

Production d'une base de données géographiques exhaustive à échelle parcellaire sur l'occupation agricole du sol :

le RPG complété

(Campagne 2014 et perspectives)

Note méthodologique

Septembre 2020

(1^{ère} version, *non définitive*)

Pierre Cantelaube, Marie Carles, Benjamin Lardot, Thomas Poméon,
Claire Séard et Camille Truche

INRAE, U.S. ODR

Auteur de correspondance : pierre.cantelaube@inrae.fr





	Financements sur projet <i>OFB</i> dans le cadre de la convention <i>OFB-INRAE</i> - Projet 11 (2019-2020)
	Financements projet spatialisation BNVD, plan d'action national ECOPHYTO II (Axe 3 action 14) (2016-2018) axe 2, action 21 du programme Ecophyto, ONEMA et du MEDDE (2014-2015)

Table des matières

Avant-propos et résumé.....	4
1. Introduction.....	5
1. Données d'entrées pour la phase 1 :.....	5
1.1 CORINE Land Cover (CLC)	5
1.2 Bd Parcellaire Express 2019 V1 (IGN)	6
1.3 Bd TOPO 2019 V3.0 (IGN)	6
1.4 Bd CARTO 2018 V3.3 (IGN)	7
1.5 RPG 2015 à 2019 (IGN)	7
1.6 Petite Région Agricole	7
1.7 Référentiel commune <i>INSEE</i>	7
1.8 Conclusion sur l'évolution des données pour la phase 1	7
2. Données d'entrées pour la phase 2 :.....	8
3. Phase 1 : procédure d'identification des parcelles ODR	8
3.1 Programme principal de traitement des <i>Parcelles ODR</i>	9
3.2 Parallélisation et mémoire RAM	12
3.3 Librairies <i>Python</i>	12
3.4 Evolution de la phase 1 : résumé et perspectives	12
3.5 Conclusion sur la phase 1 « identification des parcelles agricoles hors RPG »	13
4. Phase 2 : nouvelle méthodologie d'attribution des cultures sur les parcelles ODR	14
4.1 Affiner le fond parcellaire dédié à l'activité agricole	14
4.2 Catégories de cultures cibles.....	16
4.3 un modèle pour affecter un groupe de culture aux parcelles ODR	16
4.3.1 Variables explicatives	16
4.3.2 Données d'apprentissage pour calibrer le modèle	18
4.3.3 Modèle Logit Multinomial	19
4.3.4 Résultats obtenus via le Modèle Logit Multinomial.....	19
4.3.4 Utilisation de la SAA	20
5 Résultats et remarques par type de culture.....	21
5.1 Prairies.....	21
5.2 Grandes cultures	23
5.3 Légumes.....	24
5.4 Vignes	28
5.4 Vergers	30
5.5 Oliviers.....	34
5.6 Landes et estives	35

6. Perspectives méthodologiques pour les futurs RPG complété	35
7. Format des fichiers du RPG complété	36
8. Conclusions.....	37
Références :	38
Annexe 1 : phase 1 Préparation des <i>Parcelles ODR</i>	40
A.1.1 Préparation des données	40
A.1.2- Construction du <i>Masque 1</i>	41
A.1.3- Différence Parcelles - RPG (<i>Parcelles 1</i>).....	42
A.1.4- Différence <i>Parcelles 1 – Masque 1 (Parcelles 2)</i>	42
A.1.5- Construction du <i>Masque 2</i> (Parcelles bâties)	43
A.1.6- Différence <i>Parcelles 2 - Masque 2 (Parcelles 3)</i>	43
A.1.7- Intersection avec CLC (<i>Parcelles 4</i>)	44
A.1.8- Traitement des Vignes, Vergers (<i>Parcelles 5</i>)	44
A.1.9- Traitement des Serres Agricoles (<i>Parcelles 6</i>).....	44
A.1.10- Multi polygones vers polygones (<i>Parcelles 7</i>).....	44
A.1.11 Jointure avec les Communes PRA	45
A.1.12- Finalisation des <i>Parcelles ODR</i>	45
A.1.13- Calcul des statistiques zonales.....	46
Annexe 2 - Préparation des couches d'occupations du sol non agricole susceptibles de recevoir des traitements phytosanitaires	47
A.2.1- Cimetières	47
A.2.2- Installations sportives	48
A.2.3- Aérodrômes	48
A.2.4- Voies ferrées	48
A.2.5- Routes	49
A.2.6- Jardins familiaux	49
A.2.7- Zones d'activité ou d'intérêt.....	49
A.2.8- Bâti	50
Annexe 3. Fonctions de nettoyage.....	50
A.3.1- Nettoyage des géométries.....	50
A.3.2- Nettoyage selon la forme	50
A3.3- Nettoyage selon la surface	51
Annexe 4. Notions de cadastre et de parcellaire agricole.....	51
Annexe 5. Référence à la Statistique Agricole Annuelle	53

Avant-propos et résumé

La cartographie d'occupation du sol agricole a été élaborée à l'ODR dans le cadre du programme EcoPhyto. Le but est de compléter le RPG pour proposer une *carte de la couverture exhaustive de l'espace agricole français*. L'idée est bien de disposer, à l'échelle parcellaire, d'une représentation géographique exhaustive de la surface agricole française (métropolitaine). L'objectif peut se résumer grâce à la « formule » suivante : *RPG + RPG complété = couverture exhaustive parcellaire de la surface agricole* (La surface agricole étant définie comme l'ensemble de la superficie des terres arables, des prairies permanentes et des pâturages permanents ou des cultures permanentes, voir CCE, 2016)

La méthodologie développée permet donc de cartographier un ensemble de parcelles agricoles absentes du RPG, nommées ci-après les « parcelles ODR ». Se décomposant en 2 étapes, le travail consiste (1) à repérer ces parcelles susceptibles d'accueillir des surfaces agricoles qui n'appartiennent pas à des exploitants recevant des aides PAC surfaciques, puis (2) à déterminer les cultures présentes sur ces parcelles identifiées. De telles cartes (dénommées « RPG complété ») ont été établies pour compléter les campagnes 2012 à 2014 du RPG (ODR-Ecophyto, 2015, Cantelaube et al., 2016, Séard, 2016). Ces cartes ont été produites suivant une première méthodologie qui est ici revue et améliorée.

Initialement, dans la première version du RPG complété réalisée précédemment (pour les campagnes 2012 à 2014), l'attribution des cultures aux parcelles reposait sur un traitement statistique et ciblait la différence entre les surfaces présentes dans le RPG et les statistiques départementales sur les surfaces agricoles fournies par le service Agreste du Ministère de l'Agriculture (Statistique Agricole Annuelle, SAA) (Agreste, 2014). Un modèle probabiliste permettait d'attribuer aux parcelles ODR une ou plusieurs catégories de cultures (nous parlions alors d'« îlot ODR » plus que « parcelles ODR ») de façon à ce que la somme des surfaces par cultures ainsi obtenues, sur un département, correspondait exactement à la différence SAA – RPG. Les résultats prenaient sens à une échelle agrégée (groupes de communes, cantons, petites régions agricoles, etc.) mais n'étaient pas garantis à l'échelle de la parcelle. De plus, les biais de comparaison entre le RPG et la SAA (surfaces des exploitations du département, différences temporelles) ainsi que le niveau géographique des entrées du modèle probabiliste (principalement l'échelle de la carte CORINE land COVER (CLC, 2009) rendaient notre résultat que modérément satisfaisant et peu utilisable à une échelle fine.

Dans la version 2 du RPG complété présentée ici et élaborée dans le cadre de ce projet, cette volonté de viser absolument la référence fournie par la SAA est abandonnée. Cette référence SAA affiche en effet des biais de comparaison avec le RPG (temporalité, notion de la localisation des surfaces au siège des exploitations, etc.), et s'en détacher est rendu possible grâce à un nouvel apport d'information thématique sur l'occupation du sol à une échelle fine : l'occupation du sol fournit par l'imagerie satellitaire via la carte OSO produite par le Centre d'Etudes Spatiales de la BIOshère (Cesbio).

Le RPG complété « version 2 » présenté dans ce document vise un double objectif : d'abord produire une carte parcellaire qui complète la campagne RPG 2014¹ mais aussi constituer un premier travail pour évaluer l'opportunité d'utiliser des cartes d'occupation du sol annuelles issues de la télégestion, - sous la forme de la carte OSO - pour décrire nos « parcelles ODR ». De plus, le processus de création

¹ Et ainsi disposer d'une carte parcellaire qui complète le RPG « ancienne version » (à l'îlot, produite par l'ASP, 2006-2014), rattachée à cette dernière campagne 2014.

du fond de carte annuel « parcelles odr » est revu, pour (1) intégrer les évolutions des données d'entrée, (2) corriger des défauts remarqués sur les 1^{ères} versions et (3) être automatisé pour assurer des productions rapides les futures campagnes RPG à compléter.

1. Introduction

Le Registre Parcellaire Graphique (RPG) est produit chaque année par l'Agence de Services et de Paiement (ASP) en collaboration avec l'Institut National de l'Information Géographique et Forestière (IGN). Cette base de données géographique recense les parcelles agricoles françaises (Cantelaube et Carles, 2015). Le RPG présente l'inconvénient de ne recenser que les parcelles des exploitations agricoles sujettes à des aides de la Politique Agricole Commune (PAC, aides du 1^{er} pilier et aides surfaciques du 2nd pilier). L'ODR a construit une méthodologie pour identifier et délimiter des parcelles dites « Parcelles ODR » correspondant aux parcelles agricoles qui sont exclues du RPG. Ce travail s'inscrit dans le cadre du projet *Spatialisation de la Base Nationale des Ventes des produits phyto (BNV-d)* financé dans le cadre du plan Ecophyto (ODR-Ecophyto, 2015 ; Séard C., 2016 ; Cantelaube et al., 2016 ; Séard et al. 2017).

La combinaison du RPG et du RPG complété permet donc d'obtenir une couche exhaustive parcellaire et annuelle de la couverture du sol agricole français (hors DOM). Cette couche de données « RPG complété » est produite depuis 2014 pour couvrir les campagnes 2012 à 2014.

Le RPG complété est construit en 2 phases. La 1^{ère} phase vise à identifier les parcelles susceptibles d'accueillir les parcelles agricoles absentes du RPG². Cette étape géomatique s'effectue grâce à l'utilisation de nombreuses couches d'informations géographiques, notamment issues de l'IGN : Bd Parcellaire® (Cadastre), Bd Topo®. Cette phase a été révisée en étudiant et en intégrant les **évolutions ces données d'entrées**.

La 2nde phase attribue à ces parcelles identifiées les cultures présentes. Cette étape utilise de nombreuses informations auxiliaires, notamment la statistique agricole annuelle (SAA, Agreste) et la couche CORINE Land Cover (CLC). Cette phase a été grandement modifiée en intégrant une nouvelle donnée auxiliaire pour l'attribution des cultures sur les parcelles identifiées : la carte d'occupation du sol issue de la télédétection ; la **carte OSO** développée par le CESBIO-theia³.

Le présent document décrit la méthodologie de production de la couche de données géographiques **RPG complété campagne 2014**, suivant les évolutions des 2 phases de travail citées ci-dessus.

1. Données d'entrées pour la phase 1 :

1.1 CORINE Land Cover (CLC)

CLC est un inventaire biophysique de l'occupation du sol et de son évolution selon une nomenclature en 44 postes. CLC est mis à jour tous les 6 ans. Elle est utilisée pour attribuer aux *Parcelles ODR* le code

² Note : nous nommons « parcelles odr » les parcelles identifiées lors de la phase 1 auxquelles sont attribuée une catégorie de culture lors de la phase 2.

³ <https://www.theia-land.fr/ceslist/ces-occupation-des-sols/>

CLC d'occupation du sol correspondant (« code_18 »). La version 2018 de la carte CLC remplace la version 2012 utilisée précédemment.

Documentation : [Guide d'utilisation CLC \(CLC, 2009\)](#)

1.2 Bd Parcellaire Express 2019 V1 (IGN)

La *BD Parcellaire* « historique » n'est plus entretenue depuis son ultime édition en 2018. Elle est remplacée par la *BD Parcellaire Express*, produit régulièrement mis à jour et plus exhaustif. La *BD Parcellaire Express* est issue du PCI vecteur de la *DGFiP* et sa structure est analogue à celle de la *BD Parcellaire* (même couches et même tables attributaires), mais sans les corrections topologiques du processus *BD Parcellaire* ce qui peut entraîner parfois des problèmes de cohérence avec d'autres couches. De plus quelques départements ne sont pas (encore) intégralement couverts par cette BD (08,10,11,13,52,55,67,70).

Note : Pour les départements où la *BD Parcellaire Express* est incomplète, la *BD Parcellaire* « historique » est utilisée si toutefois elle est plus exhaustive.

Couches : *PARCELLE* et *BATIMENT*.

Documentation : [DC DL SE Comparatif IGN - BD Parcellaire VS BD Parcellaire Express \(IGN, 2019a, IGN, 2019b\)](#)

1.3 Bd TOPO 2019 V3.0 (IGN)

La *BD TOPO* est une description vectorielle tridimensionnelle du territoire et de ses infrastructures de précision métrique. La V3.0 change beaucoup de structure (thèmes, classes, attributs et valeurs d'attributs) par rapport aux versions précédemment utilisées.

La mise à jour des thèmes BATI, OCCUPATION_DU_SOL et TRANSPORT suit le cycle de mise à jour des prises de vues aériennes (de 3 à 4 ans). Le thème ADMINISTRATIF est mis à jour en continu, et fiabilisé annuellement à partir du code géographique de l'INSEE. Enfin, concernant la couche ZONE_D_ACTIVITE_OU_D_INTERET du thème SERVICES_ET_ACTIVITE, les éléments d'importance au moins départementale (établissements administratifs, sportifs, de santé, d'enseignement...) sont actualisés tous les 1 à 2 ans à partir de diverses sources partenariales et consolidées par une expertise documentaire et un passage terrain.

La V3.0 est donc la plus susceptible de se rapprocher de la vérité terrain pour les années 2015 à 2019.

Thèmes : *ADMINISTRATIF, OCCUPATION DU SOL, HYDROGRAPHIE, TRANSPORT, BATI, SERVICES ET ACTIVITE*.

Couches : *COMMUNES, DEPARTEMENT, ZONE VEGETATION, SURFACE HYDROGRAPHIQUE, TRONCONS DE VOIE FERREE, TRONCON DE ROUTE, AERODROME, EQUIPEMENT DE TRANSPORT, BATIMENT, TERRAIN DE SPORT, CIMETIERE, ZONE D'ACTIVITE ET D'INTERET*.

Documentation : [Docs IGN \(IGN, 2020a\)](#)

1.4 Bd CARTO 2018 V3.3 (IGN)

La *BD CARTO* est une représentation vectorielle bidimensionnelle du territoire et de ses infrastructures. La couche *ZONE_OCCUPATION_SOL* utilisée est élaborée à partir d'images satellitaires de 2018. Elle permet d'accéder à certaines occupations du sol que l'on ne retrouve pas dans la *BD TOPO* (Carrières, décharges, névés, glaciers, marais, rochers, broussailles ou sable).

Documentation : [Docs IGN](#) (IGN, 2019c)

1.5 RPG 2015 à 2019 (IGN)

Le *Registre Parcellaire Graphique* évolue entre 2014 et 2015 : passage des îlots (regroupement de parcelles contiguës appartenant à une même exploitation) aux parcelles culturales. Des surfaces dédiées à des éléments agro-environnementaux peuvent apparaître (zones tampons, bandes enherbées). Cette évolution peut être intégrée sans problème et sans changement dans notre procédure de création des parcelles odr.

Documentation : [Docs IGN](#) (IGN, 2018 et IGN, 2020b)

1.6 Petite Région Agricole

Les Petites Régions Agricoles (PRA, Insee, 2009) sont des régions ayant une même vocation agricole dominante ; elles ont été délimitées par l'*INSEE* en 1946. Les limites des PRA s'inscrivent dans celle des départements administratifs La dernière actualisation date de 1981. Il n'y a pas de changements pour cette source de données.

1.7 Référentiel commune INSEE

Cette table, diffusée par l'*Insee* (Insee, 2008) permet d'attribuer aux parcelles odr une commune de rattachement (pour information, pas d'autre usage). Le référentiel utilisé est le référentiel Insee 2015. Les éventuelles évolutions de ce référentiel (division ou fusion de communes) pourront être prises en compte grâce aux tables de correspondances fournies par l'*Insee*.

1.8 Conclusion sur l'évolution des données pour la phase 1

Les bases de données utilisées sont toutes diffusées sous licence ouverte. Elles proviennent de l'*IGN* (<http://www.ign.fr/>), de la base de données européenne Corine Land Cover (CLC) (<https://land.copernicus.eu/pan-european/corine-land-cover>) et du Centre d'Etudes Spatiales de la BIOsphère (CESBIO) (<https://www.cesbio.cnrs.fr/>).

Toutes les données utilisent le système de projection *Lambert-93* sauf *CLC*. Cette couche est donc re-projetée en *Lambert 93* préalablement à tout traitement. Ainsi, l'ensemble des données est superposable, interopérable. Les échelles de production des différentes données sont cohérentes et inférieures au 1:50 000 sauf pour *CLC* (1 :100 000). A titre d'exemple, la surface de la plus petite unité cartographiée dans *CLC* (seuil de description) est de 25 hectares).

	<i>IGN</i>	<i>CLC</i>	<i>CESBIO (OSO)</i>
Système de projection	<i>RGF93 / Lambert-93 (SCR EPSG : 2154)</i>	<i>Lambert II étendu</i>	<i>RGF93 / Lambert-93 (SCR EPSG : 2154)</i>
Format	<i>Shapefile</i>	<i>Shapefile</i>	<i>Tif</i>
Résolution spatiale			10 x 10 m
Echelle(s)	1 : 2 500 au 1 : 50 000 selon les BD	1 : 100 000	
Etendue	France	Europe	France

2. Données d'entrées pour la phase 2 :

Le Centre d'Etudes Spatiales de la BIOSphère (CESBIO) laboratoire mixte CNES-INRAE, produit annuellement une carte d'occupation des sols à partir d'imagerie satellitaire (séries d'images optiques multi-temporelles à haute résolution spatiale), la carte d'occupation des sols du CES OSO (pour Occupation des Sols Opérationnelle) nommée OSO ci-après. L'idée est d'utiliser cette nouvelle source d'information dans notre méthodologie d'attribution d'une catégorie de culture sur les parcelles odr. Par conséquent, celle-ci sera significativement modifiée : cette carte OSO devient la principale source d'information pour nous.

