'arriver à créer un historique des parcelles. En d'autres termes, il faut arriver à suivre, sur l'intégralité des millésimes, ce que deviennent les parcelles. Ainsi, si une parcelle A se divise, nous devons pouvoir suivre chacune de ses parties, et savoir que ces parties sont issues de A. Dans ce cadre, nous allons travailler à l'îlot, c'est-à-dire un agrégat de parcelle(s) stable sur l'intégralité des millésimes.
- **3** – À partir de ces deux éléments, nous disposons d'une donnée contenant la filiation des parcelles ainsi que leur usage. À partir de ces éléments, il est possible de calculer les flux d'artificialisation.

Le schéma à la page suivante retrace ce processus.

Principe de traitement

Il s'agit donc de travailler à une maille fine (parcelle ou ensemble de quelques parcelles), en mobilisant des techniques de calcul statistique, et éventuellement quelques méthodes géomatiques. Dans un deuxième temps, les données sont agrégées à une maille communale.

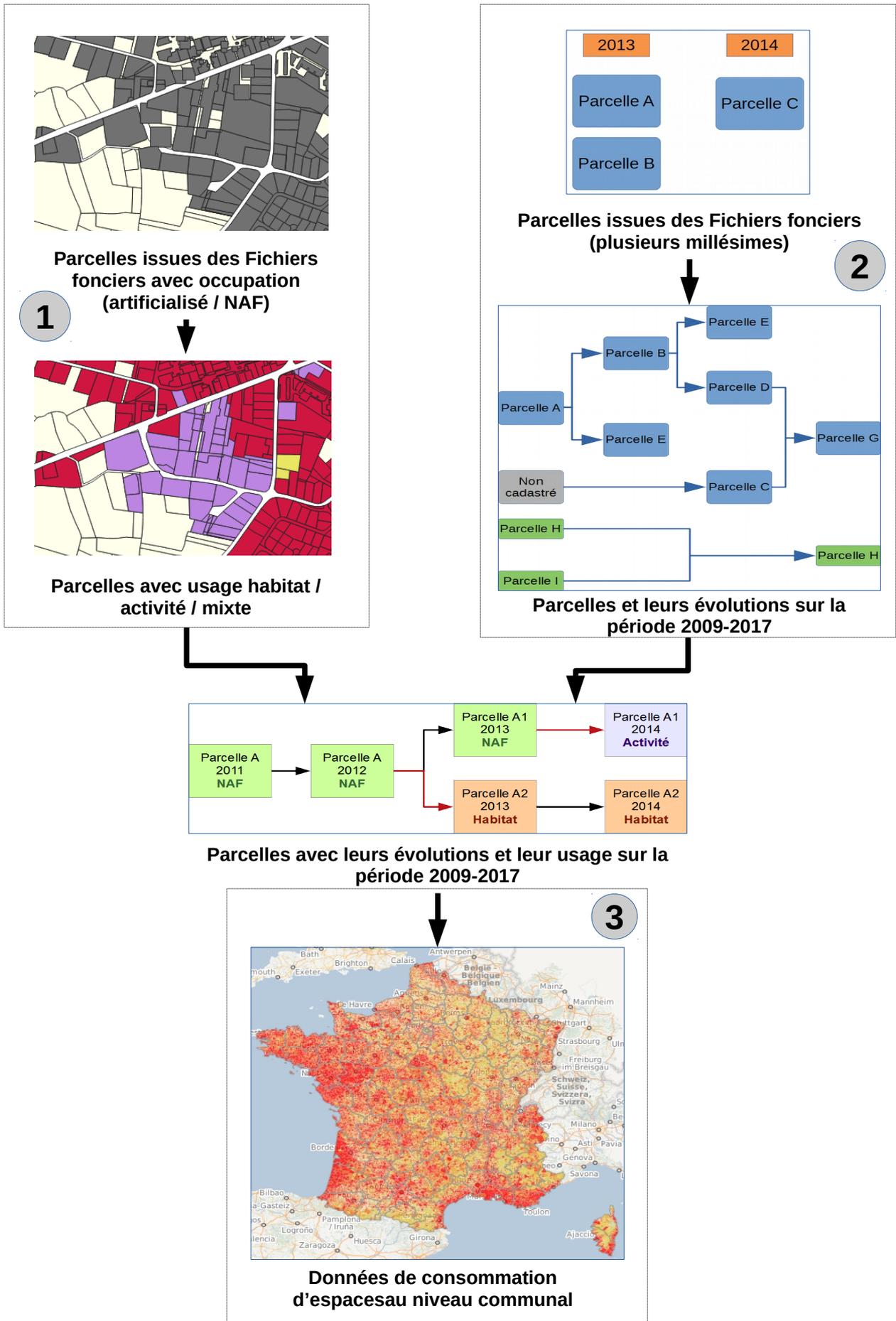
Les données sont travaillées en ne considérant que sur les entités ayant évolué. Il ne s'agit donc pas ici de travailler en stock (« 1 000 ha de cette commune sont artificialisées en 2017 »), mais bien en flux (« 10 ha de cette commune ont été artificialisées entre 2016 et 2017 »)

Une nouvelle méthode permettant de travailler à une échelle fine

La méthodologie utilisée est ici nouvelle, et remplace les précédents chiffres et données issus des Fichiers fonciers¹. Il s'agit ici d'une méthodologie plus fine, travaillant à une maille plus resserée (l'îlot / parcelle et non la commune).

Outre des données plus fines, capables de différencier habitat et activité, il s'agit aussi d'une méthodologie permettant des développements ultérieurs (croisement avec d'autres bases de données exogènes), ce qui améliorera la connaissance du phénomène.

¹ Et en particulier les rapports annuels sur « la consommation d'espaces et ses déterminants d'après les Fichiers fonciers de la DGFiP » : <https://datafoncier.cerema.fr/usages/consommation-des-espaces-et-occupation-des-sols/la-consommation-espaces-et-ses-determinants>



Résumé de la méthode d'évaluation de la consommation d'espaces à partir des Fichiers fonciers. La partie en haut à gauche concerne l'usage, en haut à droite le multi-millésime et le centre le traitement

Structuration du rapport

Le rapport contient la méthode de calcul de la consommation d'espaces. Elle s'appuie sur deux traitements préalables explicités en annexe, à savoir :

- l'affectation de l'usage de la parcelle (c'est-à-dire habitat, activité et mixte) à la parcelle, (annexe 3),
- l'établissement d'une table multi-millésime (annexe 2).

Le corps du rapport se concentre sur la transformation de données multi-millésime contenant l'usage de la parcelle. A partir de ces données, il va s'agir de :

- **A** - calculer le flux d'artificialisation, c'est-à-dire les transformations entre artificialisé, NAF et non cadastré (Étape A),
- **B** - calculer le flux usage, à savoir les transformations, au sein de l'artificialisé, entre usage d'habitat, d'activité et mixte (Étape B),
- **C** - réaliser les post-traitements (Étape C),
- **D** - produire les données finales (Étape D).

L'ensemble de ces traitements est réalisé à travers un algorithme nommé Krainbow, développé dans le cadre de cette étude. Le déroulé de l'algorithme Krainbow est présenté en Annexe 1.

Méthode de calcul de la consommation d'espaces à partir des Fichiers fonciers

Données d'entrée

Les deux étapes explicitées en annexe nous ont permis d'obtenir des données bi-millésime¹ contenant les surfaces liées aux différents usages (activité, habitat, mixte). Pour chaque îlot², nous aurons donc les informations suivantes pour le millésime N et le millésime N+1 :

- surface totale cadastrée
- surface artificialisée.

La surface artificialisée est elle-même divisée entre les éléments suivants :

- artificialisé à usage d'habitat,
- artificialisé à usage d'habitat extrapolé³,
- artificialisé à usage d'activité,
- artificialisé à usage d'activité extrapolé,
- artificialisé à usage mixte,
- artificialisé à usage mixte extrapolé,
- artificialisé non affecté

îlot	Surface parcelle 2011	Surface artificialisée 2011	Surface NAF 2011	Surf Act direct 2011	Surf Act extrapolé 2011	Surf Habitat 2011	...	Surface parcelle 2012	...
îlot1									
îlot2									
îlot3									

1 Deux millésimes des Fichiers fonciers, soient 2009-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016 et 2016-2017

2 Un îlot est un ensemble d'une ou plusieurs parcelles stables dans le temps. L'îlot est défini précisément dans l'annexe 2 (traitement multi-millésime).

3 L'usage « direct » est réservé aux parcelles artificialisées possédant un local. L'usage « extrapolé » est réservé aux parcelles artificialisées sans local : l'usage est assigné à partir de celui de ses voisins. Ce traitement est défini précisément dans l'annexe 3 (assignation de l'usage).

Artificialisé extrapolé (rappel)

L'usage de la parcelle peut être défini de deux manières. En premier lieu, il s'agit d'une affectation directe, selon les locaux présents sur la parcelle selon les Fichiers fonciers. Si cela n'est pas possible, l'affectation est définie par rapport à celle de ses voisins.

Dans ce cadre, l'affectation directe est considérée comme plus fiable que l'affectation extrapolée.

Artificialisé non affecté (rappel)

Les traitements précédents ne permettent pas d'affecter la totalité des parcelles. Pour certaines parcelles artificialisées, il est impossible d'en déterminer l'usage, même par extrapolation. Ces parcelles sont donc considérées comme « non affectées ».

Définitions

Dans la suite du rapport, nous allons utiliser deux notions :

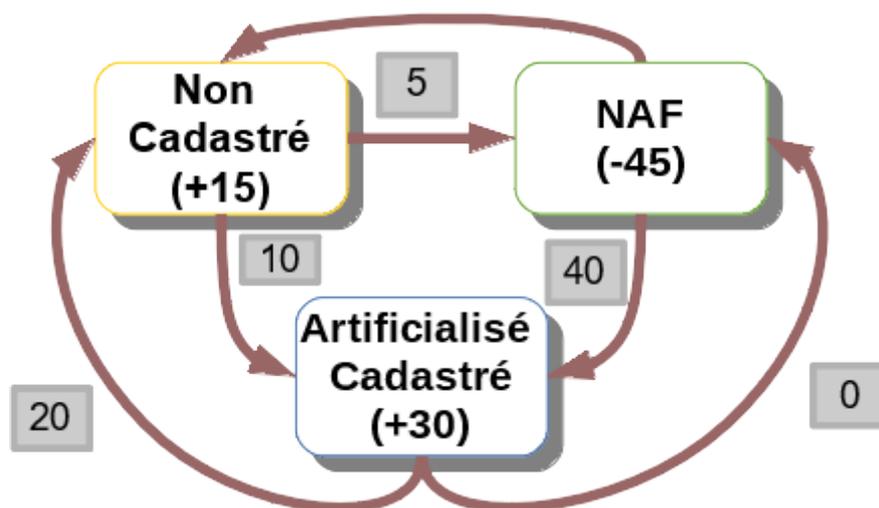
- les **deltas** sont les différences de stock entre deux millésimes. Ainsi, si la surface artificialisée vaut 10 ha en 2011 et 12 ha en 2012, le delta artificialisé 2011-2012 vaudra $12 - 10 = 2$ ha.
- Les **flux** sont les transferts entre deux grandeurs et entre deux millésimes. Ainsi le flux NAF – Arti 2011-2012 est égal au nombre d'hectares qui étaient en NAF en 2011 et en artificialisé en 2012.

Objectif de la méthode

L'objectif de ce traitement sera de transformer ces données, statiques, en flux. Ce traitement sera réalisé pour observer deux phénomènes.

Dans un premier temps, les flux entre artificialisé, NAF et non cadastré seront déterminés. Cette première sortie sera nommée « **flux d'artificialisation** ».

Dans un deuxième temps, au sein de l'artificialisé, les flux entre habitat, mixte et activité seront déterminés. Cette deuxième sortie sera nommée « **flux d'usage** ».



L'objectif sera donc de transformer les différences entre millésimes (entre parenthèses) en flux (flèches rouges), et ce pour chaque îlot.

Étape A : calcul du flux d'artificialisation

Problématique

Nous avons donc en entrée :

- la surface totale de l'îlot de l'année N,
- la surface totale de l'îlot de l'année N+1,
- la surface artificialisée de l'îlot de l'année N,
- la surface artificialisée de l'îlot de l'année N+1.

Il s'agit de définir, en sortie, 6 flux :

- Artificialisé vers NAF,
- NAF vers artificialisé,
- Artificialisé vers Non cadastré,
- Non cadastré vers Artificialisé,
- NAF vers Non cadastré,
- Non cadastré vers NAF.

Étape A-0 : traitement des parcelles problématiques

Les calculs d'artificialisation se heurtent à plusieurs problèmes :

- en 2015, les golfs ont été classés en tant que « artificialisés » par l'administration fiscale, et ce même s'ils existaient auparavant. Dans ce contexte, en prenant les données brutes, on constate une forte augmentation de l'artificialisation entre 2014 et 2015, cela étant dû à un nouveau classement fiscal et non d'un changement d'usage des sols.
- Les terrains militaires ont tendance à changer de classement cadastral entre deux millésimes, et ce sans modification d'usage du sol. Ces changements aléatoires nécessitent donc des retraitements.
- Sur certaines parcelles, la surface est mal renseignée : la surface totale des suf est ainsi différente de la surface totale de la parcelle¹.

Ces parcelles ont été ciblées pour deux raisons :

- elles concentrent les erreurs qui peuvent se trouver dans les Fichiers fonciers,
- des surfaces importantes sont en jeu. A titre d'illustration, si le camp de Canjuers (35 000 ha) changeait d'occupation des sols pour des raisons fiscales, cela reviendrait à doubler la consommation d'espaces au niveau national.

En outre, une vérification manuelle a été réalisée sur les parcelles dont l'artificialisation dépassait les 10 ha.

Traitement des golfs et des terrains militaires

Dans un premier temps, les unités foncières sur lesquelles un golf ou un terrain militaire était présent ont été repérées. Il s'agit donc des unités foncières sur lesquelles un point d'intérêt de type « Golf » ou « Terrain militaire »² était présent.

Tous les îlots dans lesquels apparaissent l'une de ces parcelles sont exclus des calculs d'artificialisation.

Traitement des surfaces mal renseignées

Les parcelles ont une surface totale, qui se divise en surface artificialisée et non artificialisée. Cependant, pour certaines d'entre elles, des problèmes de données conduisent à avoir une surface totale différente de la somme de ses composantes³.

Tous les îlots dans lesquels apparaissent l'une de ces parcelles sont exclus des calculs d'artificialisation.

Vérification manuelle

1 200 îlots présentant une artificialisation de 10 ha ont été vérifiés manuellement. Il s'agit en effet d'un petit nombre de parcelles, jouant beaucoup sur les résultats. Chacun de ces îlots a été étudié, en observant les photographies aériennes et en

¹ En d'autres termes, la surface du contenant est différente de la surface de ses parties, ce qui est clairement une erreur.

² Les golfs et terrains militaires sont déterminés à partir des points d'activités et d'intérêt (PAI) de la BD Topo produite de l'IGN.

³ En théorie, Surf arti + Surf NAF = Surface totale. Sur ces rares parcelles, ce n'est pas le cas.

faisant des recherches sur les éventuels projets de la zone.

En cas d'erreur de classement, les données liées aux îlots sont rectifiées.

Exclusion des remembrements

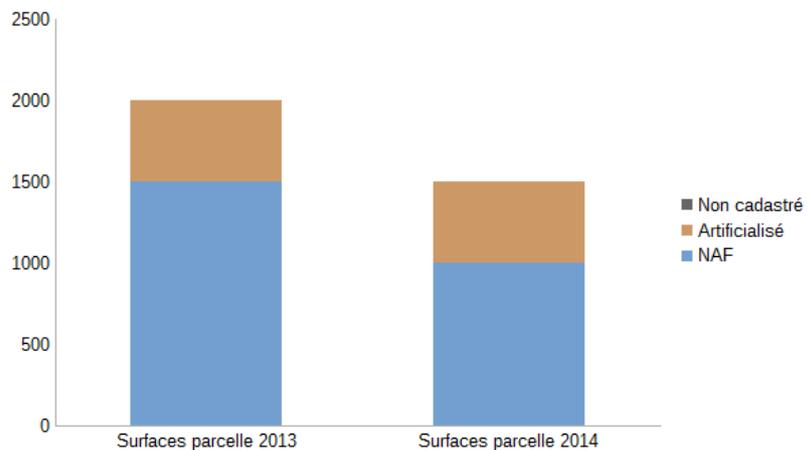
Les remembrements consistent, pour les services des impôts, à redécouper et transformer les parcelles existantes en d'autres. C'est aussi l'occasion de mieux renseigner les nouvelles parcelles. Dans ce cadre, les mises à jour peuvent provoquer des changements des surfaces artificialisées fiscales sans changement physique du sol. Les opérations de remembrement ont ainsi été exclues du calcul d'artificialisation.

En conclusion, le calcul de l'artificialisation ne prend pas en compte les golfs, les terrains militaires ni les changements dus aux remembrements.

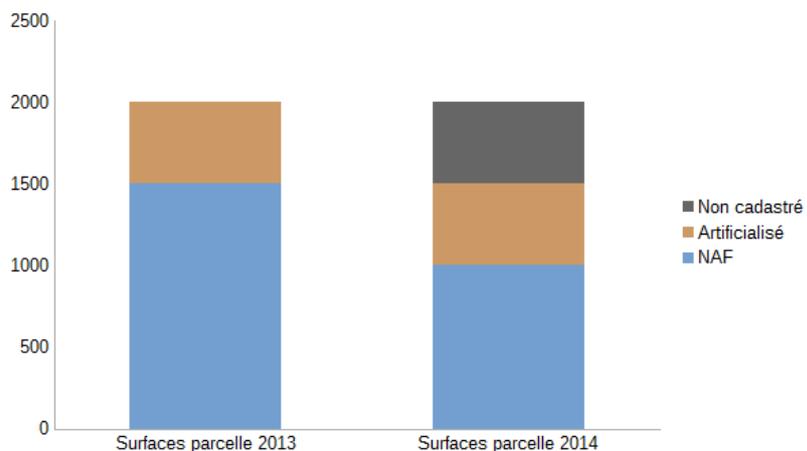
Étape A-1 : détermination des différences entre millésimes (algorithme Kirouge)

Une fois ce tri réalisé, il s'agit d'estimer les différences entre les millésimes, sur chacun des postes. Nous souhaitons travailler, entre deux millésimes, à périmètre constant¹. Dans ce cadre, les surfaces non cadastrées de l'année N ou de l'année N+1 sont complétées pour que les surfaces totales soient les mêmes entre les deux années.

Cette normalisation permet d'éviter les changements de surface dus à un réarpentage des parcelles.



Exemple de répartition avant correction



Exemple de répartition après correction

¹ Soit $Arti + NAF + Non\ Cadastéré\ (année\ N) = Arti + NAF + Non\ Cadastéré\ (année\ N+1)$. En pratique, cela signifie que la somme des deltas sera égale à zéro.

Étape A-2 : transferts entre catégories : (algorithme Kaver)

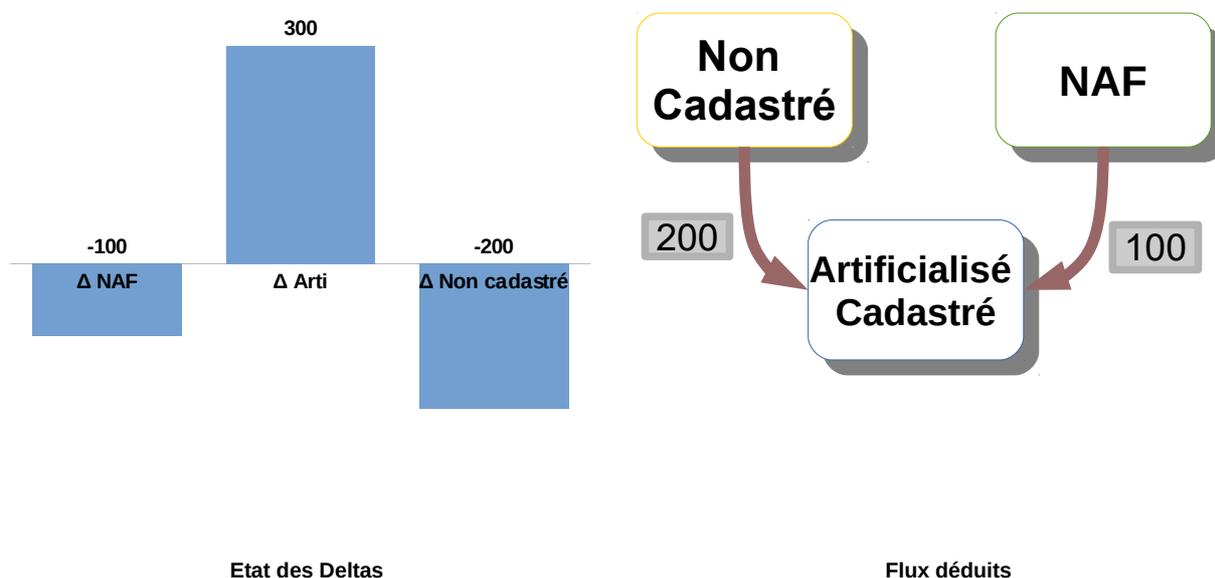
À partir de ces éléments, nous appliquons l'algorithme Kaver, qui transforme donc les 3 deltas en 6 flux.

Cet algorithme pose l'hypothèse que les flux ne peuvent être contraires. En d'autres termes, s'il existe un transfert entre espace NAF et artificialisation, on postulera que le flux contraire (artificialisé vers NAF) sera nul.

