

Marché SSP n° SSP-2013-090

**ANALYSE DU BESOIN DE PORTAGE DU FONCIER PAR DES CAPITAUX EXTÉRIEURS
VIA L'APPARIEMENT DU CADASTRE ET DU REGISTRE PARCELLAIRE GRAPHIQUE
(RPG)**

Rapport méthodologique

relatif au croisement du Registre parcellaire graphique avec les fichiers du
cadastre dans les 4 départements d'étude :

Charente-Maritime, Mayenne, Nord, Oise

Cette étude est financée par le Programme 215 du MAAF.

Ce document n'engage que ses auteurs et ne constitue pas le point de vue du
MAAF.

Décembre 2015



Agence de Services
et de Paiement



Sommaire

| | |
|---|-----------|
| Introduction : objectifs et principes généraux du croisement RPG-Cadastre | 4 |
| I. Origine et import des données dans la base | 5 |
| I.1. Données à disposition pour l'analyse de la propriété agricole et des exploitations..... | 5 |
| I.1.1. Le Registre Parcellaire Graphique (RPG) anonyme | 5 |
| I.1.2. Les données du cadastre | 6 |
| I.1.3. Données complémentaires mobilisées marginalement pour l'étude..... | 7 |
| I.1.3. Bilan des données mobilisées pour l'analyse de la propriété agricole dans les 4 départements 17, 53, 59 et 60 | 8 |
| I.2. Logiciels utilisés | 9 |
| I.3. Import des données et structuration de la base | 9 |
| I.3.1. Paramètres du serveur Windows et de la base de données..... | 9 |
| I.3.2. Unicité des tables pour les 4 départements..... | 10 |
| I.3.3. Projection géographique | 10 |
| I.3.4. Encodage | 10 |
| I.3.5. Import et corrections apportées aux données littérales..... | 11 |
| I.3.6. Import des données spatiales | 12 |
| I.3.7. Indexation sur les champs clés après import | 14 |
| I.3.8. Bilan des opérations effectuées pour réaliser l'import des tables contenant les données de base | 15 |
| II. Traitement des données : appariement du RPG et du cadastre | 17 |
| II.1. Traitement sur les comptes de propriété..... | 19 |
| II.1.1. Dédoublonnage des comptes réunissant les mêmes individus et partageant le même droit (tous types de personnes confondus) | 20 |
| II.1.2. Traitements spécifiques aux personnes physiques..... | 21 |
| II.1.3. Traitements spécifiques aux personnes morales | 22 |
| II.1.4. Nettoyage des parcelles affichant des droits de propriété non anticipés à la construction de la base..... | 23 |
| II.1.5. Typologie finale des ayants droit | 24 |
| II.2. Croisement des données spatiales au format vecteur..... | 25 |
| II.2.1. Principes généraux, opérateurs et fonctions utilisés | 25 |
| II.2.2. Croisement des couches vecteur via l'opérateur @ | 26 |
| II.2.3. Croisement des couches vecteur via l'opérateur ST_intersection..... | 26 |
| II.2.4. Nettoyage des micro-intersections dans l'échantillon obtenu | 30 |

| | |
|---|-----------|
| II.3. Sélection des exploitations dont au moins 90 % de la surface est retrouvée dans le département | 34 |
| II.4. Evaluation du croisement spatial entre localisants parcellaires et RPG | 34 |
| III. Méthodologie du croisement de 2 millésimes du RPG..... | 42 |
| III.1. Import des données littérales dans le tableur | 42 |
| III.2. Croisement de 2 fichiers du RPG de millésimes différents | 43 |
| III.2.1. Fichiers utilisés | 43 |
| III.2.2. Formats et logiciels utilisés | 44 |
| III.2.3. Système de projection..... | 44 |
| III.2.4. Jointure de chacune des couches avec les fichiers de données associées | 44 |
| III.2.5. Jointure spatiale. | 45 |
| III.2.5. Analyse du fichier résultant..... | 45 |
| IV. Principe de calcul de l'indice de dispersion spatial des îlots d'exploitation* | 47 |
| | |
| Conclusion : perspectives d'amélioration du croisement entre propriété et exploitation | 48 |
| Recommandations pour une généralisation du croisement RPG-Cadastre à la France entière | 49 |
| Lexique | 51 |
| | |
| ANNEXE1: Erreurs rencontrées lors de l'import du fichier Majic des parcelles non-bâties de la Mayenne 2013..... | 52 |
| ANNEXE 2: Extrait de la table finale obtenue par appariements successifs des bases de données pour une exploitation de Charente-Maritime | 54 |
| ANNEXE 3 : principaux cas de figure des croisements spatiaux RPG avec les localisants parcellaires et conséquences | 55 |

Ce rapport méthodologique accompagne le Rapport de synthèse de l'étude donnant les résultats du croisement des bases de données.

Introduction : objectifs et principes généraux du croisement RPG-Cadastre

L'objectif général de l'étude est d'évaluer les besoins en portage du foncier agricole et d'actualiser les connaissances sur la structure de la propriété foncière des exploitations. Cet objectif implique de relier à chaque exploitation les listes de titulaires de droits (pleine-propriété, usufruit, nue-propriété) et la surface sur laquelle chacune de ces listes exerce ses droits. Le croisement du Registre parcellaire graphique avec les fichiers du cadastre permet ainsi de cumuler sur une même surface les informations relatives à sa propriété foncière (fournies par le cadastre) et à l'exploitation agricole qui la met en valeur (fournie par le Registre parcellaire graphique).

A l'issue d'un premier travail de sélection, 4 départements de test ont été retenus, présentant des caractéristiques foncières contrastées et une vectorisation (quasi) complète du cadastre : la Charente-Maritime, la Mayenne, le Nord et l'Oise.

Dans ces 4 départements, la jointure spatiale initiale est réalisée entre les éléments de base de la BD Parcellaire (les parcelles cadastrales) et les éléments de base du Registre parcellaire graphique (les îlots de culture déclarés à la PAC par chaque exploitant). Le croisement de ces 2 couches donne un certain nombre d'intersections que des champs clés permettent ensuite de relier à un ensemble de données :

- relatives à l'exploitation : le champ clé « numéro d'îlot » permet de faire la jointure avec les informations de niveau 3 et 4 du RPG, à savoir la forme juridique de l'exploitation, sa surface totale déclarée, son identifiant anonyme,...
- relatives à la propriété, en 2 étapes successives :
 - le champ clé « code parcellaire » obtenu par concaténation, permet d'abord de faire la jointure, dans le fichier littéral du cadastre non bâti, avec le compte de propriété communal.
 - le compte communal de propriété permet ensuite de faire la jointure, dans le fichier littéral des propriétaires, avec l'identité des ayants droit de chaque compte, et avec les droits exercés par chacun d'eux.

Au cours de ces étapes, un certain nombre de dédoublonnages et de redressements sont opérés sur les différents champs pour parvenir à une table finale la plus cohérente et complète possible, tant en terme de surfaces recensées que d'identification des ayants droit.

Le présent rapport méthodologique, complémentaire du Rapport de synthèse de l'étude, présente l'ensemble de la démarche d'analyse et d'appariement de ces différentes tables, jusqu'à l'obtention du fichier final faisant la relation entre exploitations, propriétaires et surfaces détenues :

- Dans une première partie sont présentées les données à disposition pour l'étude et les procédures d'import pour la constitution d'une base de données PostgreSQL ;
- Une fois la base de données constituée, la seconde partie présente les différents traitements effectués, qu'il s'agisse de manipulations intrinsèques à certaines tables, pour expliciter les

informations qu'elles contiennent (fichier des propriétaires notamment), ou qu'il s'agisse des appariements successifs entre elles.

L'essentiel des travaux détaillés a pour point de départ le croisement effectué entre les couches vecteur des îlots de culture et les couches vecteur des parcelles. Pour compléter cette approche, une évaluation du croisement entre les îlots et les localisants parcellaires (ponctuels géolocalisant la parcelle, même si elle n'est pas vectorisée) a été effectuée (partie II.3) pour estimer les biais et envisager une généralisation du croisement à l'ensemble du territoire ;

- Enfin, en complément du croisement entre RPG et cadastre, est présentée dans la 3^{ème} partie la méthodologie utilisée pour croiser 2 millésimes du RPG, et introduire dans l'analyse des éléments de dynamique des exploitations d'une année à l'autre.

I. Origine et import des données dans la base

I.1. Données à disposition pour l'analyse de la propriété agricole et des exploitations

I.1.1. Le Registre Parcellaire Graphique (RPG) anonyme

Origine et nature du RPG

Le RPG anonyme c'est :

- Le recueil chaque année des informations de plus de 390 000 agriculteurs déclarants, sur les exploitations agricoles et leurs cultures.
- Plus de 85 % du territoire agricole représenté.
- Des registres annuels, informatisés depuis 2007:
- Des registres mis à dispositions des organismes publics selon 4 niveaux d'information
- Pour chaque exploitation le niveau 4 (maximum) fournit :
 - Un identifiant anonyme,
 - Sa forme juridique,
 - Sa surface totale,
 - Son département de rattachement,
 - Pour les exploitations individuelles la classe d'âge du chef d'exploitation.
- Pour chaque îlot de culture :
 - La forme, la surface et l'emplacement.
 - Le nom de la commune de sa localisation.
 - Les cultures qui y sont pratiquées

Il y a un fichier par département et par année. L'année est celle de la déclaration faite par les agriculteurs. Les départements utilisés pour cette étude sont ceux de la France métropolitaine.

Limites

- Il ne concerne que les exploitations consignées, c'est-à-dire faisant l'objet d'une déclaration en vue de percevoir des aides de la Politique Agricole commune (PAC) liées à la surface. Il suffit qu'une exploitation ait une seule surface éligible (c'est-à-dire de culture annuelle ou toujours en herbe), pour que l'intégralité de l'exploitation soit

consignée. Par contre les exploitations intégralement orientées vers des cultures pérennes ne sont pas enregistrées dans le RPG. La représentativité du RPG par rapport au RA sera donc très grande dans les départements quasi dépourvus de cultures pérennes et ira décroissant selon la couverture de vergers, vignes,... , des autres départements.

- Les informations venant du RPG niveau 4 donnent un nombre restreint d'informations sur l'exploitation et l'exploitant ainsi que sur les cultures pratiquées sur les îlots (pas d'information sur les productions animales par exemple).
- L'identifiant de l'exploitation étant anonyme, il ne permet pas de faire des recoupements avec des informations issues d'autres fichiers. L'identifiant d'une même exploitation est par ailleurs différent d'un millésime au suivant.

Cette dernière difficulté peut cependant être contournée. En effet, chaque mètre carré de terrain garde le même emplacement d'un millésime à l'autre. En effectuant une jointure spatiale entre deux registres d'années différentes, il est possible de recouper les informations associées aux polygones, et ainsi tracer l'évolution d'un terrain agricole dans le temps, (Cf. techniques de traitement des informations).

Cadre juridique d'utilisation

Les registres sont mis à disposition des organismes publics à travers une convention précisant l'année, le ou les départements, le niveau d'information (1 à 4).

I.1.2. Les données du cadastre

Origine et nature des fichiers

Les données cadastrales accessibles sont d'ordre cartographique d'une part (BD Parcellaire produite par l'IGN par reprojection et standardisation du plan cadastral informatisé (PCI) de la DGFIP qui est quant à lui établi par section cadastrale et littérales d'autre part (fichiers MAJIC des propriétés non bâties et des propriétaires, produits par la DGFIP).

La BD Parcellaire présente l'intérêt d'être superposable à la BD Ortho de l'IGN (les deux données font partie du référentiel à grande échelle), alors que le RPG est construit à partir de dessin sur la BD Ortho, ce qui facilite le travail de superposition des différentes couches et de rapprochement entre exploitations et propriétaires.

La principale limite de la BD Parcellaire est que la vectorisation des parcelles du plan cadastral est à des niveaux d'avancement très variables selon les départements et régions. Les dernières informations diffusées par la DGFIP indiquent qu'environ $\frac{3}{4}$ des parcelles représentant $\frac{2}{3}$ de la surface sont vectorisées à la date de réalisation de cette étude.

Si le PCI n'est pas encore vectorisé, la BD Parcellaire du même secteur ne l'est pas non plus. Par ailleurs, le délai entre vectorisation du PCI et diffusion de la BD Parcellaire vecteur qui en découle peut atteindre plusieurs années. A défaut de vectorisation, la BD Parcellaire contient dans tous les cas la couche d'information géographique des ponctuels localisants (« localisants parcellaires »). A priori, le travail de superposition de la BD Parcellaire avec le RPG donnera de meilleurs résultats avec les polygones vectorisés plutôt qu'avec les localisants parcellaires car il permet de travailler sur la

géométrie exacte des parcelles. Dans le cadre de l'étude, et en vue de comparer les 2 méthodes, il apparaît donc essentiel de choisir des départements de test entièrement vectorisés.

L'état récent d'avancement de la vectorisation du cadastre constitue donc une donnée importante dans le choix des zones d'études.

L'appariement entre la BD Parcellaire et les fichiers littéraux des propriétés non bâties peut être effectué en rapprochant les identifiants clés uniques des parcelles (référence cadastrale). La qualité des fichiers littéraux est supposée homogène sur le territoire et n'est pas spécifiquement rentrée en ligne de compte dans le choix préalable des zones d'étude.

Cadre juridique d'utilisation

La FNSafer est titulaire de droits sur toutes ces données, les 4 Safer concernées ayant donné leur accord pour l'utilisation de la BD Parcellaire.

La FNSafer a déclaré la conformité du traitement à l'autorisation unique AU-001 sur les systèmes d'information géographique après consultation de la CNIL.

I.1.3. Données complémentaires mobilisées marginalement pour l'étude

D'autres données ont été mobilisées pour l'étude, principalement pour le choix des départements de test. Elles ne sont pas rentrées en ligne de compte pour l'analyse de la propriété à partir du RPG et du cadastre, faute d'identifiant permettant leur appariement.

Données du recensement agricole

Nous avons privilégié les données du recensement le plus récent, effectué en 2010, mais pour mieux cerner l'évolution du foncier, nous avons également utilisé les données de 1988 et 2000.

Ces informations sont intéressantes à plus d'un titre :

- Le recensement est exhaustif, il concerne toutes les exploitations agricoles petites ou grandes.
- Il donne des informations à la fois sur les exploitants (âge, sexe), les exploitations, (superficie, productions), leur dimension économique, le fonctionnement de ces dernières (main d'œuvre, forme juridique, modes de faire valoir), et sur leur emplacement (commune du siège de l'exploitation).

Toutefois, faute d'identifiant commun entre deux recensements, par ailleurs très espacés dans le temps, le recensement agricole n'est en mesure ni de tracer l'évolution d'une exploitation, ni d'indiquer son devenir en matière de cession et de transmission, ou en matière de propriété foncière.

Données issues du marché foncier.

Les données du marché foncier notifié aux Safer constituent également des données indirectes pour mieux connaître la structure de la propriété. Le rapport de la taille moyenne des biens agricoles mis en vente sur la taille moyenne des exploitations est par exemple un indicateur de concentration/dispersion de la propriété par rapport à l'exploitation, phénomène variable selon les régions.

Les données de prix caractérisent également les régions sous forte pression foncière, que ce soit pour l'accès à des terres agricoles dégagant de fortes valeurs ajoutées, ou dans la perspective d'anticiper le changement d'usage des terrains, à proximité des agglomérations et du littoral notamment.

La fréquence des lots bâtis vendus libres est également une donnée à prendre en considération, en cela qu'elle renseigne des opportunités d'installation hors cadre familial pour des agriculteurs.

Compilant un certain nombre de ces indicateurs, l'article *Les marchés fonciers ruraux régionaux entre dynamiques des exploitations agricoles et logiques urbaines* (Levesque, Liorit, Pathier, 2011, Economie et Statistique n°444-445) peut fournir des éléments de synthèse en termes de grandes dynamiques foncières régionales.

I.1.3. Bilan des données mobilisées pour l'analyse de la propriété agricole dans les 4 départements 17, 53, 59 et 60

- Registre parcellaire graphique de niveau 4 2012 (déclaré courant 2012 pour diffusion en 2013 à l'extérieur de l'ASP) en projection Lambert 93 ;
- Fichiers de données du RPG au format csv (niveau 2, îlot et niveau 4, exploitation) ;
- BD Parcellaire dans le dernier état disponible fin août 2014 : livraison 2012 pour 17 et 2013 pour 53, 59 et 60. Pour chaque département, il s'agit des couches suivantes, toutes en projection Lambert 93 :
 - o couche shp surfacique des parcelles dans les zones vectorisées,
 - o couche shp ponctuelle des localisants,
 - o couche shp surfacique des limites communales.

La BD Parcellaire de l'IGN est un produit dérivé du cadastre PCI édité par la DGFIP et le délai de production et de diffusion d'une version de la BD Parcellaire peut être assez long. En conséquence, une version de la BD Parcellaire postérieure à l'année du cadastre étudié (fichiers Majic) est souvent plus proche de la situation réelle à la date voulue que la BD Parcellaire datée de l'année du cadastre.

- Données Majic III du 1^{er} janvier 2013 pour les propriétaires, texte brut (fichier à plat) ;
- Données Majic III du 1^{er} janvier 2013 pour les propriétés non bâties, au format texte brut (fichier à plat).

Cf. Documentation des fournisseurs de ces données (voir espace pro IGN et site ASP) en annexe.

I.2. Logiciels utilisés

Les données sont importées sur une station de travail ayant la configuration suivante :

- Station Dell Precision T7500 (1 processeur Xeon 4 cœurs, 12 Go de RAM, carte graphique ATI FirePro),
- Windows server 2008 R2,
- Base de données PostgreSQL 9.2/PostGIS 2.0 (mise à jour en version 2.1 en cours d'étude),
- Requêtes SQL dans Pgadmin III 1.16.1,
- SIG pour la visualisation de données : QGIS 2.0 connecté à la base PostGIS.