La carte OSO décrit l'occupation du sol grâce à une nomenclature en 17 classes (2014-2017) puis en 23 classes (à partir de 2018). Cette évolution concerne principalement les catégories liées aux activités agricoles. Ainsi, dans la première nomenclature, les surfaces agricoles (hors Prairie et autres cultures pérennes) sont décrites par 2 classes : cultures d'été et cultures d'hiver. A partir de 2018, les cultures d'hiver sont détaillées en *céréales à pailles, colza et protéagineux* et les cultures d'été distinguent *soja, tournesol, maïs, riz et tubercules ou racines*.

Il faudra donc distinguer 2 « périodes » pour le développement méthodologique de notre RPG complété (voir §6).

3. Phase 1 : procédure d'identification des parcelles ODR

La procédure d'identification des parcelles agricoles hors RPG (parcelles ODR) s'effectue grâce à une succession d'opérations/traitements géomatiques (superposition de couches géographiques, intersection, effacement, etc.). Précédemment, ceci est effectué via un « ModelBuilder⁴ » du logiciel ArcGIS (ESRI) (voir ODR-Ecophyto, 2015, Séard, 2016). Cet ensemble de tâches géomatiques est réalisé maintenant d'une manière plus cohérente avec les outils du système de gestion des données géographiques (SGBD géographique) utilisées à l'ODR (notamment l'utilisation du SGBD PostgreSQL-PostGIS) et qui facilitera l'automatisation du processus global. Le langage Python a été choisi en ce but.

⁴ « Application permettant de créer, modifier et gérer des workflows qui permettent de concaténer des séquences d'outils de géotraitement, en injectant la sortie d'un outil dans un autre outil. ModelBuilder peut être comparé à un langage de programmation visuel de création de workflows. » Voir document ESRI à l'adresse suivante : <https://desktop.arcgis.com/fr/arcmap/10.3/analyze/modelbuilder/what-is-modelbuilder.htm>

Parallèlement, cette nouvelle procédure Python permet de construire plusieurs couches communales de territoires non agricoles susceptibles de recevoir des produits phytosanitaires (voies SNCF, terrains de sport, cimetières...). Ces Couches sont nommées « *couches Ecophyto* » dans la suite du document.

Ce changement de mode opératoire vise à améliorer, automatiser et accélérer davantage les traitements. Il s'agit de produire rapidement le *RPG complété* pour les années 2015 à 2019 et, en suite, d'être en capacité de proposer les *Parcelles ODR* correspondant au RPG le plus récent le plus rapidement possible après sa mise à disposition. Ce paragraphe décrit en détail le script Python.

3.1 Programme principal de traitement des *Parcelles ODR*

Le programme principal de traitements géomatiques s'effectue ainsi :

Les parcelles ODR sont construites à partir du parcellaire cadastral de la *BD Parcellaire Express*. Les différentes étapes du traitement viennent découper et supprimer les parcelles cadastrales déjà présentes dans le *RPG* ou identifiées comme étant non agricoles (surfaces bâties, végétation, routes...). Les parcelles ou morceaux de parcelles restants sont ensuite intersectées avec *CLC* puis avec les Vignes, les Vergers, les Serres (issus de la *BD TOPO*), les communes, les PRA et enfin la donnée *OSO* pour en récupérer les contours et/ou l'information.

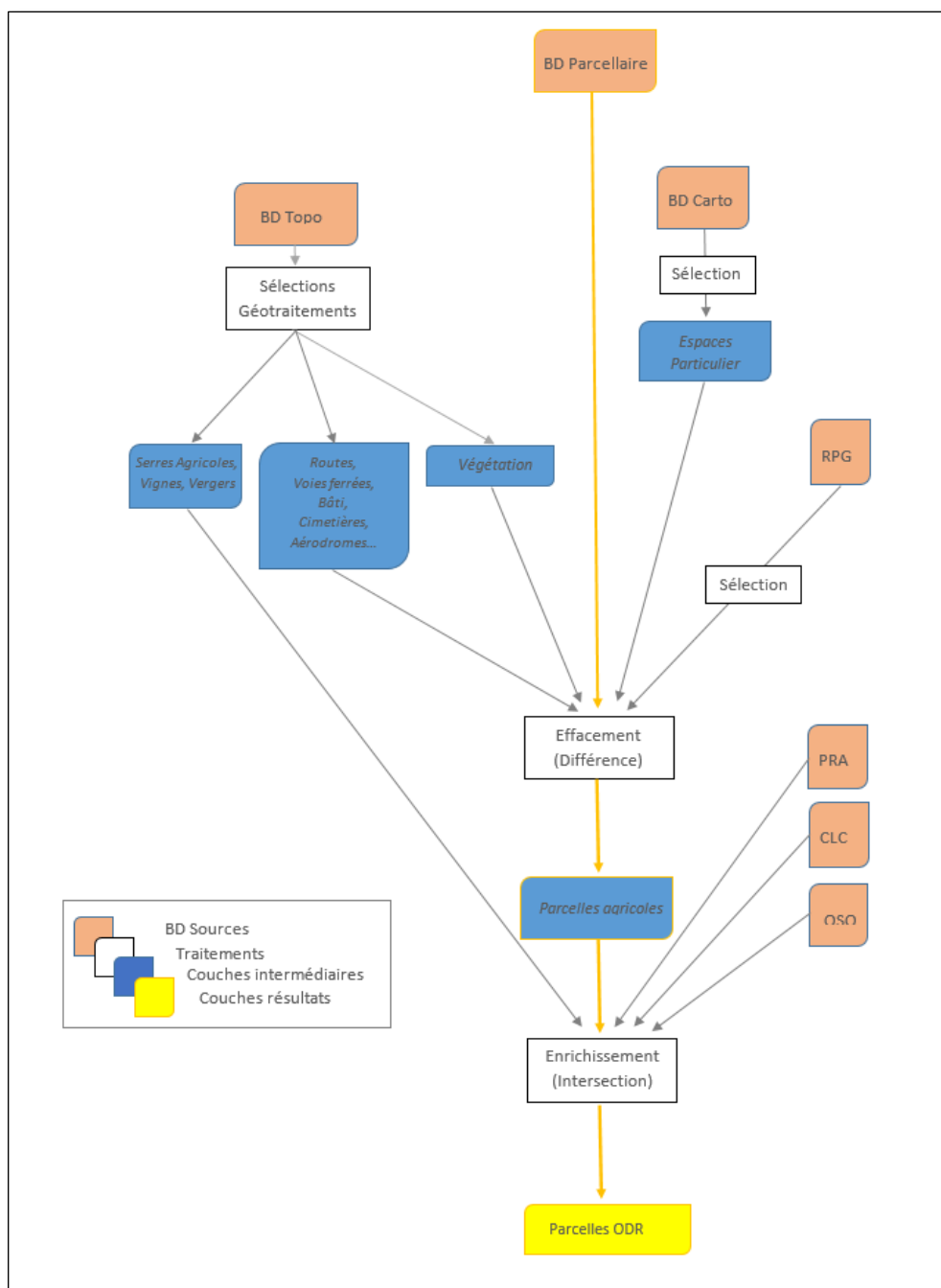


Figure 1 – Schéma global simplifié de la production des parcelles ODR.

L'ensemble du programme et des étapes successives réalisées sont décrits en annexe de ce document. Notons tout de même que les informations relatives aux surfaces dédiées aux vignes et aux vergers, ainsi qu'aux serres agricoles sont de grandes importances en vue de la 2^{ème} phase « attribution d'une culture aux parcelles ODR ». Ces informations proviennent de la Bd Topo® de l'IGN, respectivement des couches « Végétation » et « Bâti » (voir §1.3 ci-dessus). Lors des traitements, lorsque la *Végétation* rencontrée sur une parcelle est de nature « *vigne* » ou « *verger* », la parcelle n'est pas supprimée (contrairement à une parcelle de nature « forêt de conifère » par exemple) mais conservée. Il en est de même pour une parcelle constituée de « bâti » qualifiée de « Serre à usage

agricole » (contrairement à un bâti industriel ou commercial). Voir à ce sujet l'annexe 1.1 « préparation des données »).

Enfin, précisons que la généralisation des traitements à l'échelle nationale produit parfois des polygones indésirables : trop petits, géométries invalides ou vides, polygones très fins et longs (bordure de routes, de parcelles). Pour anticiper d'éventuels problèmes et en vue d'alléger les données à manipuler, 3 types de nettoyage sont appliqués après les grandes opérations spatiales du programme. Ces « nettoyages » sont décrits en annexe (annexe A.2)

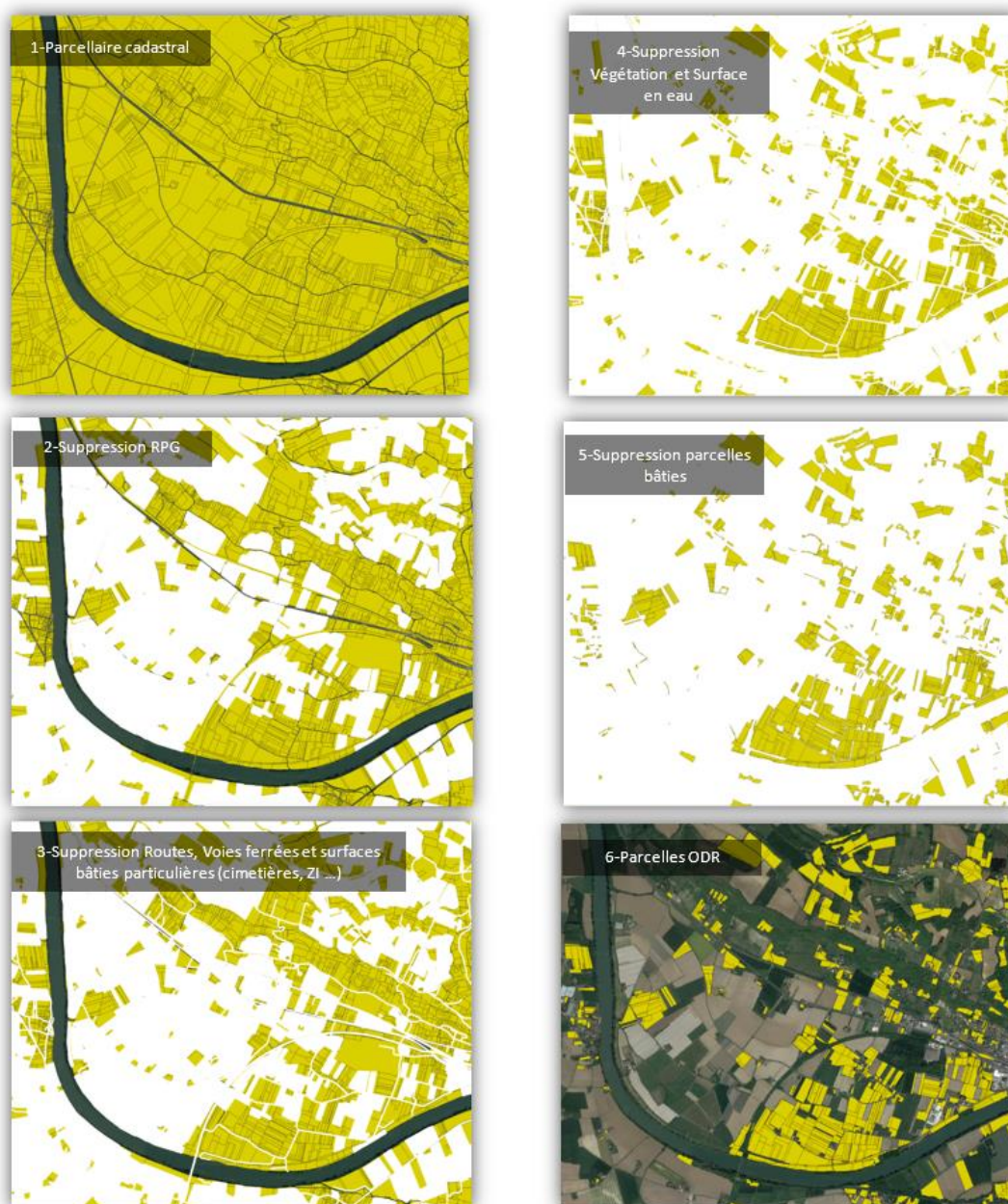


Figure 2 – Illustrations des étapes successives du processus de création du fond de carte « parcelles ODR »

3.2 Parallélisation et mémoire RAM

Afin d'accélérer la production, le programme est parallélisé (module « Multiprocessing »), permettant de répartir les traitements sur plusieurs CPU. Plusieurs départements peuvent ainsi être traités simultanément.

Le facteur limitant est la mémoire vive de la machine utilisée. Le script est assez « lourd », nécessitant de stocker de grandes quantités de données. D'autre part certaines opérations géométriques sont très coûteuses en ressources (différences et intersections). Pour gérer ce problème de mémoire vive, les données sont effacées à mesure qu'elles deviennent obsolètes par le module « Garbage Collector » qui revient régulièrement dans le script. Pour compléter ce processus de nettoyage, certaines données sont placées à « None » dès lors qu'elles ne sont plus utiles.

Une estimation des temps de calcul nécessaires peut être la suivante : la machine utilisée pour les campagnes 2015 à 2018 dispose de 16 Go de RAM et de 8 cœurs permettant de traiter 6 départements simultanément sans problème. Le temps moyen de traitement par département est de 1,8 heures (avec parallélisation), soit un peu plus de 7 jours pour l'ensemble des départements métropolitains.

3.3 Librairies Python

Nous listons ci-dessous les librairies Python utilisées

- **Pandas** et son extension **geopandas** permet de manipuler aisément de grands volumes de données avec une composante spatiale : lecture, jointure (spatiale et attributive), concaténation et opérations spatiales.
- **Time** et **Datetime** : pour mesurer les temps de traitement.
- **Glob** : pour faciliter la recherche des chemins d'accès aux données en entrée.
- **Multiprocessing** : permet la parallélisation en répartissant les traitements sur plusieurs processeurs.
- **Math** pour quelques calculs mathématiques.
- **Logging** pour créer et remplir le fichier de log.
- **Garbage collector** pour nettoyer la mémoire.
- **Rasterstats** pour calculer les statistiques zonales.
- **Numpy**.

3.4 Evolution de la phase 1 : résumé et perspectives

Comparé à l'outil de création des *parcelles ODR* précédent (ayant servi notamment à l'élaboration des *parcelles odr* 2014), la nouvelle procédure apporte de sérieuses améliorations :

- Les sources de données utilisées sont plus riches, actualisées et plus adéquates avec la temporalité du *RPG* à compléter.
- Les temps de traitement ont été divisés par 4 (1 mois => 1 semaine).
- Les *Parcelles ODR* sont maintenant enrichies de l'information sur les serres agricoles.
- Les *Parcelles ODR* sont également enrichies de l'information sur l'occupation du sol *OSO*.
- La prise en compte de l'intégralité du réseau routier permet d'éliminer certains axes qui étaient identifiés comme *Parcelles ODR* en 2014 par le model builder (routes desservant les lotissements notamment).
- Concernant les 8 couches d'occupation du sol, la Version 3 de la *BD TOPO* (largement enrichie par rapport aux versions précédentes) permet d'obtenir des résultats plus

exhaustifs.

- La procédure est presque intégralement automatisée et reproductible, grâce à la portabilité et l'utilisation de logiciel libre (*Python*).

Nous notons cependant encore certaines limites et perspectives :

La *BD Parcellaire* est encore partielle pour quelques départements et l'IGN travaille actuellement à compléter cette donnée.

La généralisation des traitements à l'échelle nationale ainsi que la succession d'opérations spatiales complexes produisent ponctuellement des polygones aux géométries problématiques. Dans quelques rares cas, ces polygones bloquent les traitements malgré les « nettoyages » appliqués. L'erreur *Python* est alors enregistrée dans le fichier de log et le script passe au département suivant. Le programme reste cependant robuste puisque pour la campagne 2015, seulement 1 département sur 94 n'a pu être entièrement traité (1 pour 2016, 2 pour 2017 et 1 pour 2018). Il pourrait être envisagé de corriger automatiquement les géométries des polygones lorsqu'ils bloquent les traitements.

3.5 Conclusion sur la phase 1 « identification des parcelles agricoles hors RPG »

Les *parcelles ODR* peuvent être vues comme des parcelles cadastrales retravaillées (voir Annexe 4) ; l'utilisation des parcelles cadastrales et leurs découpages éventuels par des éléments artificialisés (chemins, routes, etc.) et des éléments paysagers (bois, forêt, haies, etc.) permet d'approcher correctement – d'après les observations de ce que nous avons produit – le concept de parcelles culturelles et fournit des unités agricoles assez représentatives de la « réalité parcellaire » de l'occupation du sol agricole (hors RPG). Cet ensemble de parcelles identifiées est satisfaisant du point de vue des limites des « champs » ou des « prés ».

Il faut noter cependant la présence de parcelles agricoles trop découpées : la parcelle observée est représentée par plusieurs de nos parcelles, contiguës (*parcelle culturelle est constituée de plusieurs parcelles cadastrales*). L'inverse est très rare (*la parcelle culturelle est une subdivision d'une parcelle cadastrale*). Lors des observations et contrôles effectués, la fréquence d'apparition de cas où une de nos *parcelles ODR* englobe plusieurs parcelles culturelles réelles et accueille plusieurs types de culture est inférieur à 1%.