Au sein d'un îlot, on ne constatera qu'un seul sens de modification : on ne pourra donc avoir à la fois une artificialisation et une renaturation du même îlot.

Cette hypothèse est dans les faits assez crédible : chaque îlot est situé dans un espace de projet, et est constitué d'une surface plutôt faible (ordre de grandeur de la parcelle²). Il est très rare de constater deux modifications contraires sur le projet sur une année.

Enfin, l'algorithme pose comme contrainte que l'un des flux restant sera égal à 0. En d'autres termes, on postule l'absence de flux circulaires³.



Données de sortie

À ce stade, les données sont les suivantes :

îlot	Arti N → NAF N+1	NAF N → Arti N+1	NC N → NAF N+1	NAF N → NC N+1	Arti N → NC N+1	NC N → Arti N+1	...		
îlot1									
îlot2									
îlot3									

Ces données seront ensuite regroupées à l'échelle du territoire d'étude (par exemple la commune).

² Pour donner un ordre d'idée de la surface concernée, au niveau national, 80 % des parcelles font moins de 5000 m². Cela est beaucoup plus important pour les parcelles qui subissent un changement, dont la quasi-totalité fait moins de 5000 m².

³ C'est-à-dire que l'on ne peut par exemple avoir du NAF qui se transforme en artificialisé, et l'artificialisé qui se transforme en même temps en non-cadastré.

Étape B : calcul du flux d'usage

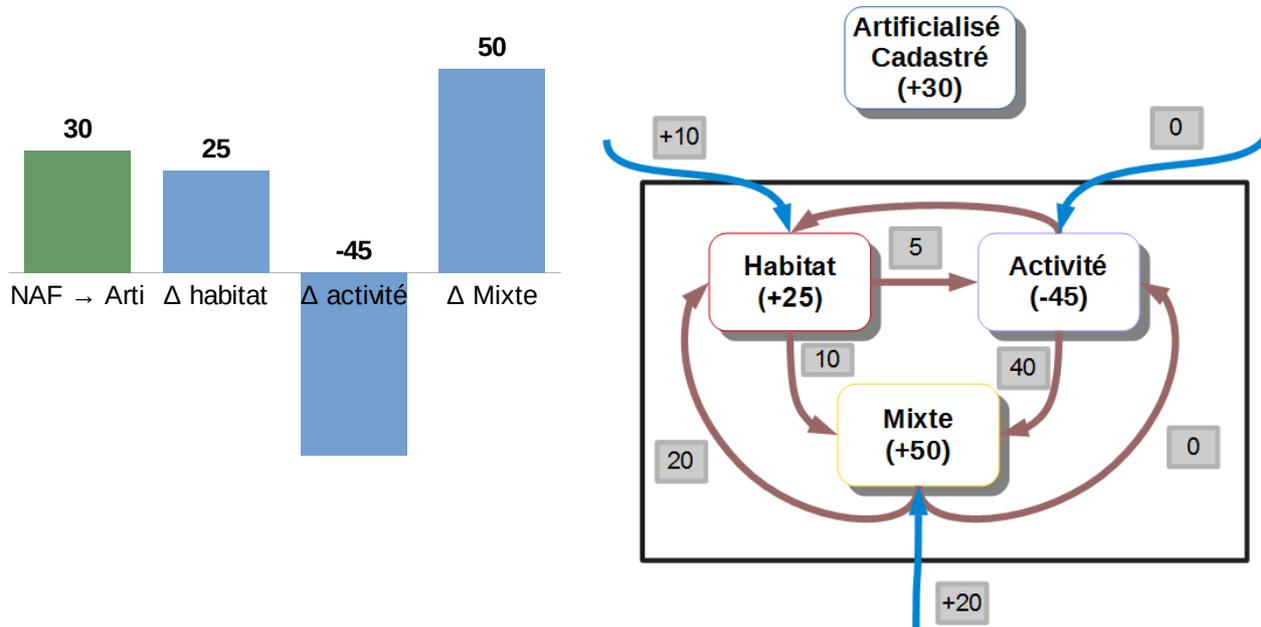
Objectif de cette partie

Nous avons donc les flux entre NAF, non cadastré et artificialisé pour chaque îlot. L'étape suivante consiste à observer les évolutions au sein de l'artificialisé entre les usages présents, c'est-à-dire :

- 1 – habitat,
- 2 – activité,
- 3 – mixte,
- 4 – « non affecté ». Cette catégorie est constituée de parcelles n'ayant pas pu être classées selon cette méthode.

L'objectif est ainsi de créer, à partir des données d'entrée, des flux de sortie entre les catégories, c'est-à-dire les données suivantes :

- 3 flux de « création », constitué de :
 - création de surface à usage d'habitat,
 - création de surface à usage d'activité,
 - création de surface à usage mixte,
- 6 flux de transferts entre habitat, mixte et activité, composé de :
 - transfert entre habitat et activité (la parcelle est à usage d'habitat l'année N, et à usage d'activité l'année N+1),
 - transfert entre activité et habitat,
 - transfert entre mixte et habitat,
 - transfert entre habitat et mixte,
 - transfert entre activité et mixte,
 - transfert entre mixte et activité.



Données en entrée

Flux souhaités en sortie : les flèches bleues sont des flux de création et les flèches rouges sont les flux de transferts. Les chiffres en encadré gris correspondent aux valeurs de ces flux

Données en entrée Sur chaque îlot, nous avons donc, pour chaque année, l'artificialisation et les surfaces affectées dans les 4 catégories d'usage (habitat, activité, mixte et « non classé »).

N° îlot	Artificialisation ¹	Habitat 2011	Activité 2011	Mixte 2011	Non classé 2011	Habitat 2012	Activité 2012	Mixte 2012	Non classé 2012
1	+ 200	100	100	100		300	100	100	
2	+ 500						300	200	100
3	0				200	200			
4	+ 350				50				400
5	0	100		200	600	600		300	
6	100	300							400
7	200	100	100		200				600

Exemple de surface en entrée. Les cases vides correspondent à des données égales à 0 (non représentées ici pour des questions de clarté)

Étape B-1 : élimination du « non affecté » (algorithme Kejaune)

Dans un premier temps, nous souhaitons limiter au maximum la présence du « non affecté » sur le territoire.

Pour chaque îlot, si au sein d'un millésime, il y a présence de non affecté et d'autres catégories, on affecte le non classé à ces catégories, au prorata du poids de ces catégories (**cas 1**). Cela se fait à enveloppe constante.

Exemple : l'îlot 5 a 200 m² de mixte, 100 m² d'habitat et 600 m² de non affecté, pour le millésime 2011. On affectera ces 600 m² aux catégories présentes, selon leur poids, c'est-à-dire 400 m² au mixte et 200 m² à l'habitat.

Une fois cette étape réalisée, si un millésime ne possède que du non affecté, et que l'autre millésime dispose d'une artificialisation dans l'une des catégories, ce non classé sera affecté au prorata des catégories de l'autre millésime (**cas 2**).

Exemple : l'îlot 6 a 300 m² d'habitat en 2011, et 400 m² de non affecté en 2012. Ce non classé de 2012 sera transformé en 400 m² d'habitat.

Ces deux étapes se font dans cet ordre, ce qui permet d'affecter les cas les plus compliqués (îlot 7)

Présence exclusive du non-affecté dans les deux millésimes

Si après ces deux affectations, on ne dispose que du non affecté dans les deux millésimes (c'est-à-dire que les surfaces d'habitat, d'activité ou de mixte sont égales à 0 dans les deux millésimes), on traite ce cas à part.

Dans ce cas, on fait la différence entre ces éléments pour créer un flux de « non affecté ».

¹ Il s'agit ici du flux des espaces NAF vers les espaces artificialisés

N° îlot	Artif ¹	Hab 2011	Act 2011	Mix 2011	Non Aff 2011	Hab 2012	Act 2012	Mix 2012	Non aff 2012	Observations
1	+200	100	100	100		300	100	100		Pas de modif
2	+500						360	240		Cas 1 : En 2012 : présence du non affecté et d'autres catégories => Transfert du non affecté 2012 vers Mix et act 2012
3	0	200				200				Cas 2 : Présence exclusive du non affecté en 2011 et d'hab en 2012 => Transfert du non classé 2011 vers habitat 2012
4	+100				300				400	Cas 3 : Présence exclusive de non cadastré dans les 2 millésimes
5	0	300		600		600		300		Cas 1 : En 2011, présence de 600 m ² de non affecté, d'hab et d'activité => transfert vers l'hab et l'act
6	100	300				400				Cas 2 : 300 m ² d'hab en 2011, et 400 m ² de non classé en 2012. Le non classé de 2012 sera transformé en 400 m ² d'hab.
7	200	200	200			300	300			Cas 1 + Cas 2 : Présence de non affecté en 2011, couplé avec de l'habitat et de l'activité. En 2012, seul le non affecté est présent. On transfère d'abord le non affecté de 2011 vers l'habitat et l'activité, puis le non affecté de 2012 est affecté selon la nouvelle répartition de 2011

Exemple de surface en après application de Kejaune. Les données d'entrée sont celles du tableau de la page précédente
Les données en gras sont celles qui ont été modifiées. Les cases vides correspondent à des données égales à 0 (non représentées ici pour des questions de clarté)

Étape B-2 : traitement des flux de changement d'usage (algorithme Korange)

Une fois ces structures modifiées, il s'agit de déterminer les flux. Au final, nous avons 3 données en entrée :

- la différence d'habitat entre deux millésimes,
- la différence d'activité entre deux millésimes,
- la différence de mixte entre deux millésimes.

Nous souhaitons, en sortie, 9 flux : 3 flux de création et 6 flux de transfert entre catégories.

De plus, on étudie la stabilité de l'enveloppe, à savoir si la surface artificialisée de l'année N est égale à la surface artificialisée de l'année N+1.

1 Il s'agit ici du flux des espaces artificialisés vers les espaces NAF

Rappel : affectation directe et affectation extrapolée

Pour chaque îlot, nous possédons à la fois la surface directe (c'est-à-dire sur la surface sur laquelle il y avait un local) et la surface extrapolée (surface artificialisée sur laquelle il n'y a pas de local, mais pour laquelle l'affectation a été réalisée par voisinage).

En d'autres termes, on différencie pour ces cas l'affectation directe, plus fiable, de l'affectation extrapolée, moins fiable.

Principes de base

Le traitement est basé sur plusieurs principes :

- Si l'enveloppe est stable, on postule que les flux de création seront égaux à 0. Les flux de transferts seront répartis selon l'algorithme Kaver (cf étape 2 et annexe).
- S'il n'y a que des surfaces extrapolées sur le premier millésime, et que l'enveloppe reste la même, on considère que tous les flux sont égaux à 0 (aucun changement interne ou externe).

En dehors de ces cas, on applique l'algorithme Korange, qui répartit :

- les flux de création en fonction du flux entre naf et arti,

les flux internes en fonction des différences de stock entre artificialisé et non artificialisé.

Le fonctionnement complet de cet algorithme est présenté en annexe.

Étape C : Post-traitements

Point d'étape

Nous avons donc tous les flux recherchés. Il s'agit, à ce stade, d'éliminer encore quelques cas problématiques et de créer une table de synthèse multi-millésime utilisable par tous.

À ce stade s'achève les travaux en bi-millésimes pour créer une table multi-millésimes.

Étape C-1 : Création d'une table multi-millésimes

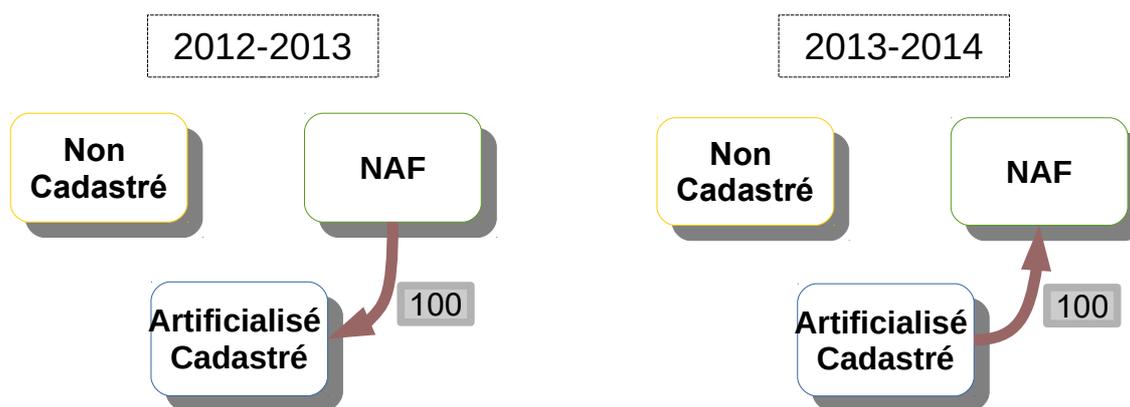
On transforme ici des tables bi-millésime en table multi-millésime, dont l'unité est l'îlot. Ce process est décrit dans l'annexe 2.

Étape C-2 : Suppression des îlots « rebond » (algorithme Kanbleu)

Il subsiste, à ce stade, de rares cas de « rebond » : un îlot est artificialisé¹ entre 2011 et 2012, puis est renaturé² entre 2012 et 2013. Il s'agit la plupart du temps de la correction d'erreurs, ou alors de projets abandonnés.

Les cas évoqués sont rares (environ 300 cas par département et par bi-millésimes), et ne correspondent pas à de grandes surfaces.

Il s'agit ainsi plutôt de corrections très locales, permettant d'utiliser ces données à une échelle communale.



Exemple de rebond

Méthode de traitement On repère ainsi les îlots qui ont fait l'objet de deux flux contraires sur 3 millésimes. Sur ces îlots, le flux naf vers arti est corrigé des flux arti vers NAF des deux millésimes suivants.

Nouveau NAF-Arti (N-N+1) = Ancien NAF-Arti (N-N+1) – Arti-NAF (N+1-N+2) – Arti-NAF (N+2, N+3)

Dans tous les cas, ce flux ne peut être inférieur à 0.

¹ C'est-à-dire que son flux naf vers arti est supérieur à zéro

² C'est-à-dire que son flux arti vers naf est supérieur à zéro. Cependant, le terme « renaturation » n'est pas ici bien utilisé, le flux arti vers naf n'étant pas à proprement parler de la renaturation (cf infra)

îlot	NAF-Arti 2011-2012	Arti-NAF 2011-2012	NAF-Arti 2012-2013	Arti-NAF 2012-2013	NAF-Arti 2013-2014	Arti-NAF 2013-2014
îlot1	+ 200	0	0	+ 100	0	+ 50
îlot2	+ 100	0	0	0	0	+ 100
îlot3	+100	0	0	0	0	+150

Îlots avec rebond avant correction

îlot	NAF-Arti 2011-2012	Arti-NAF 2011-2012	NAF-Arti 2012-2013	Arti-NAF 2012-2013	NAF-Arti 2013-2014	Arti-NAF 2013-2014
îlot1	+ 50	0	0	0	0	0
îlot2	0	0	0	0	0	0
îlot3	0	0	0	0	0	0

Îlots avec rebond après correction

Étape D : Caractéristiques du produit final

Étape D-1 : Fusion à l'échelle de rendu

À ce stade, nous regroupons toutes les données à l'échelle de travail (par exemple la commune), pour obtenir des flux à l'échelle communale.

Description du produit final

Le produit final comporte donc les résultats des flux d'artificialisation à la commune au sens des Fichiers fonciers.

Maille de restitution

Chaque ligne du tableau correspond à une commune du référentiel INSEE 2017. Il faut cependant noter les différences suivantes :

- les communes de l'île-de-Sein et de l'île-Molène¹ ne sont pas assujetties à l'impôt foncier. En conséquence, ne disposant d'aucune information sur ces territoires, ces communes ont été exclues du calcul.
- La commune de Suzan est fusionnée avec la commune de La Bastide-de-Sérou².
- Pour conserver un périmètre constant sur toute la période, les séparations de communes entre l'année 2009 et l'année 2017 ne sont pas prises en compte. En d'autres termes, si deux communes ont été fusionnées en 2012, elles seront toujours indiquées comme fusionnées. Cela concerne 6 communes³.

Millésimes

Les chiffres sont présentés de manière annuelle, sur la période 2009-2017. Cependant, nous ne disposons que des chiffres 2009-2011. Dans ce cadre, dans le document final, **les chiffres 2009-2010 et 2010-2011 sont les mêmes : ils sont égaux à la moitié des données calculées sur la période 2009-2011.**

Les flux

Les données finales comportent les flux suivants :

- 6 flux d'artificialisation,
- 9 flux d'usage,
- 1 flux d'usage sur le non-affecté.
- Ces flux sont présentés entre chaque année sur la période 2009-2017⁴.

Le tableau final contient donc 128 colonnes de flux.

Les données exogènes

Lors de la livraison du produit, des données exogènes ont été ajoutées. Il s'agit à ce stade de données INSEE, à savoir :

- données de localisation : classement ZRR, EPCI, département, région, Aire urbaine (et ses caractéristiques), Unité urbaine (et ses caractéristiques), bassin de vie.
- Caractéristiques de la commune : population, ménages et emplois des années 1999, 2010 et 2015, classe de densité
- Surface géométrique communale.

La structure complète de la table est donnée en annexe 5.

1 (codes INSEE 29 083 et 29 084)

2 La commune de Suzan (code INSEE 03904) est une enclave non délimitée au sein de La Bastide-de-Sérou (code INSEE 09042). Les services fiscaux, à l'inverse de l'INSEE, ne font pas la différence entre ces communes.

3 Avrecourt (52033) et Saulxures sont toujours considérés comme rattachés à Val-de-Meuse (52332), Chézeaux (52124) est toujours rattachée à Varennes-sur-Amance (52504), Laneuville-à-Rémy (52266) est considérée comme rattachée à La Porte du Der (52331), Lavilleneuve-Au-Roi (52278) est toujours rattachée à Autreville-sur-la-Renne (52031), Culey (55138) est toujours rattachée à Loisey (55298)

4 Soient 2009-2010, 2010-2011, 2011-2012 ... et 2016-2017, soient 8 périodes.

Quelles variables utiliser ?

Nous arrivons, à ce stade, au produit final. Celui-ci est constitué de 168 colonnes, reprenant les flux et les caractéristiques communales.

Cette base brute est mise à disposition. **Il est cependant déconseillé de l'utiliser, certaines données devant faire l'objet de précautions, que ce soient dans leur traitement ou leur interprétation** (cf infra). Ainsi, un deuxième produit simplifié est constitué, constitué des principaux indicateurs, et uniquement de données les plus fiables : il s'agit donc de la même donnée, mais avec un nombre de colonne moindre.

Fiabilité des flux

Les données de flux de NAF vers artificialisé sont fiables, ainsi que les flux de création (création d'habitat, d'activité ou de mixte). Il s'agit bien de l'artificialisation au sens fiscal.

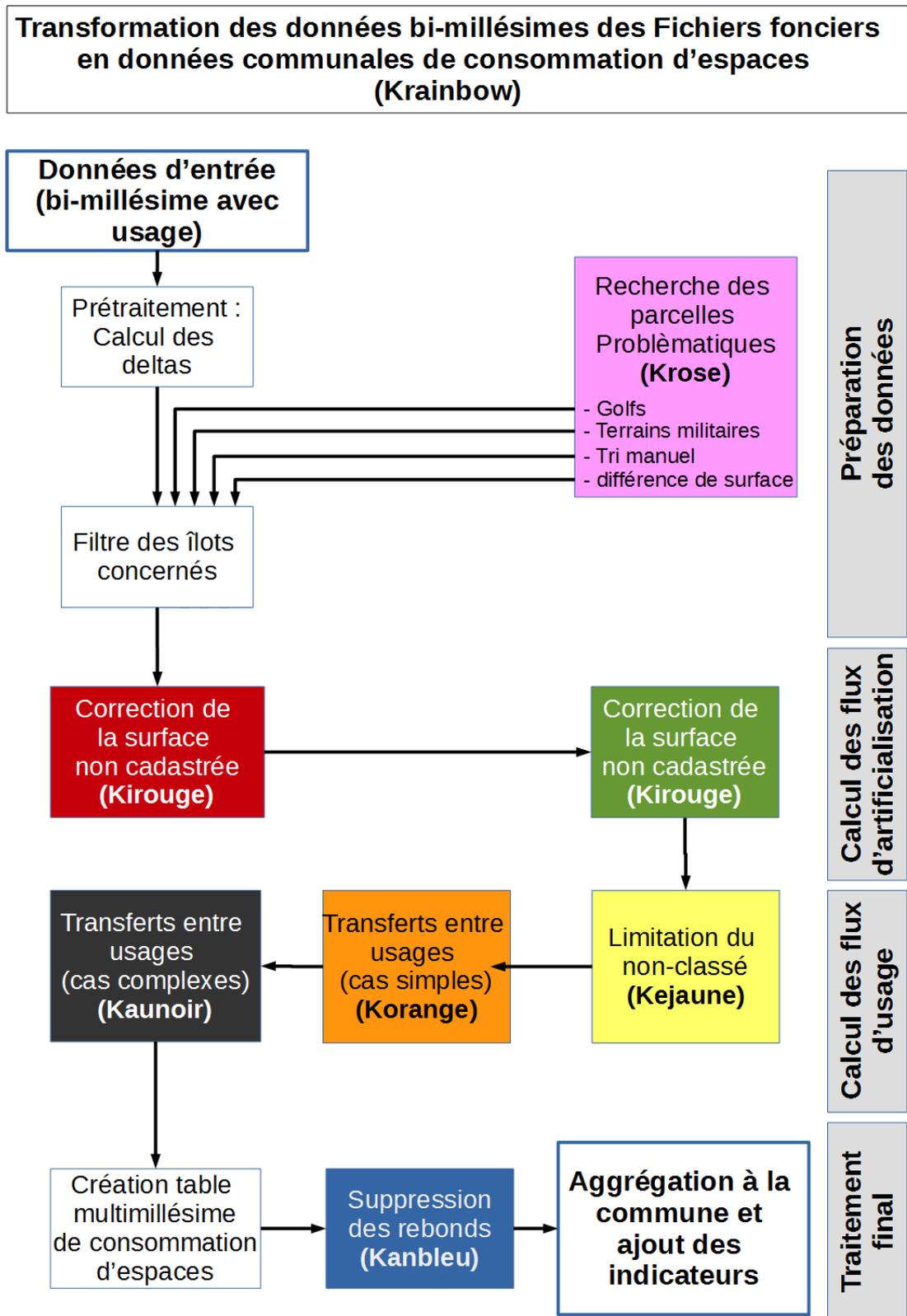
À l'inverse, les flux entre espaces artificialisés et NAF sont moins fiables. Si parmi eux on peut reconnaître de la réelle renaturation (passage effectif de l'artificialisé vers le NAF), il s'agit aussi de reconfigurations fiscales, la fiscalité des espaces NAF étant moins forte que celle des espaces artificialisés. À ce titre, le flux Artificialisé vers NAF comporte beaucoup trop d'artefacts (changements fiscaux sans changement réel de l'usage du sol) pour être considérés comme fiables. Le flux artificialisé vers NAF ne peut donc être assimilé à de la renaturation.