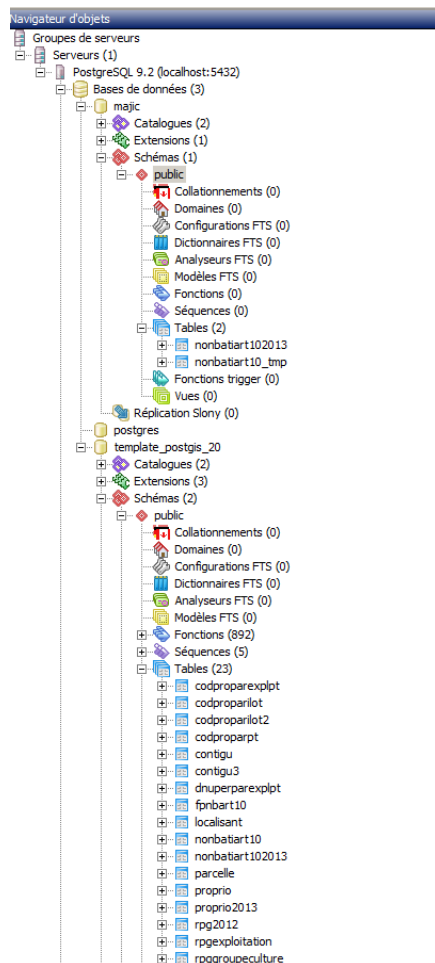
Les plugins suivants ont été utilisés :

- PostGIS shapefile and dbf loader 2.0 pour importer les données géographiques dans la base PostgreSQL, qui est installé par défaut avec Pgadmin ;
- Le plug-in « cadastre » de QGIS pour importer des fichiers Majic (testé puis abandonné).

I.3. Import des données et structuration de la base

I.3.1. Paramètres du serveur Windows et de la base de données

Toutes les données de base sont importées à partir d'un même dossier. Les données modifiées ou exportées sont également dans ce dossier.



L'utilisateur SCALCUL a tous les droits sur les fichiers et le dossier. On a activé l'option d'affichage de l'extension des fichiers (.txt, etc.).

Dans PostGIS, les données sont distribuées dans l'arborescence suivante : serveur/ base de donnée/schéma/ table. Nous avons choisi d'utiliser la base de données créée par défaut dans PostGIS, à savoir « template_postgis_20 » et le schéma créé par défaut dans cette base, appelé « public ». Les tables utilisées pour l'étude sont dans ce schéma (cf. figure 1)

Figure 1 : Arborescence de la base de données

Paramètres principaux de la base de données :

- Utilisateur : postgres
- Encodage serveur de la base en UTF-8, sachant que quelques conversions sont nécessaires pour importer Majic et RPG (voir I.4.5.) ;
- Les paramètres LC_COLLATE (ordre de tri de données alphanumériques) et LC_CTYPE (classification du jeu de caractères) sont ceux adaptés aux caractères français : FrenchFrance.1252 ;
- L'encodage client de la base de données est Win1252 (car les fichiers texte utilisés comme source sont dans cet encodage d'après les propriétés des fichiers) ;
- Les autres paramètres sont standard (tablespace : pg_default, limite de connexion : -1 (ie nombre de connexions illimité)).

Concernant les données spatiales, l'option par défaut « géométrie » dans PostGIS (coordonnées X et Y en m) est plus adaptée que l'option « géographie » (latitude et longitude en degrés, minutes et secondes).

I.3.2. Unicité des tables pour les 4 départements

Toutes les couches départementales de même type sont importées dans une même table, afin de simplifier les requêtes. Exemple : les 4 fichiers *shape* du RPG qui ont un nom de la forme SURFACES-2012-ILOTS_ANONYMES_code département_année/mois/jour de création du fichier sont importés dans une unique table spatiale rpg2012.

Toutes les tables sont dans le même schéma (public) une fois l'import réussi.

I.3.3. Projection géographique

On retient le système de projection (Srid) 2154 correspondant au Lambert93 (obligatoire en France depuis 2010) pour toutes les couches géographiques. Cela a été vérifié au préalable dans les fichiers .prj (description de la projection) accompagnant les fichiers *shape*. Il suffit de copier/coller le contenu du .prj sur le site web suivant <http://prj2epsg.org> pour se voir proposer un numéro de Srid.

Remarque : toutes les couches utilisées dans une base PostGIS qu'on souhaite croiser entre elles doivent être dans la même projection.

I.3.4. Encodage

Ce point est très important car un mauvais choix lors de l'import de données bloque l'opération. Dans d'autres cas, même si l'import semble réussir, on peut avoir des caractères spéciaux non reconnus.

On choisit un encodage serveur lors de la création d'une base de données. On peut modifier l'encodage client avant d'importer des données.

Le module d'import de fichiers *shape* dans PostGIS impose de définir l'encodage dans les options.

Remarque : UTF-8 est l'encodage par défaut de PostgreSQL car il permet de coder les caractères de la plupart des langues à alphabet latin sur un nombre réduit d'octets, Win1252 est celui utilisé par les

systèmes Windows pour les langues d'Europe de l'Ouest, LATIN1 est l'autre nom de l'ISO 8859-1 qui est également un encodage adapté aux langues d'Europe occidentale.

Le retour d'expérience montre que selon les paramètres du SGBD PostgreSQL/PostGIS, l'import direct du fichier des parcelles dans une base encodée en UTF-8 avec un encodage client en Win1252 ne fonctionne pas toujours.

La technique la plus simple est alors de créer une seconde base de données encodée en Win1252 et d'y importer les données Majic du non bâti. Pour que cela fonctionne bien, il est préférable que le schéma dans lequel on crée la table porte le même nom dans les 2 bases. On exporte ensuite la table du non-bâti :

- soit en créant un fichier *csv* qu'on importe ensuite dans la base *template_postgis_20*, en essayant si nécessaire de changer l'encodage à l'export ou à l'import ;
- soit en faisant dans Pgadmin une sauvegarde (clic droit sur la table dans l'explorateur) – on appelle cela faire un « dump » dans le jargon PostgreSQL - puis en restaurant le fichier (au format archive tar encodée en UTF-8 par exemple) dans la base de données *template_postgis_20*.

Dans le cadre de l'étude, la correction des erreurs sur les caractères qui n'auraient pas dû être présents dans l'encodage réel des fichiers a suffi, il n'y a pas eu besoin de recourir à cette technique.

I.3.5. Import et corrections apportées aux données littérales

Le code exécuté pour l'import et pour l'ensemble des traitements ultérieurs sur les tables est réalisé dans la fenêtre « édition » de Pgadmin. Ce code s'applique à la base de données sur laquelle on est connecté dans l'explorateur de Pgadmin. En cas de manipulation de plusieurs bases de données, il faut donc veiller à être connecté à la base de données voulue.

Pour les tables, on évite des problèmes en ajoutant toujours le nom du schéma dans le code SQL (ex. pour la table « localisant » dans le schéma « public » on indiquera « public.localisant »). On sauvegarde les traitements faits dans des fichiers *.sql*. Le code SQL est commenté dans ces documents, qui sont par ailleurs lisibles avec un éditeur de texte tel Wordpad.

Même si l'import peut *a priori* fonctionner car les encodages sont compatibles entre eux, PostgreSQL arrête l'import dès qu'il détecte une erreur dans le fichier texte, c'est-à-dire dès qu'une ligne présente un format ou une structure ne correspondant pas aux paramètres décrits au début du script. Dans ce cas, aucun fichier n'est créé. Par contre, le journal de l'éditeur SQL indique un code d'erreur et le numéro de la ligne du fichier qui comporte la première erreur et propose parfois une solution (*Hint*=allusion).

Si le fichier texte ou *csv* comporte plusieurs erreurs, il faut les corriger une par une en éditant le fichier texte, puis relancer autant de fois que nécessaire le script d'import jusqu'à débogage complet.

A titre d'exemple, les erreurs rencontrées lors de l'import du fichier Majic des parcelles non-bâties de la Mayenne 2013 sont listées en annexe.

Dans les données de l'étude, les erreurs d'import ne concernent que le fichier MAJIC du non-bâti. Rien à signaler pour le fichier MAJIC des propriétaires, les données littérales du RPG (format csv et dbf) et de la BD Parcellaire (dbf).

Scripts d'import des fichiers Majic

Les scripts SQL d'import de Majic sont des adaptations des scripts proposés dans les guides méthodologiques Cete et Certu publiés en 2008 et 2009 pour tenir compte des paramètres de la base de données et des évolutions des fichiers Majic de :

- 2012 : ajout du code et libellé de voie et des parcelles-mères et type de filiation pour les parcelles,
- 2013 : ajout du code de forme juridique des comptes de propriété.

Les fichiers Majic bruts ont une structure particulière. Ils contiennent dans un même fichier des données de type différent et il faut faire un tri ligne par ligne en fonction de la valeur d'un « code article » au moment de l'import en base de données. En supprimant la première ligne de chaque fichier (appelée en-tête de direction), on simplifie l'import. Pour le fichier des parcelles, on ne garde que les parcelles (code article=10) mais en ayant préalablement importé une table « non-bâti » contenant tous les types de données (parcelles, subdivisions fiscales des parcelles, exonération de taxation).

Remarque : Les formats des différents champs utilisés pour créer des clés entre tables correspondent : ils comprennent tous des 0 pour remplir les vides, sauf pour les sections cadastrales du fichier Majic des parcelles. On harmonise ce dernier avec les autres avec une fonction SQL qui remplace les vides par des 0, le but étant d'avoir un code de parcelle compatible avec celui de la BD Parcellaire.

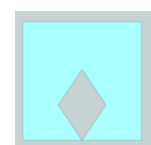
I.3.6. Import des données spatiales

Le comportement du SGBD sur les données spatiales est différent de celui observé lors de l'import des données littérales. Du moment que la donnée spatiale est conforme au format décrit (*shape* en l'occurrence), elle est importée, même si elle n'est pas conforme au standard OGC attendu par PostGIS pour faire les calculs.

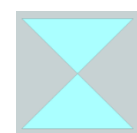
Import de polygones non-conformes

On trouve en effet des polygones importés de type :

- « *ring self-intersection* », cela correspond à un polygone à l'intérieur duquel on exclut un polygone plus petit. 356 objets de ce type dans le RPG (environ 9,2 ha de surface moyenne après redressement), soit environ 1 objet sur 1 000. Ces objets sont disséminés dans tout l'espace et correspondent souvent à des polygones complexes (probablement des îlots de cultures entourant des bâtiments d'exploitation, des surfaces boisées ou en eau) ;



- « *self-intersection* », 3 exemples dans la BD Parcellaire vecteur, soit environ 1 par million de parcelles.



La présence de polygones de ce type dans la couche importée implique que :

- les opérations sur les polygones non conformes sont fausses (notamment le calcul de la surface avec la fonction `st_area(geom)` ;
- le croisement entre couches n'aboutit pas et aucun code d'erreur ne donne d'explication. Il faut arrêter le script à la main.

NB : La couche des localisants de la BD Parcellaire image, qui n'est composée que de points, ne comporte évidemment aucune erreur de type de géométrie.

Génération d'objets non conformes par croisement de couches spatiales

Après une opération de croisement de couches spatiales, on peut obtenir des objets de type différent de ceux des objets en entrée. Ainsi, on s'attendrait à ce que le calcul de l'intersection de deux couches de polygones ou multipolygones (objets décrits par plusieurs polygones disjoints) donne des objets équivalents. Or on peut obtenir des points, des multilignes, voire des *geometrycollections* (ensemble d'objets hétérogènes comme un point et un polygone, une ligne et un polygone, etc.). La couche issue du croisement doit être nettoyée car les opérations sur les couches résultats (notamment `st_contains` ou `st_intersects`) plantent, notamment en présence de *geometrycollections*. Par contre, le journal de l'éditeur détecte ces erreurs et les explique.

Exemple, la fonction `ST_Intersects(a.geom, b.geom)` qui donne en résultat un booléen vérifiant si les géométries des objets a et b s'intersectent ou non, peut renvoyer l'erreur suivante :

```
ERREUR: Geometry could not be converted to GEOS: IllegalArgumentException: Invalid number of points in LinearRing found 3 - must be 0 or >= 4
```

```
État SQL :XX000
```

Cela signifie qu'une partie des polygones n'a pas de contour fermé.

Identification des erreurs et correction

PostGIS présente des outils utiles pour vérifier la qualité des couches géométries (ce type de champ est nommé `geom` par défaut) :

- `st_IsValid(geom)` indique si une couche géométrie est conforme au standard OGC. Eliminer les lignes qui ne remplissent pas cette condition est un premier tri ;
- `st_IsValidReason(geom)` indique si la géométrie est correcte et si ce n'est pas le cas donne une explication ;
- `st_IsSimple(geom)` permet de vérifier si une géométrie est simple : cela permet notamment de détecter les self-intersections ;
- `st_AsType(geom)` renvoie le type d'objet (POINT, POLYGON, MULTIPOLYGON, GEOMETRYCOLLECTION, etc.) du champ `geom`. Il suffit de faire un tri sur les résultats de cette fonction pour ne garder que les objets de type désiré ;
- `geometrytype(geom)` traite la suite de caractères du champ « geom » pour le formater en expression du genre POLYGON (X1 Y1 X2 Y2 etc.), Point (X Y), LINestring (X1 Y1 X2 Y2 etc.), GEOMETRYCOLLECTION (POINT X Y POLYGON X1 Y1 X2 Y2). Sa principale utilité dans le cadre de notre étude est de permettre de voir quels types d'objets sont présents dans les *geometrycollections* ;

-st_MakeValid permet de « réparer » les géométries non valides en les convertissant. Il faut vérifier les résultats de la réparation avant de les utiliser ;

Tableau 1 : Nombre d'erreurs détectées et corrections réalisées sur les couches géographiques

| Couche de données de départ | Nombre d'erreurs | Couche de données valide après vérification |
|---|--|---|
| rpg2012 325 310 lignes | 356 lignes avec erreur du type <i>Ring Self-intersection</i> | rpg2012ok : 324 954 lignes |
| parcelle2013 4 933 964 lignes | 4 erreurs : 3 avec erreur du type <i>Self-intersection</i> et 1 avec erreur du type <i>Invalid number of points in LinearRing found 3 - must be 0 or >= 4</i> » | Parcelle2013ok 4 933 960 lignes |
| localisant2013 4 956 234 lignes | 0 | Localisant2013 |
| commune | 0 erreur sur les objets, 2 recouvrements partiels (474 et 173 m ²) dans l'Oise, négligeables par rapport aux 2,38 Mha de surface totaux | Commune 2 076 lignes |

1.3.7. Indexation sur les champs clés après import

Après l'import de chaque couche, un index sur le champ clé est créé par script SQL (on peut avoir des blocages si on les crée lors de l'import, notamment lorsqu'on associe les données de plusieurs départements dans une seule table de données spatiale) :

- numéro d'îlot pour les données du RPG (îlot et exploitation),
- identifiant parcellaire et code du compte communal de propriété pour les parcelles du cadastre,
- identifiant parcellaire pour les parcelles de la BD Parcellaire,
- code du compte communal de propriété pour les propriétaires,
- géométrie pour toutes les données géographiques.

NB : une même table peut comporter plusieurs index, notamment un sur la géométrie et un ou plusieurs sur les données littérales.