Nous notons donc une adéquation satisfaisante entre la parcelle agricole (« odr ») et la parcelle cadastrale « retravaillée » (voir l'ensemble des figures présentées ci-après), avec le bémol exposés ci-dessus (plusieurs de nos parcelles contiguës forment une seule parcelle culturelle « réelle »). Ce fait revêt tout de même de l'importance lors de l'attribution des cultures (phase 2 de la méthode), car la superficie de la parcelle peut être corrélées avec la culture qu'elle accueille ; cette superficie est une des variables explicatives des modèles utilisés (voir §4.3.1). Il y a donc un risque de prédire les cultures présentes sur plusieurs parcelles de « petites » tailles au lieu d'une seule culture sur une parcelle d'une taille « plus grande ».

4. Phase 2 : nouvelle méthodologie d'attribution des cultures sur les parcelles ODR

L'utilisation de la carte d'occupation des sols OSO est un atout très significatif pour la production du RPG complété, cette carte fournit en effet une information très importante pour l'attribution des cultures aux parcelles obtenues à la fin de la phase 1.

Intégrer cette information à nos procédures et nos traitements entraîne un changement complet de méthodologie. Il est nécessaire d'étudier cette carte, d'apprendre à l'utiliser pour répondre du mieux possible à nos objectifs. Ainsi, nous allons l'utiliser pour produire à nouveau un RPG complété campagne 2014 (de plus, au début du projet, les « parcelles odr 2014 » sont les plus récentes disponibles, le fond « parcelles odr 2015 » n'a pas encore été produit).

Comme indiqué ci-dessus (§2), la nomenclature OSO évolue à partir de la campagne 2018 et devient plus détaillée en ce qui concerne les catégories d'occupation du sol agricole. La 1^{ère} campagne OSO concerne l'année 2016. C'est donc cette couche qui va être utilisée pour qualifier les parcelles ODR 2014. Cette même couche sera utilisée pour les RPG complété 2015 et 2016. La carte OSO 2017 sera utilisée pour le RPG complété 2017. A partir de 2018, les campagnes RPG complété et les cartes OSO seront contemporaines et la nomenclature OSO plus détaillée et définitive⁵.

Les fonds de carte d'occupation du sol ne sont donc pas exactement contemporains avec la campagne RPG complété pour 2014 et 2015.

Nous avons donc développé une méthodologie basée sur la nomenclature OSO où les surfaces agricoles (hors cultures pérennes et prairies) sont décrites en 2 postes et cette méthodologie sera ensuite adaptée pour l'utilisation de cartes OSO où ces surfaces agricoles seront décrites avec beaucoup plus de précisions.

4.1 Affiner le fond parcellaire dédié à l'activité agricole

Le fond de carte « parcelles ODR » utilisée pour 2014 n'étant pas celui décrit ci-dessus mais une version précédente moins précises (d'un point de vue thématique), quelques ajustements sont nécessaires en préambule : les parcelles ODR 2014 contiennent encore des surfaces non agricoles (routes, zones urbaines ; surfaces en herbe d'utilisation non agricole).

Parmi ces parcelles, certaines peuvent être repérées grâce à la carte OSO (surfaces artificialisées) alors que pour d'autres, la catégorie OSO peut ne pas correspondre à une usage agricole (*occupation du sol* : surface en herbe et *usage du sol* : terrain de sports, par exemple). Dans ce cas, c'est la carte CLC qui sera utilisée, dans la mesure du possible.

Ainsi les parcelles ODR affichant une catégorie CLC 1.24 et 1.4.2 sont écartée (Tableau 1)

OSO	CORINE Land cover	ODR
Non pris en compte	1.2.4 aéroport/aérodrome	→ parcelle écartée, qualifiée « non agricole ».
	1.4.2 sport (golf, terrain de sport, hippodrome, etc.)	

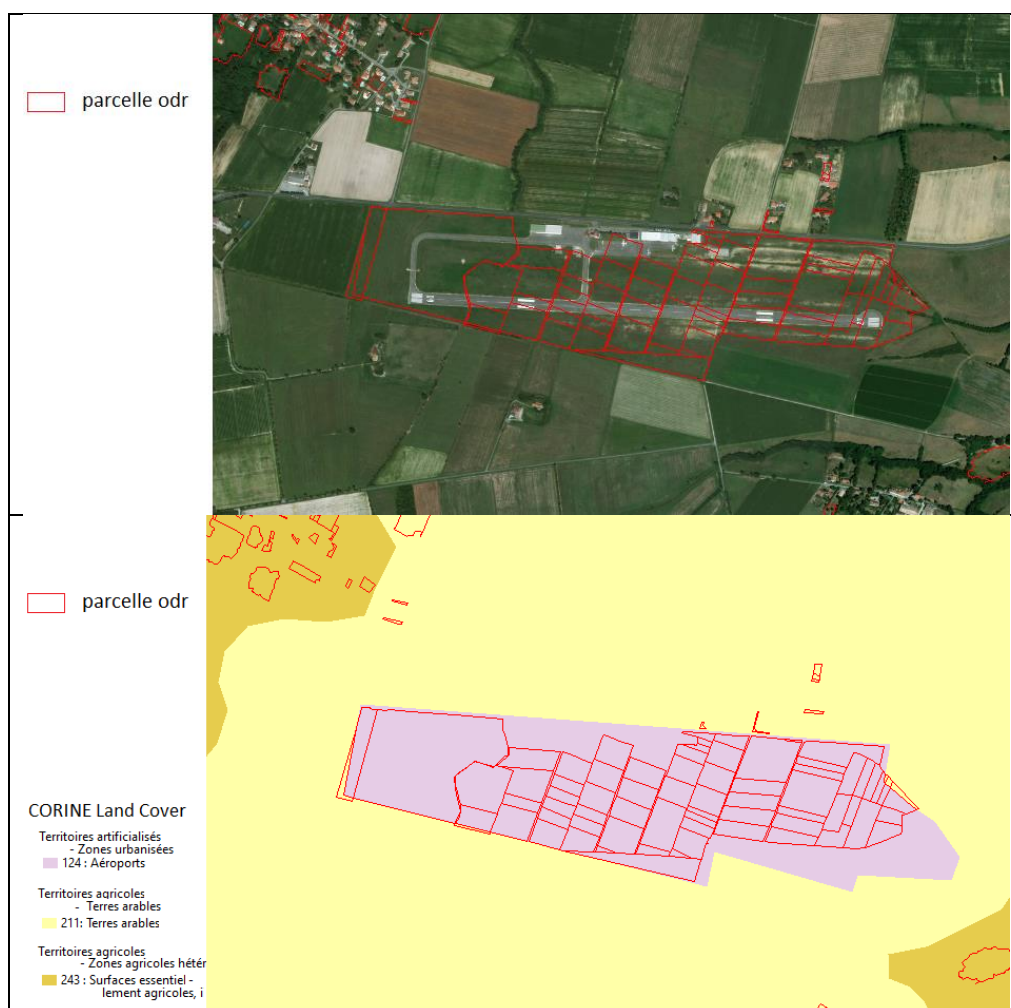
Tableau 1 – Parcelles ODR écartées

⁵ Cette nomenclature n'évoluera pas (du moins pour les catégories agricoles), information confirmée par les producteurs de la carte OSO.

D'un autre côté, nous appellerons parcelle « déduite » une parcelle non agricole dont l'occupation du sol peut être considérée comme certaine d'après les informations fournies conjointement par OSO et CLC. Ceci concerne les cas présentés dans le Tableau 2 ; de telles combinaisons semblant garantir une attribution correcte de la classe d'occupation du sol, c'est l'hypothèse que nous faisons. (Cette hypothèse peut néanmoins être localement erronée à cause de la différence temporelle entre la carte OSO et l'année ciblée 2014).

OSO(*)	CORINE Land cover	Déduction
41 urbain dense	1.1 Zones urbanisées	→ Surface artificialisée : non agricole
44 surfaces routes		
31 forêt feuillus	3.1 Forêts	→ Surface forestière/boisée : non agricole
32 forêt conifères		
51 eau	5.1 Eaux continentales	→ Surface en eau : non agricole

Tableau 2 - Parcelles ODR déduites



4.2 Catégories de cultures cibles

Le but est d'associer à chaque parcelle odr un groupe de culture de la nomenclature en 28 postes du RPG, plus une catégorie « *Non agricole* » (voir §4.1)

1	BLE TENDRE	15	LEGUMINEUSES A GRAINS
2	MAIS GRAIN ET ENSILAGE	16	FOURRAGE
3	ORGE	17	ESTIVES LANDES
4	AUTRES CEREALES	18	PRAIRIES PERMANENTES
5	COLZA	19	PRAIRIES TEMPORAIRES
6	TOURNESOL	20	VERGERS
7	AUTRES OLEAGINEUX	21	VIGNES
8	PROTEAGINEUX	22	FRUITS A COQUE
9	PLANTES A FIBRES	23	OLIVIERS
10	SEMENCES	24	AUTRES CULTURES INDUSTRIELLES
11	GEL (SURFACES GELEES SANS PRODUCTION)	25	LEGUMES-FLEURS
12	GEL INDUSTRIEL	26	CANNE A SUCRE
13	AUTRES GELS	27	ARBORICULTURE
14	RIZ	28	DIVERS

Tableau 3 – Groupes de cultures du RPG

4.3 un modèle pour affecter un groupe de culture aux parcelles ODR

L'objectif est de prédire les valeurs prises, sur chaque parcelle, par une variable catégorielle « occupation du sol » dont les classes correspondent aux cultures cibles présentés ci-dessus. Un modèle statistique est construit sur un échantillon d'apprentissage, puis est utilisé sur l'ensemble des parcelles ODR du département, afin d'estimer la classe de culture présente sur la parcelle. Nous travaillons à l'échelle du département (un modèle construit par département). Un modèle statistique vise à lier une variable à expliquer (occupation du sol sur la parcelle ODR) à l'aide de variables explicatives.

4.3.1 Variables explicatives

Un ensemble de *variables explicatives* est constitué, variables qui doivent permettre d'expliquer du mieux possible, lorsqu'elles sont prises en compte simultanément, la *variable à expliquer*, c'est-à-dire la classe de culture qui va être attribuée à chaque parcelle.

L'apport de la carte d'occupation du sol OSO est de première importance. Une parcelle odr étant constituée d'un ensemble de pixels OSO, il convient d'utiliser des indicateurs qui vont caractériser la composition de la parcelle en terme de catégorie d'occupation du sol OSO. D'autres variables sont utilisées pour qualifier nos parcelles odr.

► Variables explicatives issues de la carte OSO

Il s'agit de transférer l'information fournie par la carte OSO (ensemble de pixels) sur nos parcelles. Le premier indicateur utilisé est la catégorie d'occupation du sol majoritaire (en surface) dans la parcelle (indicateur nommé « *majorité* »). Le deuxième indicateur est le nombre de classes différentes qui apparaissent sur la parcelle (indicateur « *variété* »). Lorsque plus de 2 catégories d'occupation du sol sont présentes sur la parcelle, la médiane (indicateur « *médiane* ») de ces catégories apporte une information sur la part de la catégorie la plus présente⁶.

⁶ La valeur minimale a également été utilisée initialement, car celle-ci fonctionne sur le classement par ordre alphabétique ou alpha numérique des valeurs (ici catégories) rencontrées dans la parcelle. Parmi les catégories

→ *Perspectives*

Nous devons utiliser d'autres indicateurs ; en premier lieu des précisions sur les parts de chaque catégorie OSO qui apparaît dans une parcelle (pour remplacer *Majorité* et *Médiane*).

Et éventuellement des indicateurs qui rendent compte de la composition des parcelles en termes d'occupation du sol OSO, vraisemblablement des indicateurs de fractionnement. Nous ne pouvons pas non plus multiplier indéfiniment le nombre de variables explicatives, il s'agira donc de trouver un nombre minimum d'indicateurs qui synthétisent l'information.








Parcelle x OSO	Majorité	Variété	Médiane
	Bleu	1	Bleu
	Bleu	2	Bleu
	Bleu	2	Bleu
	Bleu	3	Bleu
	Bleu	3	Rouge
	Bleu	4	Bleu
	Bleu	4	Rouge
...

Tableau 4 - exemple de composition d'une parcelle Odr en termes de pixel OSO

de la carte OSO 2016, celles relatives aux activités agricoles (catégories d'intérêt pour nous) sont codées « 11 » ou « 12 » pour les cultures respectivement d'été et d'hiver, ou encore « 211 » (pour les prairies) et ces 3 codes sont les valeurs minimales des codes d'occupation du sol OSO (par rapport à 31 forêt, 42 urbain, etc.). Cet indicateur « minimum » permettait donc de détecter la présence d'au moins un pixel OSO « agricole » dans la parcelle. Cependant, ne constituant pas un apport significatif pour la qualité des résultats des modèles construits, l'idée de l'utiliser a été abandonnée.

► Autres variables explicatives

Il s'agit de recenser des informations sur la parcelle ainsi que de des données de contexte qui aideront à prédire la culture présente sur la parcelle. Les variables présentées dans le Tableau 5 ci-dessous ont été sélectionnées pour servir de variables explicatives :

Type	Nom variable	Code variable	Type/unité	remarques
Taille de la parcelle	Surface mesurée	Surf	Continue m ²	Attention pour les parcelles culturales trop découpées par les parcelles odr (voir §3.5)
Forme de la parcelle	Indice de Miller	IM	Continue -	Indice de circularité de la surface. IM entre 0 (=parcelle allongée, étirée) et 1 (=cercle, carré)
Caractérisation région agricole	Petite Région Agricole	PRA	Discrète -	Zonage, régions ayant une même vocation agricole dominante (voir geo.data.gouv.fr)
Caractérisation occupation du sol	CORINE Land Cover	CLC	Discrète -	Occupation du sol à grande échelle (voir eea.europa.eu)

Tableau 5 – Variables explicatives du modèle, hors carte OSO.

→ Perspectives

Il est envisagé d'intégrer une variable explicative de contexte « local agricole », plus fin que la PRA. Une piste est l'utilisation de la carte CLC et de ses unités géographiques, très utiles par exemple pour repérer un espace dédié spécifiquement à l'agriculture dans une zone plus urbanisée (nous pensons particulièrement au maraîchage).

Eventuellement ajout de l'altitude des parcelles. Peut-être plutôt comme « masque » (exclusion d'un certain type de culture au-dessus d'une altitude donnée), que comme variable explicative car n'apporterait pas nécessairement d'amélioration significative mais « coûterait » en degrés de liberté du modèle.

4.3.2 Données d'apprentissage pour calibrer le modèle

Il s'agit d'établir un ensemble de parcelles pour lesquelles nous connaissons l'occupation du sol afin de construire, calibrer le modèle.

Les parcelles déduites présentées dans le paragraphe 4.1 ci-dessus sont intégrées dans l'échantillon d'apprentissage. Ces parcelles servent à intégrer des catégories *non agricoles* dans le modèle (catégorie OSO « artificialisée » et « forestière »).

Certaines parcelles odr 2014 (donc absentes du RPG 2014) apparaissent dans le RPG 2015. : l'hypothèse est faite qu'elles pourraient être également des parcelles agricoles en 2014. Elles sont donc intégrées à l'échantillon d'apprentissage, avec la culture présente en 2015. Par la suite, sur ces

parcelles, soit l'information fournie par le RPG 2015 est conservée, soit elle est changée suivant une rotation culturale en fonction des manques observées par rapport à la SAA (voir Annexe 5).

Enfin, avoir des observations peut améliorer sensiblement la qualité prédictive du modèle. Ces observations peuvent être obtenues sur des parcelles « connues » (par exemple sur des zones d'études ou de suivies), mais cette éventualité reste très rare et nous n'avons pas utilisé de telles informations. Une autre façon d'avoir des observations est l'utilisation d'images satellites ou aériennes auxquelles sont superposées des parcelles ODR. C'est ce que nous avons fait, malgré d'éventuels biais temporels (dates des prises de vue souvent difficiles à connaître). Malgré ces biais, cette méthode permet de bien distinguer des prairies de cultures ou d'autres surfaces non agricoles, ou encore de grandes cultures annuelles de surfaces utilisées pour le maraîchage, etc. Ceci sous des hypothèses de stabilité sur quelques années.

4.3.3 Modèle Logit Multinomial

Le modèle Logit Multinomial est un modèle probabiliste non linéaire utilisé lorsque la variable à expliquer est de type quantitative à plusieurs catégories non ordonnées. Sans entrer dans les détails de la théorie statistique, un modèle Logit Multinomial vise à expliquer (« modéliser ») la probabilité d'apparition d'événements (ici apparition d'une catégorie de culture ou de « non agricole » sur une parcelle) en fonction des variables explicatives (dites également variables descriptives). Ces dernières peuvent être aussi bien quantitatives que qualitatives.

Le modèle est calibré sur l'échantillon d'apprentissage, puis utilisé pour prédire la catégorie présente sur des parcelles qui n'ont pas servis à sa calibration. Ces prédictions sont fournies sous forme de probabilité d'occurrence de chaque catégorie de la variable d'intérêt. Généralement la catégorie correspondante à la probabilité la plus haute est retenue⁷, c'est ce qui a été fait ici.

Les résultats sur cet échantillon d'apprentissage sont très bons : pour 84 départements la prédiction de la catégorie d'occupation du sol est correcte à plus de 90%, pour les autres ce taux de prédiction correcte ne descend jamais en dessous de 78%.

Ces résultats sont bien sûr à nuancer pour la prédiction sur des parcelles n'ayant pas servies à la calibration du modèle. D'abord et principalement car les parcelles « *déductions* » influencent favorablement ces résultats car leur prédiction est plutôt aisée. Si on les retire de l'échantillon d'apprentissage pour évaluer les résultats des modèles (pas pour calibrer le modèle !), les taux de prédictions correctes baissent de 10 points environ : ainsi, en reprenant la phrase ci-dessus : *pour 84 départements la prédiction de la catégorie d'occupation du sol est correcte à plus de 80%, pour les autres ce taux de prédiction correcte ne descend jamais en dessous de 70%.*

4.3.4 Résultats obtenus via le Modèle Logit Multinomial

Notons tout d'abord qu'en ce qui concerne les vignes et les vergers, la carte OSO 2014 présente de nombreuses erreurs thématiques et ces 2 catégories de cultures sont estimées uniquement à partir des données de la Bd Topo® de l'IGN, couche végétation (voir les paragraphes dédiées ci-dessous).