Les flux partant ou provenant du non-cadastré sont pour leur part fiables, mais ne recouvrent pas forcément une réalité thématique.

Enfin, les flux entre usages doivent encore être expertisés. En effet, il existe plusieurs cas où des changements théoriques d'usages ne sont en réalité que des redécoupages parcellaires. Des recherches sont en cours pour différencier les changements effectifs (requalification d'habitats en activité), des changements mineurs mais faisant changer le seuil (transformation d'un local en habitat sur les 50 de l'immeuble, ce qui fait l'intégralité de la parcelle change d'usage) ou les changements dus à la méthode.

Dans ce cadre, il est vivement recommandé de n'observer que les flux naf vers artificialisé, ainsi que leur destination (habitat, activité ou mixte).

Annexe 1 : Logigramme de la méthode générale



Logigramme présentant le traitement. Les cases en couleurs présentent les principales étapes, ainsi que les noms des algorithmes liés à cette étape.

Annexe 2 : établissement d'une table multi-millésime des Fichiers fonciers

Préalables

Objectif de l'étude

L'objectif de cette étude est de constituer une base de données reprenant toutes les informations entre les parcelles ainsi que leur devenir. Le présent rapport retrace la méthode qui permettrait de créer une table multi-millésimes.

La première étude multi-millésime

Cette présente étude vient à la suite d'une première étude réalisée en 2016 et disponible ici : <https://datafoncier.cerema.fr/etude-multi-millesimes>
Elle constitue un préalable indispensable à la bonne compréhension de la base actuelle et de ses retraitements. En particulier, il est nécessaire d'en (re)lire la partie relatives aux parcelles.

Rappel : la parcelle

La parcelle cadastrale est définie comme « une portion de terrain d'un seul tenant appartenant à un même propriétaire, située dans une même commune, une même section et un même lieu-dit. »¹. L'ensemble des parcelles constitue le plan cadastral.

La parcelle est constituée :

- d'une géométrie. Il s'agit ainsi du contour inscrit dans le plan cadastral. La majeure partie de ce plan cadastral est maintenant numérisée dans un produit nommé le PCI Vecteur. Toutes les parcelles ne sont cependant pas numérisées.
- d'informations littérales, qui sont les caractéristiques fiscales de cette parcelle. Celles-ci sont collectées dans la base de données dite « MAJIC ² », servant au calcul de l'impôt foncier.

Enfin, une parcelle possède un identifiant, codé sur 14 caractères³, composé :

- du code INSEE de la commune sur les 5 premiers caractères,
- dans le cas d'une fusion de commune, les 3 derniers chiffres du code INSEE de la commune absorbée,
- de la section cadastrale, généralement codée sur 2 lettres. Cependant, certaines sections peuvent comporter des chiffres (section « 0B » par exemple, ou sections cadastrales en Alsace-Moselle),
- du numéro de la parcelle.

Par exemple, la parcelle 59 001 000 AA 0022 :

- appartient à la commune 59 001,
- n'a pas fait l'objet d'une fusion de commune,
- appartient à la section cadastrale AA
- et porte le numéro 0022.

Rappel : les modifications de communes

L'évolution d'une commune est globalement de deux ordres :

- fusion (cas commun : en 2017, on compte ainsi 1283 communes de moins qu'en 2009),

¹ http://bofip.impots.gouv.fr/bofip/5359-PGP.html#5359-PGP_La_parcelle_44

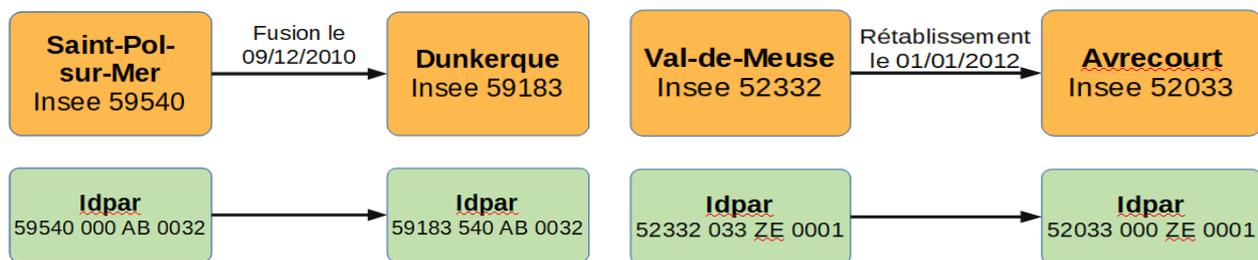
² Soit « Mise à jour des Informations cadastrales », le nom de la base de données de la Direction générale des Finances publiques

³ Le codage de la parcelle sur 14 caractères est fréquemment observée, mais n'est donc pas une convention.

- rétablissement ou séparation (cas rare : on compte ainsi 6 procédures de dé-fusion entre 2009 et 2017).

Dans le cadre d'une évolution de commune de type fusion, l'identifiant cadastral comporte à la fois le nouveau code INSEE et un suffixe de 3 chiffres reprenant l'ancien code INSEE de la commune fusionnée. Pour les communes rétablies (anciennement fusionnées), c'est ce suffixe qui devient les trois derniers chiffres du code INSEE de la commune rétablie. Autrement dit, par une comparaison multi-millésimes, il est possible de suivre une parcelle dans le temps, même avec un changement de commune. De plus, il est possible de dire si un identifiant de parcelle est modifié à cause d'une fusion ou d'un rétablissement de commune.

Par exemple, la parcelle 59 001 000 AA 0022 existe en 2012. En 2013, la commune 59 001 est absorbée par la commune 59002. La parcelle deviendra donc 59002 001 AA 0022.



Exemples de parcelles dans les communes fusionnées

Exemples de parcelles dans les communes rétablies

Les remembrements

Les remembrements sont une opération de restructuration du sol ayant pour objet de remodeler le parcellaire existant. Dans ce cadre, il s'agit d'une opération transformant de nombreuses parcelles en de nombreuses autres. Il n'existe pas de relations directes entre les parcelles.

Les bases de données utilisées

Les Fichiers fonciers Les Fichiers fonciers contiennent l'information sur les parcelles et leur état au 1^{er} janvier de l'année du millésime. Ainsi, le millésime 2012 contiendra les parcelles présentes au 1^{er} janvier 2012.

Outre le numéro de parcelle, les Fichiers fonciers contiennent toute l'information attributaire liée aux parcelles (usage, propriété...). Ces informations permettront des usages divers de la base multi-millésimes, et notamment l'évaluation de la consommation d'espaces.

Les DFI (documents de filiation informatisés) Les DFI sont une base de données reprenant l'historique des parcelles. Ce fichier est présent en open data sur data.gouv.fr¹, mis à disposition par la DGFIP depuis janvier 2018.

Le Fichier contient donc, pour chaque département, l'historique des opérations qu'ont subi les parcelles depuis 1990 (hors modifications communales ou remembrements).

Forme et contenu des DFI Les DFI sont retraités par le Cerema, afin d'obtenir une base de données contenant les colonnes suivantes :

- numéro de la transformation (appelé aussi identifiant de DFI),
- parcelle mère,
- parcelle fille,
- date,
- type de transformation.

Chaque ligne contient un couple parcelle-mère / parcelle fille. Un identifiant de DFI peut donc être présent sur plusieurs lignes. Par exemple, si la parcelle A se divise en 3 parcelles B, C et D, nous aurons les données présentes dans le tableau ci-dessous. Il s'agit donc du tableau représentatif d'un graphe (cf partie « représentation graphique »).

Par la suite, l'opération de transformation liée à un unique identifiant DFI sera nommée « filiation ».

id_dfi	Parcelle_mère	Parcelle_fille	Date	type
590010001	A	B	01/04/2012	division
590010001	A	C	01/04/2012	division
590010001	A	D	01/04/2012	division

Exemple d'une filiation

Type de transformations présentes dans les DFI

Les DFI contiennent les transformations suivantes :

- division de parcelle : une parcelle mère A se divise en plusieurs parcelles filles (relation 1 → N),
- réunion de parcelle : plusieurs parcelles mères se réunissent en une seule parcelle fille (relation N → 1),
- transfert : une parcelle devient une autre parcelle (relation 1 → 1),

¹ <https://www.data.gouv.fr/fr/datasets/documents-de-filiation-informatises-dfi-des-parcelles/>

- extraction du domaine public : une parcelle est créée à partir du non-cadastré (relation 0 → 1),
- transfert vers le domaine public : une parcelle passe en non-cadastré. Elle « disparaît » donc de la base (relation 1 → 0).

Chaque transformation est exclusive des autres. En d'autres termes, si une parcelle A est créée à partir d'une parcelle B et de non cadastré, nous aurons la présence de deux transformations (Non cadastré vers X et réunion de X et B pour créer A).

Les DFI et les fusions de communes

Lorsque deux communes fusionnent, les numéros de parcelles de la commune absorbée modifient leurs 8 premiers caractères. Ainsi, si deux communes fusionnent le 1^{er} juin 2013, on trouvera :

- l'ancien numéro de la parcelle dans les Fichiers fonciers 2013 ex : 59001000AA0025
- le nouveau numéro dans les Fichiers fonciers 2014. ex : 59002001AA0025.

A l'inverse, les DFI sont mis à jour continuellement avec les fusions de communes, et ce même si les évolutions de parcelles ont lieu antérieurement à la fusion. Admettons que la parcelle est issue d'une évolution ayant eu lieu en 2009 (soit avant la fusion de commune) : elle sera répertoriée dans les DFI en tenant compte de la fusion de commune.

On aura alors une évolution du type 59002001AA0030 → 59002001AA0025, et ce même si les parcelles n'avaient pas ce nom au moment de l'évolution.

Notation dans le rapport

Nous traiterons ainsi des deux bases de données :

- Fichiers fonciers pour obtenir les informations sur les parcelles,
- DFI pour traiter des évolutions de parcelles.

Il s'agira donc à la fois d'un stock au 1er janvier pour les FF, et d'un flux de transformation pour les DFI.

Pour faciliter la compréhension, les DFI seront présentés sous forme de flèches pleines, et les parcelles sous forme de rectangles.

Traitement général

Résumé du traitement

Le traitement s'opère en plusieurs étapes, de la plus simple à la plus complexe.

Dans un premier temps, la liaison est réalisée entre deux millésimes. Ces bi-millésimes sont ensuite fusionnées pour créer le fichier multi-millésimes.

Pour plus de facilité, nous établirons nos exemples sur le bi-millésime 2013-2014. Le traitement est cependant développé sur tous les bi-millésimes disponibles actuellement, à savoir 2009-2011, 2012-2013, 2013-2014, 2014-2015, 2015-2016 et 2016-2017.

Dès création des données 2018 des Fichiers fonciers, le traitement sera réalisé sur le bi-millésime 2017-2018.

Objectif général du traitement

L'objectif du traitement est de réaliser l'appariement entre deux millésimes, c'est-à-dire :

- de faire une liaison entre une ou plusieurs parcelles du millésime N0 avec une

ou plusieurs parcelles du millésime N1.

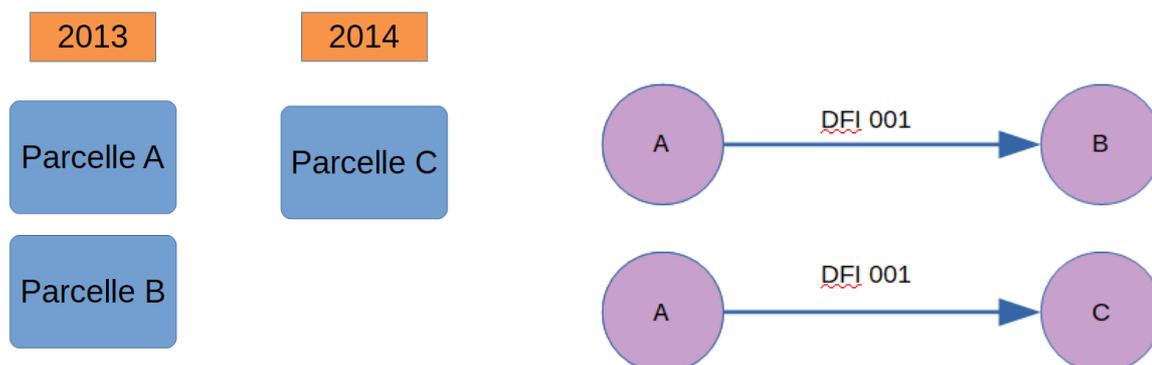
- De faire l'agglomérat le plus petit possible entre les deux entités. Par exemple, si la parcelle A reste la parcelle A, il n'y a aucune raison de la réunir avec une autre parcelle.

Forme du rendu La base de données aura donc la forme suivante :

Parcelles N0	Parcelles N1	Commentaire
[A]	[A]	La parcelle A est stable : elle est présente dans les deux millésimes
[B]	[C; D]	La parcelle B était présente sur le millésime N0. Elle s'est transformée en les parcelles C et D, présentes toutes les deux dans le millésime N1
[E ; F ; G]	[H ; I]	Les parcelles E, F et G, présentes dans le millésime N0, se sont transformées en les parcelles H et I, présentes dans le millésime N1
[K]	[Non Cadastré]	La parcelle K, présente dans le millésime N0, est passée en non cadastré
[L ; Non cadastré]	[M ; N]	Les parcelles M et N sont issues de la parcelle L et du non cadastré

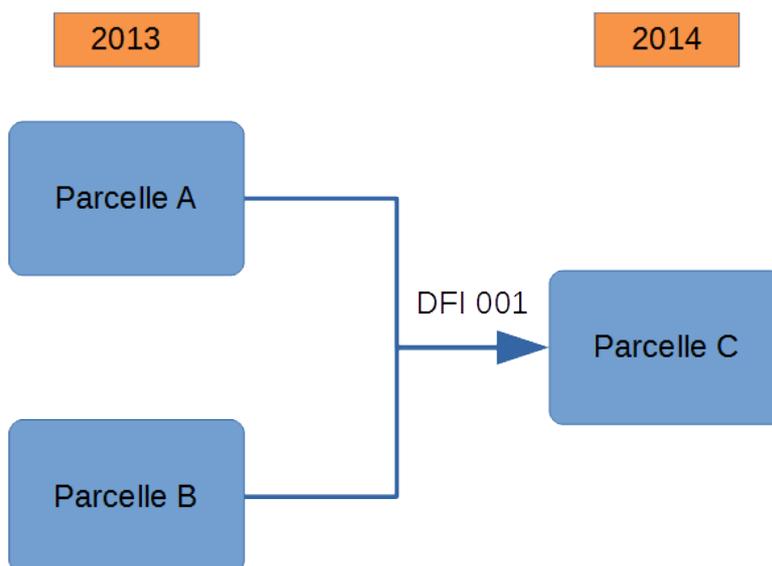
L'utilisation des DFI et des Fichiers fonciers Les Fichiers fonciers constituent un stock au 1^{er} janvier du millésime. À l'inverse, les DFI constituent un flux. Le travail consiste donc à faire coïncider les parcelles des Fichiers fonciers avec les DFI¹.

¹ De manière plus technique, il s'agit de reconstituer une structure de graphe, dont les parcelles Fichiers fonciers constituent les nœuds et les DFI des arcs (donc orientés). En particulier, le graphe résultant sera un graphe orienté acyclique, dont il s'agira d'extraire les composantes connexes. Il s'agit donc d'exploiter les apports de la théorie des graphes pour répondre au problème.



Utilisation des Fichiers fonciers

Utilisation des DFI



Résultat obtenu

Un traitement en entonnoir

Le traitement sera réalisé par étapes, chacune d'entre elles permettra de mettre en relation des parcelles de l'année N0 avec des parcelles de l'année N1. Chaque étape classe donc un certain nombre de parcelles. L'étape suivante se concentrera uniquement sur les parcelles n'ayant pas été classées précédemment.

La recherche d'un arbre pour l'opération

Le traitement revient donc à recréer un arbre, reprenant les liaisons entre les parcelles. On cherchera ensuite à repérer chaque arbre indépendant, c'est-à-dire toutes les parcelles en relations les unes avec les autres. Un exemple d'arbre indépendant, faisant la liaison entre les parcelles A et B d'un côté et C de l'autre est présenté ci-dessus.

Étape 0 : création des tables parcellaires

Pour chaque parcelle présente dans les millésimes des Fichiers fonciers, on recrée un nom post-fusion de communes. Par exemple, si une fusion a eu lieu entre la commune 59001 et 59002, on renommera toutes les parcelles commençant par 59001000 en 59002001XXX. Ce renommage permettra de réaliser les liaisons avec les DFI, qui comportent toujours les numéros de parcelles post-fusions, quelles que soient leurs dates.

Traitement des évolutions

À partir de la liste des DFI, on repère :

- communales**
- les communes stables dans le temps ;
 - les communes ayant fusionné (dont les parcelles auront un code de type 01XXXYYY, avec YYY différent de 000). Dans ce cas, on saura que la commune 01XXX a absorbé la commune 01YYY.

La reconstitution de l'historique des communes ne se fera donc qu'à partir des identifiants de parcelles présents dans les DFI.

Traitement des bi-millésimes

Étape 1 : traitement des cas simples

La grande majorité des parcelles (99,1%) ne change pas entre deux millésimes. La première étape consiste donc à traiter ces parcelles, pour lesquelles il n'y a pas d'évolution.

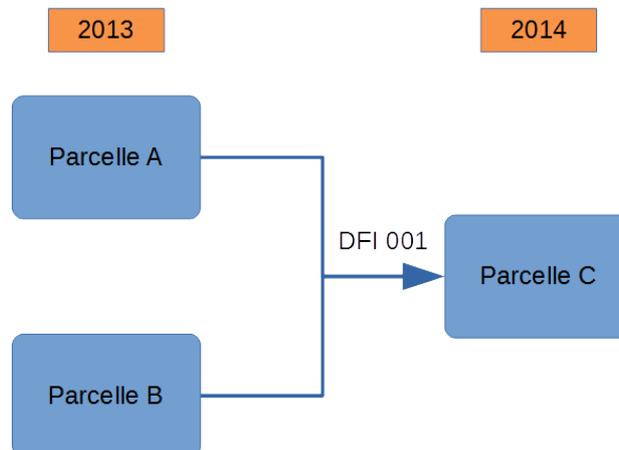
Cette étape permet en outre de récupérer la liste des parcelles existantes dans le millésime N0 et absentes de N1 et inversement, qui seront ensuite traitées dans les étapes ultérieures.

Étape 2 : traitement des évolutions simples

On traite ensuite les évolutions simples de parcelles, c'est-à-dire les parcelles dont la totalité des parcelles mères et des parcelles filles est présente dans les deux millésimes.

Il faut ainsi vérifier :

- si l'intégralité des parcelles mères est présente dans le millésime N0,
- si l'intégralité des parcelles filles est présente dans le millésime N1,
- si ces parcelles sont reliées par une seule et unique filiation.



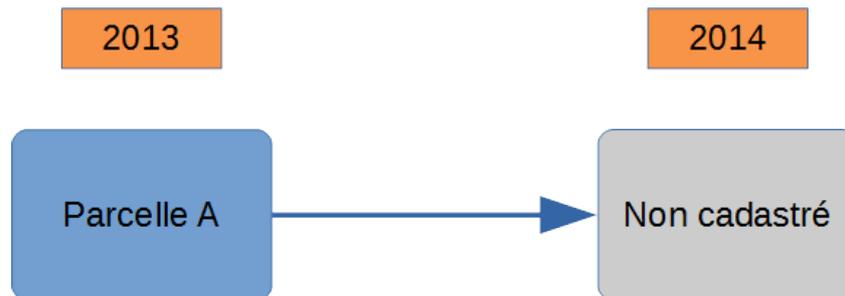
Exemple de liaison unique

Traitement du non-cadastré

Par définition, le non cadastré ne peut être présent dans les parcelles. On classe ensuite les parcelles :

- dont on retrouve la parcelle mère dans le millésime N0 et dont une filiation indique qu'elle passe en non-cadastré cette année.
- Ou dont on retrouve la parcelle fille dans le millésime N1 et dont une filiation indique qu'elle est issue du non-cadastré.