Les clés sont construites de la manière suivante :

- **code de parcelle (codpar) : depcom | | préfixe | | section | | numéro de plan**, pour pouvoir réaliser le lien entre BD Parcellaire et fichier Majic ;
- **compte de propriété communal (codpro) : depcom | | dnupro**, pour pouvoir réaliser le lien entre parcelle et compte de propriété

I.3.8. Bilan des opérations effectuées pour réaliser l'import des tables contenant les données de base

Tableau 2 : Création des tables contenant des données littérales uniquement

| Nom de la table ⁽¹⁾ | Donnée | Nom de l'algorithme SQL ou du plugin | Caractéristiques de l'import |
|---------------------------------|---|---|---|
| proprio2013 | Majic propriétaires | import Majic Ok | <ul style="list-style-type: none"> - Encodage client UTF-8 ou Win1252 - Enlever première ligne puis renommer chaque fichier départemental auparavant. - Import direct dans template_postgis_20/public |
| nonbatiart102013 | Majic parcelles (lignes ayant le code article 10 du fichier des parcelles non bâties) | import Majic Ok | <ul style="list-style-type: none"> - Encodage client : Win1252 - Enlever première ligne puis renommer chaque fichier départemental auparavant. - Import direct dans template_postgis_20/public - Import de l'ensemble des lignes hors en-tête direction dans un fichier puis sélection sur l'article 10 pour création |
| rpgexploitation2012 | csv du niveau 4 du RPG (numéro d'exploitation, surface déclarée, forme juridique, numéro d'îlot...) | mise en forme des fichiers Majic (proprio et non bâti) et csv du RPG, créations de clés et d'index ok | |
| rpggroupeculture2012 | csv du niveau 2 du RPG (numéro d'îlot, surface déclarée, groupe de culture) | mise en forme des fichiers Majic (proprio et non bâti) et csv du RPG, créations de clés et d'index ok | Importé mais pas utilisé |
| Ensemble des données littérales | | mise en forme des fichiers Majic (proprio et non bâti) et csv du RPG, créations de clés et d'index ok | Création des index, remplacement des vides par des 0 pour les champs servant pour les clés et création de clés |

(1) le SGBD est sensible à la casse, pas de majuscules

Tableau 3 : Création des tables contenant des données géographiques

| Nom de la table ⁽¹⁾ | Donnée | Nom de l'algorithme SQL ou du plugin | Caractéristiques de l'import |
|-------------------------------------|----------------------------|--------------------------------------|---|
| rpg2012ok | couche géographique du RPG | Postgis shapefile and dbf loader 2.0 | - encodage UTF-8 - élimination de 356 géométries de type <i>Ring Self-intersection</i> |
| localisant2013 | BD Parcellaire localisants | Postgis shapefile and dbf loader 2.0 | - encodage LATIN1 |
| parcelle2013ok | BD Parcellaire parcelles | Postgis shapefile and dbf loader 2.0 | - encodage LATIN1 - élimination de 4 géométries, 3 de type <i>Self-intersection</i> et un <i>polygone non fermé</i> |
| commune (ajout non prévu au départ) | BD Parcellaire communes | Postgis shapefile and dbf loader 2.0 | - encodage LATIN1 |
| Toutes les données géographiques | | Postgis shapefile and dbf loader 2.0 | Srid 2154, ne pas créer d'index spatial automatiquement, premier import avec option « create » pour le premier fichier importé, second import avec option « append » pour tous les autres fichiers du même type en une seule fois |

(1) le SGBD est sensible à la casse, pas de majuscules

II. Traitement des données : appariement du RPG et du cadastre

Une fois importées, les données peuvent être retravaillées puis croisées entre elles. L'objectif de ces traitements est d'aboutir à une table croisant les données d'exploitation et de propriété suivant les principes suivants :

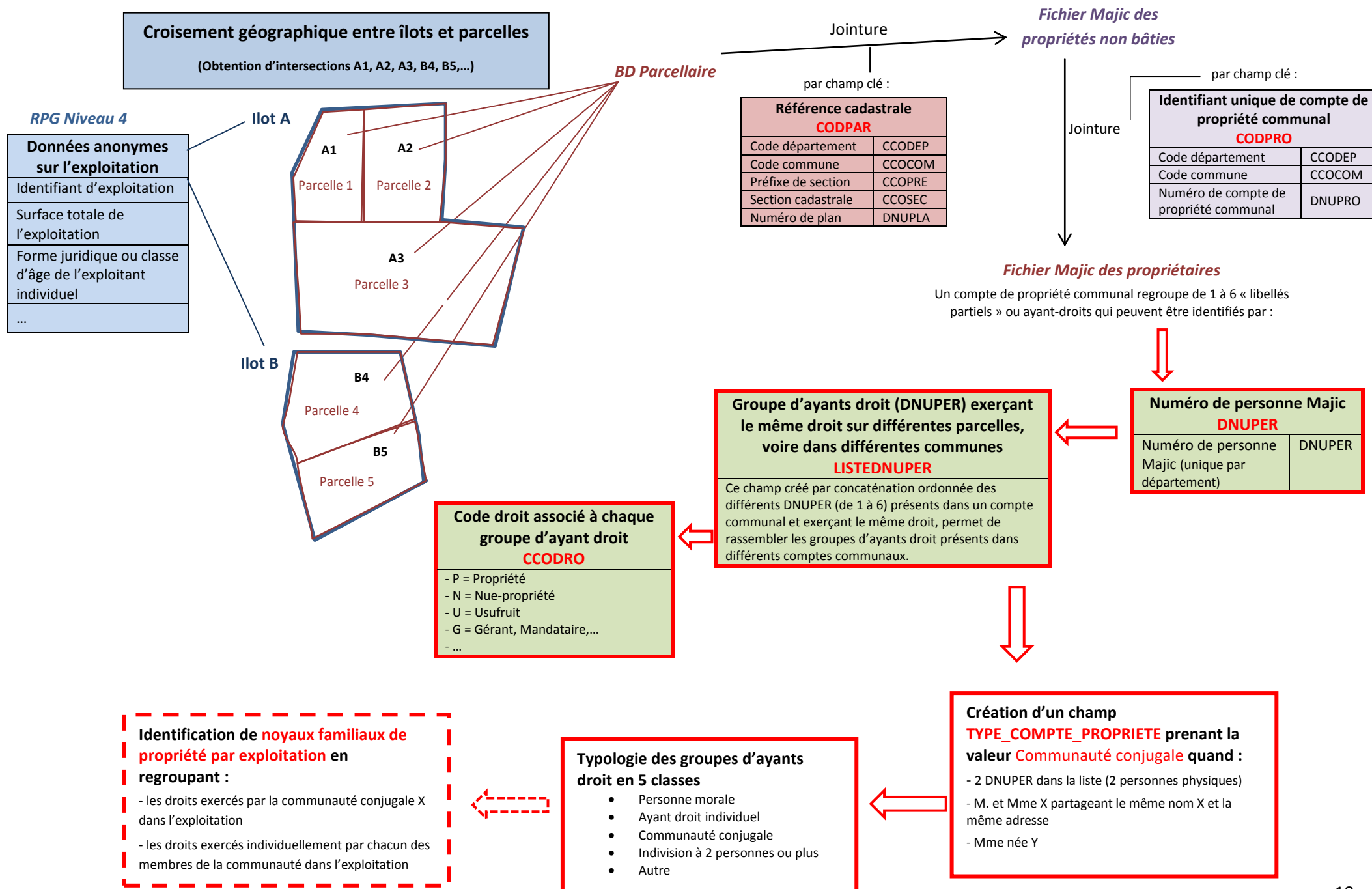
- répertorier un nombre de propriétaires agricoles (ou de listes d'ayants droit, en cas de démembrement de propriété) réaliste par rapport à la réalité. Cette condition implique de rassembler les ayants droit (ou groupes d'ayants droit, dans le cas d'indivisions) inscrits dans plusieurs comptes de propriétés communaux ;
- qualifier ces propriétaires suivant une typologie ;
- recenser le maximum de surfaces par rapport au RPG initial, que ce soit au niveau global ou par exploitation ;
- s'assurer au terme de l'appariement de la cohérence entre le nombre de propriétaires recensés par exploitation et les surfaces qu'ils détiennent (cas de micro-intersections biaisant le comptage des propriétaires par exploitation).

Le schéma global du croisement des bases de données géographiques et littérales est présenté dans la figure 2.

La structure finale de la table obtenue à l'issue des opérations de croisement est présentée en annexe 2.

Remarque : tous ces croisements sont effectués sur un modèle spatial en 2 dimensions. On considère donc la projection plane des parcelles et des îlots et non la surface réelle des objets qui tiendrait compte de la pente. Compte tenu de l'échelle très fine à laquelle on fait ce croisement et du fait que les modèles numériques de terrain facilement disponibles qui permettent de tenir compte de l'altimétrie ne sont pas très précis, il a été décidé de considérer la surface plane des polygones.

Figure 2: Schéma des différents appariements géographiques et littéraux réalisés pour mettre en relation exploitations et propriétaires



II.1. Traitement sur les comptes de propriété

Avant tout croisement avec le résultat du lien entre numéro d'îlot du RPG et les numéros de parcelles cadastrales, un certain nombre de traitements sur le fichier propriétaire sont réalisables.

La structure du fichier des propriétaires est la suivante :

- 1 **codpro (compte communal de propriété)** est lié à une liste de 1 à 6 ayants-droits ou libellés partiels (personnes physiques ou morales), qui ont chacun un code **dnuper** théoriquement unique par ayant-droit et par département ;
- chaque libellé partiel (identifié par le **dnuper**) a un droit de propriété (code **ccodro**) dans le compte communal : plein (P), usufruit (U), nue-propriété (N) ou autres (rares), qu'il peut partager dans son couple ou en indivision avec sa fratrie ou d'autres personnes (notamment un parent veuf) ;
- en cas de démembrement de propriété, on a donc une liste de nu-propriétaires et une liste d'usufruitiers (ou d'autres codes) dans le même **codpro**. Toute analyse sur les **codpro** implique de tenir compte des droits de propriété pour ne pas compter 2 fois les mêmes parcelles concernées par un démembrement de propriété ;
- le fichier comprend également un « code démembrement-indivision » (code **ccodem**) au sens de la DGFIP : une liste de personnes qui partagent le même droit dans un compte a souvent le code I (pour « indivision simple ») sur le droit P, U ou N ; la liste complète des valeurs possible est :
 - I : « Indivision simple » (de loin le plus fréquent)
 - C : « un des copropriétaires »
 - S : « succession de »
 - V : « la veuve ou les héritiers de »
 - L : « propriété en litige »

Le mode de renseignement de ce code n'est pas très clair, notamment sur la différence entre les codes S et V. Par ailleurs, on peut trouver selon les **codpro** que la même liste de personnes est considérée dans une commune en I et dans d'autres en S sans explication évidente de cette différence. De plus, ce code n'est a priori pas lié au régime matrimonial (séparation de biens, communauté de biens ou communauté réduite aux acquêts, autre) pour les couples, si bien que la notion d'indivision n'est pas toujours exacte dans ce cas. Pour toutes ces raisons, il est décidé dans cette étude de ne pas le considérer comme un des éléments de l'identifiant unique d'un compte de propriété.

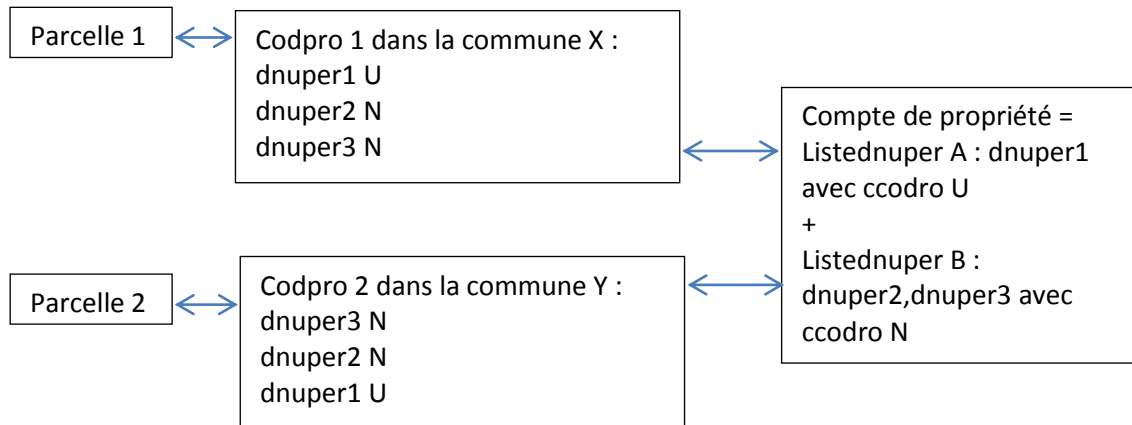
Un même **dnuper** (en supposant qu'il n'y a pas d'erreur de la DGFIP dans la suppression des doublons) peut se retrouver dans plusieurs comptes communaux parce qu'il a des droits différents, par exemple :

- pleine propriété individuelle de certains biens,
- partage de droits avec son conjoint sur d'autres biens,

- partage de droits dans une indivision (avec la fratrie ou les autres membres d'une succession) sur d'autres biens ;

Enfin, comme le cadastre est un fichier fiscal et que certains taux d'imposition sont communaux, un compte de propriété réel composé des mêmes personnes avec les mêmes droits dans plusieurs communes a autant de comptes communaux (**codpro**) qu'il y a de communes.

Figure 3 : dédoublonnement d'un compte de propriété pluricommunal



Explication : les 3 personnes dnuper 1, 2 et 3 (après vérification de l'unicité de l'identité de la personne associée à chaque dnuper) ont un bien en démembrement de propriété constitué de 2 parcelles situées dans 2 communes différentes X et Y du même département. Elles ne sont pas enregistrées dans le même ordre dans les 2 comptes communaux codpro 1 et 2 mais en faisant un tri sur l'ordre alphabétique de leur code dnuper et le droit de chaque dnuper, on peut les dédoubler si les deux codpro sont dans le même département. Finalement, les listednuper A et B avec le code de droit ccodro sont bien associées aux 2 parcelles.

II.1.1. Dédoublonnage des comptes réunissant les mêmes individus et partageant le même droit (tous types de personnes confondus)

Un dédoublonnage complet du nombre de propriétaires et une caractérisation des comptes de propriétés consiste donc :

- d'une part à éliminer les doublons de **dnuper**. Ils sont relativement nombreux pour les personnes morales, surtout quand il s'agit de gros établissements publics ou des services de l'Etat, mais concernent moins de 1% des comptes de personnes physiques ;
- regrouper les comptes de propriété pluricommunaux avec les mêmes personnes et les mêmes droits en créant une liste de **dnuper** partageant le même compte avec les mêmes droits dans plusieurs communes ou exploitations. Une difficulté supplémentaire est que la même liste de dnuper dans des codpro différents n'est pas toujours ordonnée de même manière selon les communes. De plus, les droits de propriété (surtout pour U et N puisque le problème ne se pose pas pour des propriétaires) ne sont pas toujours dans le même ordre non plus ;

La solution la plus simple est de créer un code **listednuper** qui est la liste ordonnée (pour ne pas avoir des faux doublons liés à un ordre différent des **dnuper** selon les communes) séparée par des virgules des **dnuper** ayant les mêmes droits de propriété dans le **codpro**. Pour retrouver les listes de personnes ayant des biens dans plusieurs communes, il suffit alors de chercher les groupes de personnes ayant la même **listednuper** et le même type de droit (P, U ou N) **ccodro** dans le même département.

Techniquement, ce n'est pas très simple puisque le nombre de **dnuper** varie de 1 à 6 dans un **codpro** donné. Pour créer le champ **listednuper**, on utilise les fonctions de fenêtrage (*window function*) qui permettent de faire des opérations simples (ici, concaténation de chaîne de caractères) sur les lignes d'une table partageant la même valeur pour un ensemble de champs.

Cette opération permet de réduire le nombre apparent de comptes communaux d'environ 20 %.

II.1.2. Traitements spécifiques aux personnes physiques

Redressement sur les DNUPER différents attribués à la même personne dans le même département

Les personnes physiques représentent la plus grande part en nombre des **dnuper**. Le **dnuper** ne comportant que 6 caractères, il ne suffit pas à identifier de manière unique les personnes en dehors d'un département : on trouve souvent le même code **dnuper** correspondant à des personnes totalement différentes dans plusieurs départements.

Un filtrage sur les principales données personnelles (civilité (M. ou Mme), nom d'usage (qui est celui du mari pour une femme mariée), prénom d'usage (qui est celui du mari pour une femme mariée), date de naissance et lieu de naissance,) montre que le **dnuper** est assez bien construit pour les personnes physiques dans un département donné. L'adresse de résidence n'a pas été retenue comme critère car une même personne peut avoir changé plusieurs fois de résidence ou en posséder plusieurs en même temps.

On a bien vérifié au préalable que dans un département donné, le même **dnuper** n'est associé qu'à une seule et même personne. Sur les 152 cas où on trouve deux personnes apparemment différentes, il s'agit en fait de femmes mariées qui ont le nom de jeune fille dans un compte et le nom et prénom d'épouse dans un autre (elles ont la même civilité et les mêmes dates et lieux de naissance), de personnes avec des lieux de naissance saisis différemment, ou avec un nombre de prénoms différents.

On décide de le garder comme référence après l'opération suivante :

- pour les plus de 99 % de personnes physiques qui n'ont qu'un seul **dnuper**, on le garde ;
- pour celles qui ont plusieurs **dnuper** dans le même département, on réattribue aux variantes le premier **dnuper** dans l'ordre alphabétique de l'ensemble des **dnuper** associés à la personne.

Identification des communautés conjugales parmi des groupes d'ayants droit

Depuis une révision du modèle du fichier des propriétaires en 2012, on ne peut plus lier avec certitude 2 conjoints entre eux. On fait donc l'hypothèse qu'un compte de 2 personnes physiques

(avec une civilité M et une MME, puisque le fichier date d'avant le vote de la loi sur le mariage pour tous) présentant le même nom d'usage, le même droit (P, U ou N en général), un nom de jeune fille pour MME et la même adresse est celui d'un couple marié.

Remarque : cet algorithme peut également retenir des mères et fils ayant la même adresse.

Attribution d'un âge moyen à chaque groupe d'ayants droit

La date de naissance étant renseignée et la date de validité du fichier connue, on peut calculer l'âge des propriétaires. Pour certains comptes créés depuis longtemps, aucune date de naissance n'est renseignée donc aucun âge n'est calculable. Par ailleurs, on calcule l'âge moyen de l'ensemble des personnes physiques qui partagent un droit de propriété dans un même compte.

Lien de parenté des ayants droit

En faisant des opérations sur les noms (nom d'usage et/ou nom de jeune fille pour les femmes mariées), on cherche également à typer les comptes de propriété en distinguant ceux qui rassemblent les membres d'une même famille, en supposant que les personnes de même patronyme au sein d'un même compte sont apparentées.

Pour ce faire, on a cherché si toutes les personnes d'un même compte de propriété ont :

- le même nom d'usage (cas le plus simple, notamment pour les couples mariés ou les comptes qui ne comportent que des hommes),
- si ce n'est pas le cas, si l'ensemble des personnes d'un même compte de propriété ont toutes soit leur nom de jeune fille soit leur nom d'usage qui est identique au nom le plus courant rencontré dans le compte de propriété (via une *window function*). Cela permet notamment de retrouver les liens de parenté entre les parents et leurs filles mariées ou entre les frères et leurs sœurs mariées. En effet, pour retrouver les comptes familiaux, on identifie les mères mariées et les filles non mariées par leur nom d'usage et les filles mariées par leur nom de jeune fille.

En procédant à l'échelle du compte de propriété, on peut notamment savoir si un bien démembré l'est au sein de la même famille, puisque la liste des usufruitiers et la liste des nus-propriétaires auront le même nom.