⁷ Mais ceci peut être affiné, par exemple dans le cas d'une variable binaire si les probabilités de chaque classe sont 52% et 48%.

Le Tableau 6 résume les vérifications⁸ faites sur des parcelles non utilisées pour la calibration des modèles. Il faut noter (à nouveau) que les observations peuvent être biaisées par des différences temporelles entre la campagne du RPG complété et les images satellites utilisées (souvent non datées).

<i>estimé</i>	<i>observé</i>						
	Non agricole	Cultures	Prairies	Landes	Vergers	Vignes	Divers
Non agricole	66.92	7.60	5.08	8.83	4.99	1.65	3.8
Cultures	6.22	71.08	16.01	2.87	1.27	1.49	1.06
Prairies	4.5	3.5	80.59	8.15	1.03	1.78	0.45
Landes	11.88	2.95	28.11	52.15	4.00	0.03	0.60
Vergers	3.47	0.11	3.48	1.03	90.75	0.01	1.17
Vignes	0.05	0.11	0.13	0.02	0.03	99.62	0.03
Divers	10.44	6.04	10.07	16.62	2.96	5.85	48.02

Tableau 6- tableau croisé Prédiction x Observation (surfaces) en pourcentage (100% total par ligne)

4.3.4 Utilisation de la SAA

Si dans le tableau ci-dessus toutes les grandes cultures ont été regroupées (pour une évaluation globale), les sorties du modèle Logit distinguent les catégories de cultures ; ces sorties sont en effet calées sur la nomenclature en 28 classes du RPG. Ainsi le modèle prédit des parcelles en blé, maïs, orge, tournesol, etc. Nous obtenons donc des « *surfaces RPG complété* » pour chaque type de culture

D'un autre côté, en comparant les surfaces RPG avec les surfaces indiquées par la SAA, par cultures, une estimation des surfaces absentes du RPG – par département - est réalisée. Nous calculons ainsi la part de chaque groupe de culture dans le total manquant dans le RPG pour l'ensemble de ces grandes cultures (nommons la « *répartition différentiel SAA – RPG* »).

Généralement, la *répartition des surfaces RPG complété* ne correspond pas à cette *répartition différentiel SAA – RPG*. Typiquement, nous prédisons 500ha de blé, 100ha de maïs et 100ha de Tournesol sur le département alors que, d'après la SAA, il manque 100ha de blé et 600ha de maïs alors que le tournesol est bien représenté dans le RPG. Nous changeons alors 400ha de blé prédit en maïs, ainsi que les 100ha de tournesol. Les 100ha de maïs sont conservés en l'état. Ces changements sont effectués de manière aléatoire, par groupes de communes néanmoins, pour éviter d'avoir des parcelles proches ou contigües qui alternent des cultures très différentes.

Ainsi, la référence à la SAA est encore utilisée mais est moins fondamentale que dans la première version du RPG complété (voir *Avant-propos et résumé* ci-dessus). Suivant cette référence, nous agissons sur nos sorties de modèle en effectuant une sorte de « *rotation culturale* ». N'oublions pas que le modèle est calibré sur des observations difficiles à dater avec exactitude ainsi qu'avec les

⁸ Sur 5 départements : 29, 33, 47, 50 et 84

parcelles qui apparaissent dans le RPG de l'année suivante. Cette méthode nous permet également d'obtenir des surfaces en jachère et en prairie temporaire.

Est-ce que cette façon de procéder pourrait être améliorée ? Vraisemblablement sous condition de données auxiliaires « locales » et précise thématiquement (comme indiquée, la comparaison avec la SAA peut être biaisées et de plus, nous utilisons en 2014 une carte OSO de 2016). Ou encore en suivant un raisonnement agronomique plus solide sur les pratiques de rotations culturales. Finalement, nous avons décidé de ne pas dédier plus de temps à ce sujet, considérant que l'évolution de la nomenclature de la carte OSO fera qu'à partir de 2018, ce problème ne se posera pas ou beaucoup moins (*Céréales à paille* à répartir entre *Blé*, *Orge* et *autre céréales*).

5 Résultats et remarques par type de culture

5.1 Prairies

Les prairies sont bien repérées, le RPG complété peut être considéré comme fiable pour cette catégorie d'occupation du sol. La carte OSO est très bonne pour cette classe d'occupation du sol, la difficulté pour nous vient du fait que toutes ces surfaces ne sont pas dédiées à une « utilisation agricole ». En effet, même si nous avons écarté les terrains de sport ou les aérodromes, les surfaces que nous qualifions de prairies permanentes peuvent encore être en fait situées sur jardins individuels, des parcs, ainsi que des terrains militaires⁹.

Nous notons également des surfaces qualifiées par OSO (2016) comme surface cultivées qui paraissent plus vraisemblablement être des prairies. Un autre point délicat concerne la difficulté pour évaluer le caractère temporaire des prairies. Ces 2 remarques sont liées.

La plus grande confusion se trouvent pour des parcelles qui semblent plus être des landes, dans le sens « prairie naturelle non exploitée », sans doute temporairement, ou broussailles, mais qui ne paraissent ni être pâturées, ni fauchées...

⁹ Ces terrains sont écartés également avec la nouvelle procédure pour la phase 1 développée durant ce projet, rappelons que pour 2014 nous avons utilisé les parcelles odr produites avec la procédure précédente.



Figure 4 – Exemple de parcelles du RPG complété en prairie permanente.

→ Perspectives

Pour les surfaces en herbe qui ne sont pas à usage agricole, l'amélioration de l'étape (1) de la construction du RPG complété ainsi que la disponibilité des couches IGN plus complètes et détaillé (en nomenclature thématique) devraient permettre d'améliorer leur détection.

Pour le caractère temporaire des prairies, difficile d'améliorer ce point sans un suivi des parcelles sur plusieurs années.

Pour les prairies naturelles non exploitées, leur classement en « landes » pourrait s'effectuer en augmentant le nombre d'observations qui calibre les modèles. Reste à trouver une définition plus précise de cette classe.

5.2 Grandes cultures

L'évaluation de la qualité de prédiction des grandes cultures entre elles est difficile à cause de la difficulté d'obtenir des observations fiables. Ceci même en regroupant les cultures, par exemple pour reclasser plus simplement en Céréales, oléagineux, protéagineux et autres (Cultures industrielles, semences, plantes aromatiques, etc.). Les légumes sont un peu à part (voir § ci-dessous).

Les observations sur images satellites peuvent être biaisés par les différences temporelles (RPG complété 2014 basé sur OSO 2016) et les changements des cultures sur une parcelle dues aux rotations culturales. La comparaison entre le RPG et la SAA fournit une information de première importance pour l'attribution finales d'une catégorie de grandes cultures sur les parcelles.

Remarquons également la difficulté de repérer les surfaces en jachère.

Globalement les surfaces en grandes cultures sont bien représentées dans le RPG, en pourcentage. Néanmoins, les surfaces brutes sont si grandes (maïs, blé, ...) que des absences importantes peuvent apparaître localement (voir par exemple la Figure 5 au début de ce document, où le RPG complété apporte près de 90 ha de maïs sur un ensemble de parcelles contigües).

→ Perspectives

Le fait que la nomenclature évolue pour devenir bien plus détaillée pour les cultures dans les cartes OSO à venir (à partir de 2018) devrait constituer un apport très important pour le RPG complété (la présente version est effectuée à partir d'une nomenclature en 2 postes : cultures d'été et *cultures d'hiver*).

Cette évolution nous incite à ne pas consacrer trop de temps ici à cette problématique de la qualification « précise » des grandes cultures.



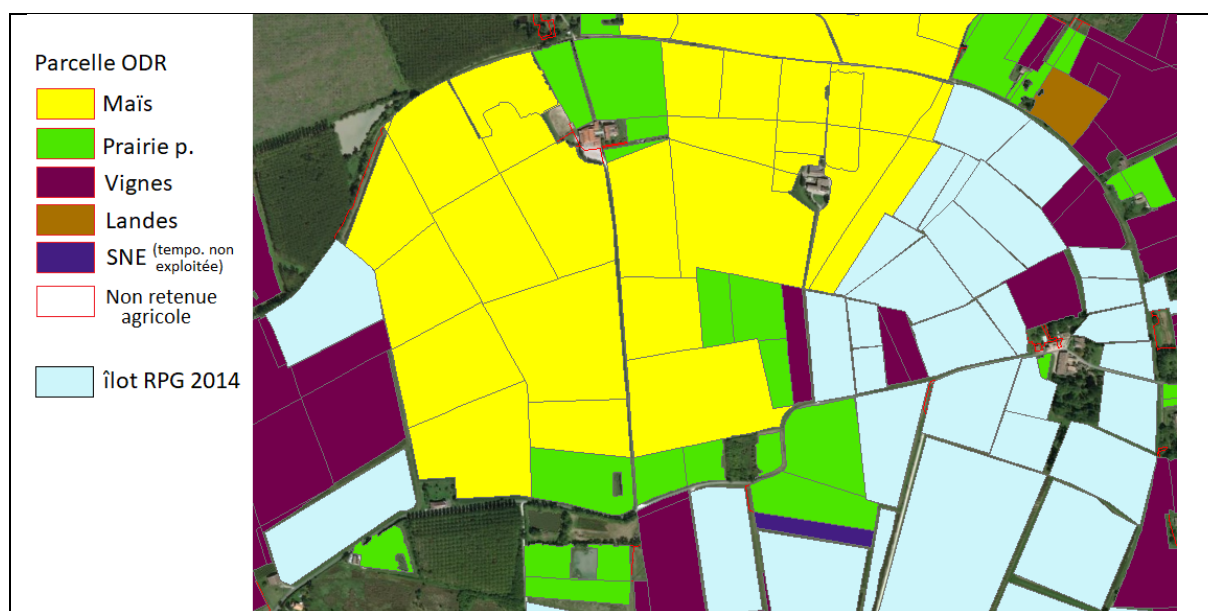


Figure 5 – Exemple de parcelle du RPG complété en grande culture (Maïs en l'occurrence)

5.3 Légumes

La difficulté vient du classement théorique des légumes selon OSO dans les classes *culture d'été* ou *culture d'hiver*. Cependant, nous notons un nombre significatif de parcelles visiblement dédiées au maraîchage qui sont classées dans une des catégories OSO « urbaine ». Ceci pour les cultures sous serre mais également pour les cultures de plein champ. Ce fait est d'autant plus problématique que des zones de maraîchage absentes du RPG sont fréquemment situées près de zones urbaines (particulièrement en Bretagne, en Provence) ou encore « englobée » dans des métropoles. Par conséquent, même en augmentant les observations, le risque de surestimer les surfaces en maraîchage est bien présent.

Notons que de nombreuses parcelles du RPG complété 2014 apparaissent dans le RPG 2015, déclarée en *légumes* ; et le modèle Logit parvient alors à prédire correctement cette classe de culture.



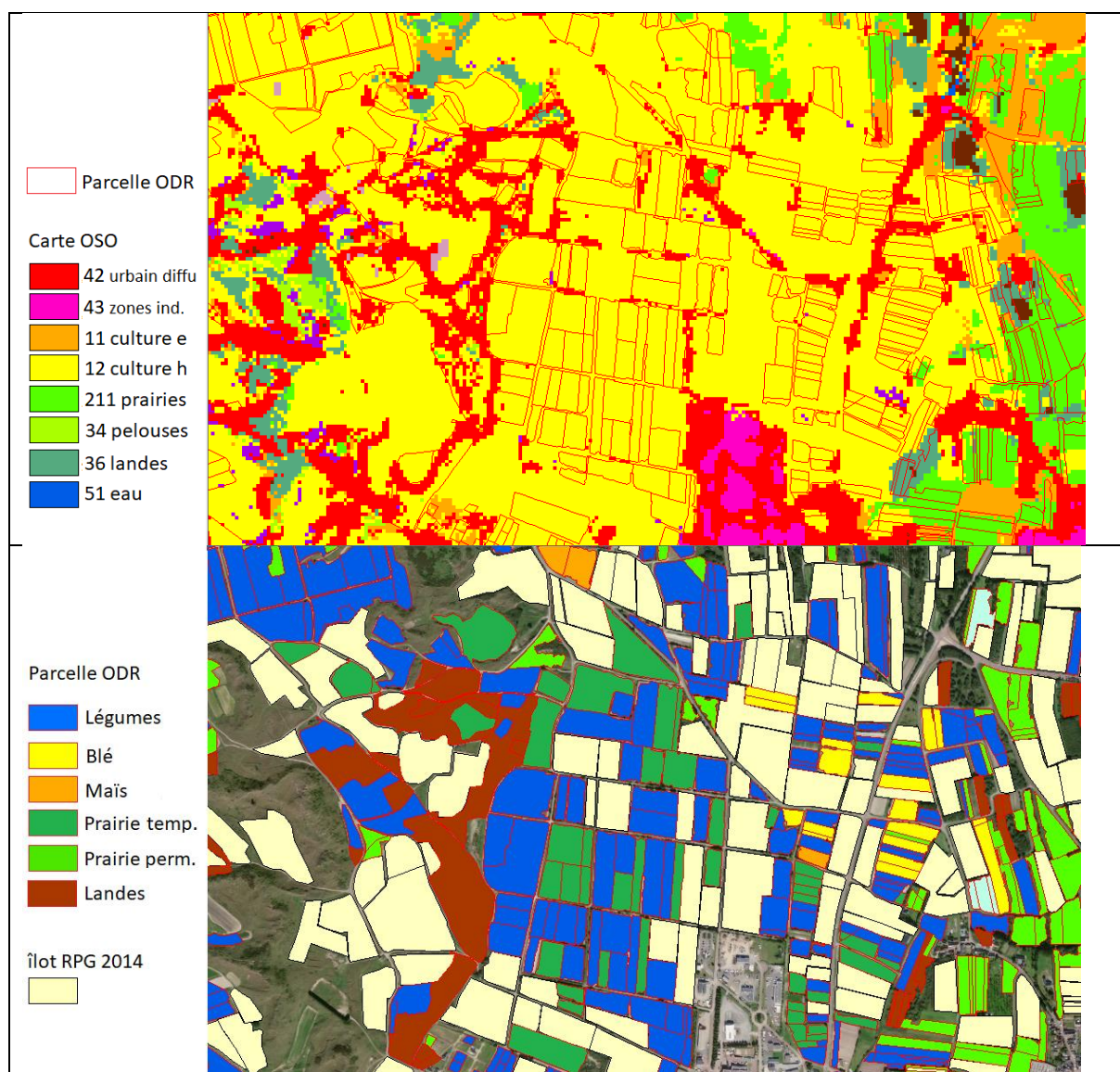


Figure 6 – Exemple de parcelles dédiées au maraîchage : parcelles ODR (en haut), carte OSO 2016 (milieu) et RPG complété 2014 (en bas)

Une autre difficulté réside dans la différenciation des légumes et des fleurs (tous deux classés dans la même catégorie de la nomenclature en 28 postes du RPG), voire de certaines cultures industrielles (par exemple les plantes aromatiques sur la côte d'azur).

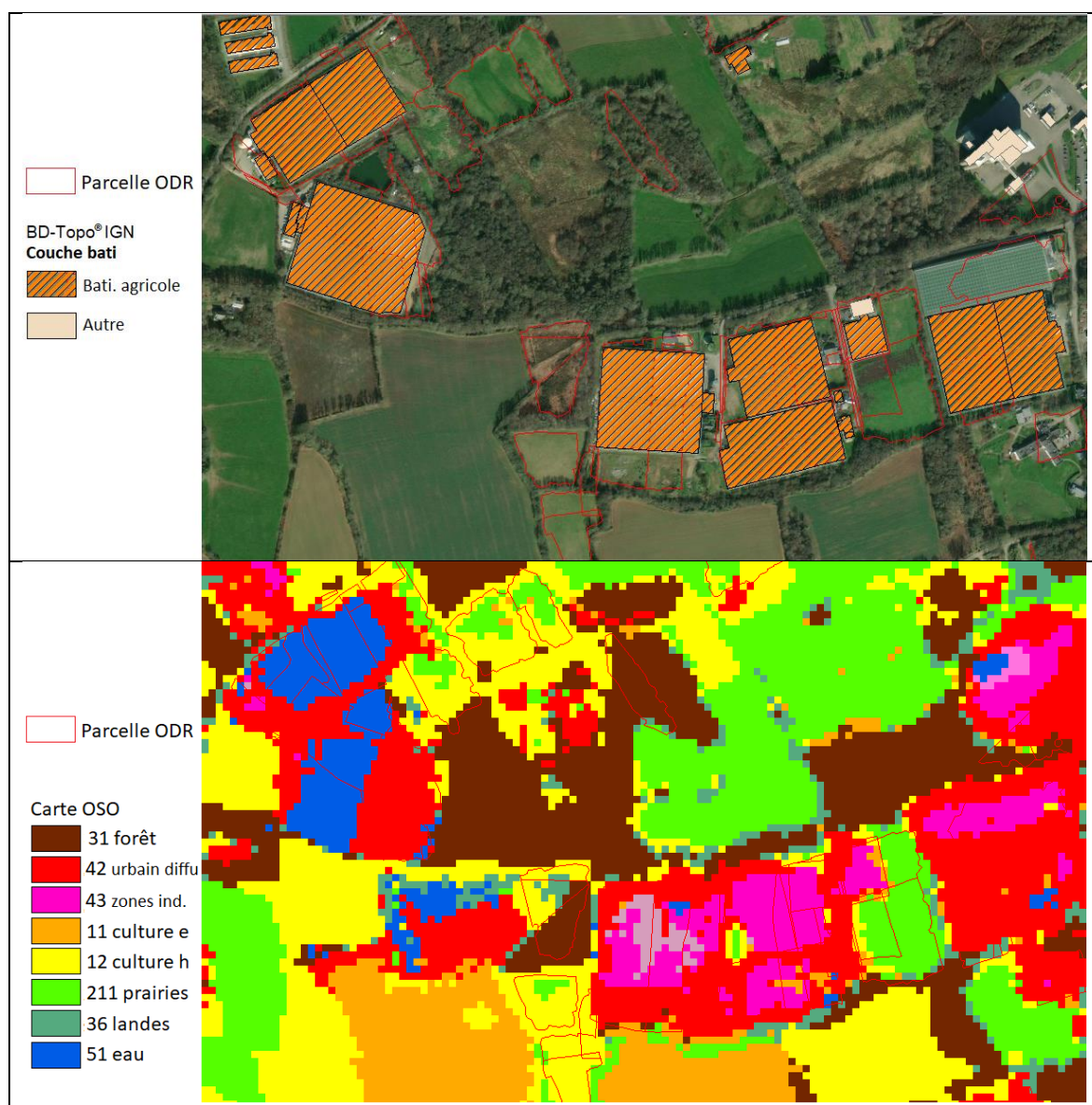


Figure 7 – Exemple de parcelles accueillant des serres/bâtiments agricoles pour la culture de légumes

Une autre source d'information très utile est fournie par la couche Bati de la Bd-Topo®, « bati industriel » de nature agricole ; *bâtiment agricole* ou *Serres* (Par exemple en Bretagne). Lorsque de tels bâtiments sont présents sur une parcelle odr, celle-ci est qualifiée « Légumes » (ceci en amont du modèle Logit, ces parcelles ne rentrent pas dans le processus de modélisation). Cette information de la Bd-Topo® est d'autant plus importante que la carte OSO voit ces surfaces comme des zones urbanisées (diffus, zones industrielle). En réalité il peut s'agir également de fleurs, d'horticulture. Le code « groupe de culture » est néanmoins le même que pour les légumes (gc25). Dans le nom de la culture, la mention « bâtiment » est ajoutée (« Légumes (bâtiment) », par défaut, éventuellement « Fleurs (bâtiment) » si vérifié). Il faut noter qu'une autre confusion est possible : ces bâtiments peuvent aussi accueillir des animaux (Porcherie, etc.). Une façon d'améliorer notre carte et de lever ces incertitudes serait de se procurer de l'information via les adresses de tous ces types de établissements agricoles, mais cela reste difficile à mettre en place à ce jour.