Dans les deux cas, il s'agit des parcelles qui apparaissent pour la première fois dans les millésimes.

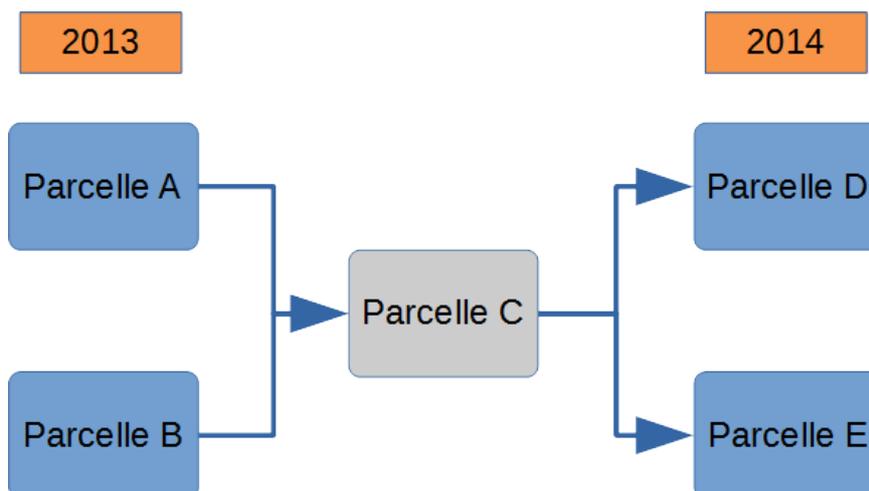


Exemple de passage en non-cadastré

Étape 3 : traitement des évolutions complexes

Il peut potentiellement se passer plusieurs évolutions de DFI entre deux millésimes des Fichiers fonciers. Dans ce cadre, il est nécessaire d'arriver à identifier l'ensemble des parcelles-mères et des parcelles-filles.

Dans l'exemple ci-dessous, les parcelles A et B sont présentes en 2013, et les parcelles D et E en 2014. Elles sont cependant liées par deux filiations qui passent par une parcelle C, qui n'aura pas le temps d'apparaître ni dans le millésime 2013, ni dans le millésime 2014 des Fichiers fonciers.



Exemple d'évolution complexe

Mode de traitement Il est ainsi nécessaire de repérer les parcelles étant en relation les unes avec les autres. Le travail est ainsi réalisé sur les DFI : on reprend, toutes les filiations comprises entre deux dates déterminées¹. Il s'agit ainsi d'une liste des liaisons entre parcelles.

Le détail de cet algorithme est présenté en annexe.

On considère ainsi que les DFI sont liés entre eux à partir du moment où ils possèdent au moins un élément en commun².

Parcelle N0	Parcelle N1	Identifiant DFI
A	B	001
A	C	001
B	D	002
B	E	002
Non Cadastré	F	003
D	G	004
F	G	004
H	J	005
I	J	005

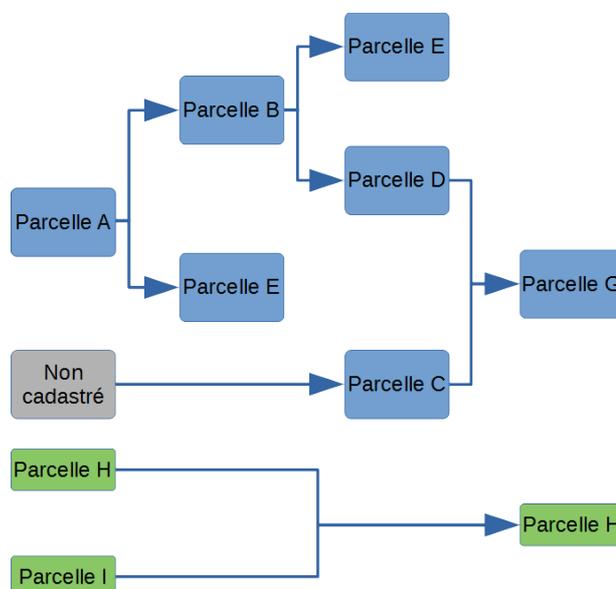


Tableau des filiations, tel que présents dans la base

Correspondance, sous forme de graphe, du tableau. Toutes les parcelles se verront donc attribuer la même « couleur », témoignant de leur appartenance à un même lot. À ce stade, on ne sait cependant pas si ces parcelles sont présentes dans les Fichiers fonciers

Traitement de l'arbre Une fois que l'on a retracé les liaisons entre les arbres, on regarde les extrémités de chacun d'entre eux, à savoir les parcelles n'ayant pas de parcelle-mères ou pas de parcelles filles. Dans notre exemple précédent, il s'agit des parcelles A, H et I d'un côté, et E, G et J de l'autre.

On regarde ensuite si l'intégralité de ces extrémités sont présentes dans le millésime N0 ou N1 des Fichiers fonciers.

Dans notre exemple précédent, si les extrémités sont présentes, on insérera les lignes suivantes dans notre résultat :

Parcelle N0	Parcelle N1
[A ; Non cadastré]	[E ; G]
[H ; I]	[J]

1 La date de début correspond à la première date à laquelle on retrouve une correspondance entre un DFI et une parcelle. La date de fin est la dernière date.

2 En reprenant la structure de base, c'est-à-dire celle d'un graphe orienté, on applique sur celui-ci un algorithme permettant de repérer les composantes connexes du graphe.

Étape 4 : élagage des arbres

Lorsque les extrémités des arbres ne sont pas présentes dans les millésimes, on réalise un traitement pour « élaguer » les arbres, à savoir couper les branches inutiles.

Dans notre cas précédent, nous ne retrouvons pas la parcelle « G » dans le millésime N1. Cependant, nous retrouvons les parcelles D et F : il faut donc supprimer la dernière liaison (D+F → G).

Étape 5 : traitement des sauts de millésimes

À ce stade, nous avons traité la majeure partie des parcelles (99,85%). Il nous reste cependant, dans les deux millésimes N et N+1, des parcelles n'ayant pas réussi à s'unir, notamment pour des questions de « sauts de millésime ».

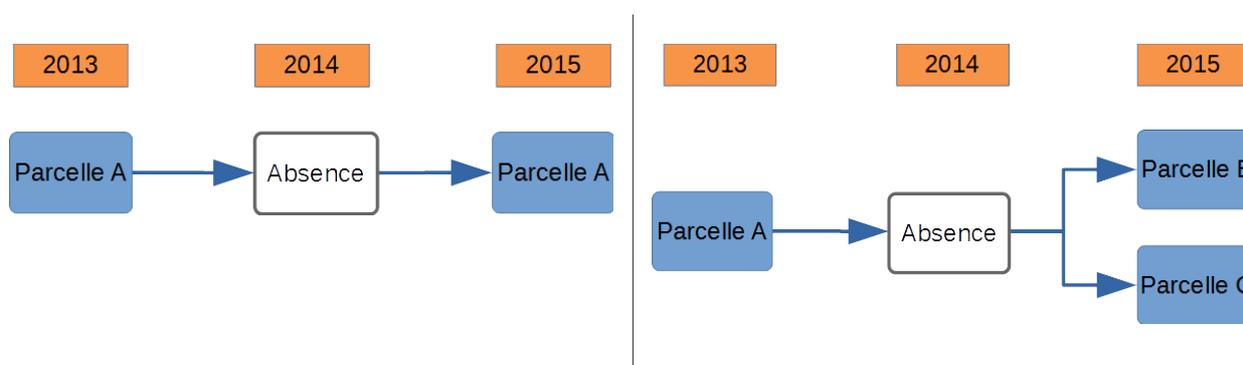
La problématique des sauts de millésime

Dans le peu de parcelles restantes, on observe parfois le phénomène suivant. Une parcelle A est présente en 2011, absente en 2012, puis à nouveau présente, sous la même forme, en 2013.

Ce phénomène peut se présenter sous diverses variantes :

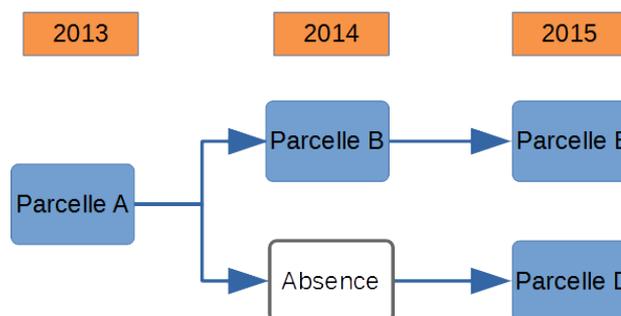
- saut simple : une parcelle A est présente en 2011, absente en 2012 puis la parcelle A réapparaît en 2013.
- saut avec modification : selon les DFI, la parcelle A est censée se séparer en B et C. On retrouve A dans le millésime 2011, et B et C dans le millésime 2013. On ne retrouve cependant ni A, ni B, ni C dans le millésime 2012.
- Saut partiel avec modification : Nous avons, dans le millésime 2011 la parcelle A, qui est censée se séparer en B et C. Nous retrouvons B en 2012. Cependant, il faut attendre 2013 pour retrouver C.

Il est à noter que les exemples précédents ont été présentés avec un seul saut de millésime. On peut cependant (rarement) retrouver des sauts de plusieurs millésimes. Ce phénomène reste toutefois très rare.



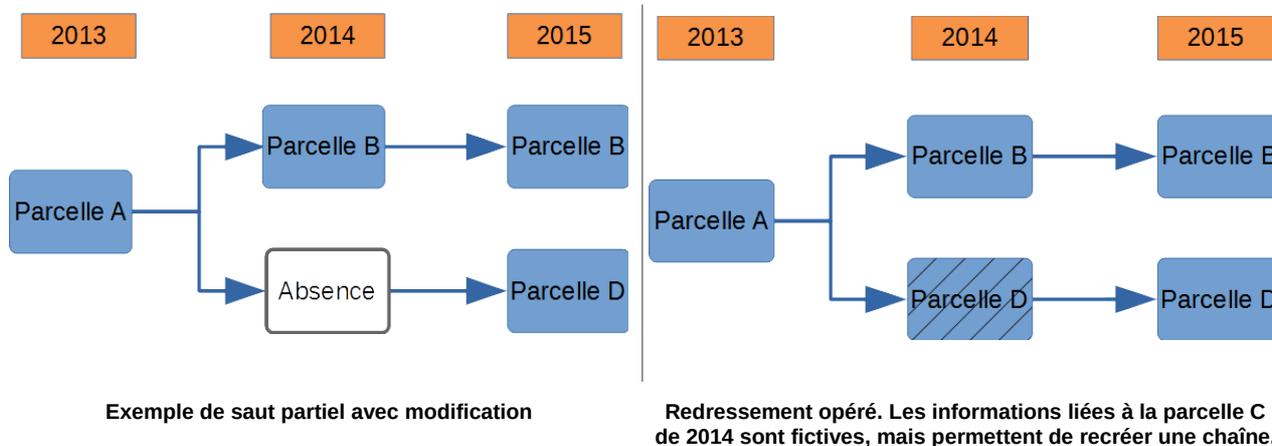
Exemple de saut « simple »

Exemple de saut « avec modification »



Exemple de saut « partiel avec modification »

Méthode de traitement Lors de notre traitement, nous repérons la parcelle manquante, et nous la reportons artificiellement, avec ses caractéristiques, dans un autre millésime. Il a été fait le choix méthodologique de systématiquement reprendre les caractéristiques de la parcelle figurant dans le millésime suivant.



Étape 6 : fusion des éléments restants

À ce stade, il reste plusieurs éléments :

- les évolutions complexes avec sauts de millésimes
- des disparitions ou fusions de parcelles pour lesquelles on ne retrouve pas de filiation dans les DFI. Ces disparitions peuvent concerner tout type de modification.
- Des remembrements.

Il faut noter que les remembrements concernent la quasi-totalité des parcelles restantes (soit 0,15%).

Dans l'étape 6, toutes les parcelles restantes de l'année N d'une commune sont mises en relation avec toutes les parcelles restantes de l'année N+1 de cette commune.

Contributions de chacune des étapes

Au niveau national, sur l'intégralité des millésimes, cela revient à traiter 630 043 600 parcelles, réparties comme suit :

Étape	Pourcentage de parcelles traitées
Étape 1 (stabilité)	99,14 %
Étape 2 (lien direct avec les DFI)	0,66 %
Étapes 3 et 4 (liaisons complexes)	0,03 %
Étape 5 (sauts de millésimes)	0,01 %
Étape 6	0,15 %

Cependant, il est nécessaire de rappeler que les phénomènes que l'on souhaite observer sont situés dans les parcelles qui suivent une évolution. Par exemple, les étapes 3 et 4 sont peu nombreuses par rapport au stock global des parcelles, mais concernent majoritairement des projets d'aménagement, pour lesquels l'observation est importante.

Il est donc nécessaire de pondérer les chiffres présents, en rappelant que les phénomènes intéressants à observer sont situés dans les étapes 2 et ultérieures.

Réalisation de la table multi-millésime

Réalisation de la table finale

A ce stade, nous avons donc les liaisons entre les tables bi-millésimes. Il s'agit à ce stade de recréer une table unique multi-millésime. L'identifiant de cette table sera nommé idilot. Ces îlots constitueront ainsi un périmètre stable au cours de tous les millésimes.

Lors de l'arrivée d'autres millésimes des Fichiers fonciers, la totalité de ces îlots sera donc recalculée.

Forme de la table finale La table finale aura ainsi la structure suivante, après ajout des champs thématiques (cf infra) :

2009	2011	2012	nblocal2012	nblocal2013
[A]	[A}	[B ; C]	0	0
[F]	[F]	[F]	2	7
[G ; H]	[I ; J ; K]	[I ; J ; K]	3	3

Méthode pour créer la table finale

À partir des tables bi-millésimes, nous reprenons les liaisons donnant lieu à modification. En d'autres termes, il s'agit des liaisons établies dans les étapes 2 et suivantes.

Mise en place de la table des liaisons entre parcelles

On transforme ces liaisons en une base de deux colonnes, mettant en lieu deux parcelles. Une liaison entre les parcelles [G; H] et [I;J;K] sera ainsi transformée en le tableau suivant :

Colonne1	Colonne2
G	I
G	J
G	K
G	I
H	I

Toutes ces liaisons, dans la totalité des bi-millésimes, est insérée dans ce tableau à deux colonnes.

Définition et détermination des lots

On réalise sur ce tableau multi-millésime l'algorithme de recherche de graphes connexes (cf annexe), de la même manière que lors de l'étape 3. Un indicateur, commun à toutes les parcelles qui partagent une liaison, et ce quel que soit le millésime, sera créé.

Colonne1	Colonne2	Bi-Millésime concerné	Identifiant du lot
G	I	2012-2013	lot1
G	J	2012-2013	lot1
G	K	2012-2013	lot1
G	I	2012-2013	lot1
H	I	2012-2013	lot1
I	L	2013-2014	lot1
I	M	2013-2014	lot1

Création de la table de correspondance

À partir de cette table, on crée une table de correspondance, qui décrira à quel lot appartient chaque parcelle. En d’autres termes, il s’agira de la table suivante, qui reprend la totalité des parcelles présentes pour la totalité des millésimes :

idpar	lot
G	lot1
H	lot1
I	lot1
J	lot1
K	lot1
L	lot1
M	lot1

Création de la table finale

À partir de la table précédente, on regroupe les tables bi-millésimes selon l’appartenance à ces lots. On réalise une jointure, sur l’identifiant du lot, entre tous ces bi-millésimes pour obtenir la table finale, qui sera donc du type :

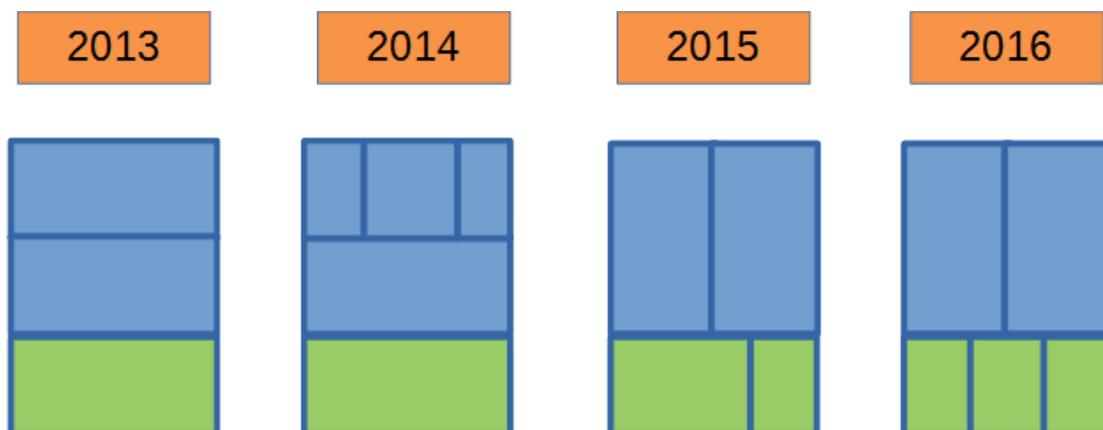
Idlot	2009	nblocal 2009	2011	nblocal 2011	2012	nblocal 2012	2013	nblocal 2013	geometry
lot1	[A]	0	[A]	0	[B ; C]	0	[D ; E]	0	
lot2	[F]	0	[F]	2	[F]	2	[F]	7	
lot3	[G ; H]	0	[I ; J ; K]	0	[I ; J ; K]	3	[I ; J ; L ; M]	3	

Signification thématique de l’îlot

L’îlot est le plus petit périmètre stable au cours de la période considérée. En d’autres termes, ce périmètre global ne change pas : tout au plus ses subdivisions changent. Pour une parcelle sans changement sur toute la période, l’îlot sera égal à la parcelle.

L’îlot est donc un espace de projet. Il s’agit de l’agglomération de parcelles se modifiant. C’est à partir de cet îlot que l’on pourra donc tenter de repérer les projets d’urbanisme, à savoir les lotissements ou zones d’activités.

Il faut toutefois rappeler que le contour de l’îlot est refait chaque année : le l’îlot du multi-millésime 2009 – 2016 ne sera ainsi pas forcément le même que le l’îlot du multi-millésime 2009 – 2017.



Identification des îlots. Toutes les parcelles en bleu appartiennent au lot 1, et toutes les parcelles en vert appartiennent au lot 2.

Conclusion et intérêt de l'étude

Une table en cours de création

Ce rapport ne présente ainsi que la méthodologie générale, celle-ci pouvant évoluer par la suite.

La méthodologie présentée ci-dessus permet ainsi de suivre l'évolution d'un ensemble de parcelles stables tout au long de leur vie. Au cœur d'un même îlot, on peut ainsi observer les compositions / recompositions de parcelles.

À ce titre, il serait ainsi possible de suivre, sur chacun de ces îlots, les évolutions physiques telles que présentées dans les Fichiers fonciers. À ce titre, il pourrait être possible de suivre des opérations de construction au sein de ces mêmes îlots.

Le premier usage de cette table serait le calcul de la consommation d'espaces à partir des Fichiers fonciers. Cette méthodologie particulière sera donc développée dans un autre rapport.

Ajout des champs thématiques

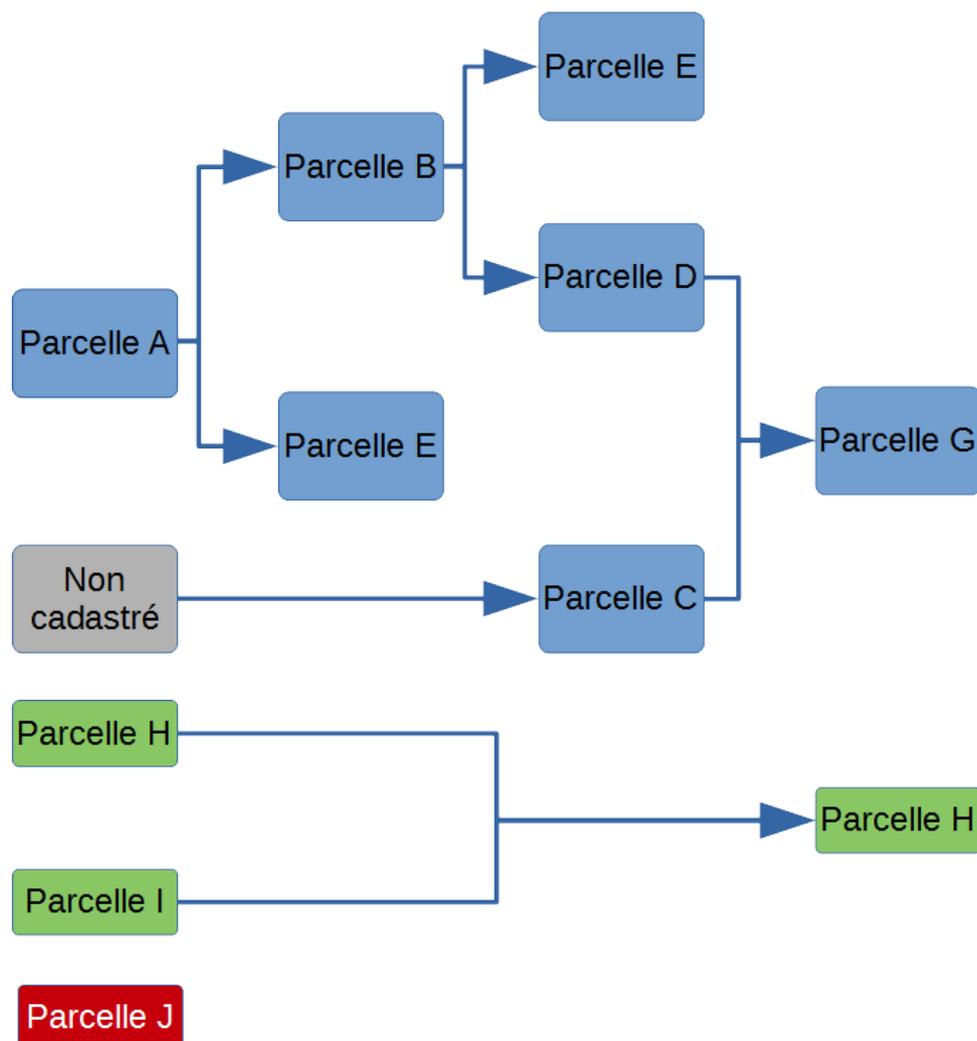
En parallèle de ce traitement, des champs thématiques présents dans les Fichiers fonciers pourront être inclus.