II.1.3. Traitements spécifiques aux personnes morales

Les personnes morales sont décrites par des champs particuliers. Une analyse détaillée en comparant le code **dnuper**, le code **gtoper** (pour distinguer personnes physiques et morales), le code de groupe de personne morale (**ccogrm**), la forme juridique (**dforme**) ou le code de forme juridique (**dformjur**) et le libellé montre plusieurs choses :

- le **gtoper** est fiable : aucune erreur évidente n'a été détectée dans l'ensemble des comptes des 4 départements d'étude ;
- la quasi-totalité des personnes morales sont seules à détenir leur droit (P, U ou N et plus rarement les autres codes) dans un **codpro** donné. Il n'y a en effet pas de logique fiscale ou managériale à partager un droit de propriété entre plusieurs sociétés (il vaut mieux créer une société spécifique dans ce cas), les résultats sont donc cohérents sur ce plan. On trouve quand même quelques dizaines de **codpro** dans lesquels plusieurs personnes morales ou une

ou deux personnes physiques et une personne morale partagent leur droit de propriété (voir II.1.4) ;

- les personnes morales publiques comme l'État ou les gros établissements publics ont souvent plusieurs **dnuper** dans le même département, il semble plus opérationnel de créer des catégories pour simplifier l'analyse ;

- le code **ccogrm** qui est censé permettre de distinguer les différentes échelles de personnes publiques (de l'État à la commune) et les personnes morales privées n'est pas très bon, on ne le retient donc pas pour créer des catégories ;

- un des principaux défauts du **ccogrm** est que la catégorie « non définie » comporte de nombreuses personnes morales qu'il est pourtant facile de typer grâce au champ dénomination. Il y a notamment beaucoup d'EARL, GAEC, GFA et SCI (le libellé commence par le sigle ou par « groupement foncier agricole », etc.), formes juridiques très importantes pour analyser la propriété agricole, donc on crée un code de forme juridique redressée **dforme2**.

Dans notre analyse, nous nous limiterons à bien différencier les formes juridiques agricoles, les types de collectivités et les divers établissements ou organismes potentiellement propriétaires d'espaces non bâtis (conservatoire d'espace naturels, CELRL, EPF, etc.).

II.1.4. Nettoyage des parcelles affichant des droits de propriété non anticipés à la construction de la base

D'après la documentation officielle des fichiers Majic, plusieurs droits peuvent coexister sur la même parcelle : ainsi on trouvera un **listedenuper** avec un **ccodro** « U » systématiquement associé à un autre **listedenuper** avec un **ccodro** « N ». De même, mais de manière très marginale dans l'échantillon des parcelles associées à un îlot PAC, les codes B (bailleur à construction) et R (preneur à construction). Dans les sommes de surfaces réalisées par la suite il importe donc de ne prendre en compte que l'un des 2 codes pour ne pas faire de double compte.

Or il arrive que des codes en théorie non jumelés coexistent sur la même parcelle, ce qu'il convient de corriger. Pour repérer ces cas problématiques sont créées la variable **listeccodro** qui concatène les valeurs prises par le champ **ccodro** sur une même parcelle (et donc un même compte communal). Dans le cadre de l'étude, on s'est attaché à ne corriger que les cas les plus significatifs en nombre de comptes et en surface, éventuellement de manière arbitraire, pour obtenir *in fine* un décompte optimal des surfaces. Les principaux redressements effectués concernent la coexistence d'un propriétaire, d'un usufruitier et d'un nu-propriétaire (1 000 comptes pour 3 900 ha, dans l'échantillon à l'issue du croisement). On retiendra dans ce cas uniquement les codes N et U en partant du principe que le code P est par défaut et que si la DGFIP a saisi un démembrement de propriété, c'est que celui-ci a été déclaré.

Par ailleurs, le code **CCODRO=G** (« gérant, mandataire, gestionnaire ») qui apparaît dans 1 460 comptes pour 12 000 ha à l'issue du croisement est en fait systématiquement associé à un autre code (P, ou le doublet N,U), ce qui n'est pas précisé dans la documentation du cadastre. Il convient donc de ne pas le comptabiliser pour faire les totaux de surface.

Enfin les modalités de construction de la variable **listednuper** entraînent la répétition de certaines lignes, dans le cas précis où une ou plusieurs personnes physiques partagent la pleine propriété avec une ou plusieurs personnes morales, ce qui constitue une bizarrerie juridique. Ce cas de « compte mixte » concerne 200 comptes pour 850 ha. Pour simplifier ce cas de figure non anticipé à la construction de la table, on ne retient que la ou les personnes physiques du compte, mais un traitement particulier pourrait être envisagé.

II.1.5. Typologie finale des ayants droit

Tableau 4 : Typologie des groupes d'ayants droit établie à l'issue des regroupements et redressements de la table des propriétaires

| | | |
|---|--|--|
| <p>Personnes morales Ces personnes sont identifiées par la variable GTOPEP prenant la valeur 2</p> | <p>Forme juridique DFORME : basée sur la nomenclature INSEE à 4 caractères alphanumériques : elle permet d'identifier une bonne partie des personnes morales agricoles, ainsi que les personnes publiques (Communes, département,...).</p> | <p>La forme juridique DFORME a été consolidée à partir du champ Dénomination (DNOM) : recherche sur les caractères de type « GFA », « GAEC », « Société anonyme », etc. quand DFORME n'était pas renseigné</p> |
| <p>Personnes physiques Ces personnes sont identifiées par la variable GTOPEP prenant la valeur 1</p> | <p>Ayants droit individuels</p> | <p>Ce sont les ayants droit que l'on retrouve seul dans un ou plusieurs comptes communaux</p> |
| | <p>Communautés conjugales</p> | <p>Il s'agit de 2 personnes physiques réunies dans un ou plusieurs comptes et présentant les caractéristiques suivantes :</p> <ul style="list-style-type: none"> - M et Mme - même nom d'usage - même droit (P, U ou N en général), - un nom de jeune fille pour Mme - même adresse |
| | <p>Indivisions</p> | <p>Il s'agit de 2 personnes ou plus réunies dans un ou plusieurs comptes communaux, qui ne sont pas des communautés conjugales</p> |
| <p>Autres</p> | <p>Comptes mixtes</p> | <ul style="list-style-type: none"> - personnes morales et physiques dans le même compte de propriété - plusieurs personnes morales dans le même compte de propriété |

II.2. Croisement des données spatiales au format vecteur

Le principe général est de construire une requête spatiale dans laquelle on cherche tous les polygones de parcelles de la BD Parcellaire inclus ou intersectant des îlots de cultures du RPG. Cette opération est effectuée sur les 4 départements, à l'exception de 9 communes de Mayenne, où le cadastre n'est pas encore vectorisé dans la BD Parcellaire. En sortie, on cherche donc à créer des doublets (**codpar, num_ilot**) auxquels on liera ensuite les données littérales de l'exploitation et des propriétaires pour réaliser des comptages.

II.2.1. Principes généraux, opérateurs et fonctions utilisés

Dans le cadre de l'étude, 2 croisements géographiques sur les couches vecteur sont testés :

- l'opérateur @ (qui travaille sur l'enveloppe des objets) pour identifier les parcelles incluses dans les îlots
- l'opérateur *ST_intersection* (qui travaille sur la géométrie réelle des objets) entre les parcelles au format vecteur et les îlots, après avoir exclu les données non conformes (Cf. II.2.3.).

Principe des opérateurs de croisement spatiaux dans PostGis

PostGIS permet de faire tous les croisements, et notamment les croisements spatiaux, dans une seule base de données : d'une part, on peut utiliser des opérateurs spatiaux (de type « est inclus dans », «intersecte ») de la même manière que les opérateurs sur les nombres ou les chaînes de caractères. D'autre part, le champ géométrie (*geom*) des données spatiales peut être utilisé comme tous les autres champs.

Plusieurs types de fonctions ou opérateurs sont disponibles :

- certains sont des booléens, qui renvoient à un résultat vrai ou faux. On peut les utiliser pour créer des conditions dans une requête. Par exemple, *st_intersects(a.geom, b.geom)* retournera « vrai » si des objets des couches a et b s'intersectent ;
- d'autres renvoient un résultat, notamment une géométrie. Par exemple, *st_intersection(a.geom, b.geom)* renvoie la couche des géométries des intersections des objets des couches a et b ;
- un point important est de vérifier si l'opération porte sur :
 - o l'enveloppe des objets (ou *bounding box*, *bbox* ou cadre limite) : dans ce cas le calcul est rapide mais le résultat approximatif.
 - o ou sur la géométrie réelle des objets : le temps de calcul est beaucoup plus long mais le résultat plus précis). C'est précisé dans la documentation des fonctions.

Note : pour rendre plus lisibles les algorithmes, on peut donner des alias aux tables utilisées dans la requête. On choisit par exemple d'utiliser P pour propriétaire, NB pour le fichier Majic du non-bâti, BDP pour la BD Parcellaire et RPG pour le fichier du RPG.

Remarque : Une analyse sur la couche du RPG a permis de révéler un grand nombre de chevauchements d'îlots entre eux, ce qui vient augmenter artificiellement les surfaces répertoriées par rapport à la réalité physique. Sur les 325 310 îlots étudiés, on dénombre 44 511 superpositions représentant 349 ha (sur 1,5 million d'ha environ), soit 0,02 %. Ces chevauchements n'ont a priori rien de surprenant, le RPG étant construit sur la base des déclarations de surfaces de chaque exploitation, chaque déclarant dessinant lui-même le contour de ses îlots. La plupart des chevauchements repérés ont une forme de bande de quelques mètres de large, à la jointure de 2 parcelles exploitées par 2 exploitations différentes. Dans quelques cas plus rares, ce sont des parcelles entières qui sont déclarées 2 fois : sur les 4 départements on dénombre 82 superpositions d'îlots mesurant plus de 1 ha. Dans le cadre de l'étude, comme il est impossible de trancher a priori sur le rattachement d'une parcelle à l'une ou l'autre des exploitations, et compte tenu des faibles surfaces concernées, aucun redressement n'a été effectué, et ces parcelles ont éventuellement été attribuées à l'une et l'autre des exploitations concernées.

II.2.2. Croisement des couches vecteur via l'opérateur @

Le principe est de chercher les parcelles incluses dans des îlots et non l'inverse, car les îlots sont en moyenne beaucoup plus étendus (4,8 ha d'après calcul sur les shp) que les parcelles (0,46 ha d'après Majic et la BD Parcellaire).

L'opérateur @ aboutit à un résultat relativement rapidement, mais n'est pas très fiable : sur les 4 départements, on n'arrive à affecter que 2/3 des 1,6 millions d'ha du RPG à des parcelles cadastrées, ce qui est insuffisant pour les besoins de l'étude. Deux explications à ce croisement médiocre :

- d'une part, on sous-estime le nombre et la surface des intersections puisqu'on exclut par construction les parcelles qui ne sont pas totalement incluses dans les îlots ;
- d'autre part, il y a des erreurs dues au fait que l'opérateur @ travaille sur « l'enveloppe » ou « cadre limite » (bounding box ou bbox) et non sur la géométrie réelle des objets.

II.2.3. Croisement des couches vecteur via l'opérateur ST_intersection

Echec du croisement brut entre l'ensemble des parcelles et l'ensemble des îlots

En seconde approche est testée l'intersection spatiale entre parcelles (au format vecteur) et îlots, afin de tenir compte de la surface des parcelles réellement à l'intérieur d'un îlot, que l'intersection soit complète ou partielle.

Ce croisement apparaît en principe beaucoup plus précis, mais le croisement brut entre l'ensemble des parcelles et l'ensemble des îlots, compte tenu du nombre de combinaisons possibles, n'aboutit pas, et ce même en restreignant l'analyse à un département.

Succès du croisement grâce au fractionnement communal des parcelles et des îlots

Pour réduire le nombre de combinaisons possibles entre parcelles et îlots, une solution consiste à effectuer une recherche sur les codes de communes, avant de calculer l'intersection à l'intérieur de chaque commune. En procédant ainsi, l'ordre de grandeur de la durée totale de calcul est ramené à quelques heures pour 4 départements.

Comme les îlots n'ont aucune raison d'être intégralement contenus dans une seule commune, et qu'à la marge on observe également quelques débordement de parcelles à l'extérieur de l'enveloppe

communale de la BD Parcellaire, il est impératif de créer deux couches intermédiaires, obtenues après intersection avec la couche « commune » de la BD Parcellaire :

- **ilot_com** qui correspond aux fractions infracommunales des îlots. Le code de commune attaché aux données du niveau 4 pour chaque îlot n'est qu'indicatif. Il est quand même valable à plus de 90 % dans les 4 départements considérés ;
- **bdp_com** qui correspond aux fractions infracommunales de parcelles. Une parcelle est en théorie intégralement rattaché géographiquement à une commune. Mais comme le PCI vecteur qui sert à élaborer la BD Parcellaire sous forme de polygone n'est pas jointif en bordure de division cadastrale (section ou groupe de sections cadastrales) ou en limite communale, on peut avoir des bandes de quelques mètres d'une parcelle qui dépassent de la limite communale. Le critère de choix logique est donc de ne retenir que les fractions de parcelles dont le code de commune est identique à celui de la couche « commune » qui les intersecte ;

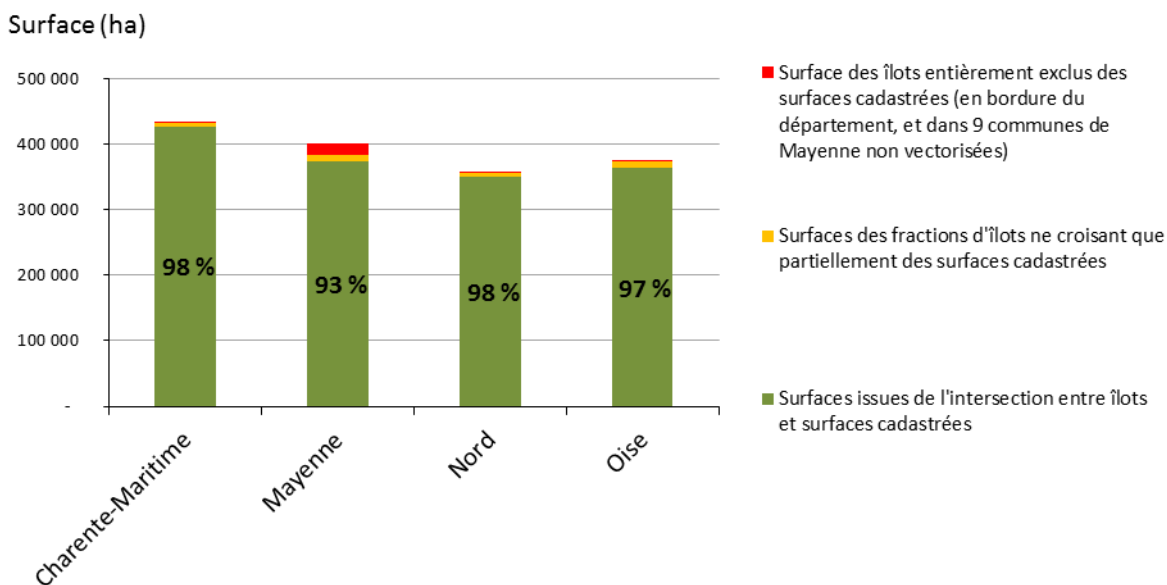
On croise alors ces deux couches infracommunales entre elles pour récupérer des objets correspondant à leur intersection. Les objets en sortie comportent trois champs : **num_ilot**, **codpar** et **igeom** (géométrie de l'intersection).

Une vérification systématique (recherche des objets de l'intersection finale dont le code commune de parcelle est incohérent avec la couche commune de la BD Parcellaire) montre que le résultat est très fiable : on ne trouve qu'un seul objet (éliminé par la suite) ayant un code commune de parcelle différent de la commune de BD Parcellaire, cette erreur étant due au fait que pour compenser les erreurs du PCI vecteur, 2 communes de l'Oise ont des limites communales partiellement superposées.

Bilan des surfaces retenues à l'issue du croisement

Figure 4: Part des surfaces initiales du RPG retenues ou perdues à l'issue du croisement avec le cadastre

Part des surfaces retenues à l'issue du croisement sur les 4 départements : 96,5 %



Le résultat du croisement s'avère très bon car on affecte un numéro de parcelle et une géométrie à la plus grande part du RPG. On ne perd qu'environ 55 000 ha par rapport aux 1,5 millions d'ha du RPG de départ (soit 3,5 %), pertes qu'on explique de la manière suivante :

- environ 16 000 ha sont perdus qui correspondent à 9 communes de Mayenne n'ayant pas encore de BD Parcellaire vecteur, alors que le département est complètement vectorisé. Cela s'explique par le décalage entre la création du PCI vecteur par la DGFIP et la mise à jour de la BD Parcellaire ;
- environ 8 000 ha sont constitués d'îlots localisés en limite de département et n'intersectant pas du tout la couche commune de la BD Parcellaire (représentés en vert dans la figure 4) ;
- environ 31 000 ha sont constitués de fractions d'îlots n'intersectant que partiellement des parcelles cadastrées (représentées en bleu dans les figures 4 et 5). Ces fractions sont localisées :
 - soit en limite de département : une partie seulement de l'îlot est incluse dans l'enveloppe départementale de la BD Parcellaire, d'où la perte du reste de l'îlot.
 - soit à l'intérieur des communes, lorsque l'îlot déborde sur un espace non cadastré. Les fractions d'îlot ainsi perdues présentent le plus souvent des formes linéaires : la comparaison de quelques secteurs avec des couches image (Ortho ou Scan25) sur le géoportail IGN révèle qu'il s'agit de bordures de routes, de cours d'eau ou d'espaces boisés. Une partie de ces fractions peut aussi correspondre à d'anciens chemins vicinaux (restés dans le domaine non cadastré) et désormais mis en culture. Ces

fractions d'îlot non retenues par le croisement dessinent un réseau dense mais représentent une surface totale relativement faible car les linéaires en question sont très étroits (cf. figure 5).

Remarque : Le croisement a permis également de repérer (et nettoyer) la répétition dans la BD Parcellaire d'une vingtaine de parcelles dans le Nord, exactement superposées.