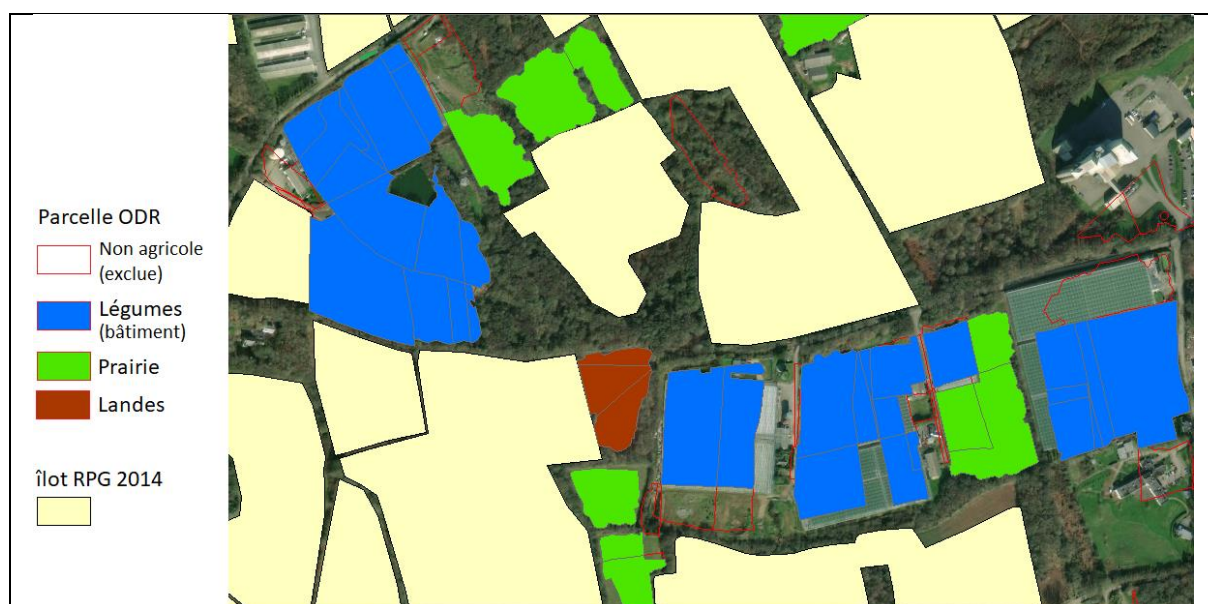


Figure 8 – RPG et RPG complété 2014 sur l'exemple de la figure précédente.

→ Perspectives

Ajouter aux variables explicatives une variable de contexte « local agricole » à une échelle plus fine que la PRA devrait être particulièrement intéressant pour les légumes.

De plus, la nouvelle nomenclature OSO en 23 classes inclue une catégorie « racine/tubercule » qui nous aidera significativement pour la localisation de surfaces dédiées aux légumes, au maraîchage.

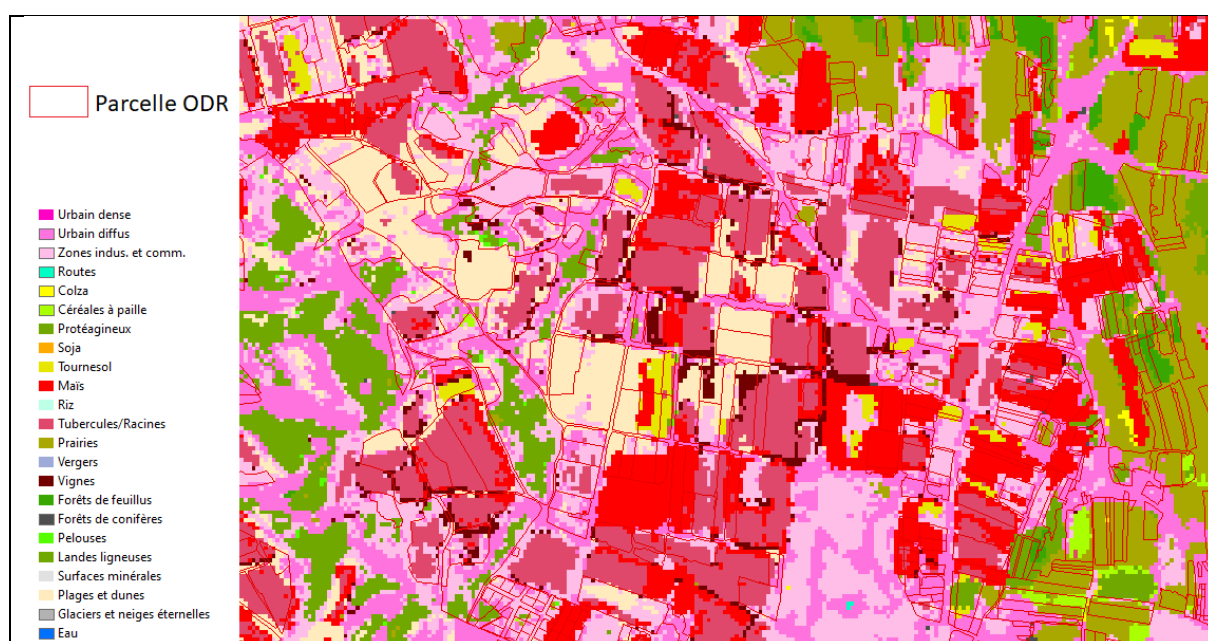


Figure 9 – Carte OSO 2018 sur l'exemple de la figure précédente

5.4 Vignes

Nous rappelons qu'à ce jour, les surfaces en vignes sont issues de la couche Végétation de la Bd Topo®.

Cette méthode pourra être modifiée lorsque nous construirons le RPG complété des campagnes plus tardives (post 2016), sous réserve de changements dans la couche OSO sur cette culture. A ce jour, le *RPG complété* pour cette culture est donc basé sur une hypothèse de stabilité des vignes et représente plus une photographie de la période de plus ou moins 2,3 ans autour de 2014. (La couche Végétation de la Bd Topo® n'est pas mise à jour uniformément sur l'ensemble des départements).

Autres classes reportées du RPG 2015 (sur parcelles ODR 2014 présentes dans le RPG 2015) :

- restructuration du vignoble (RVI)
- raisins de cuve (VRC)
- raisins de table (VRT)

Pour cette culture, les résultats de surfaces qui complètent le RPG correspondent plutôt bien à la différence entre la SAA et le RPG (Tableau 7), sauf quelques exceptions (Var, Marne). Ce tableau montre l'importance des surfaces non présentes dans le RPG pour cette culture : dans la Marne (51), le RPG complété contient plus de 85% de la surface totale départementale en vigne (elle-même peut être sous-estimée, voir phrase précédente).

Rappelons que la donnée SAA concerne les surfaces des exploitations qui ont leur siège social dans le département. Les résultats de la SAA sont le fruit de consolidations et d'arbitrages, les surfaces en vigne sont obtenues à partir du RPG, du casier viticole informatisé des Douanes et d'informations rassemblées auprès des correspondants agricoles locaux, des organismes professionnels, des agriculteurs, etc. Attention donc au biais de comparaisons !

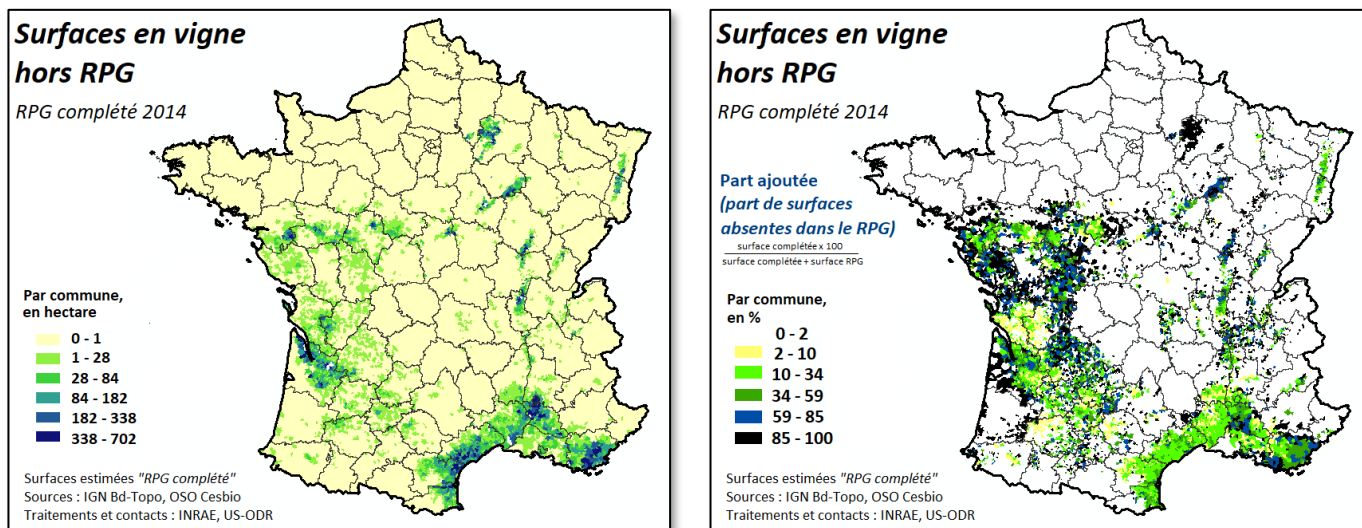


Figure 10 – Surfaces en vignes dans le RPG complété 2014. En surfaces brutes (gauche) et en part du total RPG +RPG complété (droite), équivalent de la part ajoutée.

Département	RPG (en ha)	RPG complété (en ha)	% RPG complété vs [total]	[Total] = RPG + RPG complété	SAA (en ha)	Diff [Total] - SAA (en ha)	Diff [Total] - SAA en %
33	74 592	46 412	38,35	121 004	120 109	-895	-0,74
34	61 846	21 210	25,53	83 056	85 799	2 743	3,19
11	56 703	12 210	17,71	68 913	67 893	-1 020	-1,5
30	45 936	11 471	19,98	57 407	55 470	-1 937	-3,49
84	32 436	19 396	37,42	51 832	51 260	-572	-1,11
16	37 840	3 714	8,93	41 554	41 191	-363	-0,88
17	37 688	3 470	8,43	41 158	40 491	-667	-1,64
83	18 142	14 320	44,11	32 462	28 843	-3 619	-12,54
51	3 051	18 209	85,64	21 260	24 453	3 193	13,05
66	15 633	8 386	34,91	24 019	23 891	-128	-0,53
32	18 936	702	3,57	19 638	20 005	367	1,83
49	15 690	4 744	23,21	20 434	19 676	-758	-3,85
69	9 118	8 102	47,05	17 220	17 378	158	0,91
26	13 487	4 031	23,01	17 518	16 873	-645	-3,82
71	6 295	5 359	45,98	11 654	13 250	1 596	12,04
24	10 067	2 637	20,75	12 704	12 307	-397	-3,26
44	8 331	4034	32,62	12 365	11 861	-504	-4,25
07	8 721	1735	16,59	10456	11108	652	5,87
13	8 598	2006	18,91	10604	10455	-149	-1,42
37	5 633	3852	40,61	9485	9 963	478	4,80
68	5 265	3645	40,90	8910	9 197	287	3,12
89	2 945	3781	56,21	6726	7 660	934	12,19
10	1 172	6160	84,01	7332	7 233	-99	-1,37
81	6 553	434	6,21	6987	6 943	-44	-0,63
67	4 372	2322	34,69	6694	6 911	217	3,14
41	5687	1280	18,37	6967	6640	-327	-4,92
47	5836	840	12,58	6676	6455	-221	-3,42
28	4014	1030	20,42	5044	5780	736	12,73
46	4364	799	15,48	5163	5262	99	1,88
18	1474	2135	59,16	3609	4290	681	15,87
82	2455	544	18,14	2999	2900	-99	-3,41
64	2337	213	8,35	2550	2735	185	6,76
02	251	1238	83,14	1489	2518	1029	40,87
39	856	893	51,06	1749	2262	513	22,68
40	1589	200	11,18	1789	2041	252	12,35
73	956	748	43,90	1704	1940	236	12,16
31	1478	243	14,12	1721	1612	-109	-6,76
58	412	1055	71,92	1467	1570	103	6,56
86	1163	365	23,89	1528	1476	-52	-3,52
85	644	656	50,46	1300	1339	39	2,91
42	420	378	47,37	798	1235	437	35,38
79	714	200	21,88	914	951	37	3,89
2A	833	80	8,76	913	946	33	3,49
63	300	192	39,02	492	819	327	39,93
01	339	386	53,24	725	789	64	8,11
04	705	68	8,80	773	714	-59	-8,26

Tableau 7 – Surface en vigne dans le RPG complété 2014, somme avec les surfaces déjà présentes dans le RPG 2014 et comparaison avec les données SAA 2014

→ Perspectives

A l'avenir, une meilleure correspondance entre les campagnes RPG à compléter et la référence temporelle des couches IGN *Végétation* utilisées doit permettre une estimation encore plus fine des surfaces en vigne absentes du RPG, notamment en repérant des changements (abandon, « jachère », restructuration, etc.). A évaluer également la « qualité » des cartes OSO à venir (voir § sur les vergers ci-dessous).

5.4 Vergers

La méthode d'estimation des surfaces des vergers est dépendante des observations suivantes :

- Les vergers dans la couche Végétation de la Bd Topo® semblent globalement surestimés : nous notons en effet des confusions avec des bois (plantation d'arbres alignés), des prés comprenant quelques arbres fruitiers (voir Figure 13), etc.. De plus, les vergers justement rencontrés dans cette couche de l'IGN peuvent correspondre non seulement à des vergers (au sens arbres fruitiers) mais aussi des oliviers, des truffières (ces 2 cas dans le sud-est) ou encore des pépinières.
- Il est important de noter que la couche OSO utilisée (2016) sous-estime largement ce type de culture (voir Figure 11). Cependant, dans les versions ultérieures de cette même carte OSO, la représentation des vergers semblent bien meilleure (Figure 12).

Ainsi un traitement spécifique est appliqué aux parcelles ODR qui sont couvertes par la classe « Verger » de la couche Végétation de la Bd Topo®.

Végétation Bd Topo®	OSO	→ Résultat
Verger	221 Vergers	Verger
	Autre classes	Modèle Logit spécifique

► Modèle Logit spécifique aux parcelles ODR qualifiées de Verger par la Bd-Topo®.

Ce modèle est construit et calibré sur les parcelles ODR présentes dans le RPG 2015 et sur des observations. A nouveau, plus le nombre d'observations est élevé, plus le modèle est robuste et performant (sauf si ces observations constituent un échantillon non représentatif...).

Catégories du RPG rencontrées	20	VER	Vergers
	23	OLI	Oliviers
	22		Fruits à coques
	28	TRU	Truffières (chênaie de plants mycorhizés)
	28	SBO	Surface boisée sur une ancienne terre agricole
	18		Prairies
	17		Landes
	17	BOP	Bois pâturés
			Non agricole

Autres classes reportées du RPG 2015 (sur parcelles ODR 2014 présentes dans le RPG 2015) : *Petit fruit rouge (PFR)*. Cette culture n'est pas incorporée dans le modèle (car trop marginales, engendrerait une surestimation de la culture en question).

Nous présentons ci-dessous un tableau (Tableau 8) récapitulatif des surfaces additionnées du RPG et du RPG complété, comparées avec les surfaces de l'inventaire Agreste¹⁰ des vergers 2013 (Agreste, 2013). Ce tableau illustre également les difficultés de trouver des références pour évaluer nos résultats. Parfois les surfaces RPG seules sont supérieures à la référence Agreste ; à cause de différences temporelles ? de différence dans les définitions des classes des nomenclatures ? de la présence de « vergers familiaux » (hors exploitation) ?