En particulier, il est possible d'inclure des champs relatifs à la consommation des espaces naturels, agricoles et forestiers.

Annexe 2.1 : algorithme de recherche de composantes connexes d'un graphe

Lien avec le problème

Notre base de données est constituée de nombreux liens entre parcelles, constitués par les filiations. Nous souhaitons retrouver les îlots, c'est-à-dire étiqueter toutes les parcelles ayant au minimum un lien entre elles.



Exemple de base de données : nous souhaitons ainsi distinguer le groupe des parcelles bleues (qui ont un lien, même indirect, entre elles), le groupe des parcelles vertes et le parcelle J, isolée. Nous aurons donc 3 îlots.

Formulation mathématique En théorie des graphes, retrouver les îlots indépendants est équivalent à la recherche des composantes connexes d'un graphe non orienté.

Structuration de la base de données Nos données sont stockées dans une base de données relationnelles, ce qui entraîne un stockage de l'arbre sous forme de liste d'adjacence. Dans l'exemple précédent, ce sera donc sous la forme suivante :

Début	Fin
A	B
A	C
B	D
B	E
Non Cadastéré	F
D	G
F	G
H	J
I	J

Exemple de liste d'adjacence

Implémentation classique de l'algorithme

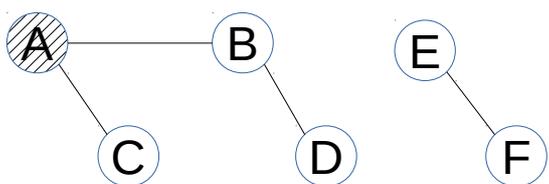
De manière classique, pour rechercher les composantes connexes du graphe, un simple parcours en profondeur du graphe suffit. En d'autres termes, il s'agit de l'algorithme suivant :

1. choisir au hasard un sommet non marqué
2. marquer ce sommet. Cette marque ou « couleur » constituera le nom (ou la « couleur » de l'îlot).
- 3.

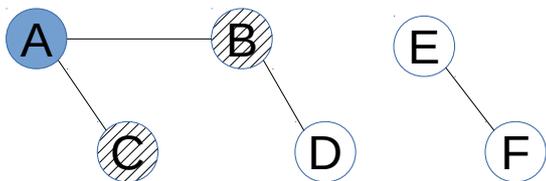
Si ce sommet a des liens avec des sommets non marqués, les marquer et rechercher leurs descendants.

Si ce sommet n'a pas de lien avec des sommets non marqués, arrêter.

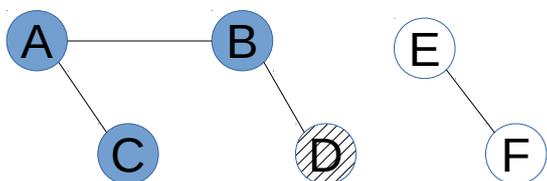
S'il existe, dans l'ensemble de la base, des sommets non marqués, reprendre à l'étape 1 et choisir une autre couleur. Sinon, stopper l'algorithme.



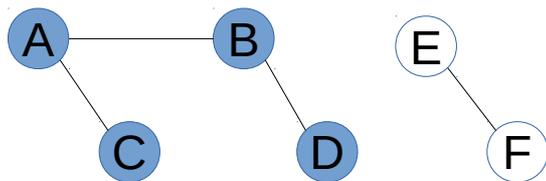
On choisit au hasard le sommet « A »



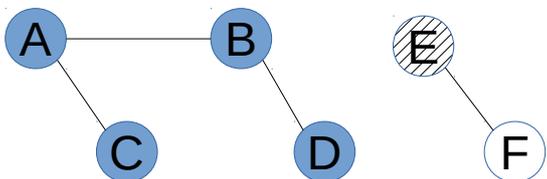
On marque (ici en bleu) le sommet A et on recherche ses descendants (ici B et C)



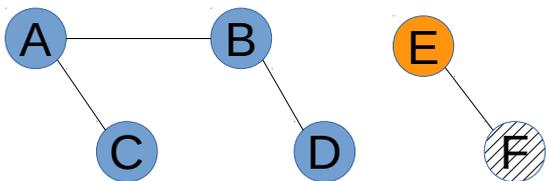
On applique l'algorithme aux descendants (B et C), et on les marque dans la même couleur.



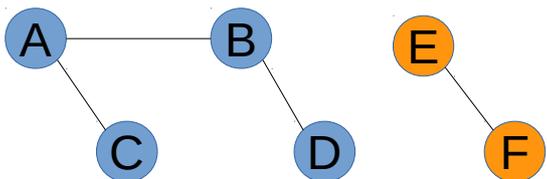
Le sommet D est marqué mais n'a pas de descendants.



On recherche donc au hasard un autre sommet parmi les sommets non marqués. Ce sera donc le sommet E



On colore le sommet E en orange, et on lui applique la même méthode



Tous les sommets sont colorés : l'algorithme s'arrête

Implémentation actuelle de l'algorithme

Notre graphe a plusieurs propriétés :

- il est constitué de très nombreux graphes connexes. En d'autres termes, il est très éparpillé, et il résultera de cet algorithme un très grand nombre de composantes connexes.
- Notre graphe est stocké dans une base de données relationnelles, qui rend plus difficile le parcours systématique en profondeur. En particulier, le parcours du graphe par récurrence s'avère plus complexe et moins rapide.

En conséquence, il est nécessaire d'adapter l'algorithme.

Principe de base

L'algorithme parcourt donc la liste d'adjacence, de la première à la dernière ligne. Il faut rappeler ainsi que chaque ligne de la liste d'adjacence correspond à un arc (liaison entre deux sommets)

1. L'arc considéré n'a pas de couleur (toujours vrai par construction)

2. Rechercher si l'arc est relié à un sommet déjà étiqueté.

Si ce n'est pas le cas, lui assigner une nouvelle couleur (cas 1)

Si l'arc est relié à un sommet déjà étiqueté, lui assigner la couleur du sommet étiqueté (cas 2)

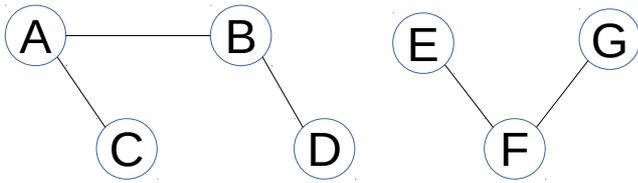
Si l'arc est relié à deux sommets étiquetés de deux couleurs différentes (ex : bleu et rouge) (cas 3) :

Étiqueter l'arc avec une des couleurs (ex : bleu)

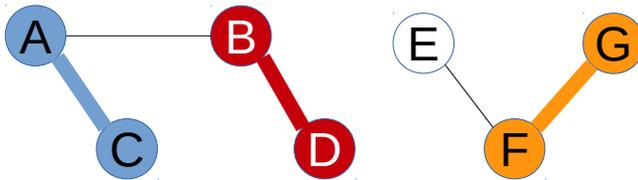
Rechercher, dans toute la base, les arcs de couleur rouge et les transformer en arcs de couleur bleue.

3. Passer à l'arc suivant

4. S'il s'agit de la dernière ligne, stopper l'algorithme.

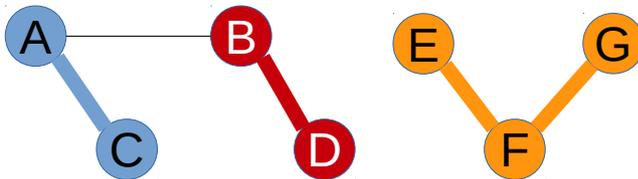


Arbre de départ : on sélectionne un arc au hasard (en réalité, il s'agit du premier arc présent dans la liste)

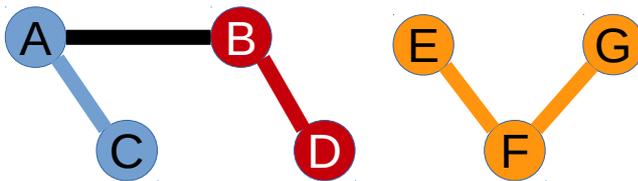


Si cet arc n'est pas relié à un sommet étiqueté, colorier l'arc (et les sommets correspondants) d'une nouvelle couleur. Par exemple, à gauche, les arcs A-C, B-D, F-G bénéficient d'une nouvelle couleur (cas 1).

Dans les faits, ces 3 arcs seront examinés successivement

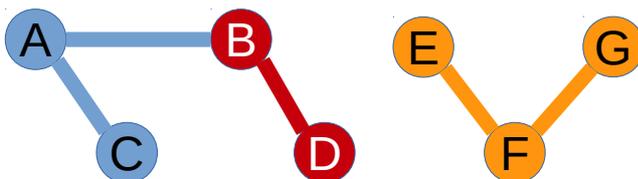


L'arc E-F est relié à un sommet déjà colorié (ici le sommet E, en orange). Il reprendra donc la couleur du sommet (cas 2)

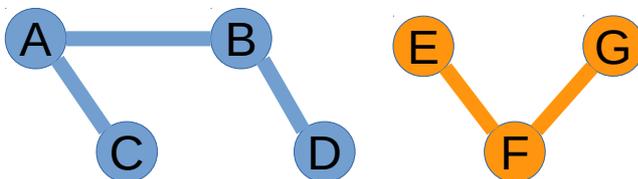


L'arc A-B est relié à deux sommets déjà coloriés (en bleu et en rouge).

Dans ce cas,



1. On colorie l'arc A-B avec l'une des couleurs (ici bleu)



2. On « repeint » la couleur rouge par la couleur bleue (on transforme la couleur 2 en couleur 1)

Tous les sommets sont coloriés : l'algorithme s'arrête

Annexe 3 : Affectation de l'usage aux parcelles des Fichiers fonciers

Contexte et présentation de l'étude

Présentation de l'étude

Ce rapport vise à établir, pour chaque parcelle cadastrale considérée comme artificialisée, l'usage de cette parcelle. En d'autres termes, l'objectif est de classer chaque parcelle dans une des catégories suivantes :

- habitat,
- activité,
- mixte.

Contexte de l'étude

Pour analyser les déterminants de la consommation d'espaces, il est nécessaire de pouvoir faire la différenciation entre l'artificialisation destinée à l'habitat et celle destinée à l'activité, et ce sur l'intégralité des millésimes des Fichiers fonciers¹. Dans ce cadre, et afin d'être comparable dans le temps, cette différenciation doit pouvoir se faire de la même manière sur tous les millésimes, actuels et à venir.

Objectif L'objectif unique de cette méthodologie est de permettre le calcul de cette artificialisation.

En particulier, cette assignation a pour objectif de travailler en flux, c'est-à-dire de mettre en avant les changements entre catégories. En d'autres termes, si une parcelle à usage d'activité est considérée comme à usage d'habitat, mais n'évolue pas entre les millésimes, cela n'aura pas d'influence.

Études précédentes La présente méthodologie se base sur des travaux antérieurs du Cerema, qui visaient à enrichir une occupation des sols, et en l'occurrence l'occupation des sols grande échelle de l'IGN.

Cet enrichissement a été repris et adapté pour permettre le classement des parcelles.

En préalable de ce rapport, il est conseillé de lire les documents relatifs à ces deux travaux, à savoir :

- Cerema Nord-Picardie, Perrine Rutkowski, Martin Bocquet, *Qualification de l'usage des zones US 235 de l'OCS GE par les Fichiers fonciers – Étude exploratoire à l'échelle de la commune*, février 2017²
- Cerema Nord-Picardie, Perrine Rutkowski, Martin Bocquet, *Qualification de l'usage des zones US 235 de l'OCS GE par les Fichiers fonciers - Étude d'approfondissement à l'échelle de la commune*, mars 2018³

Principe de base et présentation du rapport

L'objectif est donc d'assigner un usage (habitat, activité ou mixte) à des parcelles artificialisées⁴. Cet usage se fera en fonction des locaux présents sur cette parcelle : une parcelle comportant un local à usage d'habitat sera classée elle-même comme « habitat » ((Étape1).

1 Soient les millésimes 2009, 2011, 2012, 2013, 2014, 2015, 2016 et 2017

2 <https://datafoncier.cerema.fr/usages/consommation-des-espaces-et-occupation-des-sols/enrichissement-ocsge-par-les-fichiers-fonciers>

3 <https://datafoncier.cerema.fr/usages/consommation-des-espaces-et-occupation-des-sols/enrichissement-ocsge-ign-fichiers-fonciers-methode-approfondie-sur-grand-territoire>

4 On considère, dans ce cadre, que les parcelles artificialisées sont celles classées comme artificialisées au sens des Fichiers fonciers, et donc dont l'attribut dcntarti est supérieur à 0. Pour plus d'information sur ce que cette réalité recouvre, il est possible de se référer à l'étude suivante : Mesure de l'artificialisation à l'aide des Fichiers fonciers : définition, limites et comparaison avec d'autres sources, présent sur <https://artificialisation.biodiversitetousvivants.fr/>

Cependant, certaines parcelles, artificialisées, ne portent pas de locaux. Il s'agit notamment ce que l'on nommera des « espaces associés » (parkings, jardins, voiries...). On assignera donc un usage à ces parcelles, et ce en fonction de l'usage des parcelles adjacentes (Etape2).

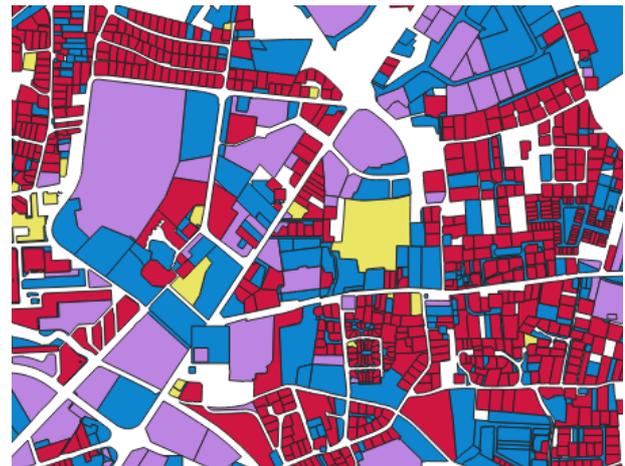
Structuration du rapport

La première partie du rapport portera donc sur l'assignation initiale : que considère-t-on comme « artificialisé » au sens des Fichiers fonciers, et comment définit-on une parcelle à usage « habitat » ou « activité » (Etape1).

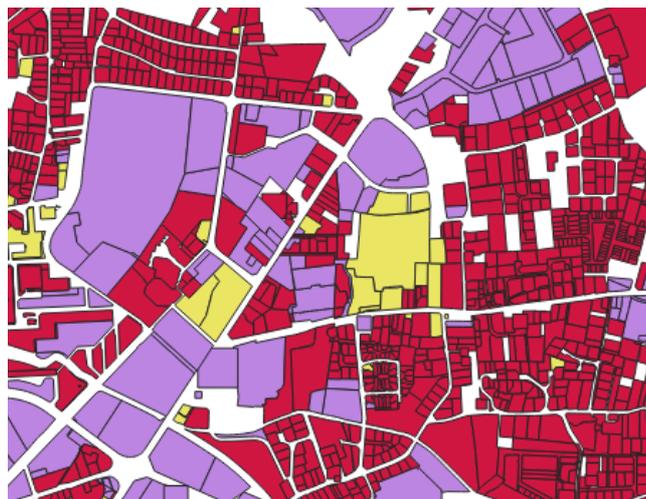
La deuxième partie se concentrera sur le processus d'affectation des espaces associés (Etape2).



Préparation : recueil des parcelles artificialisées



Étape 1 : assignation des parcelles avec locaux selon leur usage. Les parcelles bleues sont les parcelles artificialisées, mais sans locaux (donc sans affectation)



Étape 2 : assignation de l'usage aux parcelles artificialisées ne portant pas de locaux

Étape 1 : Affectation initiale des parcelles

Bases de données utilisées

Ce traitement repose en grande partie sur la base de données des Fichiers fonciers¹, issue des données de la taxe foncière. Le traitement est réalisé sur tous les millésimes disponibles².

Ces millésimes étant comparés par la suite, il est donc nécessaire de réaliser le traitement de manière comparable.

Lors du traitement, certaines corrections sont réalisées à partir de la BDTopo de l'IGN.

Liste des parcelles à traiter

Nous travaillerons sur les parcelles artificialisées au sens fiscal, c'est-à-dire dont l'attribut « dcntarti » est supérieur à 0.

Nous ne travaillons ici qu'à l'échelle de la parcelle, et non à l'échelle de la TUP³, celle-ci n'étant pas présente dans les premiers millésimes des Fichiers fonciers.

Affectation des parcelles⁴

On assigne à chaque parcelle une typologie en fonction de l'usage déclaré des locaux qu'elle contient. La typologie peut prendre 4 valeurs :

- résidentiel,
- activité,
- mixte (parcelles contenant à la fois de l'activité et de l'habitat),
- non affecté (parcelles artificialisées, mais ne contenant pas de local. L'objectif de cette méthode sera de minimiser le nombre de parcelles appartenant à cette catégorie).

Chaque parcelle aura une et une seule affectation.

Principe de la méthode

Pour chaque entité, on regarde le nombre de locaux à usage de logements ou d'activités. On distingue quatre cas :

- L'entité ne contient que des locaux d'un seul type. Il s'agit du cas majoritaire : dans ce cas, il n'y a aucune ambiguïté pour assigner un usage.
- L'entité possède des locaux à usage industriel.
- L'entité possède un commerce avec boutique.
- Les autres cas.

Pour chacun des cas, si un des usages (activité ou résidentiel) est largement prépondérant sur l'autre, l'entité est affectée ainsi.

Étape 1.0 : préparation de la table

Pour rappel, on travaille sur la table des parcelles. On crée une nouvelle table, à l'échelle de la parcelle pour fournir, sur chaque entité :

- le nombre de locaux à usage d'habitation⁵,

1 <https://datafoncier.cerema.fr/>

2 Soient 2009, 2011, 2012, 2013, 2014, 2015, 2016 et 2017

3 La TUP est un ensemble de parcelles adjacentes de même propriétaire. L'intérêt de la TUP par rapport à la table parcellaire est présenté dans le guide de prise en main des Fichiers fonciers :

<https://datafoncier.cerema.fr/ressources/fichiers-fonciers>

4 Ce paragraphe est repris des deux études réalisées pour l'enrichissement d'une occupation des sols

5 On utilise ainsi la variable « dteloc »

- le nombre de locaux à usage d'activité,
- la surface corrigée des locaux d'activités : on ne compte dans celle-ci que les locaux dont la surface est supérieure à 20 m²,
- la surface corrigée des locaux d'habitation : on ne compte dans celle-ci que les locaux dont la surface est supérieure à 20 m²,
- la valeur locative¹ des locaux à usage d'habitation,
- la valeur locative des locaux à usage d'activité.

Ces différentes informations permettront ainsi de bâtir les critères permettant de repérer l'affectation dominante.

Exclusion de certains locaux d'activités

Dans le décompte précédent, on exclut les locaux de type suivant :

- chantiers,
- transformateurs,
- antennes téléphoniques².

En effet, la présence de ces locaux peut fausser l'affectation. En particulier, ils peuvent être présents sur de très grandes parcelles, qui sont ainsi affectées, à tort, comme de l'activité.

Étape 1.1 : détermination des cas

L'objet de cette première étape est de déterminer l'affectation de chaque parcelle.

En premier lieu, s'il y a uniquement des locaux résidentiels /d'activités, la parcelle est affectée en résidentiel/activités. Il en est de même si la parcelle ne possède pas de local (affectation en « non-bâti »). Il s'agit du cas 0, majoritaire.

S'il y a des parcelles avec des locaux résidentiels et d'activités, on distingue alors trois cas :

- cas 1 : il y a au moins un local à usage industriel³,
- cas 2 : il y a au moins un commerce avec boutique,
- cas 3 : le reste.

Ces cas sont évalués de manière croissante : si une parcelle possède un local industriel et un commerce, elle rentrera dans le cas 1.

Étape 1.2 : calcul des scores

Lors de cette étape, on calcule des scores permettant de déterminer si la parcelle est plutôt à usage d'habitation ou à usage d'activités. Ce calcul est réalisé pour tous les cas de la même manière. Par contre, l'utilisation qui en sera faite sera différente selon les cas.

Pour chaque parcelle, on détermine deux types de score :

- **le score « normal »** correspond à un faisceau d'indices permettant de savoir si la parcelle est plutôt à usage d'activités. Il s'agit de seuils relativement larges (ex : la surface d'habitation est supérieure à 4 fois la surface d'activités).
- **la prédominance** permet de savoir si une variable est très présente par rapport aux autres. Il s'agit de seuils très discriminants. (ex : la surface d'habitation est supérieure à 20 fois la surface d'activités).