Remarque : le domaine public non cadastré est constitué des terrains exemptés de la taxe foncière des propriétés non bâties (source : BOFIP). Il est constitué par l'emprise des :

- *voies publiques : rues, places publiques, routes nationales et départementales, voies communales et chemins ruraux ;*
- *eaux : cours d'eaux qu'ils soient domaniaux, non domaniaux ou mixtes ;*
- *rivages de la mer ;*
- *lacs s'ils appartiennent au domaine public ;*
- *canaux de navigation de l'Etat non concédés ;*
- *de certains immeubles particuliers comme les aérodromes et dépendances qui remplissent simultanément les conditions suivantes : appartiennent à l'État, au département ou à la commune, ne sont pas ouverts à la circulation aérienne publique, sont affectés à un service d'utilité générale et improductifs de revenus.*

Figure 5: Fractions d'îlot et îlots entiers non retenus dans le croisement en limite de département

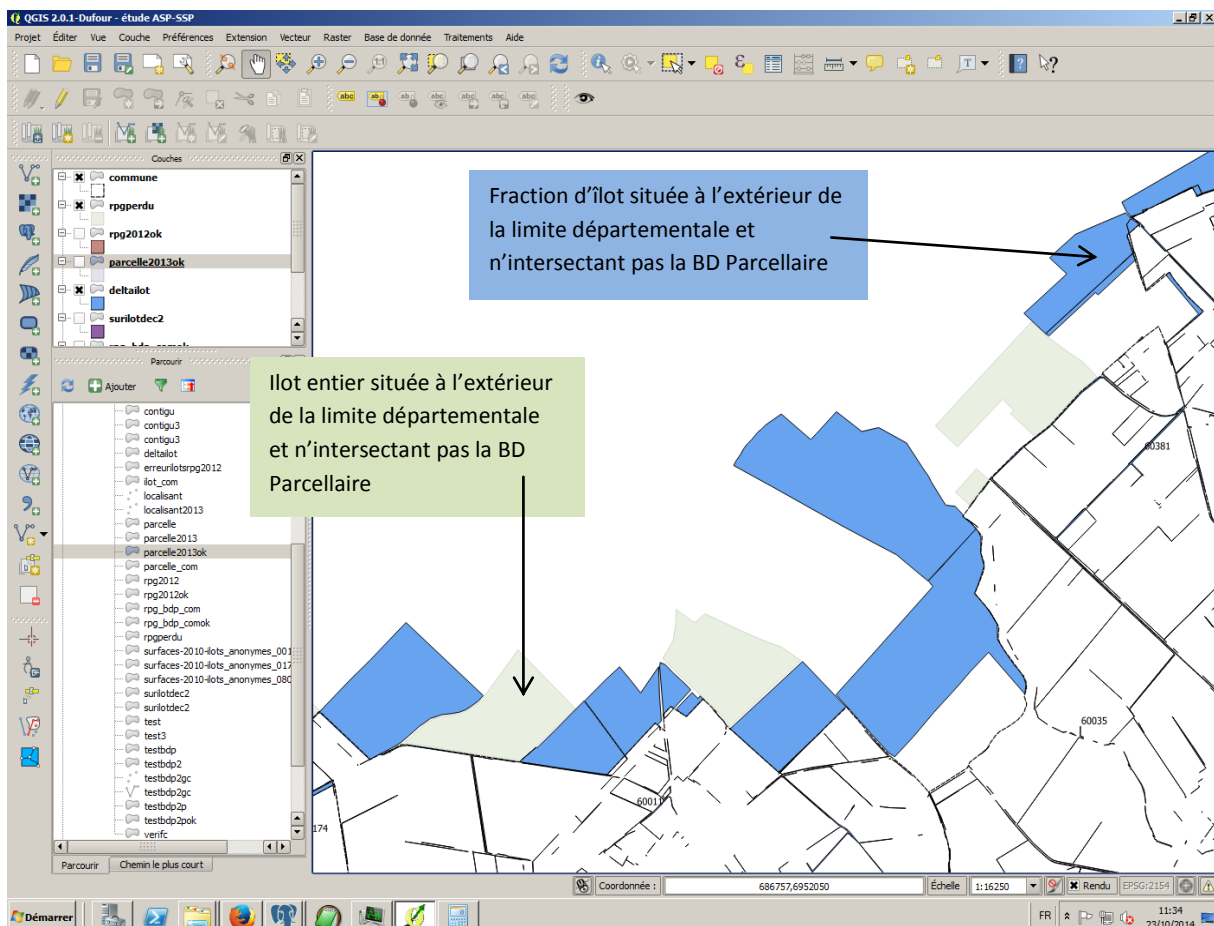
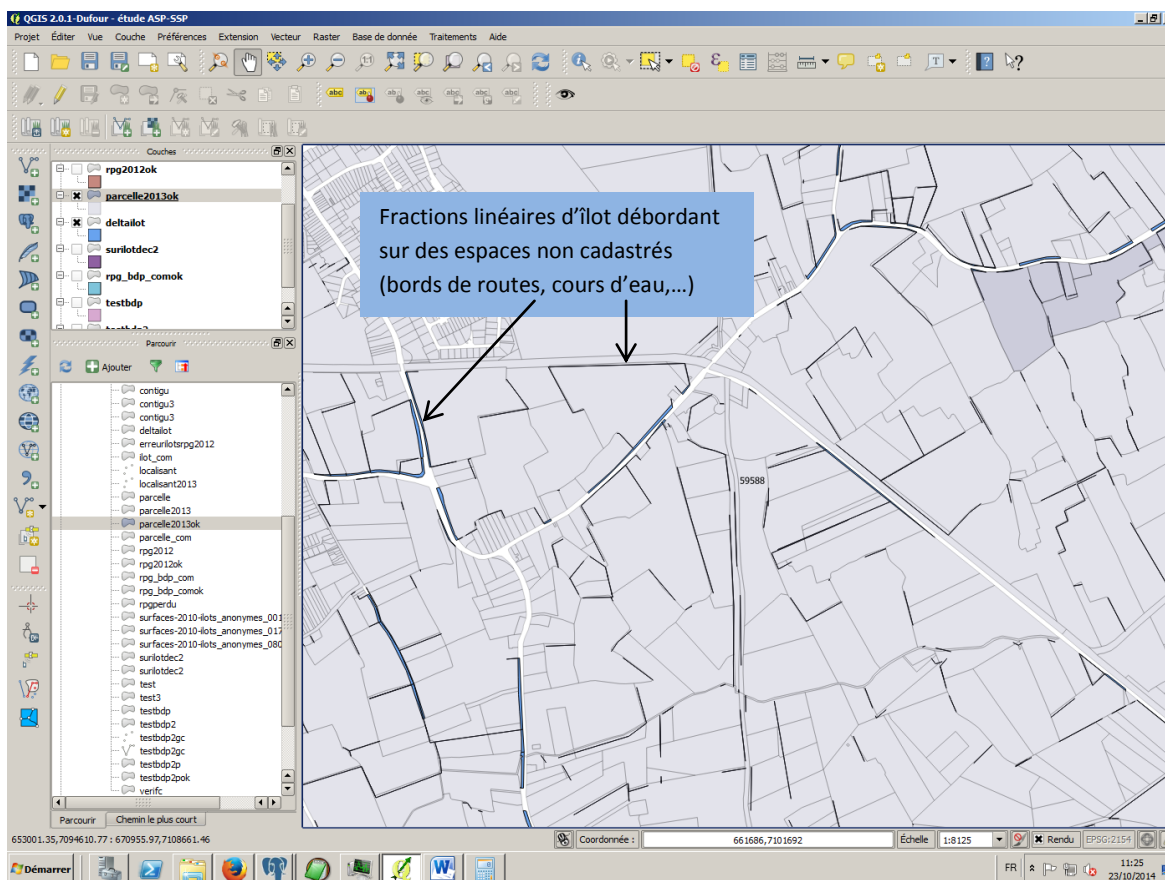


Figure 6: Fractions d'îlot non retenues dans le croisement au sein des communes



II.2.4. Nettoyage des micro-intersections dans l'échantillon obtenu

Le croisement effectué avec l'opérateur *st_intersection* est extrêmement précis car aucun filtre n'est appliqué lors du calcul, en dehors de la vérification de la présence des 2 géométries dans la même commune.

Une conséquence positive de cette précision dans le croisement est de conserver dans l'échantillon l'intégralité des surfaces du RPG croisant des surfaces cadastrées. Une conséquence indirecte moins satisfaisante est la génération de très nombreuses intersections, en particulier de très petite taille (quelque m² à centaines de m²). Ces micro-intersections s'expliquent par plusieurs facteurs:

- une différence dans l'échelle de précision des couches : le cadastre est précis à une échelle moyenne de l'ordre du 1/2 000^e qui varie selon les sections cadastrales (héritage de l'époque du cadastre papier où le principe le plus fréquent était que chaque section soit représentée sur une seule feuille quelle que soit sa surface totale) alors que l'ASP annonce une précision du RPG au 1/5 000^e ;
- les bords d'îlots ne coïncident pas souvent avec des limites de parcelles car les bornes n'existent pas toujours ou ne se repèrent pas facilement sur le terrain et encore moins lors du dessin dans TéléPAC ;

- le calage n'est pas parfait sur les bords de chemin, de haies, de bois, etc. entre la BD Ortho (sur laquelle sont dessinés les îlots) et la BD Parcellaire, notamment en bordure de section cadastrale ou de commune à cause des reprojections faites par l'IGN pour pallier la discontinuité entre sections et entre communes du PCI. On a d'ailleurs souvent des micro-intersections sous forme de bandes de quelques mètres de large ;

Du fait de la superficie et du nombre de ces micro-intersections, on peut s'interroger sur leur conservation dans l'échantillon car elles introduisent plusieurs biais dans les indicateurs générés ultérieurement :

- un certain nombre de micro-intersections sont le résultat d'un îlot dont le contour déborde légèrement sur une parcelle non-agricole (forêt, terrain d'agrément d'un particulier), ce qui introduit éventuellement, par croisement avec le fichier des propriétaires, des groupes d'ayants droit en trop dans l'échantillon (s'ils ne possèdent pas de terrains agricoles par ailleurs).
- un certain nombre de micro-intersections sont le résultat d'un îlot débordant légèrement sur une autre parcelle, exploitée pour l'essentiel par une autre exploitation agricole. Son propriétaire apparaît par conséquent dans les 2 exploitations, et pour l'une d'entre elle pour seulement quelques m², ce qui vient fausser l'indicateur de type « Nombre de propriétaire par exploitation ».

Il convient donc de réaliser un nettoyage de ces micro-intersections sur les principes suivants :

- éliminer un maximum de comptes de propriété relatifs à du micro-parcellaire peu significatif par rapport à la taille moyenne des exploitations.
- ne pas évacuer une part trop significative du total des surfaces totales recensées
- ne pas amputer un nombre trop important d'exploitations d'une part significative de leur surface, pour ne pas fausser les indicateurs produits par exploitation.

Dans la méthodologie employée pour le nettoyage de ces micro-intersections, le 2^{ème} principe est vérifié a posteriori.

Filtres employés à la parcelle

On crée 2 ratios attribués à chaque intersection :

- **ratio_sumigeom_parcelle_tot** est la part de la somme des intersections relatives à la parcelle rapportée à la surface totale de la parcelle d'appartenance : ce ratio permet de filtrer les intersections appartenant des parcelles minoritairement agricoles.
- **ratio_igeom_agri** est la part de l'intersection dans la somme des intersections relatives à la parcelle : ce ratio permet de filtrer les intersections participant faiblement à la surface agricole de la parcelle.

Malgré le filtrage rendu possible grâce à ces 2 variables, il n'apparaît pas judicieux de les appliquer sur des intersections de grande taille (supérieure à 0,5 ha) qui ne sauraient relever d'artefact de croisement entre les couches. Le seuil de 0,5 ha a été déterminé pour présenter un bon compromis entre le nombre de comptes de propriété éliminés et la part totale de surface conservée. En conséquence, en éliminant les intersections de moins de 0,5 ha pour lesquelles

ratio_sumigeom_parcelle_tot est inférieur à 10 % ou **ratio_igeom_agri** est inférieur à 10 % (voir figures 6 et 7), on élimine 1 % de la surface totale de l'échantillon de croisement mais 34 % des propriétaires (test effectué uniquement sur les surface en pleine propriété, pour simplifier la relations entre comptage des ayants droit et comptage des surfaces).

```
CREATE TABLE public.synthese_prop_expl_2013okccV3liste_net_10pct as
select * from public.synthese_prop_expl_2013okccV3liste_m3
where not (((ratio_igeom_agri < 0.1 ) or (ratio_sumigeom_parcelle_tot < 0.1)) and
surintersection/10000 < 0.5);
```

NB : conversion en ha car comme l'unité de longueur est le m, l'unité de surface des intersections est le m²

Figure 7: Principe du filtrage sur **ratio_igeom_agri** : nettoyage des micro-intersections représentant moins de 10 % de la fraction agricole de la parcelle.

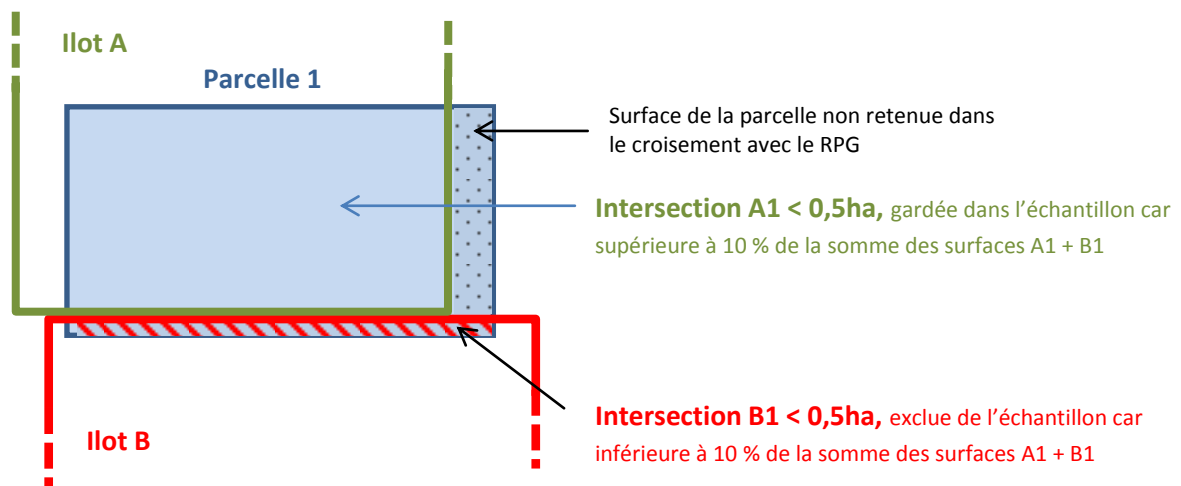
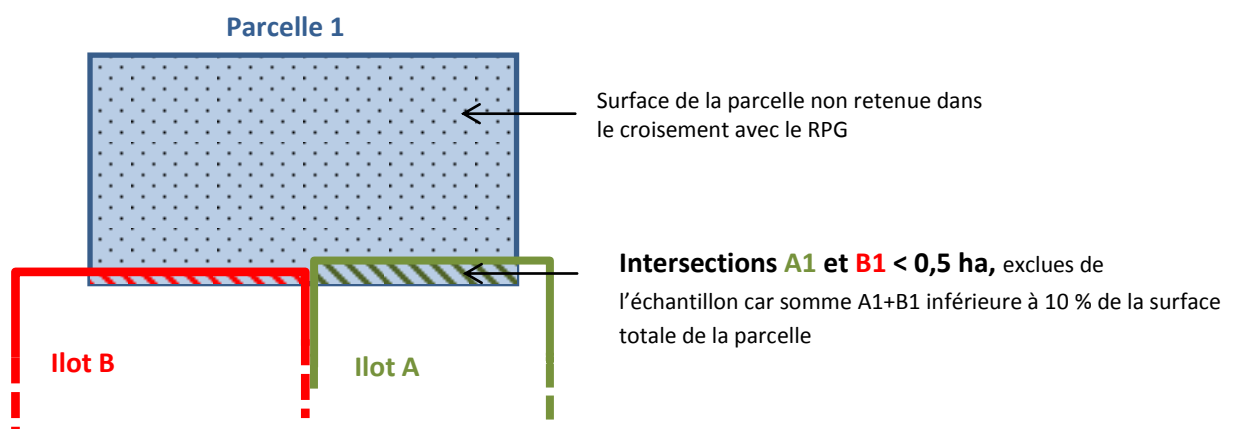


Figure 8: Principe du filtrage sur **ratio_sumigeom_parcelle_tot**: nettoyage des micro-intersections dont la somme des surfaces représente moins de 10 % de la surface totale de la parcelle.