¹⁰ Sur 8 espèces : pomme, poire, pêche, prune, abricot, cerise, kiwi et agrumes. Concerne uniquement les exploitations agricoles professionnelles. De plus, seules les surfaces des *exploitations ayant plus de 0.5 ha* sont comptabilisées pour les poires et cerises et les surfaces des *exploitations de plus de 1 ha* pour les autres espèces.

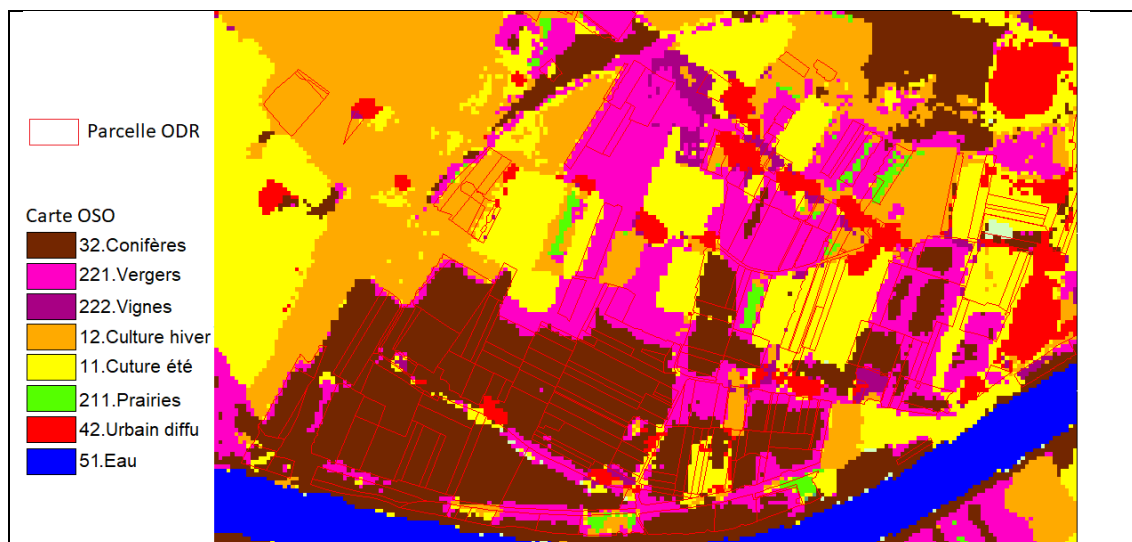


Figure 11 – Illustration, parcelles odr sur des vergers (Lot-et-Garonne). Couche IGN bd Topo Végétation (milieu) et carte OSO 2016 (bas)

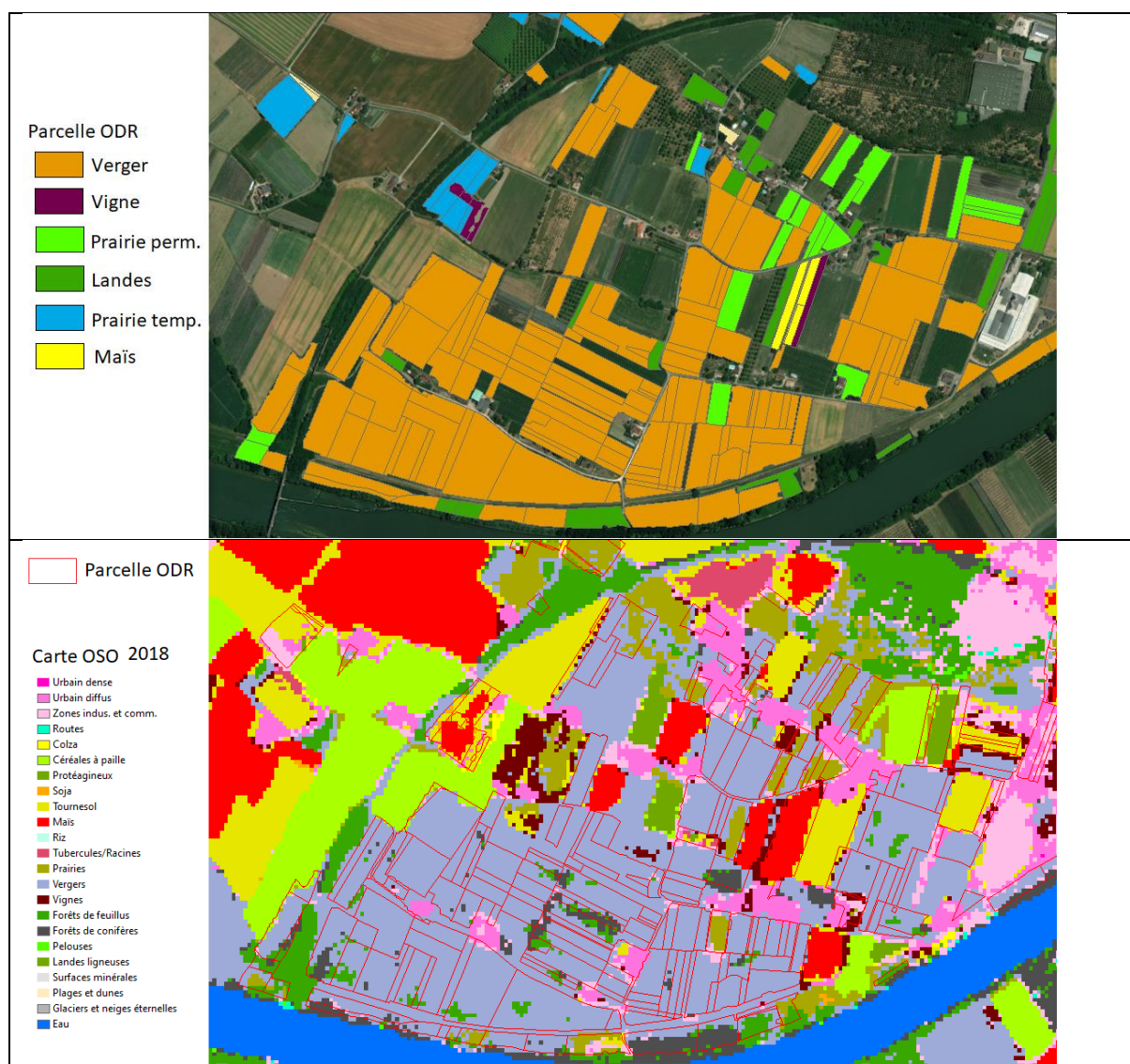


Figure 12 – (suite) RPG complété 2014 (haut) et Carte OSO 2018 (bas) sur le même site.



Figure 13 - Illustration, parcelles odr sur des zones qualifiées de verger par la Bd-Topo® de l'IGN (région Normandie)

→ Perspectives

La Figure 12 suggère une amélioration significative de la catégorie « Verger » dans les cartes OSO plus récentes, cela laisse augurer une utilisation plus massive et plus simple de cette information OSO pour détecter de telles cultures, ainsi que la possibilité de les suivre dans le temps (parcelles non exploitées temporairement, arrachage, etc.). C'est un point très important pour la production du RPG complété, tant pour sa qualité que pour son intérêt d'utilisation (capacité à détecter les évolutions, les changements).

Région	Surface RPG (A)	Surf. RPG complété (B)	Surface Totale (A)+(B)	% RPG complété vs total	Agreste 2013	SAA 2014 (avec Sec. Stat)
Paca	12 894	12 153	25 050	48.51	18 334	18 818
Rhône-Alpes	14 107	4 438	18 517	23.97	16 240	
Aquitaine	15 404	1 907	17 285	11.03	17 319	
Midi-Pyrénées	10 351	3 680	13 988	26.31	10 814	
Languedoc-Roussillon	7 858	1 079	8 932	12.08	12 787	8 571
Limousin	1 736	262	1 990	13.16	2 262	
Pays de la Loire	6 424	2 203	8 599	25.61	5 834	
Haute Normandie	1 034	617	1 644	37.53	429	2 182
Basse Normandie	4 513	2 423	6 929	34.97	416	5 275
...	
<i>France (hors DOM)</i>	89 535	30 461	119 606	25.46	96 465	

Tableau 8 – Surfaces occupées par les vergers, par région, RPG et RPG complété (surfaces en hectare).

5.5 Oliviers

Les oliviers sont vus comme un cas particulier de verger et sont estimés de la même façon que ces derniers (voir § ci-dessus). Nous présentons ici quelques éléments de comparaison.

Département	Surface RPG	Surf. RPG complété	Surface Totale	(total régional)	% RPG complété vs total	Agreste (avec Sec. Stat)
04	915.20	414.62	1 329.82	(Paca) 10 659.66	31.18	1 319
06	465.90	1 019.89	1 485.79		68.64	982
13	3 346.60	548.01	3 895.61		14.07	4 100
83	1 339.25	1 256.59	2 595.84		48.41	2 560
84	784.37	568.23	1 352.60		42.01	1 100
11	567.00	32.67	595.67	(Occitanie) 3 705.61	5.48	565
30	1 090.24	162.55	1 252.79		12.98	1 355
34	1 153.84	239.63	1 393.47		17.20	1 480
66	457.86	5.82	463.68		1.26	440
07	169.78	110.86	280.64	(Rh-Alpes) 1 312.42	39.50	250
26	276.84	754.84	1 031.78		73.16	950
2A	568.89	16.04	584.93	(Corse) 1 769.57	2.74	850
2B	910.33	274.51	1 184.64		23.17	1 260
Total	12 046.10	5 404.26	17 447.26		30.97	17 211

Tableau 9 – Surfaces occupées par la culture d'oliviers, RPG et RPG complété (surfaces en hectare).

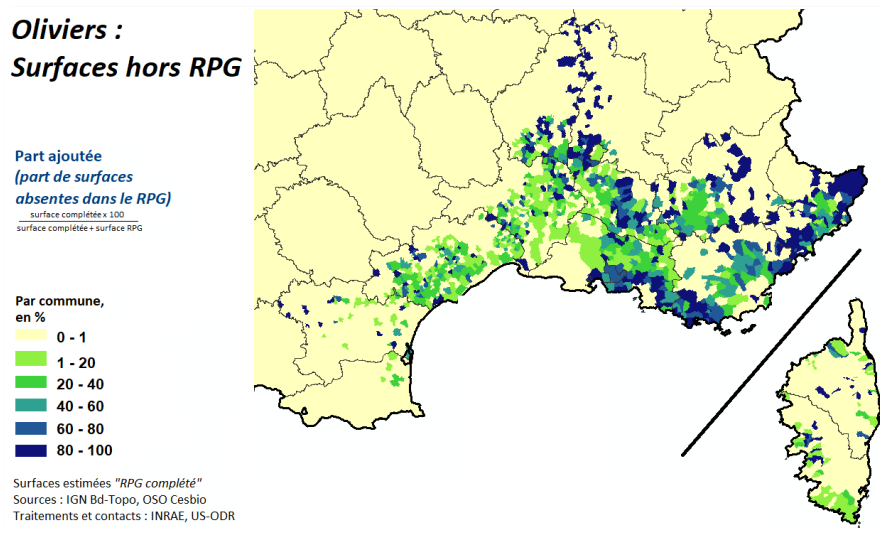


Figure 14 –part des surfaces ajoutées au RPG (2014), niveau communal.

5.6 Landes et estives

Les landes peuvent parfois être confondues avec les prairies (et vice-versa)

Comme dans le RPG (gc 17 Estives/landes), nous distinguons 2 types de « landes » : d'un côté des surfaces agricoles herbeuses mais non pâturées ; non exploitées, non fauchées, incluant notamment des broussailles, voire ronces, etc. De l'autre côté, des territoires d'estives (sur certains départements uniquement), des surfaces pastorales.

Dans les départements de montagne, notre RPG complété contient de larges zones de type « estives » (*Surfaces pastorales à dominantes herbagères* (SPH) ou *ligneuses* (SPL) ; catégories apparaissant grâce aux parcelles odr présentes dans le RPG 2015. Dans les faits, il y a sans doute des raisons pour que certaines de ces aires ne soient pas présentes dans le RPG (en altitude : trop de rochers, etc.), nous les surestimons sans doute (cela serait plutôt des alpages). En tout cas, nous ne pouvons pas en déduire, par exemple, qu'elles pourraient constituer d'éventuelles estives supplémentaires.

6. Perspectives méthodologiques pour les futurs RPG complété

Suivant le changement de nomenclature de la carte OSO à partir de 2018 (voir §2), il faut distinguer 2 périodes, selon la nomenclature proposée. De plus, la première version d'OSO cartographie l'occupation du sol de l'année 2016. Les RPG complété 2015 et 2016 seront établies sur la même carte OSO que celle utilisée pour le RPG complété 2014.

Pour les campagnes 2014 à 2016, seule l'information fournie sur les surfaces départementales fera évoluer les cultures du RPG complété et ceci uniquement pour les cultures que nous avons soumis à une éventuelle rotation culturale : grandes cultures, légumes, prairies temporaire, ainsi que les jachères. Par conséquent, ces RPG complété 2014 à 2016 seront identiques pour les cultures pérennes, prairies permanentes et landes/estives, sauf dans les cas suivant :

Entre la campagne n et la campagne n+1 : soit (1) la parcelle apparaît dans le RPG de la campagne n+1 (alors soit la culture prédite l'année n est inchangée, soit la culture change ; nous passons d'une culture prédite à une culture déclarée). Soit (2) la parcelle n'est plus présente parmi les parcelles ODR campagne n+1 (par exemple la parcelle a été utilisée pour accueillir une construction).

Campagne RPG complété	Carte OSO utilisée		SAA
	Année	Nomenclature	
2014	2016	17 classes	2014
2015	2016		2015
2016	2016		2016
2017	2017		2017
2018	2018	23 classes	2018
20yy	20yy		20yy

Tableau 10 – temporalité des données pour les sources OSO et SAA des RPG complété 2014 à 2018

A partir de la campagne 2018, la carte OSO comportera plus de classes d'occupation du sol agricoles - ce qui est évidemment un très bon point pour le RPG complété - et la méthodologie devra être modifiée ; le modèle sera vraisemblablement utilisé uniquement pour les cas d'incertitudes entre les nomenclatures (exemple : OSO « céréales à paille » ↔ RPG complété « Blé » ou « orge » ou « autre céréale ») et l'information OSO sera gardée lorsque la catégorie de culture correspond à une classe que nous cherchons (Maïs, Tournesol, etc.). Il restera alors à comparer nos estimations avec les statistiques de la SAA.

7. Format des fichiers du RPG complété

Le RPG complété est construit et distribué par département. Il consiste, pour chaque département donc, en un fichier de forme, format Shape-file® ESRI. Un polygone correspond classiquement à une parcelle, et est associé à une ligne du fichier de données attributaires. Pour rappel, à chaque parcelle correspond une culture et une seule.

Description des colonnes du fichier attributaire :

Champ	Description	Unité	Remarque
ID_PARCELL	Identifiant de la parcelle	-	
SURF_HA	Surface	hectare	Superficie de la parcelle = surface de la culture
CULTURE	Nom de la culture	-	Parfois nom dans la nomenclature du RPG 2015
GC_RPG	Code du groupe de culture	-	Dans la nomenclature en 28 postes du RPG
COM_INSEE	Commune INSEE de parcelle	-	
CODE_PRA	Code de la PRA	-	

Tableau 11 – Champ des tables attributaires du RPG complété

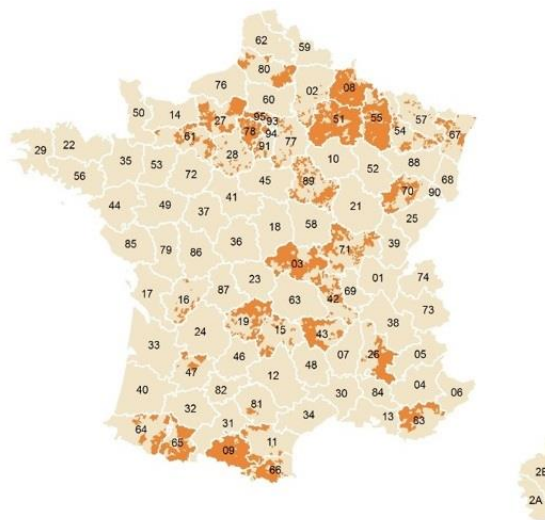
La projection des fichiers géographiques est le RGF93/Lambert 93 (SCR EPSG : 2154) et sa résolution est celle des données IGN utilisée en entrée (Bd Parcellaire, Bd Topo®).

Remarque :

Attention, les parcelles ODR pour 2014 sont donc créées à partir de la Bd-Parcellaire disponible en 2015. Tous les départements n'étaient pas alors disponibles (les communes en orange sur la carte ci-contre n'étaient pas disponibles).

Par conséquent, le RPG complété 2014 sera absent sur ces communes, excepté pour la vigne qui a été rajoutée hors parcelle cadastrale (BD Topo® IGN, couche végétation, voir §5.4 ci-dessous). Les surfaces en vigne sont donc supposées être *exhaustives* sur le RPG complété 2014

L'utilisation de la Bd Parcellaire express (§1.2 ci-dessus) garantit à présent une exhaustivité sur tout le territoire français métropolitain.



8. Conclusions

Etablir cette couche d'occupation du sol RPG complété 2014 a permis de mettre à disposition cette information parcellaire exhaustive (associée au RPG) de l'occupation du sol agricole sur la dernière campagne du RPG « ancienne version » (2007-2014).

L'exercice a validé la pertinence d'utiliser de la carte OSO. Il a montré également l'intérêt d'une première intégration de cette source d'information à notre méthodologie, malgré le décalage temporel entre le RPG complété 2014 et la carte OSO 2016, et une nomenclature OSO qui va devenir obsolète. Il était important de manipuler largement cette nouvelle information, pour bien la comprendre ; interprétation des classes, atouts/limites de certaines d'entre elles (priorité à donner par rapport à des informations issues de l'IGN, par exemple pour les vignes). La façon de l'utiliser au mieux (changement d'échelle pixels vers parcelle unique) pour notre projet a pu également être étudiée.