Chaque parcelle aura ainsi un score entre 0 et 3 permettant de déterminer l'affectation. Si ce score normal ne suffit pas à déterminer de manière certaine

- 1 Il s'agit d'une donnée fiscale, servant de base au calcul de l'impôt. À localisation identique, une forte valeur locative traduit souvent une grande surface. Il s'agit donc d'un indice pour savoir si la parcelle est essentiellement à usage résidentiel ou d'activités
- 2 Pour cela, on retire de la table des locaux les locaux dont le champ « cconlc » est égal à « AT », « UE » ou « CH », ainsi que les locaux dont « cconlc » est égal à « U » et « cconac » = '3512Z' (antennes relais)
- 3 On utilise pour cela le code d'évaluation ccoeva égal à « A ».

l'affectation, on vérifiera le score de prédominance pour permettre quand même l'affectation.

Calcul du score normal

Pour chaque parcelle, on calcule les éléments suivant :

- test de densité : la densité d'habitation de la parcelle est faible (c'est-à-dire inférieure à 10 logements par ha)¹,
- test de surface : la surface des locaux professionnels est prépondérante (la surface professionnelle représente plus de 80 % de la surface totale),
- test de valeur locative : la valeur locative liée aux locaux professionnels est prépondérante (la valeur locative professionnelle représente plus de 80 % de la valeur locative totale).

Chaque parcelle aura ainsi un score compris entre 0 et 3, selon le nombre de critères qu'elle remplit, 3 signifiant une forte probabilité de prédominance de l'activité, 0 une faible probabilité. Les scores de 3 vont déboucher ainsi sur un classement de la parcelle en « activités », et de 0 en « mixte » ou « habitat » selon les cas.

Calcul de la prédominance

Pour certains cas, les données présentées pour les activités économiques sont moins fiables. De même, les locaux d'activités peuvent ne correspondre qu'à une très faible portion de l'espace.

La parcelle est ainsi classée en « prédominance forte de l'activité » si elle répond **au moins un** des critères suivants :

- test de densité : la densité d'habitat de la parcelle est très faible (c'est-à-dire inférieure à 1 logement par ha),
- test de surface : la surface des locaux professionnels est fortement prépondérante (la surface professionnelle représente plus de 95 % de la surface totale),
- test de valeur locative : la valeur locative liée aux locaux professionnels est fortement prépondérante (la surface professionnelle représente plus de 95 % de la surface totale).

La parcelle est ainsi classée en « prédominance forte de l'habitat » si elle répond **au moins un** des critères suivants :

- test de surface : la surface des locaux résidentiels est fortement prépondérante (la surface professionnelle représente plus de 95 % de la surface totale),
- test de valeur locative : la valeur locative liée aux locaux résidentiels est fortement prépondérante (la surface professionnelle représente plus de 95 % de la surface totale).

Les scores normaux et de prédominance sont ainsi utilisés et combinés en fonction des différents cas déterminés à l'étape 2.

Étape 1.3-cas1 : gestion des parcelles avec des locaux industriels (cas 1)

Les locaux qualifiés d'industriels, c'est-à-dire le cas 1, sont évalués par les impôts par méthode comptable². Sans rentrer dans les détails, cela concerne les grands établissements industriels présents sur de grandes étendues. Par nature, les parcelles comportant ce type de locaux ont vocation à être classées à usage d'activités.

Cependant, les informations sur les locaux évalués par méthode comptable ne sont pas toujours fiables. Par exemple, la surface du local n'est pas utile au calcul de l'impôt : elle est donc souvent inscrite comme égale à 0.

On présume donc que la parcelle est par nature à usage d'activités, sauf si

1 Ce seuil peut paraître haut, mais il faut rappeler qu'il n'est calculé que sur les zones sur lesquelles il y a au moins un local d'activités et un local professionnel.

2 Pour plus d'information : <http://bofip.impots.gouv.fr/bofip/1443-PGP.html>

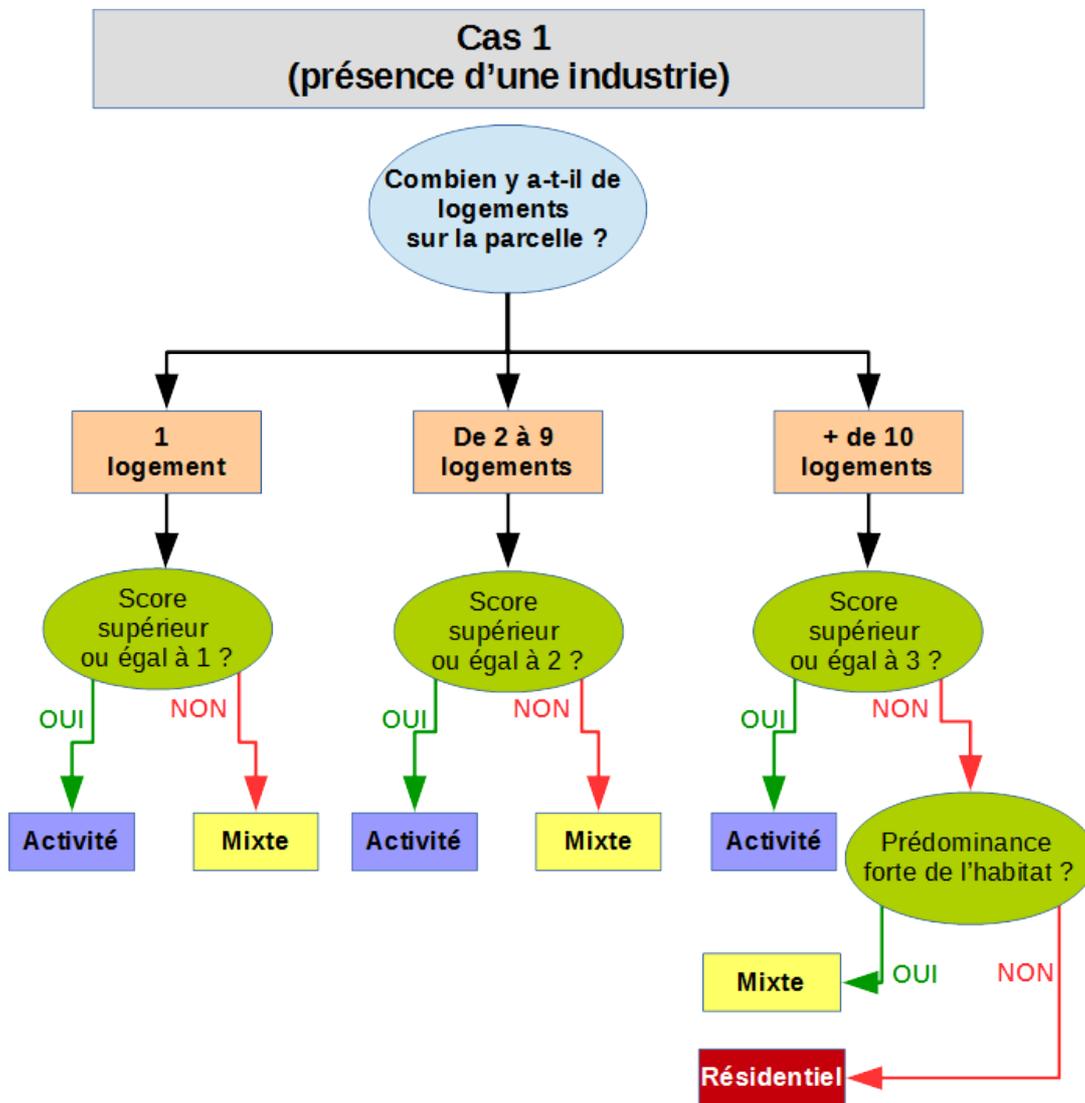
d'autres critères prouvent le contraire.

Un croisement du score et du nombre de locaux

Le score normal est croisé avec le nombre de locaux d'habitation présents sur la parcelle :

- si la parcelle ne possède qu'un seul local d'habitation, il s'agit en général d'une loge de gardien. Dans ce cas, on classe en « activités »,
- si la parcelle possède beaucoup de locaux d'habitation (plus de 10), on peut être sur une parcelle « mixte ». Dans ce cas, on ne classe en activités que si l'on est certain de la prédominance de l'activité (score maximal de 3). De plus, s'il y a beaucoup de logements, il se peut que l'usage soit résidentiel. Dans ce cas, on regarde s'il y a une forte prédominance de l'habitat. Si c'est le cas, la parcelle est classée en « habitat ». Il y a en effet la possibilité (minime) que des locaux soient considérés à tort comme industriels,
- dans les cas intermédiaires (de 2 à 9 logements), il est nécessaire d'avoir une forte présomption de la prédominance d'activité (score de 2 ou plus) pour la classer en « activité ».

Dans les autres cas, la parcelle est classée en « mixte ».



Récapitulatif de l'affectation pour le cas n°1

**Étape 1.3-cas2 :
gestion des
parcelles avec un
commerce avec
boutique (cas 2)**

Les commerces avec boutique ont vocation à s'installer plutôt au sein d'un tissu mixte. On présume donc que la parcelle est par nature à usage mixte, sauf si d'autres critères montrent que l'activité est prépondérante.

**Un croisement des
scores**

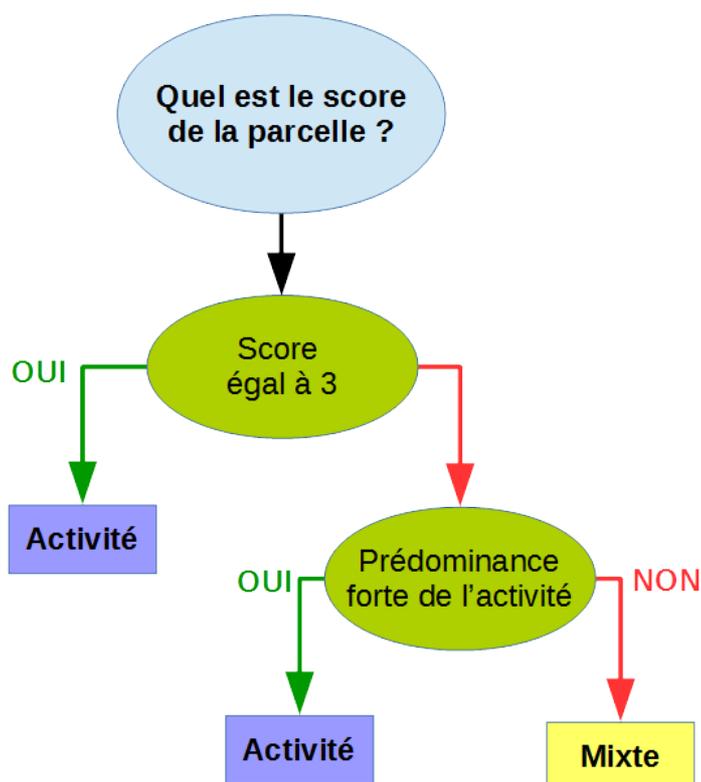
Si le score normal est de 0 (très faible probabilité de prédominance de l'activité), la parcelle est classée en mixte. Si le score est de 3 (très forte probabilité de prédominance de l'activité), la parcelle est classée en « activité ».

Pour les cas intermédiaires, on regarde si on a une prédominance forte de l'activité (c'est-à-dire si l'activité correspond à plus de 95 % de la surface et/ou de la valeur locative). Si c'est le cas, la parcelle est classée en « activités ». Sinon, elle est classée en « mixte ».

En d'autres termes :

- on classe en « activités » s'il y a prédominance forte de l'activité ou si le score est égal à 3,
- on classe en « mixte » sinon.

**Cas 2
(présence d'un commerce avec boutique)**



Récapitulatif de l'affectation pour le cas n°2

Étape 1.3-cas3 :

Le cas 3 correspond aux autres parcelles. Dans les faits, il s'agit souvent de

gestion des autres cas ambigus (cas 3)

parcelles avec des bureaux ou des locaux tertiaires. Cependant, il peut s'agir aussi d'un local d'habitation ou d'activité isolé dans un grand immeuble.

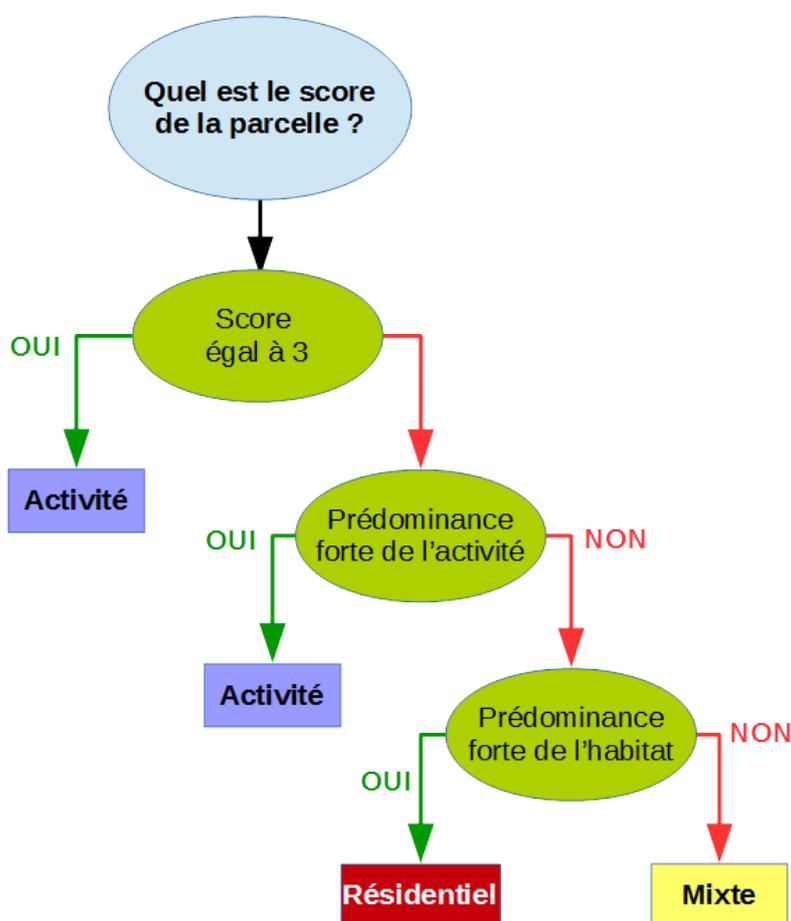
Il est donc nécessaire de tester 2 hypothèses :

- la parcelle est essentiellement à usage d'activité, et le logement n'est qu'un cas isolé (loge de gardien, logement de fonction, etc.),
- la parcelle est essentiellement à usage de logements, et le local d'activité n'est qu'un cas isolé (profession libérale dans un grand immeuble, présence d'un garage classé comme à usage professionnel au sein d'une résidence, etc.).

Dans les autres cas, la parcelle sera considérée comme « mixte ».

Par rapport au cas 2, pour lequel le classement sera toujours au moins mixte, on étudie ici la possibilité que la parcelle soit aussi résidentielle.

Cas 3 (autres cas)



Récapitulatif de l'affectation pour le cas n°3

Correction des affectations pour le public

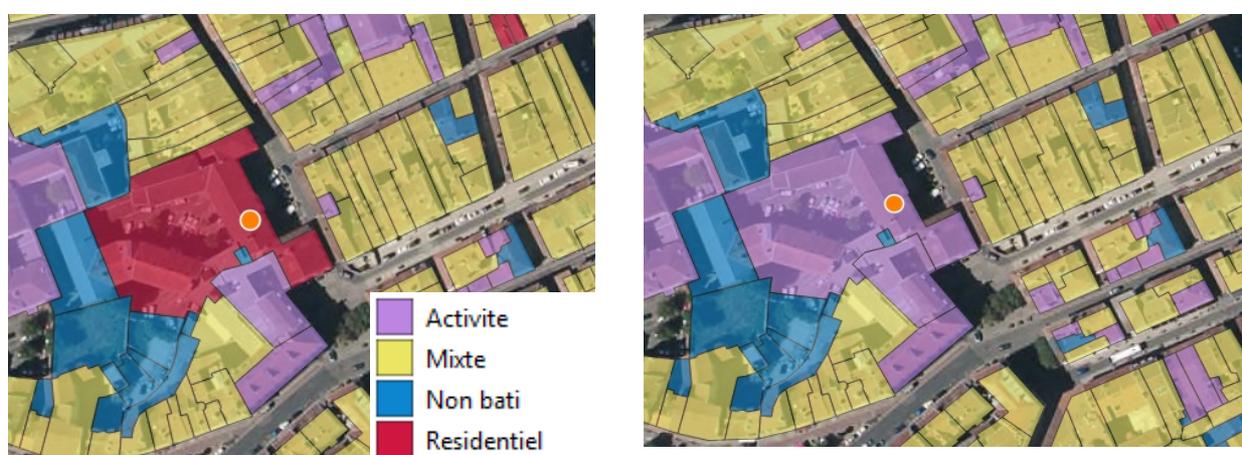
Les Fichiers fonciers ne contiennent qu'un nombre réduit d'information sur les bâtiments et équipements publics. À ce titre, il est nécessaire de redresser les affectations initiales à l'aide de sources de données extérieures.

Les premières étapes d'amélioration de l'affectation ont été réalisées à l'aide des seuls Fichiers fonciers. Cependant, pour cette étape, on ne peut se passer de données externes.

Étape 1.4 : Croisement entre la parcelle et la BD Topo

Les bâtiments publics ne sont pas toujours présents dans les Fichiers fonciers. Par exemple, une parcelle où il existe un collège public et des logements peut ne faire apparaître que les logements. Dans ce cas, la parcelle sera classée, à tort, comme « résidentielle », alors que sa vocation est plutôt tertiaire.

Pour le redressement, on croise la table obtenue dans les étapes précédentes avec différentes informations contenues dans la BD Topo¹.



Passage de la parcelle portant la mairie de « Résidentiel » (à gauche) à « Activité » (à droite). Le point rouge correspond à la localisation de la mairie dans le PAI administratif de la BDTopo.

1 <http://professionnels.ign.fr/doc/DC-BDTopo-2-2.pdf>

Liste des couches utilisées On utilise ainsi les couches BD Topo suivantes. Certaines modalités, reliées à des équipements qui n'ont que peu d'influence sur l'affectation de la parcelle, sont toutefois exclues.

Nom de la couche BDTopo	Modalités des champs
PAI_ADMINISTRATIF_MILITAIRE	Tous à l'exception de la modalité « Maison forestière »
PAI_SCIENCE_ENSEIGNEMENT	Toutes
PAI_SANTE	Toutes

Après un certain nombre de tests, les lieux de cultes n'ont pas été utilisés. En effet, ils reprenaient aussi un certain nombre de petites chapelles, de croix, etc. La prise en compte de ces petits éléments créait ainsi artificiellement des zones tertiaires.

Croisement On considère ainsi toutes les parcelles¹, qui croisent² un point d'intérêt de la BD Topo. Si ces parcelles ont plus de 10 logements, elles sont classées en « Mixte ». Sinon, elles sont classées en « activité ».

Pour les Points d'Activité ou d'Intérêt (PAI) « administratif / militaire », on n'applique ce traitement que si les parcelles ont au moins un propriétaire public, ou si le PAI est un bureau de poste.

1 On utilise pour cela le champ TYPPROP des Fichiers fonciers. On considère comme publique une parcelle ayant au moins l'un des locaux appartenant à un propriétaire public (Typprop contenant '01', '02', '03', '04', '05', '09')

2 C'est-à-dire dont le point est situé à l'intérieur d'une parcelle

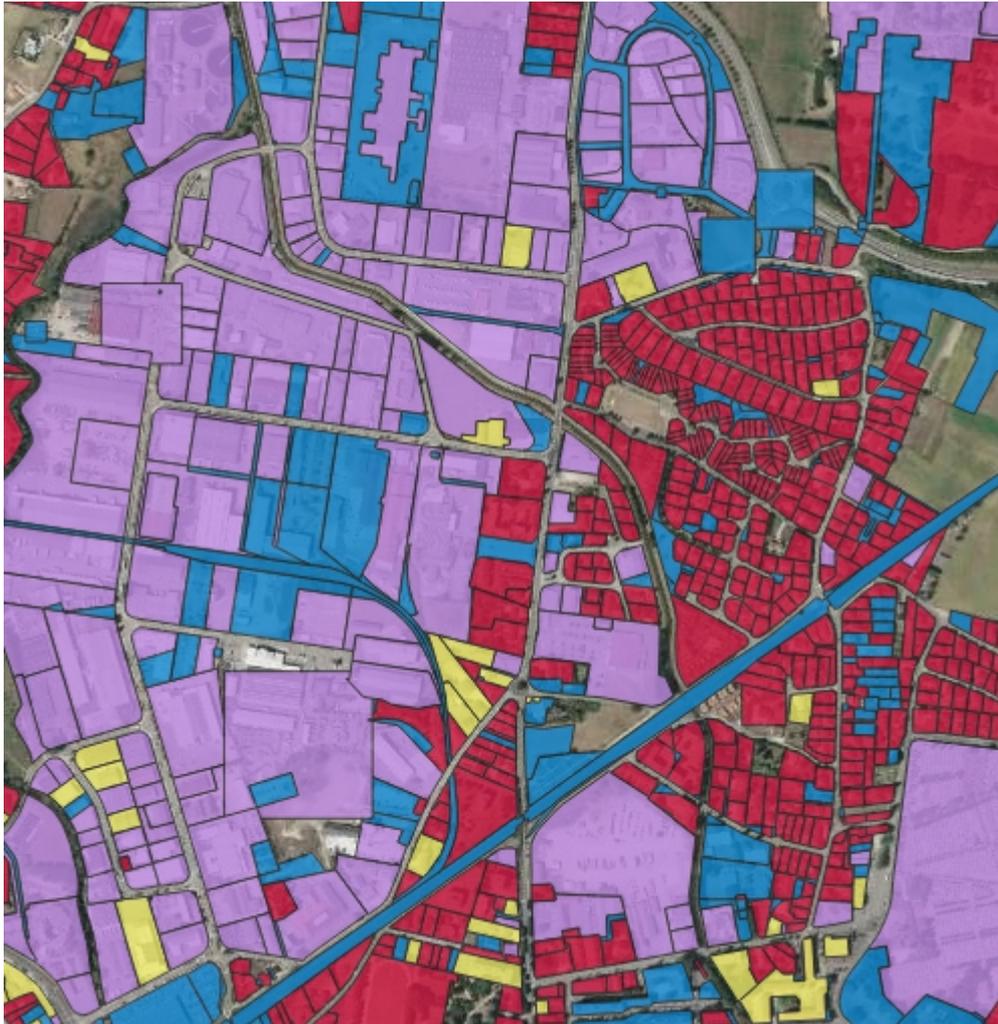
Affectation des parcelles artificialisées non bâties

Point d'étape

À ce stade, nous possédons donc une liste de parcelles. Au sein de celles-ci :

- une partie bénéficie d'une affectation,
- une partie est encore classée en non-bâti.