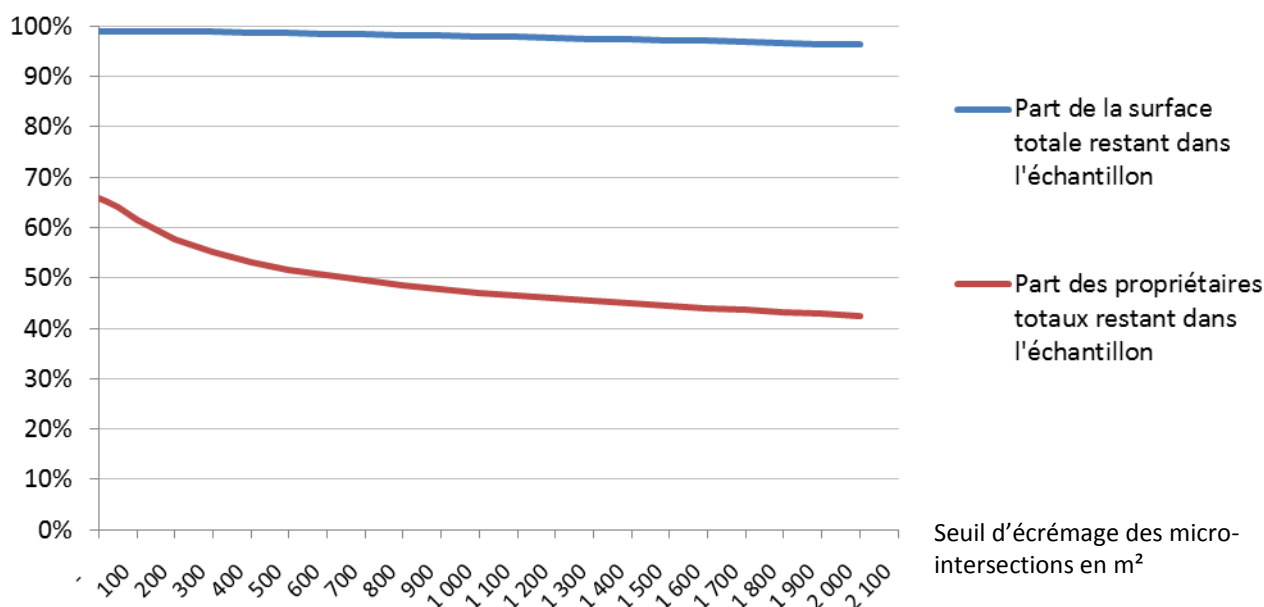


Filtres sur la surface absolue des micro-intersections

A l'issue de ce premier filtrage, il subsiste encore un grand nombre d'intersections de quelques centaines de m², ne présentant en cumulé qu'une petite part des surfaces totales mais un nombre important de comptes de propriété. Un deuxième filtrage sur la surface absolue des intersections est donc pratiqué sur l'échantillon (voir figure 8).

On vérifie auparavant que même en consolidant par exploitation l'ensemble des intersections associées au même compte de propriété (une liste de personne avec le même droit de propriété), il reste beaucoup de micro-intersections de quelques centaines de m². Autrement dit, il est très fréquent que la réunion de plusieurs micro-intersections relatives au même compte de propriété constitue encore une part très marginale d'une exploitation donnée.

Figure 9: Bilan des surfaces en pleine propriété et des propriétaires restants selon le seuil d'écrémage des parcelles.



La courbe affiche une diminution importante du nombre total de compte de propriété quand on écrème les intersections de quelques centaines de m². Au-delà de 500 m², la diminution s'avère plus linéaire, et ce dans les 4 départements de test, tandis qu'à ce seuil la perte de surface par rapport à l'échantillon de départ s'avère encore très mineure : 98,6 %.

Ce seuil de surface brute d'écrémage de 500 m² n'est toutefois appliqué qu'aux exploitations de plus de 1 ha de manière à ne pas amputer trop sévèrement les petites structures.

```
CREATE TABLE public.synthese_prop_expl_2013okccV3liste_net_10pct_500m2 as
select * from public.synthese_prop_expl_2013okccV3liste_net_10pct
where not surintersection<500 and surpar_expl>1);
```

Bilan des surfaces finalement retenues par exploitation

Une fois ces filtrages effectués, il est important de s'assurer que chacune des exploitations a été faiblement impactée par les opérations. Il apparaît que les exploitations ayant été amputées de plus de la moitié de leur surface parcellaire au cours du traitement sont au nombre de 3. Ce nombre passe à 16 pour plus de 25 % de la surface perdues et à 187 pour plus de 10 % de surface perdues. Au regard des 19 500 exploitations de l'échantillon, ces pertes apparaissent relativement peu significatives pour le calcul ultérieur des indicateurs.

II.3. Sélection des exploitations dont au moins 90 % de la surface est retrouvée dans le département

Une fois le croisement réalisé, on a vérifié pour chacune des exploitations le ratio de sa surface totale déclarée (c'est un des attributs du niveau 3 du RPG) qu'on retrouve en faisant la somme des surfaces des intersections avec la BD Parcellaire vecteur. Il est plus exact de se référer à la surface calculée des polygones des intersections qu'à la surface déclarée par groupe de cultures pour chaque îlot (données de niveau 2) d'après les tests effectués.

Pour la plupart des exploitations, ce rapport est égal ou très proche de 100 % puisque le non cadastré ne représente que de l'ordre de 1 % de la surface des îlots, sauf dans deux cas :

- une partie significative de l'exploitation est située dans un autre département ;
- certains îlots sont situés en bordure de département mais en dehors des limites de la couche des communes de la BD Parcellaire si bien qu'ils n'ont pas été retenus dans notre croisement (voir explication de la méthode de croisement).

Pour avoir des résultats correspondants à des exploitations entières ou quasi-entières, il a été décidé après analyse de la distribution de ce ratio dans chacun des 4 départements de retenir celles dont le ratio dépasse 90 % (les 19 500 déjà citées), soit 80 % des exploitations de départ.

II.4. Evaluation du croisement spatial entre localisants parcellaires et RPG

Cette partie propose de comparer la précédente méthode (que nous appellerons méthode « vecteur »), nettoyage des micro-intersections compris, avec un croisement spatial entre les localisants parcellaires de la BD Parcellaire Raster de l'IGN et les îlots du RPG (que nous appellerons méthode « localisant »). Pour rappel, dans la BD Parcellaire Raster de l'IGN, un localisant parcellaire est un objet ponctuel, le plus souvent localisé au barycentre de la parcelle, dont les données attributaires correspondent aux références cadastrales de la parcelle qu'il matérialise (aucune valeur superficielle n'est associée, il faut donc la récupérer par ailleurs). Les localisants parcellaires de la BD Parcellaire Raster de l'IGN seront dénommés simplement localisants parcellaires.

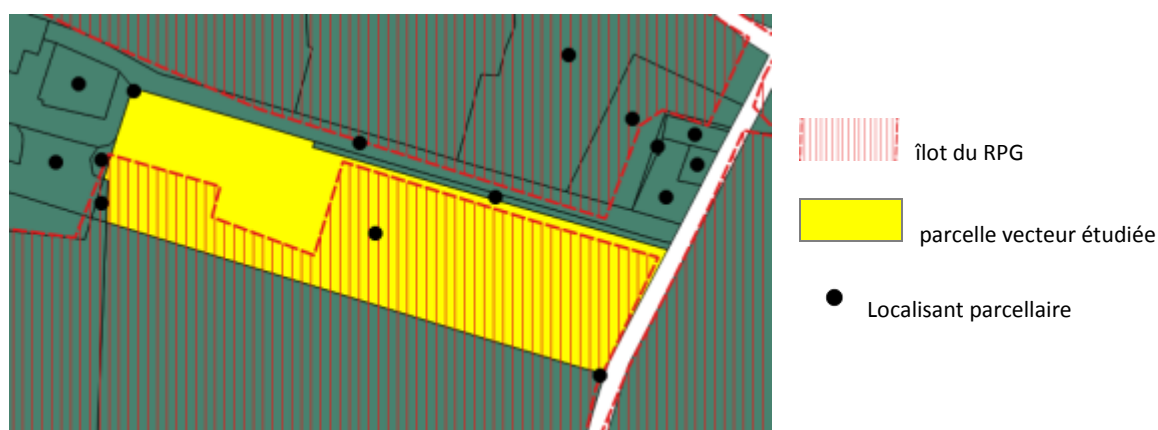
La comparaison des 2 méthodes permettra de connaître la fiabilité des traitements sur des départements dont le cadastre n'est pas encore vectorisé, et où seuls les localisants parcellaires permettent le croisement avec le RPG.

Principe et constats

Pour illustrer la différence entre des deux méthodes, la figure 10 montre une parcelle (référéncée 531800000E0725) au format vecteur en jaune, en partie recouverte par un îlot du RPG (zone hachurée rouge) et les localisants parcellaires sous forme de points noirs.

Dans cet exemple, la parcelle jaune n'est pas entièrement hachurée de rouge. En effet, certaines parties de parcelles (bois, sols...) ne font pas l'objet d'une activité agricole et ne sont pas déclarées à la PAC. Il est important de préciser que cette situation ne reflète pas la majorité des parcelles, il s'agit d'un cas particulier choisi pour bien illustrer la différence entre les deux méthodes.

Figure 10 : comparaison d'une intersection d'un îlot avec le polygone et le localisant d'une parcelle



Quand un croisement spatial entre le RPG et la parcelle vecteur est effectué, il est possible de calculer précisément la superficie de la partie jaune hachurée rouge (= la partie agricole de la parcelle). Dans le cas d'un croisement spatial entre le RPG et le localisant parcellaire (point au centre de la parcelle), il est juste possible de savoir si le localisant se situe à l'intérieur (il sera sélectionné) ou à l'extérieur de l'îlot (il ne sera pas sélectionné).

Tableau 5: Comparaison des informations recueillies selon les 2 méthodes pour le croisement d'une parcelle avec un îlot (Cf. figure 10)

| | Méthode « vecteur » | Méthode « localisant » |
|-------------------------------------|---------------------|---|
| Surface de l'objet vecteur parcelle | 1,5640 ha | ND (c'est un point) |
| Surface cadastrale de la parcelle | 1,5286 ha | 1,5286 ha |
| Résultat du croisement spatial | 1,0894 ha | Oui (le localisant est à l'intérieur de l'îlot) |

| | | |
|--|-----------|-----------|
| Surface de la parcelle affectée à l'îlot | 1,0894 ha | 1,5286 ha |
| % de la surface de la parcelle affectée à l'îlot | 70% | 100% |

Dans le cas de la figure 10, que la méthode localisant va surévaluer, au sein de l'exploitation agricole qui met en valeur l'îlot, la superficie affectée au propriétaire de la parcelle 531800000E0725.

Remarque : une différence sensible (354 m² soit 2,3%) entre la superficie du polygone vecteur (calculée) et la superficie cadastrale de la parcelle (mesurée sur le terrain). En observant ces différences sur toutes les parcelles étudiées, elles ne sont pas toujours en faveur des superficies calculées puisque, sur les 4 départements, 43 % des parcelles étudiées ont une superficie cadastrale supérieure à la superficie du polygone vecteur. Les superficies cadastrales sont au total 0,24 % inférieures aux superficies calculées à partir des polygones ce qui tout à fait marginal.

Les principaux cas de figure rencontrés avec la méthode « localisant » et leurs conséquences en comparaison avec la méthode « vecteur » sont détaillés en Annexe 3. D'après les observations réalisées, la tendance est à une légère surévaluation des superficies par la méthode « localisant » en raison d'une grande majorité cas de figures (cas 2, 3 et 4) pour lesquels les méthodes sont équivalentes (à la différence entre superficie cadastrale et superficie calculée près) additionnés des cas 1 et 6, fréquemment rencontrés, et pour lesquels l'intégralité de la superficie cadastrale est affectée à des parcelles partiellement concernées par un îlot du RPG.

Préparation des données et traitement

Comme pour le croisement de la BD Parcellaire vecteur avec les îlots du RPG, le croisement des localisants parcellaires de la BD Parcellaire Raster avec les îlots est réalisé après fractionnement communal (couche **ilot_com** générée précédemment), afin de gagner en temps de calcul

Préparation de la couche des localisants parcellaires :

Avant de procéder au croisement, il est d'abord vérifié que chaque localisant parcellaire se situe bien à l'intérieur du contour de la commune à laquelle la parcelle qu'il matérialise appartient. Après examen des localisants parcellaires sur les 4 départements (1,2 million de parcelles), 1 053 d'entre eux se situent hors de leur commune. Après avoir rapatrié depuis MAJIC les superficies des parcelles correspondantes, il se trouve que :

- La superficie totale des parcelles ayant des localisants parcellaires hors commune représente moins de 65 ha (sur les 2,2 millions d'ha cadastrés des 4 départements étudiés),
- Seules 9 de ces parcelles présentent une superficie supérieure à 1 ha avec une superficie moyenne de 1,56 ha

Ces parcelles ont été écartées au vu des superficies en jeu.

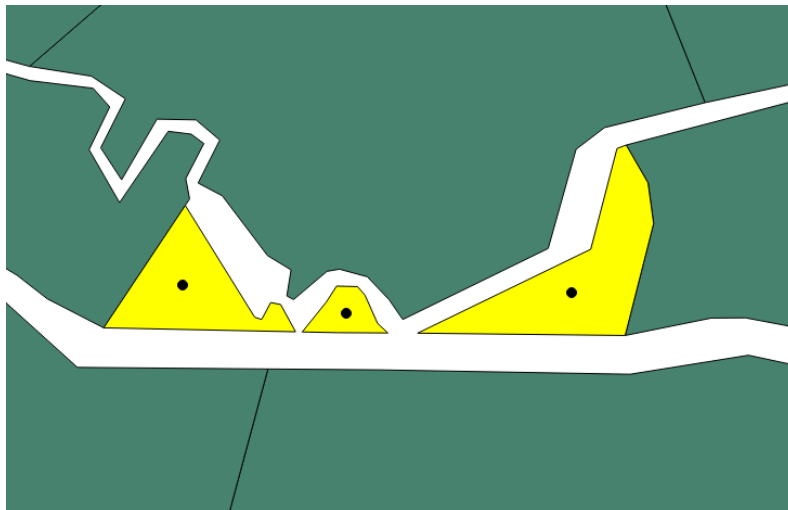
Toujours pour optimiser le temps de calcul et vérifier la qualité des localisants parcellaires, il est créé une table ne contenant que les localisants parcellaires « agricoles » (une table superficie agricole

déclarée est créée en fusionnant les îlots PAC par commune grâce à la commande ST_Union que nous intersectons ensuite avec la table des localisants à l'aide de la commande ST_Contains).

10 parcelles agricoles présentent plus d'un localisant parcellaire : pour 7 d'entre elles, il s'agit de doublons des localisants parcellaires (il s'agit du défaut de la BD Parcellaire de l'IGN pour quelques parcelles en double dans le département du Nord déjà évoqué dans le croisement « vecteur »). Ces doublons ont été supprimés.

Pour les trois autres parcelles, il s'agissait de parcelles particulières en plusieurs parties et qui présentent donc plusieurs localisants parcellaires (voir exemple ci-après). Les localisants parcellaires correspondant ont été conservés.

Figure 11 : parcelle composée de plusieurs polygones avec plusieurs localisants



La parcelle référencée 17024000ZS0001 est composée de 3 polygones (en jaune) dans la BD Parcellaire vecteur, et de 3 localisants (points noirs) dans la BD Parcellaire Raster.

La couche des localisants parcellaires ainsi préparée est appelée : **loc_par_agri_unique**

A ce stade, sont retenues 1 214 745 parcelles (1 214 748 localisants parcellaires) contre 1 684 748 dans le croisement du parcellaire vecteur (micro-intersections comprises).

*Croisement spatial des couches **loc_par_agri_unique** et **ilot_com***

Lors de l'appariement spatiale entre les localisants parcellaires agricoles (points) et les îlots du RPG (polygones), quand un localisant se situe à l'intérieur d'un îlot, il est sélectionné et l'identifiant de l'exploitation correspondante lui est affecté.

La table de croisement entre les couches **loc_par_agri_unique** et **ilot_com** contient 1 215 171 localisants : 423 localisants sont donc sélectionnés plusieurs fois. Il s'agit d'endroits où des îlots de RPG se superposent entre eux (comme expliqué précédemment) et avec un localisant parcellaire :

- soit quand des parcelles entières sont déclarées deux fois,
- soit quand les limites de deux îlots voisins se superposent légèrement mais au niveau d'une parcelle de petite taille ou de forme particulière, d'où sélection de son localisant).

Enrichissement de la table après croisement spatial :

Afin de poursuivre les traitements, les **superficies cadastrales** – issues de la base Majic – sont associées aux doublets îlots-numéros de parcelle des parcelles. La jointure permet de renseigner les superficies de 1 208 870 localisants parcellaires : 6 301 localisants (soit 0,5 % du nombre de localisants) n'ont pas été retrouvés dans MAJIC (il n'était donc pas possible de récupérer leur surface ni de les raccorder à un compte de propriété) : ils ont donc été écartés de l'analyse.

Les **propriétaires** liés aux parcelles sont ensuite intégrés à partir de la base MAJIC. La jointure permet d'affecter un code propriétaire unique à chaque parcelle (ou presque puisque de 40 parcelles appartenant à 14 comptes de propriété n'affichaient pas de propriétaires dans la base MAJIC). Les 40 localisants parcellaires correspondants ont donc été écartés de l'analyse.

La table affiche au final 1 208 830 localisants parcellaires agricoles avec pour chacun d'eux, une surface cadastrale, le numéro de l'exploitation du RPG, un compte unique de propriété (le même que pour la méthode vecteur).

Exploitation des résultats

Une table contenant pour chaque exploitation, les superficies détenues par chaque compte unique de propriété, pour la méthode « localisants » et pour la méthode « vecteur », est ensuite générée.

Cela permet de comparer pour chaque exploitation, pour les deux méthodes, le nombre de propriétaires en commun, mais aussi les superficies en commun pour chacune d'elles. Ces résultats sont ensuite agglomérés pour chaque département étudié **et pour l'ensemble de l'étude.**

Tableau 6: Comparaison du nombre moyen de propriétaires par exploitation suivant les 2 méthodes de croisement

| | | | | |
|-----------|-----------|-----------|-----------|--------------|
| 17 | 53 | 59 | 60 | Total |
|-----------|-----------|-----------|-----------|--------------|

| | | | | | | |
|--|-------------------------------|--------|-------|--------|--------|---------------|
| Nombre d'exploitations de l'étude (celles dont plus de 90% de la surface est cadastrée dans le département suite au croisement «vecteur») | | 5 433 | 6 313 | 5 208 | 2 553 | 19 507 |
| Rapport du nombre de propriétaires méthode «vecteur» / méthode «localisants» | | 103,4% | 99,5% | 103,8% | 100,5% | 102,6% |
| Nombre moyen de propriétaires par exploitation | Méthode « vecteur » (V)* | 15,5 | 5,5 | 22,3 | 18,9 | 14,7 |
| | Méthode « localisants » (L) | 15 | 5,6 | 21,6 | 18,8 | 14,4 |
| | Communs aux deux méthodes (C) | 13,4 | 4,6 | 20,1 | 16,3 | 12,8 |

*Obtenu après nettoyage des micro-intersections

La méthode « vecteur » affiche un nombre de propriétaires supérieur, ce qui pourrait laisser penser que le nettoyage des micro-intersections de la méthode « vecteur » est encore insuffisant sauf pour la Mayenne, où le nettoyage a un impact plus important du fait du plus petit nombre de propriétaires par exploitation dans ce département.