Ce double objectif produire un RPG complété 2014 et étudier l'intégration de la carte OSO dans notre méthodologie a pu être atteints ; le RPG complété 2014 produit affichant globalement une certaine qualité et des éléments d'améliorations pour l'utilisation d'OSO en vue des futures campagnes du RPG complété ont pu être établies.

Ainsi, des perspectives intéressantes apparaissent. La prochaine étape consistera à réviser la méthodologie pour la production du RPG complété campagne 2018, c'est-à-dire une méthodologie adaptée à la nouvelle nomenclature OSO. Associé à une plus grande automatisation des processus, cela assurera la possibilité de compléter les futurs RPG dès leurs mises à disposition.

Références :

Agreste, 2013. Inventaire des vergers 2013. Agreste Chiffres et Données Agriculture - n° 227 - février 2015. Publication du Ministère de l'Agriculture, Service de la statistique et de la prospective.

Agreste, 2014. La Statistique Agricole Annuelle : présentation générale. Méthodologie, définitions et concept. Note méthodologique, Ministère de l'Agriculture et de l'Alimentation.
http://agreste.agriculture.sg-ppd.maaf.ate.info/IMG/pdf_methosaa.pdf

Cantelaube Pierre, Carles Marie, Séard Claire et Truche Camille, 2016. Méthodologie d'élaboration de la carte d'occupation du sol « RPG complété » version 1 (INRA ODR 2013-2017). Note méthodologique ODR, septembre 2016.

CCE Cour des Comptes Européenne, Rapport spécial, 2016, *Le système d'identification des parcelles agricoles: un instrument utile pour déterminer l'admissibilité des terres agricoles malgré une gestion encore perfectible*. Luxembourg: Office des publications de l'Union européenne, ISBN 978-92-872-5942-4, ISSN 1831-0850, doi:10.2865/416661.

CES "Occupation des SOIs" – OSO. <http://osr-cesbio.ups-tlse.fr/~oso/>

CLC, 2009, CORINE Land Cover France : Guide d'utilisation. Document technique du Commissariat général au développement durable Service de l'observation et des statistiques. Ministère de l'environnement, European Environment Agency. Février 2009

de Rincquesen Hubert, Julien Vincent , Lécuyer Charlotte, Levesque Robert, Liorit Dimitri, Melin Maxence, Pathier Guillaume et Quelin Claire, 2016, Structure foncière des exploitations agricoles. Appariement entre le registre parcellaire graphique et le cadastre pour construire une typologie des exploitations françaises d'après leur structure foncière. *Notes et Études Socio-Économiques (NESE) du Centre d'Etudes et de Prospective (CEP)*, n° 41, Décembre 2016, pp. 51-82.

IGN, 2018. RPG Version 1.0 – Descriptif de contenu et de livraison – Juillet 2018

IGN, 2019a. BD PARCELLAIRE® Version 1.2 – Descriptif de contenu – Juillet 2019

IGN, 2019b. Comparatif Parcellaire Express (PCI) / BD PARCELLAIRE® – Avril 2019

IGN, 2019c. BD CARTO® Version 3.3 – Descriptif de contenu – Juin 2019

IGN, 2020a. BD TOPO® Version 3.0 - Descriptif de contenu - Mai 2020

IGN, 2020b. RPG Version 2.0 – Descriptif de contenu et de livraison – Août 2020

INSEE, 2008, Code Officiel Géographique, www.insee.fr

INSEE, 2009, Les petites régions agricoles, www.insee.fr

ODR-Ecophyto, 2015. « *Mieux suivre spatialement l'usage des pesticides, en particulier sur les bassins versants, en s'appuyant sur un observatoire des ventes détaillé au code postal de l'utilisateur final de*

produit » Rapport réalisé dans le cadre de l'axe 2, action 21 du programme Ecophyto, pour le compte de l'ONEMA et du MEDDE. Septembre 2015. INRA (US ODR, INRA Sadapt)

Séard Claire, 2016. Mise à jour de la cartographie d'occupation du sol, Note Méthodologique ODR, INRAE, septembre 2016.

Thines Georges et Lempereur Agnès, 1975, Dictionnaire général des sciences humaines, Paris, Éditions universitaires, 1034 pages.

Annexe 1 : phase 1 Préparation des *Parcelles ODR*

Cette partie présente les différentes étapes du programme par ordre chronologique.

Note : Les fonctions utilisées sont indiquées en vert et les couches *Shapefile* nécessaires en orange. Les noms des attributs utilisés sont indiqués entre guillemets doubles (ex. « NATURE ») et les valeurs d'attribut entre guillemets simples (ex. 'Agricole').

A.1.1 Préparation des données

Les principales données nécessaires sont préparées ici. Il s'agit de sélectionner les données relatives au département à traiter et de mettre en forme leurs tables attributaires. Sept fonctions sont mobilisées pour ce travail.

➤ Préparation du **département traité** :

`def geomDep(numDep) - BD TOPO – DEPARTEMENT`

Sélection du département à traiter.

➤ Préparation des **communes** :

`def geomCom(numDep) - BD TOPO – COMMUNE`

Sélection des communes du département. Les communes sont utilisées dans la deuxième partie du script pour produire les 8 couches d'occupation du sol qui sont à l'échelle communale.

➤ Préparation de **CLC** :

`def clc_decoup(numDep) - CLC - CLC 2018`

Découpage de CLC sur l'emprise du département. Le résultat de cette fonction permettra d'ajouter aux *Parcelles ODR* le code CLC d'occupation du sol par intersection.

➤ Préparation des **parcelles cadastrales** :

`def parcelles (numDep) - BD PARCELLAIRE EXPRESS - PARCELLE`

- Sélection des parcelles cadastrales du département.
- Calcul du champ « *surfinitm2* » (surface initiale de la parcelle avant les découpages).
- Création du champ « *CODE_INSEE* » : concaténation des colonnes « *CODE_DEP* » et « *CODE_COM* ».
- Création du champ « *num_par* » (identification de la parcelle) : concaténation des colonnes « *CODE_DEP* », « *CODE_COM* », « *FEUILLE* », « *SECTION* », et « *NUMERO* ».

➤ Préparation du **RPG** :

`def rpg (numDep,annee) - RPG`

Sélection des parcelles *RPG* pour le département concerné et l'année étudiée.

➤ Préparation des **serres agricoles** :

`def select_serres (numDep) - BD TOPO - BATIMENT`

Sélection sur les champs « NATURE » = 'Serre' et « USAGE1 » = 'Agricole'.
Les serres seront intersectées avec les *Parcelles ODR* pour en conserver les contours et l'information.

➤ Préparation des vignes et vergers :

def select_vignes (numDep) - BD TOPO - *ZONE_DE_VEGETATION*

Sélection sur le champ « NATURE » = 'vignes', 'vergers', 'bananeraie', 'canne à sucre', 'houblonnière' et 'rizière'.

Les Vignes et Vergers seront intersectées avec les *Parcelles ODR* pour en conserver les contours et l'information.

A.1.2- Construction du *Masque 1*

Quatre fonctions successives permettent de sélectionner des territoires (à partir des *BD TOPO* et *CARTO*) dont on sait qu'ils ne sont pas à vocation agricole. Les objets sélectionnés par ces fonctions sont alors regroupés pour former le *Masque 1*. Celui-ci sera ensuite retranché du parcellaire cadastral par différence géométrique.

➤ Voies ferrées :

def voies_ferrees (numDep) - BD TOPO - *TRONCON_DE_VOIE_FERREE*

La géométrie linéaire de cette donnée impose de la transformer en polygones. Nous estimons qu'une voie de chemin de fer fait 5 mètres de large. On applique donc un buffer de (2,5 mètres x « NB_VOIES ») aux tronçons sauf ceux 'En projet'. (L'attribut « NB_VOIES » est obligatoire dans la *BD TOPO* et est donc toujours renseigné).

➤ Les routes :

def routes (numDep) - BD TOPO - *TRONCON_DE_ROUTE*

De même, la géométrie linéaire de cette donnée impose de la transformer en polygones.

(Sélection des tronçons 'Autoroute', 'Départementale', 'Nationale', 'Route européenne', 'Route intercommunale' (attribut « CL_ADMIN »). Cette étape est supprimée. Cela permet de prendre en tous les tronçons de route pour la construction du masque et donc d'affiner les résultats.

Sélection de tous les tronçons de route dont l'attribut « NATURE » est différent de 'Escaliers' et de 'Bac ou liaison maritime'.

Ces tronçons sont bufferisés selon un champ « demilarg » créé et rempli à partir de l'attribut « LARGEUR ». Lorsque le champ « LARGEUR » n'est pas renseigné, on attribue au champ « demilarg » les valeurs suivantes : 'Type autoroutier':10, 'Bretelle' :1.5, 'Route à 2 chaussées':1.5, 'Route à 1 chaussée':1.5, 'Route empierrée':1.5, 'Chemin':1, 'Rond-point':2.5, 'Piste cyclable':0.5, 'Sentier':0.5. Les largeurs utilisées ici sont tirées de la note méthodologique de Seard C. (septembre 2016) « *Mise à jour de la cartographie d'occupation du sol* ».

Note : Dans la *BD TOPO*, les routes à une chaussée sont matérialisées par une ligne et celles à deux chaussées par deux ; on leur attribue donc la même largeur (1.5 m).

➤ Une sélection de bâti de la *BD TOPO* :

def select_bati1 (numDep)

Cette fonction sélectionne les données dans différentes couches de la *BD TOPO* et les concatène.

- *BD TOPO* - *CIMETIERE* : Tous les cimetières.
- *BD TOPO* - *TERRAIN_DE_SPORT* : Tous les terrains de sport.
- *BD TOPO* - *AERODROME* : Tous les aéroports dont la géométrie est non fictive (attribut « FICTIF » = 'Non').
- *BD TOPO* - *ZONE_D_ACTIVITE_OU_D_INTERET* : Toutes celles dont l'attribut « NATURE » n'est pas 'Divers agricole' et dont la géométrie est non fictive (attribut « FICTIF » = 'Non').
- *BD TOPO* - *EQUIPEMENT_DE_TRANSPORT* : Tous ceux dont la géométrie est non fictive (parkings, aires de triage, aires de repos...).

➤ Une sélection de non bâti issue des *BD TOPO* et *CARTO* :

def select_non_bati1 (numDep)

- *BD TOPO* - *ZONE_DE_VEGETATION* : Sélection des 'Bois', 'Forêt fermée de conifère', 'Forêt fermée de feuillus', 'Forêt fermée mixte', 'Forêt ouverte', 'Haie', 'Lande herbacée' (présent seulement en Guyane), 'Lande ligneuse', 'Mangrove', 'Peupleraie', et 'Zone arborée' du champ « NATURE ».

Note : La couche *ZONE_DE_VEGETATION* de la *BD TOPO* est désormais « multi-thème » pour l'ensemble des départements métropolitain, c'est-à-dire que tous les types de couverts végétaux sont différenciés. Le « mono-thème (Attribut « NATURE » = 'Zone arborée') ne concerne plus aucun département. D'autre part, 'Lande Ligneuse' peut avoir une utilisation agricole mais nous considérons qu'il s'agit d'un biais mineur.

- *BD TOPO* - *SURFACE_HYDROGRAPHIQUE* : Tout.
- *BD CARTO* - *ZONE_OCCUPATION_SOL* : « NATURE » = 'Carrière – décharge', 'Glacier – névé', 'Marais', 'Rocher – éboulis', 'Sable – gravier', 'Mangrove', 'Broussailles'.

A.1.3- Différence Parcelles - *RPG* (*Parcelles 1*)

On retire des parcelles cadastrales celles incluses dans le *RPG* de l'année de référence par différence géométrique ; les parcelles résultantes sont nommées *Parcelles 1*.

def overlay (parcelles, rpg, how="difference")

Puis 3 nettoyages successifs :

- Nettoyage des géométries.
- Nettoyage selon la forme (*Miller* < 0.015).
- Nettoyage selon la taille (>50m²).

A.1.4- Différence *Parcelles 1* – *Masque 1* (*Parcelles 2*)

On retire ensuite des *Parcelles 1* les territoires compris dans le *Masque 1* par différence géométrique ; les parcelles résultantes sont nommées *Parcelles 2*.

def overlay (parcelles1, masque1, how="difference")

Puis 3 nettoyages successifs :

- Nettoyage des géométries.
- Nettoyage selon la forme (*Miller* < 0.015).
- Nettoyage selon la taille (>50m²).

Note : L'identifiant des parcelles construit lors de la préparation des données (« num_parc ») ne s'avère pas unique. On ajoute donc à cette étape un identifiant « vraiment » unique aux *Parcelles 2* (« id_parc » = « num_parc » + incrémentation). Il sera nécessaire pour la suite.

A.1.5- Construction du *Masque 2* (Parcelles bâties)

Le *Masque 2* est une sélection des parcelles cadastrales qui contiennent du bâti. Ces parcelles n'ont pas de vocation agricole et seront donc retirées.

Deux précautions s'imposent :

- 1) Ne pas écarter les parcelles dont les bâtiments sont des Serres Agricoles.

Pour cela, les bâtiments (*BD PARCELLAIRE EXPRESS* - *BATIMENT*) qui intersectent des bâtiments de « NATURE » 'serre' et à « USAGE » 'agricole' (*BD TOPO* - *BATIMENT*) sont préalablement sélectionnés et écartés (ainsi ils ne participeront pas à la construction du *Masque 2*).

- 2) Ne pas écarter les parcelles contenant une petite surface de bâtiments par rapport à la surface totale de la parcelle (cas du petit bâtiment au milieu d'un champ de vigne...).

Pour cela, on calcul la surface totale de bâti par parcelle : les bâtiments de la *BD PARCELLAIRE* (tous, sauf ceux qui sont des serres donc) sont intersectés avec les *Parcelles 2*. La surface de chaque bâtiment est alors calculée puis sommée par parcelle. Seules les parcelles dont le ratio (surface de bâti / surface de parcelle restante) > 0,025 sont retenues dans le *Masque 2*. Ce seuil est obtenu empiriquement après de nombreux tests.

Nettoyage :

- Nettoyage des géométries.

A.1.6- Différence *Parcelles 2* - *Masque 2* (*Parcelles 3*)

On retire des *Parcelles 2* les territoires compris dans le *Masque 2* par différence géométrique; les parcelles résultantes sont nommées *Parcelles 3*.

```
def overlay (parcelles2, masque2, how="difference")
```

Puis 3 nettoyages successifs :

- Nettoyage des géométries.
- Nettoyage selon la forme (*Miller* < 0.015).
- Nettoyage selon la taille (>50m²).

Note : A cette étape, on calcul les surfaces restantes de chaque parcelle (« surfm2 »). Cette surface sera nécessaire pour la fonction de nettoyage final des résidus ([A.1.12- Finalisation des Parcelles ODR](#)).

A.1.7- Intersection avec CLC (*Parcelles 4*)

Les *Parcelles 3* sont intersectées avec CLC. Cela permet de les découper selon l'occupation du sol CLC et de capitaliser le « code_18 ».

```
def overlay (parcelles3, clc_dpt, how="intersection")
```

Nettoyage :

- Nettoyage des géométries.

A.1.8- Traitement des Vignes, Vergers (*Parcelles 5*)

Les *Parcelles 4* sont intersectées avec les *Vignes - Vergers* (donnée préparée en 6.1) pour en récupérer les contours et l'information. Les parcelles résultantes sont nommées *Parcelles 5* et sont enrichies d'un champ « NATURE1 » avec pour valeurs possibles 'vigne', 'verger', 'bananeraie', 'canne à sucre', 'rizière', 'houblonnière' ou 'NULL'.

Note : Le champ « NATURE1 » a la valeur 'NULL' quand l'intersection n'a pas donné de résultat.

L'intersection se déroule en 3 étapes afin de conserver tous les enregistrements :

- Intersection (*Parcelles 4 - Vignes, Vergers*) = *Parcelles 5a*
- Différence (*Parcelles 4 - Vignes, Vergers*) = *Parcelles 5b*
- Concaténation (*Parcelles 5a + Parcelles 5b*) = *Parcelles 5*

A.1.9- Traitement des Serres Agricoles (*Parcelles 6*)

Comme pour les Vignes et Vergers, les *parcelles 5* sont intersectées avec les *Serres* (donnée préparée en 6.1) pour en récupérer les contours et l'information. Les parcelles résultantes sont nommées *Parcelles 6* et sont enrichies d'un champ « NATURE2 » avec pour valeurs possibles 'serre' ou 'NULL'.

L'intersection se déroule en 3 étapes afin de conserver tous les enregistrements :

- Intersection (*Parcelles 5 - Vignes, Vergers*) = *Parcelles 6a*
- Différence (*Parcelles 5 - Vignes, Vergers*) = *Parcelles 6b*
- Concaténation (*Parcelles 6a + Parcelles 6b*) = *Parcelles 6*

Nettoyage :

- Nettoyage des géométries.

A.1.10- Multi polygones vers polygones (*Parcelles 7*)

Avec les géo-traitements successifs, la géométrie de certaines parcelles est devenue « multi parties ». La fonction « explode » permet de séparer les différents « morceaux » composant une parcelle. Les parcelles résultantes sont nommées *Parcelles 7*.

```
def explode (parcelles6)
```

A.1.11 Jointure avec les Communes PRA

Une fonction découpe préalablement la donnée *Communes PRA* sur l'étendue du département traité:

```
def pra_decoup (numDep).
```

Certains codes *INSEE* des communes PRA étant obsolètes, nous passons par une table de passage de l'*INSEE* pour attribuer à chaque commune PRA (et chaque *Parcelles 7*) le code géographique de l'année 2020 (« CODGEO_2020 »).

Une jointure attributaire des *Communes PRA* sur les *Parcelles 7* via ce champ permet alors d'attribuer aux parcelles le « code PRA » correspondant.