L'objectif est donc d'arriver à affecter ces parcelles non-bâties mais artificialisées.



Exemple d'affectation directe des parcelles artificialisées. On constate encore la présence d'espaces associés (en bleu dans le lotissement de droite et dans la zone d'activité à gauche)

Problématique : les parcelles artificialisées sans local

Les parcelles artificialisées sans local représentent plusieurs cas, dont des exemples sont donnés ci-dessous. L'objectif de cette partie est d'arriver à attribuer un usage à ces parcelles.

Il faut toutefois rappeler que l'usage des parcelles s'appréciera, par la suite, en dynamique. En d'autres termes, on observera, in fine, que les variations entre les espaces. Ainsi, une parcelle déjà artificialisée en 2009 ne rentrera pas dans les calculs.



Dans cet exemple, les espaces associés sont constitués par les jardins de propriétés



Ici, les espaces associés sont constitués d'une route, de places et d'espaces communs



Le parcellaire reprend ici qu'une partie du bâtiment. Le local étant placé sur cette parcelle, les autres parcelles apparaissent comme non bâties, mais artificialisées



On peut parfois constater que des locaux publics ou des locaux agricoles isolés répertorient leur parcelle de référence comme artificialisée, sans disposer d'information sur leur local. Ce cas reste toutefois relativement rare



Les espaces associés sont aussi constitués d'un parcellaire en cours d'aménagement, mais pas totalement construit



Enfin, de grands espaces artificialisés ne possèdent par nature pas de locaux. Il s'agit ici de la piste d'un aéroport



Ou d'une carrière



Ou enfin d'un champ photovoltaïque

Affectation des polygones « non affecté » par voisinage

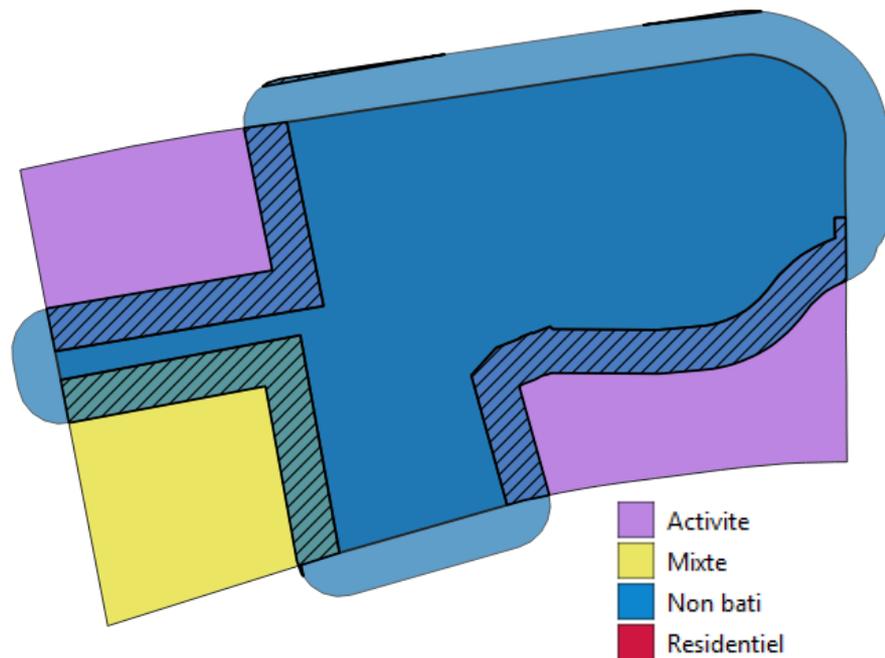
De nombreuses zones artificialisées sans local selon les Fichiers fonciers peuvent être affectées en étudiant les polygones affectés qui les entourent. Par exemple, on va pouvoir classer en « Résidentiel » une parcelle non affectée entourée de parcelles résidentielles, correspondant au jardin d'une maison ou de la résidence. De la même manière, on pourra classer en « Activité » le parking d'une zone commerciale grâce au repérage des locaux commerciaux sur les parcelles environnantes.



Figure 6 : Exemple de polygone « non affecté » qui pourra être affecté en « Résidentiel » grâce à l'importante part de périmètre touché par cette typologie.

Repérage des voisins Pour chaque polygone de type « non affecté », un buffer de 0,5 m est appliqué afin de repérer géomatiquement les polygones affectés immédiatement voisins intersectés par des polygones, tout en s'affranchissant des problèmes de topologie imparfaite. Les polygones voisins ont été préalablement groupés par typologie commune, si bien que pour un polygone non affecté étudié, il va être possible de hiérarchiser l'influence des différents types le voisinant. Par exemple sur la figure 6, le fait d'avoir groupé les différents polygones « Résidentiel » voisins va permettre de s'apercevoir qu'ils sont majoritaires par rapport au voisin « Activité ».

Influence des voisins affectés L'influence des typologies voisines du polygone « non affecté » est caractérisée par la longueur du contour commun entre chaque typologie et le polygone à affecter. Plus la longueur du contour commun entre le polygone et le type voisin est grande, plus le type voisin a de l'importance pour influencer la typologie du polygone « non affecté » à affecter. Sur le plan géomatique, cela revient à mesurer la surface d'intersection entre le buffer du polygone « non affecté » et chaque ensemble de polygones voisins de typologie identique.



Le polygone non affecté (en bleu) intersecte ici les zones à usage d'activité (en violet) et à usage mixte (en jaune). La surface intersectée (hachuré) d'activité est supérieure : le polygone sera considéré comme à usage d'activité

Aff1 : Critère des 65 % Un polygone « non affecté » est affecté à l'aide d'un type voisin si (et seulement si) ces deux points sont vérifiés :

- Le type voisin occupe plus de 65 % du contour du polygone « non affecté » **touché** par des polygones affectés.

ET

- Le type voisin occupe plus de 20 % du contour **total** du polygone « non affecté »

OU

Le type voisin occupe plus de 20 m du contour du polygone.

Le processus est mené **de façon itérative** jusqu'à ce qu'il n'y ait plus de polygones à affecter de la sorte.

NB : pour le premier point, il s'agit bien d'un seuil sur le contour touché par des typologies voisines, et non du contour total du polygone. Dans le cas d'un polygone en bordure d'une zone US235, et donc possédant peu de voisin, ce seuil peut être très dur à atteindre. Le second point vise à s'assurer que le type voisin occupe effectivement une part substantielle du contour total du polygone.

Aff2 : Critère des 40 % pour le mixte

Dans un second temps, on considère que dans des cas limites (seuil de 65 % pas tout à fait atteint), l'affectation en « Mixte » peut être privilégiée.

On effectue donc une deuxième affectation pour la seule typologie « Mixte ». On affecte un polygone « non affecté » en « Mixte » dès lors que le voisin de type « Mixte » occupe entre 40 et 65 % du contour du polygone « non affecté » touché par des polygones affectés.

Les critères à vérifier sont donc les deux suivants :

- Le type voisin occupe entre 40 et 65 % du contour du polygone « non affecté » **touché** par des polygones affectés.

ET

- Le type voisin occupe plus de 20 % du contour **total** du polygone « non

affecté »

OU

Le type voisin occupe plus de 20 m du contour du polygone.

Le processus est mené **de façon itérative** jusqu'à ce qu'il n'y ait plus de polygones affectés de la sorte.

NB : pour le premier point, il s'agit bien d'un seuil sur le contour touché par des typologies voisines, et non du contour total du polygone. Dans le cas d'un polygone en bordure d'une zone, et donc possédant peu de voisin, ce seuil peut être très dur à atteindre. Le second point vise à s'assurer que le type voisin occupe effectivement une part substantielle du contour total du polygone.

Le tableau ci-dessous reprend le nombre et les surfaces de polygones par typologie après cette deuxième étape d'affectation.

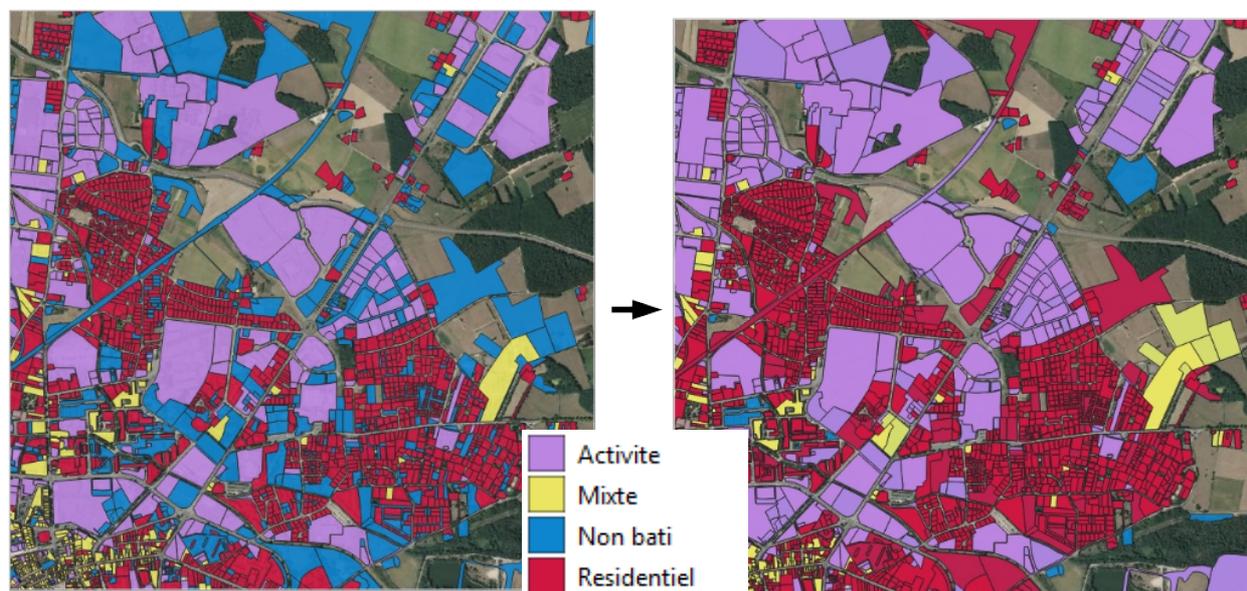
Aff3 : Forçage de l'affectation des polygones « Indéterminé »

Une dernière tentative d'affectation est réalisée afin d'assigner une classification à des polygones indéterminés, même lorsque le précédent tour d'affectation ne l'a pas permis. Il s'agit d'une affectation moins fiable, pour lesquels les seuils d'intersection (et donc les précautions d'usage liées) ont été amoindries.

Après cette étape, les polygones indéterminés restants ne seront que des polygones isolés (c'est-à-dire sans voisin).

Résultat de l'étape 5

Visuellement, on peut apprécier la disparition des surfaces « non affecté » au profit d'une typologie « logiquement » affectée, comme le montre la figure ci-dessous.



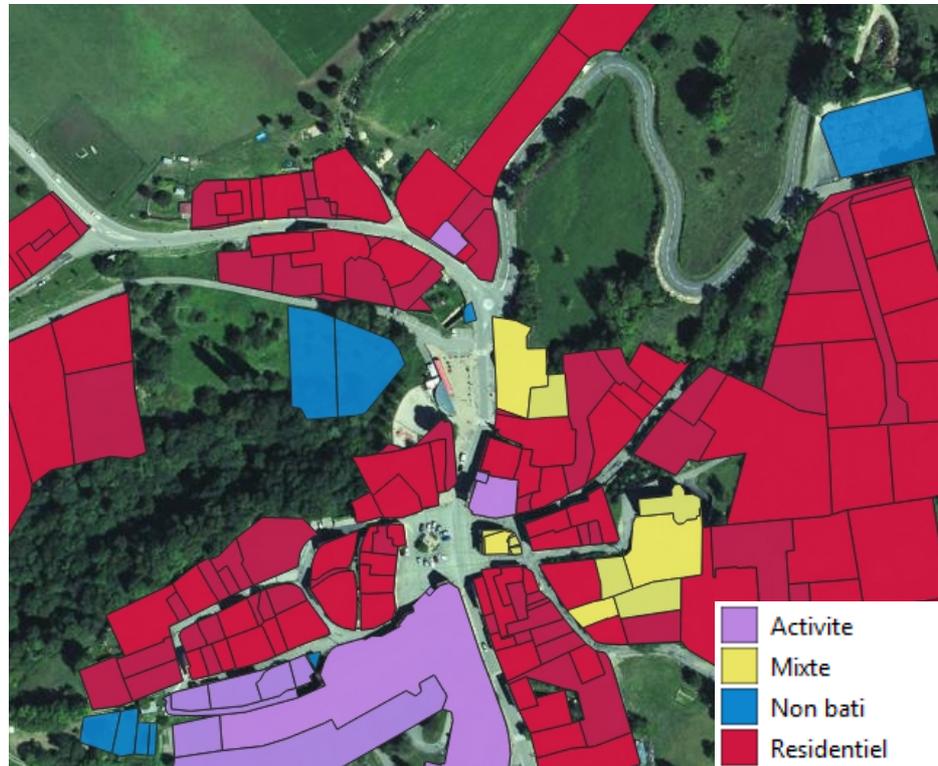
Comparaison de la carte avant l'étape 5 (à gauche) et après l'étape 5 (à droite).

Résultat intermédiaire

A ce stade, une partie des parcelles est affectée. Dans les faits, il s'agit des parcelles immédiatement adjacentes à des parcelles affectées.

Cependant, nous avons choisi de ne travailler que sur les parcelles artificialisées,

en excluant les autres. Dans les faits, cela donne un aspect très fragmenté aux polygones traités. Il n'est en effet par rare qu'un chemin (non cadastré) sépare deux parcelles.



Exemples de polygones artificialisés non affectés avec un buffer de 0,5 m.

Affectation plus lointaine

A ce stade, le taux d'affectation est encore bas. Il est donc nécessaire d'être plus permissif sur l'affectation, notamment en incluant une influence des parcelles plus lointaines.

La première affectation (ci-dessus) était réalisée en prenant un buffer de 0,5 m.

Le traitement est à réaliser plusieurs fois, avec des buffers de :

- 10m,
- 20m,
- 50m.

Ces affectations restent toutefois beaucoup plus rare que l'affectation par voisinage direct.



Affectation après un buffer de 10 m

Résultat et conclusion

Résultat final

Le traitement est réalisé sur la totalité des millésimes¹, et sur la totalité du territoire national. Chaque millésime et chaque département fait l'objet d'une table séparée. Les tables de sorties ont donc la structure suivante :

Identifiant parcellaire	Surface totale	Affectation	Fiabilité
Parcelle1	1000	Mixte	0
Parcelle2	2000	Habitat	10
Parcelle3	500	Habitat	0,5
Parcelle4	600	non affecté	

Fiabilité Chaque parcelle se voit assigner un score de fiabilité, égal au buffer utilisé :
 Les parcelles avec une fiabilité de « 0 » sont celles bénéficiant d'une affectation directe (étape1)
 Les autres nombres correspondent au buffer utilisé : une fiabilité de « 20 » (affectation en prenant des buffers de 20 m) est donc moindre qu'une affectation de « 0,5 » (utilisation d'un buffer de 0,5 m, soit une affectation à l'aide d'une parcelle adjacente).
 Un indicateur permet en outre de savoir à quelle étape la parcelle a été affectée.

¹ Soit 8 millésimes à la date de l'étude (2009, 2011, 2012, 2013, 2014, 2015, 2016 et 2017)

Temps de traitement Au vu du nombre important de parcelles, le traitement est très gourmand en temps machine. Avec un serveur performant, il faut compter environ 250 heures de temps machine pur (soit un peu plus de 10 jours).

Contribution de chaque étape

Sur la totalité des millésimes et des départements, les parcelles sont affectées de la manière suivante :

	Étape d'affectation	Nombre de parcelles concernées	Surface concernée (en ha)	Pourcentage des parcelles	Pourcentage de la surface
Étape 1	Affectation directe (présence d'un local)	162 045 595	26 557 975	57,4 %	53,2 %
Étape 2 (buffer 0,5 m)	Affectation à l'aide des parcelles adjacentes	100 268 255	17 120 828	35,5 %	34,3 %
Étape 2 (buffer 10 m)	Affectation à l'aide des parcelles situées à moins de 10 m	8 905 849	2 204 359	3,2 %	4,4 %
Étape 2 (buffer 20 m)	Affectation à l'aide des parcelles situées à moins de 20 m	2 140 234	636 586	0,8 %	1,3 %
Étape 2 (buffer 50 m)	Affectation à l'aide des parcelles situées à moins de 50 m	2 206 205	672 271	0,8 %	1,3 %
	Parcelles restant non affectées	6 788 669	2 692 105	2,4 %	5,4 %
	Total	282 354 807	49 884 124	100 %	100 %

Résultats en termes d'affectation

Les résultats en termes d'affectation sont les suivants. On retrouve ainsi une majorité de parcelles à usage résidentiel. De son côté, le mixte est peu représenté. Il faut toutefois noter qu'une part encore importante des parcelles reste non affectée. En pratique, il s'agit souvent de grandes activités sans locaux en dehors du tissu urbain (carrières, parc photovoltaïque...).

Étape d'affectation	Nombre de parcelles concernées	Surface concernée (en ha)	Pourcentage des parcelles	Pourcentage de la surface
Mixte	10 301 187	1 219 676	3,6 %	2,4 %
Activité	17 056 724	7 515 166	6,0 %	15,1 %
Habitat	245 208 227	38 457 175	87,9 %	77,1 %
Non affecté	6 788 669	2 492 105	2,4 %	5,4 %
Total	282 354 807	49 884 124	100 %	100 %

Conclusion

À ce stade, nous obtenons donc une liste des parcelles ainsi que de leur affectation. Cette liste permettra, dans un second temps, d'obtenir des informations sur l'affectation des parcelles urbanisées.

Annexe 4 : description des algorithmes Kaver et Korange

Annexe 4.1 : description de l'algorithme Kaver

Données de base et hypothèses

Nous avons en entrée 3 deltas, dont la somme est égale à 0. Nous souhaitons, en sortie, disposer de 6 flux.

Ces flux sont tous positifs.

Formulation mathématique

Ce problème correspond au système d'équation suivant, Delta Arti, Delta NAF et Delta NC étant connues, les Flux étant les inconnues

$$\begin{cases} \Delta \text{ Arti} + \Delta \text{ NAF} + \Delta \text{ NonCad} = 0 \\ \forall \text{ Flux}, \text{ Flux} \geq 0 \text{ (1)} \\ \forall i \in [\text{ Arti}, \text{ NAF}, \text{ NonCad}], \Delta i = \sum_j \text{ Flux } j, i - \sum_k \text{ Flux } i, k \text{ (2)} \end{cases}$$

La première équation indique que tous les flux sont positifs. La deuxième équation indique que pour chaque delta, la somme des flux sortants moins les flux entrants est égale à ce delta. Par exemple, Delta Arti est égal à la somme des flux NAF-Arti, NonCad-Arti, moins les flux Arti-NAF, et Arti-NonCad.

Il s'agit ainsi d'un système d'équations linéaires sous contraintes typique d'un problème d'optimisation linéaire.

Insuffisance de la définition : hypothèse de simple flux

À ce stade, nous avons cependant un nombre d'inconnues trop important. Cela conduit à une infinité solutions : il est ainsi nécessaire d'ajouter des hypothèses pour aboutir à une solution unique.

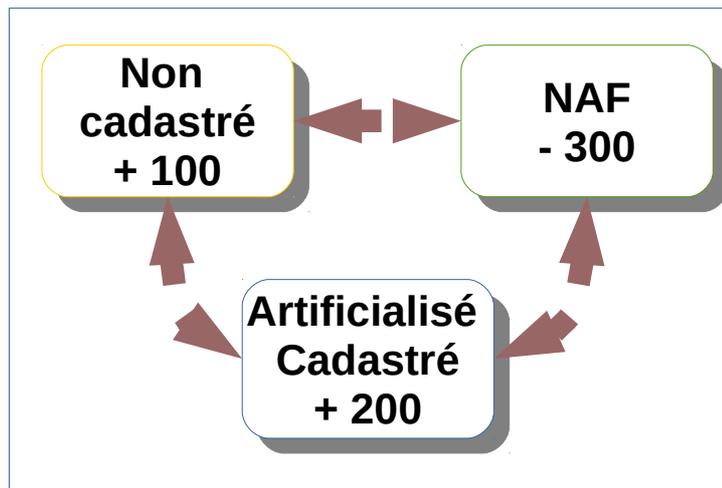
On postule ainsi que **pour un îlot donné, il ne peut y avoir deux flux contraires**. En d'autres termes, au sein d'un îlot, il ne peut y avoir à la fois un changement de NAF vers arti et d'arti vers NAF. Cela conduit à diviser par deux le nombre d'inconnues à rechercher, la moitié d'entre elles au moins étant systématiquement nulle.