87 % (= C/ V) des propriétaires de la méthode vecteur ont été retrouvés par la méthode localisant ; ce pourcentage est assez satisfaisant. Les 13 % (= V/C) de propriétaires « en trop » avec la méthode vecteur correspondent à ceux pour lesquels le localisant n'a pu être affecté à une exploitation alors que l'intersection issue de la méthode « vecteur » était au-dessus des seuils du nettoyage – cas 5 et 7 de l'annexe 3. Pour les 11 % (= L/C) de propriétaires « en trop » avec la méthode « localisant », il s'agit de ceux pour lesquels la/les parcelle(s) ont été l'objet du nettoyage des micro-intersections alors que leur localisant a été sélectionné – cas 8 de l'annexe 3.

Tableau 7: Part des exploitation présentant les même propriétaires suivant les 2 méthodes de croisement

| | | 17 | 53 | 59 | 60 | Total |
|---|--|-------|-------|-------|-------|--------------|
| Rapport des surfaces méthode « vecteur » / méthode « localisants » | | 96,7% | 97,1% | 97,7% | 97,2% | 97,1% |
| Exploitations présentant exactement les mêmes propriétaires pour les deux méthodes | Nombre d'exploitations | 836 | 2 267 | 851 | 301 | 4 255 |
| | Part des exploitations étudiées | 15,4% | 35,9% | 16,3% | 11,8% | 21,8% |
| | Rapport des surfaces méthode « vecteur » / méthode « localisants » | 95,2% | 96,3% | 96,3% | 96,2% | 96,1% |
| Exploitations avec un pourcentage de propriétaires de la méthode "vecteur" retrouvés par la méthode "localisant" strictement entre 90% et 100% | Nombre d'exploitations | 975 | 272 | 1703 | 655 | 3605 |
| | Part des exploitations étudiées | 17,9% | 4,3% | 32,7% | 25,7% | 18,5% |
| | Rapport des surfaces méthode « vecteur » / méthode « localisants » | 96,5% | 96,8% | 97,5% | 96,6% | 96,9% |

Comme pressenti lors de l'analyse des cas de figure des croisements spatiaux, la méthode « localisant » surévalue toujours les surfaces (mais légèrement : + 2,9 %) par rapport à la méthode « vecteur ». En effet, la surface vecteur est calculée à partir du croisement exact de l'îlot et de la parcelle (surface allant de 0 à 100 % de la parcelle mais très souvent strictement moins de 100 %) et a subi un nettoyage alors que la surface obtenue par la méthode « localisant » est soit 0 % (il s'agit assez souvent de petites parcelles à la marge des îlots) soit 100 %. La règle du « tout ou rien » donne très souvent le « tout »...

Plus d'une exploitation sur 5 présente exactement les mêmes propriétaires pour les deux méthodes et 40 % des exploitations présentent plus de 90 % des propriétaires en commun.

Tableau 8: Rapport de surfaces obtenues selon les 2 méthodes par rapport aux surfaces initiales du RPG

| | 17 | 53 | 59 | 60 | Total |
|--|--------|--------|-------|--------|--------|
| Rapport surface méthode « vecteur » / surface RPG* | 97,2% | 97,8% | 97,3% | 97,6% | 97,5% |
| Rapport surface méthode « localisants » / surface RPG* | 100,5% | 100,7% | 99,6% | 100,4% | 100,3% |
| Rapport de la surface des propriétaires en commun pour les deux méthodes / surface méthode « vecteur » | 98,4% | 98,5% | 97,6% | 97,7% | 98,1% |
| Rapport de la surface des propriétaires en commun pour les deux méthodes / surface RPG* | 95,6% | 96,3% | 95,0% | 95,4% | 95,6% |

* il s'agit de la surface calculée des îlots (ou parties d'îlots) situés strictement dans le département pour les 19 507 exploitations étudiées

Les valeurs des différents rapports de surface présentés au tableau ?? sont comparables quel que soit le département.

Concernant le rapport surface méthode « vecteur » / surface RPG, il est de 97,5 % (les 2,5 points de différence étant dus pour 0,9 point à la présence de non cadastré dans les îlots - cas 2 de l'annexe 3 - et pour 1,6 point au nettoyage des micro-intersections - cas 1, 4 et 8 de l'annexe 3).

Comme attendu, le rapport surface méthode « localisants » / surface RPG est supérieur à 100, mais avec une valeur de 100,3 %, cette surévaluation est loin d'être excessive (nous pourrions l'évaluer à 0,3 % d'excédent + 0,9 % de non cadastré = 1,2 %).

Le rapport de la surface des propriétaires en commun pour les deux méthodes et de la surface en méthode « vecteur » permet d'affirmer que **plus de 98 % des surfaces de la méthode « vecteur » sont retrouvées par la méthode « localisant »**.

Enfin, pour mesurer la surface géométrique des îlots bien renseignée par la méthode « localisant », le rapport de la surface des propriétaires en commun pour les deux méthodes par rapport à la surface du RPG, donne 95,6 %, ce qui est très satisfaisant, surtout en considérant que 0,9 % des surfaces déclarées à la PAC sont non cadastrées : **sur la partie cadastrée le ratio s'élève donc à 96,5 % de surfaces affectées aux bons propriétaires par la méthode « localisant »**.

En matière de pistes d'améliorations de la méthode « localisant », notamment l'atténuation du phénomène de surévaluation des surfaces, un nettoyage des petits comptes de propriété pourrait être appliqué comme nous l'avons fait pour la méthode « vecteur ».

III. Méthodologie du croisement de 2 millésimes du RPG

Afin d'introduire des données dynamiques dans l'évolution des exploitations, le croisement de 2 millésimes du RPG a été effectué en marge de l'étude. Ce croisement permet de pallier partiellement le changement d'identifiant d'exploitation qui intervient d'un millésime à l'autre. Le croisement géomatique des couches de 2 millésimes successifs (ou distants de quelques années) permet d'évaluer l'évolution d'une exploitation, que ce soit du point de vue de la forme juridique ou de la surface totale.

III.1. Import des données littérales dans le tableur

Les données littérales associées aux couches **géographiques** sont fournies en format **.csv**. **C'est le numéro d'îlot qui constitue le champ clé permettant de relier les objets géographiques à ces données littérales. Le tableur utilisé pour leur mise en tableau est le logiciel Excel. Il faut veiller à la bonne correspondance des types de données** (caractère, ou numérique), notamment les données numériques décimales (avec points ou virgules).

Les données associées sont en 2 tableaux que l'on peut relier par le numéro identifiant de l'îlot :

- Les données de surface avec une ligne pour chaque culture de chaque îlot de cultures.
- Les autres données.

Création d'une variable synthétique Forme juridique de l'exploitation ou tranche d'âge. Les exploitations individuelles ont tous la même forme juridique. Pour elles la tranche d'âge de l'exploitant est indiquée : A moins de 40 ans, B de 40 à 50 ans, C de 50 à 60, D plus de 60 ans. On a créé un champ unique qui indique :

- la forme juridique pour les exploitations sociétaires,
- la tranche d'âge pour les exploitations individuelles.

Création de champs « surface dans l'îlot » propre à chaque culture. Le tableau des données de surface permet d'établir un tableau croisé, avec une ligne par îlot de culture et une colonne par type de culture. La jointure entre ce nouveau tableau et celui des autres données permet de calculer pour chaque îlot, sa surface cultivée par type de sol (terres arables, sols à culture pérenne, surfaces toujours en herbe).

Utilisation : comparaison des superficies comptées par le RA 2010 et comptées par le RPG 2010 selon les différents types d'usage du sol.

Regroupement des données d'une même exploitation. On ne peut sommer directement les surfaces déclarées des exploitations dans un fichier ou ces surfaces sont indiquées plusieurs fois (autant de fois qu'il y a d'îlots dans l'exploitation ; il faut donc :

- Trier les données par numéro d'exploitation (clé 1) et par îlot (clé 2)
- Numéroté de 1 à n les n îlots au sein d'une même exploitation.
- Créer des tableaux croisés avec un filtre sur ce « numéro d'îlot au sein d'une même exploitation »,

On peut alors obtenir des paramètres :

- Concernant les îlots : Nombre total d'îlots, Nombre moyen par exploitation, surface totale des îlots, surface par type de sol, par forme juridique ou âge de l'exploitant. (aucun filtre)
- Concernant les exploitations : Surface totale ou moyenne des exploitations, surfaces selon le département de rattachement de l'exploitation, selon la forme juridique. (filtre à « 1 »)

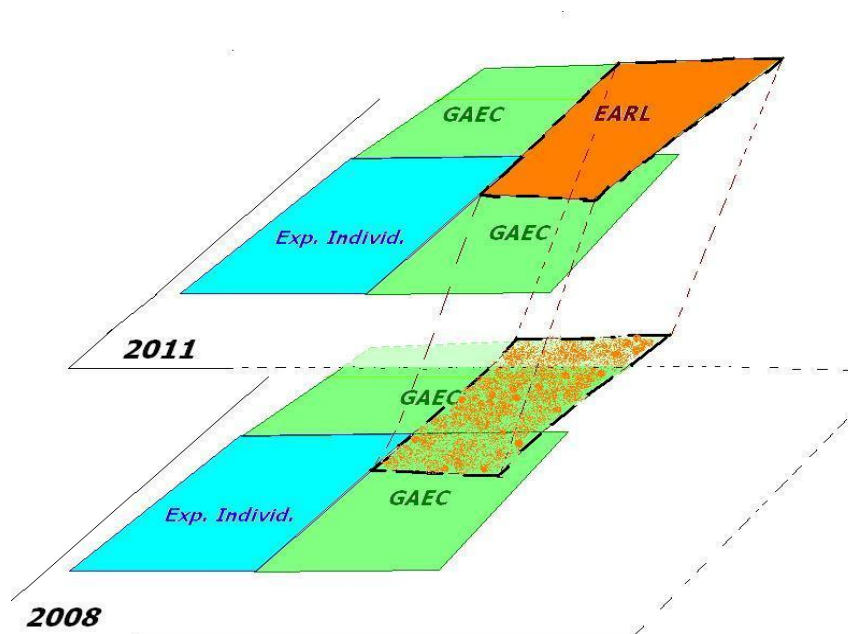
Utilisation : Comparaison sur deux années des superficies déclarées, et des nombres d'exploitations. Comparaison selon la forme juridique (ou la classe d'âge pour les exploitations individuelles).

III.2. Croisement de 2 fichiers du RPG de millésimes différents

L'objectif du croisement de 2 fichiers du Registre parcellaire graphique de millésimes différents est de cumuler sur une même surface, les informations relatives à celles de 2 années.

Les éléments des fichiers du RPG sont des îlots de culture. Si ces 2 éléments coïncident sur une même surface, le polygone que trace leur intersection a la même forme et le même emplacement pour les deux fichiers, on parle de « géométrie » commune aux deux fichiers. Cette géométrie commune va être utilisée pour opérer une jointure entre ces deux fichiers ; On parle alors de jointure spatiale entre 2 fichiers.

Figure 12: Exemple de jointure spatiale entre 2 millésimes du RPG



En opérant une jointure spatiale entre les fichiers cartographiques des deux années 2008 (en bas) et 2011 (au-dessus), on peut identifier un changement de forme juridique pour un terrain situé au départ dans un GAEC (en vert) et ensuite dans une EARL (en orange). On peut alors connaître la surface concernée par un changement et les informations (superficie totale, forme juridique) des exploitations concernées par ce changement.

III.2.1. Fichiers utilisés

Les paires d'années utilisées sont soit 2008 et 2011, soit plus souvent 2007 et 2012.

III.2.2. Formats et logiciels utilisés

Les données associées sont fournies en format .csv. Elles ont été traitées, pour leur mise en tableaux, par le logiciel Excel (voir paragraphe précédent, traitement des données associées).

Les données images sont des fichiers *shape*, Ces fichiers ont deux champs un identifiant d'îlot et un champ contenant les données vectorielles de l'image (il est nommé *obj* dans les tables de Mapinfo et des fichiers avec le logiciel PostGis).

La lecture de ces fichiers, la jointure avec les fichiers de données associées, l'analyse carto-statistique et la jointure spatiale entre 2 fichiers ont été réalisés avec le logiciel MAPINFO.

III.2.3. Système de projection

Il faut veiller à respecter le système de projection :

- Avant 2010 : NTF_Lambert_II_étendu ;
- A partir de 2010 : projection Lambert 93

III.2.4. Jointure de chacune des couches avec les fichiers de données associées

Une fois importée ces données vecteur, il faut opérer une jointure entre ce fichier et celui des données littérales associées.

Pour faciliter le croisement ultérieur de 2 fichiers et pour pouvoir automatiser les opérations, il est indispensable que les champs soient dans le même ordre et avec le même début de nom. Pour distinguer les 2 colonnes d'un même type d'information mais provenant de fichiers de 2 années différentes il faut rajouter l'année en suffixe à chaque nom de colonne : ex « Superficie îlot 2007 » et « Superficie îlot 2012 » au lieu de 2 champs « Superficie îlot » indiscernables l'un de l'autre.

La jointure vecteur / données associées se fait grâce à l'identifiant d'îlot commun. Il faut veiller à ce qu'il soit de même type dans les deux fichiers. On obtient des fichiers correspondant à un département et une année

Exemple : « Image et Donn Ass 017 2012 » pour le département n°17 et pour l'année 2012.

III.2.5. Jointure spatiale.

La jointure spatiale entre 2 fichiers An_A et An_B portant sur 2 années différentes s'opère par une requête SQL de la forme :

```
SELECT Infos du fichier An_A ; Infos du fichier An_B ; Areaoverlap(An_A.obj, An_B.obj)
From An_A; An_B
WHERE An_A.obj intersects An_B.obj
SORT BY : Infos du fichier An_A ; Infos du fichier An_B
```

On obtient un fichier contenant pour chaque intersection entre les îlots des 2 fichiers **An_A et An_B**:

- les infos du fichier An_A;
- celles du fichier An_B ;
- la surface de cette intersection.

Par exemple si l'information sur chaque fichier est la forme juridique on aura des lignes du type :

| F Juridique an A | F Juridique an B | Surface concernée |
|------------------|------------------|-------------------|
| GAEC | EARL | 0, 56 |
| GAEC | GAEC | 0, 01 |
| GAEC | GAEC | 0, 02 |

Ce qui s'interprète ainsi : il y a 0, 56 km² de GAEC qui se retrouvent dans un EARL et en tout 0, 03 km² de GAEC qui sont restés dans un GAEC

III.2.5. Analyse du fichier résultant.

A partir de ce fichier on peut établir un tableau à 2 entrées, celle des catégories issues des champs de l'an_A et celle des catégories issues des champs de l'an_B, il permet de connaître les superficies mises en jeu dans tous les types de changements de situation entre les années A et B

| Situation An A | | Exploitation Individuelle | | | | Forme sociétaire | | |
|---------------------------|----------|---------------------------|-------|-------|-----|------------------|------|-------|
| | | < 40 ans | 40-50 | 50-60 | >60 | EARL | GAEC | Autre |
| Exploitation Individuelle | < 40 ans | | | | | | | |
| | 40-50 | | | | | | | |
| | 50-60 | | | | | | | |
| | >60 | | | | | | | |
| Forme sociétaire | EARL | | | | | | 56 | |
| | GAEC | | | | | | 03 | |
| | Autre | | | | | | | |

Interprétation : dans l'exemple du tableau ci-dessus, on peut dire que 56 ha de GAEC sont maintenant en EARL, soit par transformation juridique du GAEC en EARL, soit par cession de l'exploitation.

Utilisations :

- Changement de forme juridique d'un même terrain. Flux de foncier exploité entre les différentes formes juridique.
- Devenir des exploitations . Devenir des GAEC lorsqu'ils sont dissous. (Formes juridiques et superficies des exploitations des repreneurs)

IV. Principe de calcul de l'indice de dispersion spatial des îlots d'exploitation*

*Variable employée pour établir la typologie des structures foncières d'exploitation, Cf. Rapport de synthèse de l'étude

L'indice de dispersion spatiale d'une exploitation est la somme des distances entre ses îlots pris 2 à 2 et pondérées par leur importance relative (en surface) au sein de l'exploitation.

Il peut être interprété comme la distance moyenne parcouru par un exploitant qui visiterai les différents îlots de son exploitation selon une fréquence liée à leur taille.