A.1.12- Finalisation des *Parcelles ODR*

➤ Élimination des résidus 1 : def Nettoyage (fil, how='residu')

Les découpages successifs créent des résidus de parcelle dont la taille peut être importante mais qu'il est nécessaire d'éliminer du résultat. Ne travailler que sur la surface de la parcelle ne suffit pas à éliminer ces résidus. Ainsi, on travaille sur une combinaison entre la surface finale et le rapport entre surface après les découpages et surface initiale.

Cette fonction effectue une sélection des *Parcelles ODR* qui respectent la formule :

$$((\text{surffinale} > 1500) \& ((\text{surfm2}/\text{surfinitm2}) > 0.4) \& ((\text{surfm2}/\text{surfinitm2}) < 0.8)) \text{ or } ((\text{surffinale} > 100) \& ((\text{surfm2}/\text{surfinitm2}) > 0.7))$$

Avec :

- « *surfinitm2* » : surface initiale de la parcelle en m².
- « *surfm2* » : surface de la parcelle en m² après le découpage des masques (*RPG*, végétation, surface bâtie...) mais avant la séparation des polygones (multi polygones vers polygones).
- « *surffinale* » : surface finale en m² de chaque polygone.

Note : Les bornes utilisées ici sont tirées de la note méthodologique de Seard C. (septembre 2016) « *Mise à jour de la cartographie d'occupation du sol* ».

Par ailleurs, cette fonction conserve tous les « morceaux » de serres, vignes ou vergers (qui peuvent être de petites dimensions) dès lors qu'ils font plus de 50 m² (contre 100 m² pour les autres) :

$$(\text{NATURE} == \text{'Verger'}) \& (\text{surffinale} > 50) | (\text{NATURE} == \text{'Vigne'}) \& (\text{surffinale} > 50) | (\text{NATURE} == \text{'Bananeraie'}) \& (\text{surffinale} > 50) | (\text{NATURE} == \text{'Canne à sucre'}) \& (\text{surffinale} > 50) | (\text{NATURE} == \text{'Rizi\`ere'}) \& (\text{surffinale} > 50) | (\text{NATURE} == \text{'Houblonni\`ere'}) \& (\text{surffinale} > 50)$$

➤ Élimination des résidus 2 : def Nettoyage (fil, how='miller')

Le choix a été fait d'éliminer les objets dont l'indice de circularité (*Miller*) est inférieur à 0.12. Nous avons conservé la valeur utilisée dans le *model builder* initial.

- Calcul de la surface finale des *Parcelles ODR*: « *surffinale* ».
- Ajout d'un identifiant unique des *Parcelles ODR*: « *id_fin* ».
- Calcul de l'indice de circularité de Miller des *Parcelles ODR* : « *Ind_circ* »

surfinitm2	CODE_INSEE	code_18	NATURE1	NATURE2 ▼	code_pra	surffinale	id_fin	Ind_circ
5365,480...	47210	222	Verger	Serre	47399	1271,7...	47_210_1_0...	0,382...
18955,32...	47210	211		Serre	47399	3783,5...	47_210_1_0...	0,748...
4712,603...	47210	211	Vigne		47399	862,57...	47_210_1_0...	0,780...



Figure 15 - Parcelles ODR détectés dans le Lot-et-Garonne (47) et table attributaire associée

A.1.13- Calcul des statistiques zonales

Les *Parcelles ODR* sont maintenant enrichies de l'information *OSO*. Le module *python Rasterstats* permet de calculer les statistiques zonales pour chaque *Parcelle ODR* à partir du raster *OSO*.



Figure 16 - Parcelles ODR (délimitées en rouge) superposées aux pixels OSO

Ainsi, 23 champs sont créés (« cl1 », « cl2 » ... « cl23 ») correspondant aux 23 classes d'occupation du sol de OSO. Pour chacune d'elle le nombre de pixels intersectant la parcelle est compté. *Rasterstats* calcul également le nombre total de pixel des parcelles (« count »), la valeur de pixel la plus petite (« min »), les valeurs de pixel en minorité et en majorité (« minority » et « majority »), la déviation standard (« std ») ainsi que la valeur médiane (« median »).

Ensuite, la variété (« variety ») est calculée en comptant le nombre de classes différentes présentes dans chaque parcelle.

Enfin 9 champs de regroupement sont créés à la faveur de la nomenclature emboîtée d'OSO :

- « urbain » = Urbain dense + Urbain diffus + Zones industrielles et commerciales + Surfaces routes
- « cultures » = Colza + Céréales à pailles + Protéagineux + Soja + Tournesol + Maïs + Riz + Tubercules/Racines
- « prairies » = Prairies
- « vergers » = Vergers
- « vignes » = Vignes
- « forets » = Forêts de conifères + Forêts feuillus
- « pelouse » = Pelouse
- « landes ligneuses » = Landes ligneuses
- « divers » = Surfaces minérales + Plages et dunes + Glaciers et neiges + Eau + Valeurs nulles

cl1	cl2	cl3	cl4	cl5	cl6	cl7	cl8 ^	cl9	cl10	cl11	cl12	cl13	cl14	cl15	cl16	cl17	cl18	cl19	cl20	cl21	cl22	cl23
		6,0...																				
		7,0...																				
count	min	minority	majority	std	median	variety	urbain	cultures	prairies	vergers	vignes	forets	pelouses	landes	divers							
6	3,00...	3,0000...	3,00000...	0,000...	3,00000...	1	6,0000...	0,000000...				0,00000...			0,0000...							
7	3,00...	3,0000...	3,00000...	0,000...	3,00000...	1	7,0000...	0,000000...				0,00000...			0,0000...							

Figure 17 - Champs correspondants aux statistiques zonales ajoutés à la table attributaire des Parcelles ODR

Les *Parcelles ODR* sont alors exportées au format *Shapefile* en intégrant dans le nom des fichiers le numéro du département ainsi que l'année du RPG pris en compte (Ex. 01_Parcels_ODR_rpg_2017.shp).

Annexe 2 - Préparation des couches d'occupations du sol non agricole susceptibles de recevoir des traitements phytosanitaires

A.2.1- Cimetières

- **BD TOPO - CIMETIERE** : Sélection de tous les cimetières.

- Intersection avec les communes : `def overlay (cimetiere, commune, how="intersection")`

- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Note : L'identifiant de chaque enregistrement est celui issu de la *BD TOPO*. Cet identifiant est donc dédoublé quand un cimetière intersecte plusieurs communes.

A.2.2- Installations sportives

- *BD TOPO - TERRAIN DE SPORT* : Sélection sur le champ « NATURE » = 'Grand terrain de sport' ou 'Petit terrain multisports' ou 'Piste de sport' ou 'Terrain de tennis' (on écarte les bassins de natation).
- *BD TOPO - ZONE D'ACTIVITE OU D'INTERET* : ZAI dont la géométrie est réelle (« FICTIF » = 'non'), « CATEGORIE » = 'Sport', et « NATURE » = 'Centre équestre' ou 'Centre de tir' ou 'Golf' ('Autre équipement sportif', 'Hippodrome', 'Sports mécaniques' et 'Stade' sont déjà compris dans les TERRAIN_DE_SPORT).
- Concaténation de ces deux données.
- Intersection avec les communes : `def overlay (sport_tot, commune, how="intersection")`
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Note : L'identifiant de chaque enregistrement est celui issu de la *BD TOPO*. Cet identifiant est donc dédoublé quand un terrain de sport ou une zone d'activité intersecte plusieurs communes.

A.2.3- Aérodrômes

- *BD TOPO – AERODROME* : Sélection de tous les aérodrômes dont la géométrie est réelle (« FICTIF » = 'non').
- Intersection avec les communes : `def overlay (aerodrome, commune, how="intersection")`
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Note : L'identifiant de chaque enregistrement est celui issu de la *BD TOPO*. Cet identifiant est donc dédoublé quand un aéroport intersecte plusieurs communes.

A.2.4- Voies ferrées

- *BD TOPO – EQUIPEMENT DE TRANSPORT* : Sélection sur le champ « NATURE » des 'aires de triage' dont la géométrie est réelle (« FICTIF » = 'non').
- *BD TOPO - TRONCON_DE_VOIE_FERREE* : Utilisation de la couche générée lors de la construction du *Masque 1* (voie ferrées buferisées en fonction du nombre de voies).([A.1.2- Construction du Masque 1](#))
- Concaténation de ces deux données.
- Dissolve des géométries.

- Intersection avec les communes : `def overlay (voies_ferrees, commune, how="intersection")`
- Création d'un identifiant unique.
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

A.2.5- Routes

- **BD TOPO - TRONCON_DE_ROUTE** : Utilisation de la couche générée lors de la construction du *Masque 1* (routes buferisées en fonction de leur « CL_ADMIN » - [A.1.2- Construction du Masque 1](#))
- Dissolve des géométries.
- Intersection avec les communes : `def overlay (routes, commune, how="intersection")`
- Création d'un identifiant unique.
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

A.2.6- Jardins familiaux

- **BD TOPO - ZONE D'ACTIVITE ET D'INTERET** : Sélection des ZAI dont la géométrie est réelle (« FICTIF » = 'non') et dont la « NAT_DETAIL » = 'Jardins familiaux'.
- Intersection avec les communes : `def overlay (zai_jardins, commune, how="intersection")`
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Note : L'identifiant de chaque enregistrement est celui issu de la *BD TOPO*. Cet identifiant est donc dédoublé quand un jardin intersecte plusieurs communes.

A.2.7- Zones d'activité ou d'intérêt

- **BD TOPO - ZONE D'ACTIVITE ET D'INTERET** : Sélection de toutes les « CATEGORIE » sauf 'Sport' (déjà présent dans les terrains de sport) dont la géométrie est réelle (« FICTIF » = 'non') et dont la « NAT_DETAIL » est différente de 'Jardins familiaux'.
- Intersection avec les communes : `def overlay (zai, commune, how="intersection")`
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Note : L'identifiant de chaque enregistrement est celui issu de la *BD TOPO*. Cet identifiant est donc dédoublé quand une zone d'activité intersecte plusieurs communes.

A.2.8- Bâti

- *BD PARCELLAIRE EXPRESS* - **BATIMENT** et **PARCELLE** : Utilisation de la couche générée lors de la construction du *Masque 2* (= sélection des parcelles cadastrales qui contiennent du bâti - [A.1.5- Construction du Masque 2 \(Parcelles bâties\)](#))
- Multi polygones vers polygones : `def explode (bati)`
- Suppression des polygones < 20 m²
- Intersection avec les communes : `def overlay (bati, commune, how="intersection")`
- Dissolve des géométries.
- Création d'un identifiant unique.
- Calcul des surfaces des entités.
- Export au format *Shapefile*.

Annexe 3. Fonctions de nettoyage

La généralisation des traitements à l'échelle nationale produit parfois des polygones indésirables : trop petits, géométries invalides ou vides, polygones très fins et longs (bordure de routes, de parcelles). Pour anticiper d'éventuels problèmes et en vue d'alléger les données à manipuler, 3 types de nettoyage sont appliqués après les grandes opérations spatiales du script :

A.3.1- Nettoyage des géométries

- **Buffer 0** sur la colonne géométrie :
`parcelles ["geometry"] = parcelles.buffer(0)`
- On écarte les **géométries vides** :
`parcelles = parcelles.loc[~parcelles.geometry.is_empty]`
- On écarte également les rares entités avec des **géométries invalides** :
`parcelles = parcelles[(parcelles['geometry'].is_valid)]`

A.3.2- Nettoyage selon la forme

L'indice de circularité de *Miller* est utilisé pour écarter certains polygones particuliers :

$$\text{Indice de Miller} = (4 * \pi * \text{Area}) / \text{périmètre}^2$$

Il s'agit du rapport de la superficie d'une entité, à celle d'un cercle de même périmètre. Cette indice varie de 0 à 1. Il est de 0 pour une ligne et de 1 pour un cercle.

Il permet d'écarter les polygones très fins et longs (résidu fréquent après les opérations de différences).

Le choix a été fait d'éliminer les polygones dont l'indice de circularité < 0.015 après les trois principales opérations spatiales du script. Les tests effectués ont montré que ce seuil (très bas) éliminait uniquement des polygones indésirables s'apparentant à des lignes.



Figure 18 - Nettoyage avec l'indice de Miller (< 0.015)

Note : l'indice de *Miller* est également utilisé pour le nettoyage final des *Parcelles ODR* avec un autre seuil (0.12).

A3.3- Nettoyage selon la surface

On calcul la surface des entités puis on élimine celles inférieurs à 50 m². Ces parcelles seront de toute façon aussi éliminées en fin de procédure (voir A.1.12 ci-dessus).

Annexe 4. Notions de cadastre et de parcellaire agricole

Depuis le XIX^{ème} siècle, un plan parcellaire représente l'ensemble des parcelles d'une commune (le territoire de la commune est divisé en sections, chaque section donne lieu à un plan parcellaire). La « parcelle » est une portion de terrain appartenant à un même propriétaire et présentant une même nature (les plans peuvent faire apparaître des propriétés bâties - maisons et immeubles - et non bâties - prés, vignes, champs, etc). Au cours du XX^{ème} siècle, des remaniements du cadastre sont intervenus dans le but de permettre des révisions des évaluations foncières servant de base à l'imposition. Aujourd'hui, les plans cadastraux sont régulièrement mis à jour, et sont consultables sous format numériques¹¹. Réalisée à partir de l'assemblage de ces plans cadastraux dématérialisées, La BD Parcellaire® de l'IGN fournit l'information cadastrale numérique, géo-référencée et continue de l'ensemble du territoire français (IGN, 2019). De première importance, cette base de données va constituer la donnée d'entrée sur laquelle repose notre carte « RPG complété ».

¹¹ cadastre.gouv.fr

Cependant, la parcelle agricole (ou culturale) et la parcelle cadastrale présentent des caractéristiques différentes : la 1^{ère} est définie via l'utilisation agricole dont elle fait l'objet alors que la 2^{nde} est issue des informations relatives à sa propriété foncière.

En effet, la parcelle, selon Thines et Lempereur (1975), est « *une pièce de terre d'un seul tenant formant un tout au regard de la propriété, de l'exploitation ou de la culture; la parcelle est dite cadastrale lorsqu'elle est considérée sous l'angle juridique de la propriété, elle est qualifiée de culturale lorsqu'elle est envisagée sous l'angle de la culture qui la caractérise; elle est appelée parcelle d'exploitation, lorsqu'elle dépend d'un même exploitant.* »

Une information qui serait utile à notre projet est la nature des parcelles (terres agricoles, bois, constructibles, etc.). Cette classification est effectuée sur avis de la commission des impôts des communes et est disponible « uniquement » à ce jour sur demande aux mairies, d'après les numéros de parcelles. Elle n'est donc pas accessible pour une étude à l'échelle du territoire national comme la nôtre.

Les délimitations des parcelles culturales sont généralement moins stables dans le temps que celles des parcelles cadastrales. La parcelle culturale (un champ, un pré, une plantation, etc.) peut correspondre à un ensemble de parcelles cadastrales contiguës, entières ou partielles, portant généralement une seule culture, limitées par des éléments facilement repérables et permanents (tels qu'un chemin, un bois, un ruisseau, etc.). Mais des parcelles culturales peuvent également découler d'une subdivision de la parcelle cadastrale. Alors qu'en est-il de nos « *parcelles ODR* » ? Sont-elles représentatives d'une réalité agricole sur le terrain ?

Annexe 5. Référence à la Statistique Agricole Annuelle

La SAA reste l'unique source de statistiques annuelles sur les surfaces agricoles à l'échelle du département¹². Il s'agit de données « établies », selon Agreste, par les services¹³ du Ministère de l'agriculture. « Etablies » dans le sens où elles ne sont pas issues d'une enquête spécifiques ni issues de données de suivis ou réseaux ; mais « estimées » à partir de résultats d'enquêtes « périodiques » et locales, d'informations obtenues auprès d'acteurs locaux, d'organisations professionnelles et des données de déclaration PAC (comprendre : le RPG). « Les résultats de la SAA sont le fruit d'arbitrages » (Agreste, 2014).

Comme indiqué plus haut, la comparaison de ces statistiques avec le RPG est sujette à des biais importants : biais de temporalité, de nomenclature. De plus, la SAA d'un département se réfère aux surfaces agricoles des exploitations du département en question (si nous pouvons effectuer un calcul des surface identique côté RPG, il reste le problème de surfaces à compléter, qui sont – elles – par définition situées dans le département étudié). De plus, le problème du secret statistique appliqué aux données de la SAA ne facilite pas la production de données exhaustives.

Il semble également que l'homogénéité entre les départements n'est pas garantie dans la SAA ; par exemple sur certains départements nous notons des arrondis à l'hectare, quand pour d'autres les surfaces semblent arrondies au millier d'hectares.

Il est à noter également que les surfaces du RPG, pour une culture donnée, sont fréquemment plus élevées que celles affichées dans la SAA. De Rincquesen et al. (2016) note par exemple que « *selon les territoires, un déficit de surfaces (cultures permanentes non enregistrées dans le RPG) ou un excédent (surfaces en herbe collectives non prises en compte par le RA, ou déclaration différenciée) peut être constaté entre le RPG et le RA* ».

Nous entendons alors utiliser les données de la SAA pour repérer les cultures les plus manquantes dans le RPG, par département, parmi l'ensemble des grandes cultures. La SAA nous permet de répondre à ce type de question : manque-t-il, dans le département *dd* et la campagne *yy*, plus de blé que de maïs ? (ou bien : les deux cultures sont-elles globalement sous représentées dans le RPG) ? des protéagineux ? des légumes ? etc.

¹² Le département constitue l'échelle à laquelle nous voudrions confronter la SAA et le RPG.

¹³ Services régionaux de l'information statistique et économique des directions régionales de l'agriculture, de l'alimentation et de la forêt. La coordination de l'opération, la centralisation des données et leur exploitation à l'échelon national sont assurées par le SSP, bureau des statistiques sur les productions et compatibilités agricoles.