Dans notre utilisation, et au vu de la granulométrie fine des îlots, cette hypothèse est peu contraignante. Cependant, poser cette hypothèse n'aurait pas pu être possible dans le cas de travaux à la communes, pour lesquelles il peut y avoir une partie de la commune qui s'artificialise pendant qu'une autre partie se renature.

Cela simplifie le problème ainsi

$$\begin{cases} \Delta \text{ Arti} + \Delta \text{ NAF} + \Delta \text{ NonCad} = 0 \\ \forall i \in [\text{ Arti}, \text{ NAF}, \text{ NonCad}], \Delta i = \sum_j \text{ Flux } j, i \text{ (2)} \end{cases}$$

Nous perdons donc la contrainte (1), mais diminuons le nombre d'inconnues.



Exemple avec Delta NonCad = +100, Delta NAF = -200 et Delta Arti = -100

Insuffisance de la définition : hypothèse de la simplicité des flux

A ce stade, les équations sont cependant encore liées. Nous avons donc encore deux équations pour 3 inconnues. A titre d'exemple, le système représenté sur le schéma ci-dessus s'écrit (attention aux signes !) :

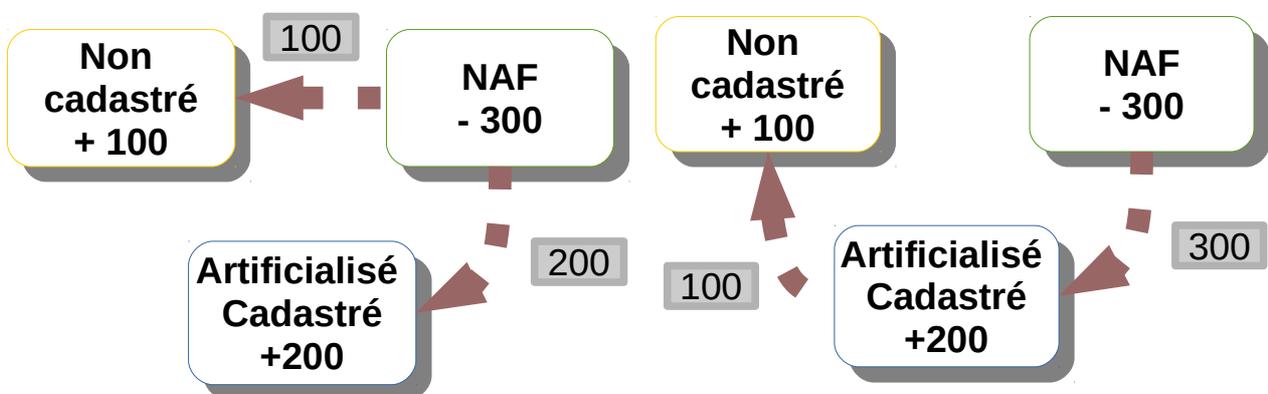
$$\begin{cases} -Flux\ ArtiNonCad + Flux\ NAFArti = 200 & (1) \\ Flux\ NonCadNAF - Flux\ NAFArti = -300 & (2) \\ Flux\ ArtiNonCad - Flux\ NonCadNAF = 100 & (3) \end{cases}$$

L'équation (3) est donc clairement égale à $-(1) - (2)$.

Il est nécessaire de postuler encore une nouvelle hypothèse pour obtenir une solution unique, à savoir que seuls des flux de même sens peuvent arriver dans une catégorie¹. En conséquence seuls deux flux seront différents de zéro (ou que l'un des flux sera égal à zéro)².

De manière plus thématique, il s'agit de dire qu'une catégorie qui gagne de la surface ne peut en perdre en même temps : il ne peut à la fois y avoir un transfert de NAF vers Arti, et de Arti vers le non cadastré.

Une autre manière formulation consiste à dire qu'entre plusieurs hypothèses de transformation, on privilégie la plus simple, à savoir que la parcelle ne change pas d'usage (ou se modifie le moins possible).



Type de schéma possible

Type de transfert rendu impossible par l'hypothèse de simplicité des flux

¹ Il semblerait que cette condition soit équivalente à minimiser la somme des valeurs absolues des flux. Cette affirmation mériterait cependant d'être démontrée.

² Attention, cette condition est la conséquence du postulat, mais n'y est pas équivalent.

Conséquences et mise en place de l'algorithme

Ces éléments étant posés, on calcule le système précédent en distinguant 3 cas :

- si les trois deltas sont nuls, tous les flux seront nuls (cas trivial)
- si un des deltas est égal à 0, on considère uniquement le flux entre les deux deltas non nuls.
- Si aucun des deltas est égal à 0, nous aurons deux deltas de signe positif, et l'un de signe négatif³. Dans ce cas, nous considérerons deux flux, partant tous deux du delta négatif.

En conséquence, l'algorithme Kaver calculera, en sortie, 6 flux, dont au moins 4 seront égaux à 0.

3 Ou alors un positif et deux négatifs. Cela est une conséquence directe du fait que la somme des deltas est égale à 0.

Annexe 4.2 : description de l'algorithme Korange

Principe de l'algorithme Korange

L'algorithme Korange est une version plus complexe de Kaver : nous avons en entrée 3 Deltas, mais leur somme n'est pas égale à 0. Ainsi, Korange est utilisé pour déterminer les transferts entre habitat, activité et mixte : il peut y avoir une augmentation de l'artificialisation de chacun des postes, ainsi que des transferts entre ceux-ci.

Dans ce contexte, nous devons déterminer 9 flux : 6 flux internes, sur le même modèle que Kaver, mais aussi 3 flux externes.

Formulation mathématique

À l'image de Kaver, Korange vise à répondre au problème suivant :

$$\begin{cases} \forall Flux, Flux \geq 0 \text{ (1')} \\ \forall i \in [Arti, NAF, NonCad], \Delta i = \sum_j Flux_{j,i} - \sum_k Flux_{i,k} + Flux_{ext,i} \text{ (2')} \end{cases}$$

Postulats

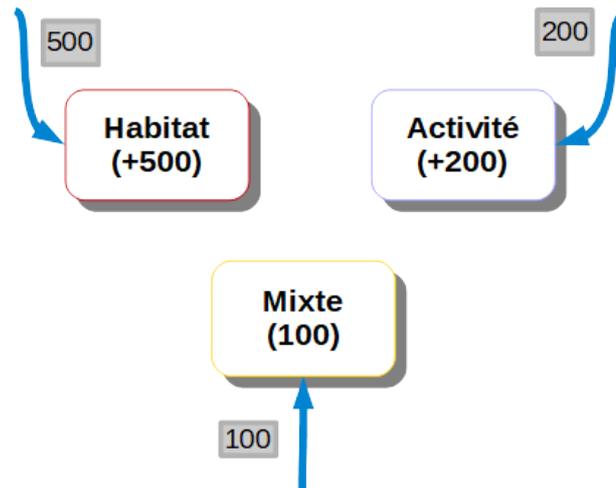
On réalise ici les mêmes postulats que pour Kaver, à savoir :

- deux flux contraires ne peuvent exister,
- hypothèse de simplicité des flux : on minimise le nombre de flux, et on postule qu'un usage ne peut avoir que des flux du même signe (tous les flux sont entrants ou tous les flux sont sortants).

Gestion des cas simples

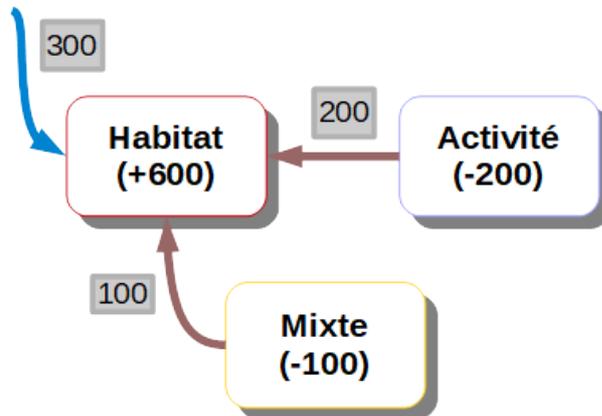
L'application des hypothèses supplémentaires de Kaver permet d'obtenir des solutions uniques dans les cas suivants :

- Cas 1 : tous les deltas sont du même signe : dans ce cas, on ne retrouve que les deltas externes. Si Delta Act = 200, Delta Hab = 500 et Delta mix = 100, on aura donc le schéma suivant ci-dessous. On constate ainsi l'absence de flux internes (transferts entre catégories). Ce cas est de loin le plus répandu.



Exemple de cas simple : tous les deltas sont de même signe

Cas 2 : un des deltas est positif, les deux autres sont négatifs, et le total est positif. Dans ce cas, l'apport d'artificialisation est assigné au delta positif. On applique ensuite Kaver pour déterminer les flux internes.



Exemple de cas simple : un des deltas est positif

**La gestion des cas complexes :
algorithme Kaunoir**

Les deux premiers cas peuvent se résoudre de manière non ambiguë avec les postulats de Kaver. Cependant, il existe un troisième cas pour lequel :

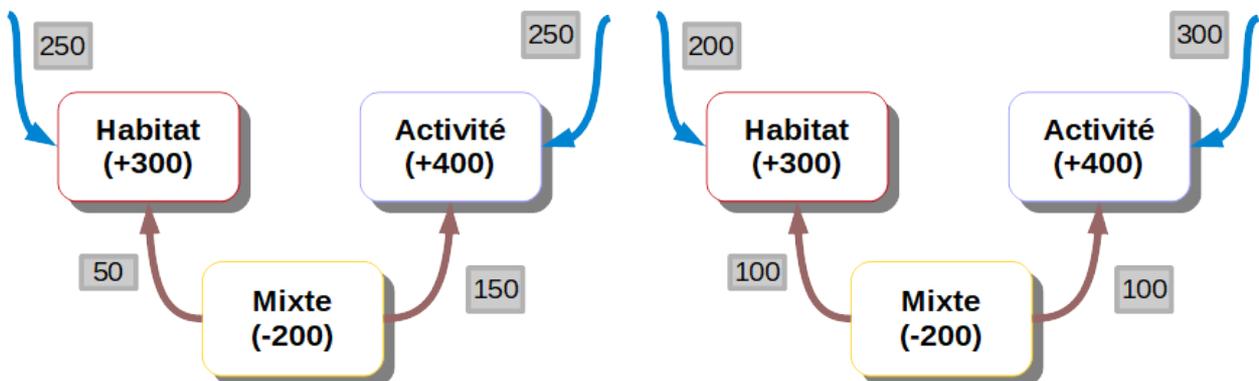
- deux deltas sont positifs,
- un delta est négatif,
- la somme des deltas est positive¹.

Pour ce cas, il existe plusieurs solutions pouvant répondre au problème, tout en respectant les contraintes.

Exemple de cas ambigu

Il peut notamment d'agir du cas suivant :

- Delta Hab = 300
- Delta Act = 400
- Delta Mix = -200.



Ces deux solutions sont deux exemples d'un résultat répondant aux postulats de Kaver.

¹ Cela est aussi valable pour le cas opposé.

Ajout d'une hypothèse supplémentaire

Ce cas doit donc être traité à part. Dans ce cadre, il est souhaitable d'ajouter des contraintes. Plusieurs solutions peuvent exister :

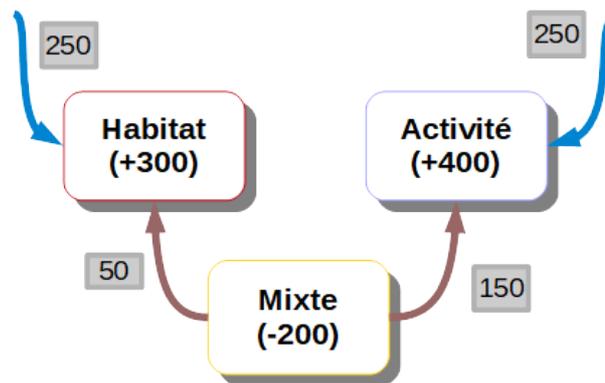
- il est possible de privilégier une catégorie parmi les autres : par exemple, on pourrait considérer qu'il serait souhaitable de maximiser le flux de mixte, puis le flux d'activité.
- De la même manière, il serait possible de privilégier certains flux internes par rapport à d'autres.

Cependant, pour des raisons thématiques, il n'y a aucune raison de privilégier une catégorie ou un flux par rapport à d'autres. Dans ce cadre, ces solutions ne seront pas mises en place. Il est donc souhaitable de mettre en place une solution symétrique, par exemple selon les flux suivants :

- le flux interne est réparti équitablement entre les autres postes (schéma de droite),
- le flux externe est réparti équitablement entre les postes (schéma de gauche).

Au final, les flux externes seront mis en avant, et resteront plus importants que les flux internes. Dans ce contexte, il est préférable de privilégier une approche équitable des flux externes.

L'algorithme prendra donc la somme des deltas, qu'il répartira équitablement entre les flux externes.

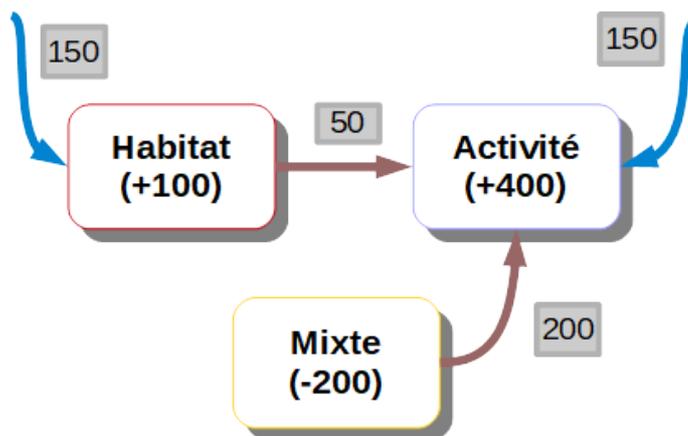


Solution choisie : le total des deltas est égale à 500, qui sont répartis équitablement entre les deux flux positifs (Habitat et activité)

Traitement des derniers cas particuliers

Dans certains cas, il n'est pas possible de répartir équitablement entre les catégories : c'est notamment le cas lorsque l'une des catégories est bien supérieure à l'autre.

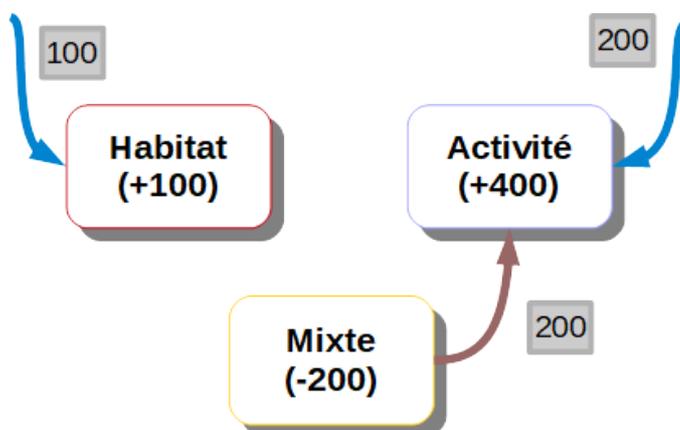
Par exemple, si dans l'exemple précédent l'habitat est égal à +100, si l'on répartit de manière équivalente les flux entrants, on obtient le résultat suivant :



Répartition égale des flux entrants. La composante habitat dispose ainsi à la fois de flux d'entrée et de sortie, ce qui ne répond pas à nos hypothèses.

Cette répartition viole l'un de nos postulats, à savoir que chaque poste ne devait avoir que des flux du même signe : ici, le flux extérieur abonde l'habitat, et un flux interne vient le ponctionner.

On considère donc que ces flux externes ne peuvent excéder le delta de la catégorie. Dans ce cas, la répartition sera donc la suivante :



Répartition finale des flux entrants

Annexe 5 : structure et contenu des données

Contenu des données

Les données sont séparées selon les 4 blocs suivants.

Idcom et idcomtxt : Identifiant communal

Chaque ligne de la donnée correspond à une commune. Cette liste de commune correspond aux communes INSEE 2017, idcom correspondant au code INSEE, idcomtxt au nom de la commune.

Il faut cependant noter les différences suivantes :

- les communes de l'île-de-Sein et de l'île-Molène¹ ne sont pas assujetties à l'impôt foncier. En conséquence, ne disposant d'aucune information sur ces territoires, ces communes ont été exclues du calcul.
- La commune de Suzan est fusionnée avec la commune de La Bastide-de-Sérou².
- Pour conserver un périmètre constant sur toute la période, les séparations de communes entre l'année 2009 et l'année 2017 ne sont pas prises en compte. En d'autres termes, si deux communes ont défusionnées en 2012, elles seront toujours indiquées comme fusionnées. Cela concerne 6 communes³.

Données d'identification de la commune

Pour chaque commune, nous disposons de leur appartenance aux zonages administratifs classiques (EPCI, département, région...), ainsi qu'aux zonages INSEE (unité urbaine...).

Ces éléments sont issus de données INSEE.

Données communales

En outre, les données contiennent les éléments liés à la population, aux ménages et aux emplois de la commune pour les années 1999, 2010 et 2015.

Données de consommation d'espaces

Enfin, les données de consommation d'espaces sont présentes, à savoir :

- les 6 flux d'artificialisation,
- les 9 flux d'usage,
- un flux sur le non affecté.

Chacun de ces flux est décliné entre deux années.

Deux produits en diffusion

Les données de consommation d'espaces sont disponibles en deux produits :

- le produit complet, comportant la totalité des champs,
- le produit de base les 57 colonnes les plus importantes. La liste de ces colonnes est disponible page suivante.

De manière générale, il est fortement recommandé d'utiliser le produit de base.

1 (codes INSEE 29 083 et 29 084)

2 La commune de Suzan (code INSEE 03904) est une enclave non délimitée au sein de La Bastide-de-Sérou (code INSEE 09042). Les services fiscaux, à l'inverse de l'INSEE, ne font pas la différence entre ces communes.

3 Avrecourt (52033) et Saulxures sont toujours considérés comme rattachés à Val-de-Meuse (52332), Chézeaux (52124) est toujours rattachée à Varennes-sur-Amance (52504), Laneuville-à-Rémy (52266) est considérée comme rattachée à La Porte du Der (52331), Lavilleneuve-Au-Roi (52278) est toujours rattachée à Autreville-sur-la-Renne (52031), Culey (55138) est toujours rattachée à Loisey (55298)

Données et nom des champs

Nom du champ	Description
idcom	identifiant INSEE 2017 de la commune
idcomtxt	Nom de la commune
idreg	identifiant de la région
idregtxt	nom de la région
iddep	identifiant du département
iddeptxt	nom du département
epci	identifiant de l'EPCI 2017
epcetxt	nom de l'EPCI 2017
au10	Identifiant de l'aire urbaine 2010 de la commune
typpopau10	tranche de population de l'aire urbaine
typau	typologie de la commune dans l'aire urbaine
uu	identifiant de l'unité urbaine
naf09art10	Flux entre NAF et artificialisé, sur la période 2009-2010
art09act10	Flux NAF vers artificialisé destiné à l'activité sur la période 2009-2010
art09hab10	Flux NAF vers artificialisé destiné à l'habitat sur la période 2009-2010
art09mix10	Flux NAF vers artificialisé destiné au mixte sur la période 2009-2010
naf10art11	idem sur la période 2010-2011
...	idem sur les autres périodes
nafart0917	total des flux entre naf et artificialisé sur la période 2009-2017
artact0917	Flux NAF vers artificialisé destiné à l'activité sur la période 2009-2017
arthab0917	Flux NAF vers artificialisé destiné à l'habitat sur la période 2009-2017
artmix0917	Flux NAF vers artificialisé destiné au mixte sur la période 2009-2017
artcom0917	Flux artificialisé divisé par la surface communales : % de surface communale convertie en artificialisée
pop10	Population 2010 (source INSEE)
pop15	Population 2015 (source INSEE)
pop1015	Variation de population entre 2010 et 2015
men10	Ménages 2010 (source INSEE)
men15	Ménages 2015 (source INSEE)
men1015	Variation des ménages entre 2010 et 2015
emp15	Emploi 2010 (source INSEE)
emp1015	Emploi 2015 (source INSEE)
emp1015	Variation de l'emploi entre 2010 et 2015
artpop1015	Artificialisation 2010-2015 par population supplémentaire
artmen1015	Artificialisation 2010-2015 par ménage supplémentaire
artemp1015	Artificialisation 2010-2015 par emploi supplémentaire
surfcom17	Surface communale



Cerema Nord-Picardie -44 Ter , rue Jean BART - CS 20275 - 59019 LILLE Cedex

Tél : 03 20 49 60 00 – Fax : 03 20 53 15 25 – Courriel : DTerNP@cerema.fr

Siège social : Cité des Mobilités – 25, avenue François Mitterrand – CS 92 803 – F-69674 Bron Cedex – Tél : +33 (0)4 72 14 30 30

Établissement public – Siret : 130 018 310 00313 – TVA Intracommunautaire : FR 94 130018310 – www.cerema.fr