Ainsi si une exploitation de superficie S contient n îlots de surface S1, S2,...Sn , d'importance relative S1/S, S2/S,.....Sn/S, et distants entre eux des distances d(1,2) d(1,3), d(1,n), d(2,1), d(2,2), ...l'indice de dispersion spatiale sera

$$\sum_{i=1}^{i=n} (\sum_{j=1}^{j=n} d(i,j) \cdot R_j) \cdot R_i$$

Les coefficients Ri, Rj sont des rapports de surface. Ce sont donc des nombres sans dimension, l'unité de cette indice est donc la même que celui des distances (km par exemple).

| Libellé | Exploitation | llot | Surface de l'exploitation | Surface de l'îlot | Importance de l'îlot dans l'exploitation | distance avec l'îlot a | distance avec l'îlot b | distance avec l'îlot c | distance avec l'îlot d | ...//..... | distance avec l'îlot n | distance pondérée (DP) de l'îlot i aux autres îlots | contribution de l'îlot i à l'indice | Valeur cumulée de l'indice | indice de dispersion de l'exploitation |
|-----------------------|--------------|------|---------------------------|-------------------|--|------------------------|------------------------|------------------------|------------------------|------------|------------------------|--|-------------------------------------|--|--|
| Notation dans formule | e | i | Se | Si | Ri =Si/Se | d(i,a) | d(i,b) | d(i,c) | d(i,d) | | d(i,n) | DPi = d(i,a)*Ra +d(i,b)*Rb +d(i,c)*Rc +.....+ d(i,n)*Rn | DPi*Ri | Dp1*R1+DP2*R2 + +DPi*Ri | |
| | 1 | a | 8 | 6 | 0,75 | 0 | 4 | | | | | 1 | 0,75 | 0,75 | |
| | 1 | b | 8 | 2 | 0,25 | 4 | 0 | | | | | 3 | 0,75 | 1,5 | 1,5 |
| | 2 | a | 10 | 5 | 0,5 | 0 | 1 | 4 | | | | 1,1 | 0,55 | 0,55 | |
| | 2 | b | 10 | 3 | 0,3 | 1 | 0 | 5 | | | | 1,5 | 0,45 | 1 | |
| | 2 | c | 10 | 2 | 0,2 | 4 | 5 | 0 | | | | 3,5 | 0,7 | 1,7 | 1,7 |

Conclusion : perspectives d'amélioration du croisement entre propriété et exploitation

Les résultats obtenus montrent la faisabilité d'une analyse poussée de la propriété des exploitations avec les fichiers considérés. On peut lister les perspectives d'amélioration ou de complétion dans l'ordre décroissant :

La plus importante : identifier le ou les exploitants parmi l'ensemble des listes d'ayants-droit associés à l'exploitation.

Pour ce faire, le plus rapide serait de travailler sur le RPG de niveau 5 qui donne en plus le nom ou la raison sociale de l'exploitant (cf. circulaire de 2007 sur la diffusion du RPG). Cela serait particulièrement utile pour les 70 % d'exploitations individuelles, mais encore insuffisant pour les exploitations sociétaires puisqu'il manquerait l'identité des associés ou des gérants qui servirait à faire le lien avec la propriété de ceux-ci dans l'exploitation.

On irait plus loin en utilisant également les données MSA pour obtenir la liste nominative des exploitants d'une même exploitation extraite du fichier des cotisants non salariés.

Tout ceci nécessiterait une autorisation de la CNIL voire de la CADA pour le RPG de niveau 5. Obtenir l'autorisation de la CNIL serait plus facile pour le SSP en faisant valoir son statut de service statistique ministériel au sens de la loi de 1951 lors de sa déclaration de responsable du traitement.

Explorer les liens familiaux entre les premiers propriétaires/usufruiteurs de l'exploitation (identification du noyau des parents et de celui des enfants), pour mettre en évidence les stratégies de transmission familiale. On a déjà vu dans l'échantillon que le démembrement entre personnes physiques est très souvent fait entre le ou les parents qui ont l'usufruit, et un ou les enfants qui ont la nue-propriété.

Apporter des éléments de dynamique de la structure foncière en calculant des indicateurs d'évolution annuelle, sur l'exemple de la partie III du rapport de synthèse.

Analyser en détail la composition de la petite voire très petite propriété agricole dans les exploitations en comparant la réalité physique et la réalité sociale (existence ou absence de bail). Il faudrait peut-être compléter cette analyse par un travail de terrain.

Généraliser l'analyse à l'ensemble du territoire. Au-delà des recommandations techniques ci-après, l'étude à l'échelle nationale de données en grande partie structurées par département ouvre des questions méthodologiques :

- le croisement à partir des localisants de la BD Parcellaire est réalisable sur la France entière avec de bons résultats sur les propriétaires ; de meilleurs résultats peuvent être obtenus sur les surfaces en utilisant les données vecteur pour les 2/3 du territoire dans lesquels cette forme est disponible.
- il convient d'optimiser les requêtes pour limiter le temps de calcul (on arrive à l'échelle de 1 à 100 millions d'objets à l'échelle nationale par table selon le type de données d'où la nécessité de décomposer les calculs et d'écrire des algorithmes de jointure les plus efficaces possibles)

- pour s'affranchir du code de département, une nouvelle étape est à franchir dans le dédoublement des ayants droit ou groupes d'ayants droit (cela nécessite de réunir la propriété de personnes enregistrées dans différents départements donc avec des identifiants DNUPER différents).
- il faut également pouvoir également traiter le cas des exploitations à cheval sur plusieurs départements, sachant que l'identifiant n'est pas unique pour une exploitation multidépartementale.
- il faudrait également traiter des données sur les cultures pérennes, notamment en mobilisant le casier viticole pour couvrir la plus grande partie possible de la surface agricole.

Enfin, il faudra rapprocher le résultat de ces nouvelles analyses aux objectifs poursuivis concernant le besoin de portage du foncier agricole :

- **Selon le niveau de priorité donnée aux installations** hors cadre familial ou dans le cadre familial
- **Lien avec les politiques publiques** : efficacité et mise en application du contrôle des structures (encadrement des marchés par la sélection des candidats) et analyse des montages sociétaires (notamment holdings) dans l'optique d'une éventuelle régulation de l'optimisation fiscale

Recommandations pour une généralisation du croisement RPG-Cadastre à la France entière

Logiciels :

- Serveur Linux plutôt que Windows : pour cette étude de faisabilité, un serveur Windows a été retenu en raison de la connaissance par l'administrateur réseau de la FNSafer de cet OS. Cependant, en vue de la généralisation, l'utilisation d'une base de données Postgre installée sur un OS Linux pourrait être intéressante pour plusieurs raisons :
 - o PostgreSQL est considéré par les spécialistes comme étant plus performant d'environ 10 à 15 % et plus facile à installer sur Linux que sur Windows, ce qui devient non négligeable quand on traite de gros volumes de données. Il en est de même pour tous les plugins et QGIS. Cela s'explique par le fait que ces outils sont souvent développés en premier sur cet OS puis portés sur Mac OS (ce qui n'est pas trop compliqué puisque Mac OS est un système Unix) et Windows (ce qui nécessite plus de travail) ;
 - o On trouve plus facilement sur le web de la documentation sur des scripts en ligne de commande Linux que sur des scripts en langage Windows ;
 - o Il n'y a pas de coût de licence de l'OS ;
- Prendre toujours la dernière version stable de PostGIS car cette extension évolue encore beaucoup, notamment sur l'optimisation en temps de calcul des requêtes spatiales;
- Utiliser éventuellement un logiciel ETL (extract, transform, load) configuré pour les données produites par les administrations et établissements publics français pour importer les

données dans Postgre, plutôt que la moulinette d'import pour les shp et des scripts pour Majic. Point de vigilance : essayer de disposer d'une option de détection des erreurs et exclusion ou de correction des données non conformes.

Traitement sur l'ensemble des données:

- Ne jamais supposer que les données fournies sont conformes. Pratiquement tous les types de données testés dans cette étude comportaient des erreurs plus ou moins graves :
 - Erreurs spatiales dans le RPG : objets non conformes à la norme OGC, intersections partielles d'îlots d'exploitations différentes
 - Erreurs spatiales dans la BD Parcellaire : quelques objets non conformes à la norme OGC, des intersections partielles d'objets, 25 parcelles en double (totalement identiques, géométrie comprise) dans une commune de Nord ;
 - Erreurs d'encodage dans les fichiers Majic du non bâti.
 - Erreurs ou incohérences dans les fichiers de propriété qui supposent des vérifications systématiques (plusieurs droits de propriété, comptes mixtes de personnes physiques et morales, erreurs sur les noms, adresses, etc.)
 - Absence de renseignements de certains champs parmi les données associées de niveau 4 du RPG, comme celui de la surface des îlots. Cette donnée peut toutefois être récupérée avec les données de niveau 2 (surfaces par culture) ou calculée dans PostGIS (calcul d'aire)
- Systématiser les « plans de requête » (option explain et explain analyze en tête des requêtes) avant de lancer les calculs nationaux : le principe est de voir si des index supplémentaires doivent être créés pour accélérer le calcul et détecter dès le départ les calculs trop complexes pour les décomposer si possible.
- Utiliser les tables les plus légères possibles : au besoin créer des extraits de tables avec le nombre de colonnes et le nombre de lignes strictement nécessaire, quitte à pratiquer des jointures après.

Traitement des données spatiales :

Même si un premier contrôle est fait avec l'ETL, vérifier les géométries après import avec les opérateurs ST_IsValid et en cas d'erreur les explications avec ST_IsValidReason pour savoir comment corriger.

Après un croisement spatial, filtrer systématiquement la couche en sortie pour ne retenir que des objets de type attendu. Eliminer en particulier les *geometrycollections*, quitte à récupérer dans une couche spécifique les polygones qui en font partie ;

Pour avoir de bons résultats, donner la priorité aux opérateurs travaillant sur les géométries réelles par rapport aux opérateurs travaillant sur les enveloppes. La contrepartie est que cela oblige à découper le croisement si le nombre d'objets en entrée est important.

Lexique

csv : comma separated value, fichier de données littérales. La différence avec le format texte est que le séparateur est une virgule ou un point-virgule et non une tabulation ;

encodage (de caractères) : table normalisée qui indique la correspondance entre la valeur des octets d'un fichier et les caractères attendus. Il en existe plusieurs variantes, dont certaines permettent de traiter les caractères spéciaux et accentués ;

index : information d'une table de base de données avec une structure particulière permettant d'accélérer le tri des lignes. Il est donc logique de créer un index sur l'identifiant d'une table littérale et sur la couche géométrie de données spatiales ;

Majic : « mise à jour des informations cadastrales », fichier du cadastre diffusé par la DGFIP sous forme de fichiers à plat, variante d'un fichier texte. Ces fichiers à plat sont également appelés « fichiers fonciers standards » par la DGFIP ;

OGC : *Open Geospatial Consortium*, organisme qui édite des normes sur le traitement spatial des données. Des données spatiales peuvent avoir une structure correcte par rapport aux spécifications de l'éditeur du format, mais ne pas être conformes aux standards OGC, ce qui fait que les requêtes ne se terminent jamais ou donnent des résultats faux ;

Schéma : dans une base de données PostgreSQL, niveau intermédiaire entre la base de données et la table ;

Shp (ou *shape* ou *shapefile*) : format de données géographiques développé par ESRI pour les produits ArcGIS, ArcView, etc. Une couche de données au format shp comprend plusieurs fichiers ayant le même nom et une extension différente, au minimum un .shp (la donnée géographique proprement dite qui peut être un ensemble de points, de lignes ou de polygones), un .shx (index de la géométrie) et un .dbf (table de données littérales). On peut aussi trouver entre autres un .prj qui indique la projection géographique utilisée. Ce format est un standard de fait car il est bien documenté, ce qui fait que de nombreux logiciels libres savent le lire ou le créer.

ANNEXE 1 : Erreurs rencontrées lors de l'import du fichier Majic des parcelles non-bâties de la Mayenne 2013

1^{ère} erreur : liée aux accents dans les chaînes de caractères

COPY public.nonbati_tmp FROM 'D:/etude/NBAT_530.txt' ;

ERREUR: le caractère dont la séquence d'octets est 0x90 dans l'encodage « WIN1252 » n'a pas d'équivalent dans l'encodage « UTF8 »

Solution : remplacer dans un éditeur de texte un é majuscule dans la ligne par un E majuscule, trouvé dans un libellé de voie (« rue de Sablé » en majuscules sachant que le é majuscule n'est pas géré en Win1252 qui est pourtant bien l'encodage du fichier txt).

Le message équivalent avec la séquence d'octets 0x00 a été corrigé en remplaçant tous les caractères vides d'une ligne du fichier txt par des espaces. Il semble que ce code est lié au caractère NULL (vide).

2^{ème} erreur : liée aux retours chariot

ERREUR: retour chariot trouvé dans les données

HINT: Utilisez « \r » pour représenter un retour chariot.

Solution : éditer le fichier texte en appuyant sur entrée à la fin prévue de la ligne (après le 200^{ème} caractère dans le cas du non-bâti puisque la ligne comprend 200 caractères d'après la documentation)

3^{ème} erreur : valeur dépassant le format prévu initialement car intégrant des données innattendues

ERREUR: valeur trop longue pour le type character varying(200)

CONTEXT: COPY nonbati_tmp, ligne 3365744, colonne tmp : « 530265275 C0086
10000006912AP0003401011975 0 1 00498B4994... »

État SQL :22001

Quelques problèmes sont dus à la présence d'informations d'un article 10 (descriptif d'une parcelle) et 30 (exonération de taxation) d'une même parcelle dans une seule ligne.

Solution : Correction manuelle pour avoir les données d'une parcelle en se fiant à la documentation et aux parcelles voisines (voir le code et le libellé de voie par exemple ou laisser des vides après les champs bien codés).

4^{ème} erreur : valeur alphanumérique pour un entier

ERREUR: syntaxe en entrée invalide pour l'entier : « 000008AH0 »

État SQL :22P02

Solution : l'entier en question est la surface de la parcelle qui a été déclarée comme un nombre à 9 chiffres dans le script d'import. Une ligne n'avait par erreur que 7 chiffres aux emplacements attendus et les 2 caractères des variables alphanumériques suivantes étaient pris à la place.

Solution : mettre 2 « 0 » en plus aux numéros de cases attendus pour la surface de la parcelle.

ANNEXE 3 : principaux cas de figure des croisements spatiaux RPG avec les localisants parcellaires et conséquences (1/2)

| | |
|----------|--|
| <p>1</p> | <ul style="list-style-type: none"> - Cas fréquent - Parcelle concernée par 2 îlots - Surface Localisant = 3,32 ha - Surface Vecteur îlot A= 3,12 ha - Surface Vecteur îlot B = 0 car la surface de l'intersection (2134 m²) est inférieure à 10 % de la surface vecteur de la parcelle (3,33 ha)* <p>Conséquences : faibles</p> <ul style="list-style-type: none"> - Parcelle affectée au même îlot pour les 2 méthodes - Surface Localisant légèrement surévaluée (ici + 6,4%) |
| <p>2</p> | <ul style="list-style-type: none"> - Cas fréquent - Surface Vecteur = Surface Localisant (ici 5,37 ha) - Particularité : la Surface déclarée est légèrement supérieure à la surface cadastrale (du « non cadastré » a été déclaré) <p>Conséquences : nulles</p> <ul style="list-style-type: none"> - Parcelle affectée au même îlot pour les 2 méthodes - Surface Localisant = Surface Vecteur |
| <p>3</p> | <ul style="list-style-type: none"> - Cas fréquent - Surface Vecteur = Surface Localisant (ici 1,18 ha) <p>Conséquences : nulles</p> <ul style="list-style-type: none"> - Parcelle affectée au même îlot pour les 2 méthodes - Surface Localisant = Surface Vecteur |
| <p>4</p> | <ul style="list-style-type: none"> - Cas assez fréquent - Surface Vecteur = 0 car la surface de l'intersection (1310 m²) est inférieure à 10 % de la surface de la parcelle (2,08 ha)* - Surface Localisant = 0 <p>Conséquences : nulles</p> <ul style="list-style-type: none"> - Parcelle non affectée à l'îlot pour les 2 méthodes |



îlot du RPG



parcelle vecteur étudiée



Localisant parcellaire

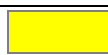
* voir règles de nettoyage des micro-intersections de la méthode « vecteur »

ANNEXE 3 (suite): principaux cas de figure des croisements spatiaux RPG avec les localisants parcellaires et conséquences (2/2)

| | |
|----------|---|
| <p>5</p> | <ul style="list-style-type: none"> - Cas assez fréquent - Parcelle concernée par 2 îlots - Surface Localisant = 24408 m² - Surface Vecteur îlot A = 16432 m² - Surface Vecteur îlot B = 7860 m² <p>Conséquences : parfois importantes</p> <ul style="list-style-type: none"> - Parcelle intégralement (24408 m²) affectée à l'îlot A par la méthode « Localisant » alors qu'elle est affectée aux îlots A (16432 m²) et B (7860 m²) par la méthode « vecteur » - Peut occasionner un propriétaire de moins pour la méthode « Localisant » (mais presque toujours pour la plus petite partie de la parcelle divisée) |
| <p>6</p> | <ul style="list-style-type: none"> - Cas assez fréquent - Surface Vecteur = 2,83 ha - Surface Localisant = 3,91 ha <p>Conséquences : parfois importantes</p> <ul style="list-style-type: none"> - Parcelle affectée au même îlot pour les 2 méthodes - Surface Localisant surévaluée (ici + 38%) |
| <p>7</p> | <ul style="list-style-type: none"> - Cas peu fréquent - Il s'agit en général de parcelles de petite surface - Surface Vecteur = 1787 m² - Surface Localisant = 0 <p>Conséquences : importantes si parcelle grande</p> <ul style="list-style-type: none"> - Parcelle non détectée par la méthode « Localisant » - Peut occasionner un propriétaire de moins pour la méthode « Localisant » |
| <p>8</p> | <ul style="list-style-type: none"> - Cas peu fréquent - Il s'agit en général de parcelles de petite surface - Surface Vecteur = 0 car la surface de l'intersection (438 m²) est inférieure à 500m²* - Surface Localisant = 757 m² <p>Conséquences : limitées</p> <ul style="list-style-type: none"> - Parcelle – en général de petite taille – conservée par la méthode « Localisant » alors qu'elle est éliminée par la méthode « Vecteur » - Peut occasionner un propriétaire de plus pour la méthode « Localisant » |



îlot du RPG



parcelle vecteur étudiée



Localisant parcellaire

* voir règles de nettoyage des micro-intersections de la méthode « vecteur »

