



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

le RPG complété.

Partie 2 : Attribution des cultures aux parcelles identifiées comme susceptibles d'accueillir les surfaces agricoles hors RPG.

*Pierre Cantelaube et Benjamin Lardot, février 2021,
Note méthodologique, version 3 revue Septembre 2022*



La note méthodique décrivant la création et la construction du RPG complété est constituée de 2 documents :

1. Production de la couche géographique des parcelles susceptibles d'accueillir les surfaces agricoles hors RPG (Lardot et al., 2022, hal-03647500)
- 2. Attribution des cultures aux parcelles identifiées**

Auteur de correspondance : pierre.cantelaube@inrae.fr

Pour citer ce document :

Cantelaube P. et Lardot B., 2021, Construction d'une base de données géographiques à échelle fine exhaustive sur l'occupation agricole du sol : Le « RPG complété ». *Partie 2 : Attribution des cultures aux parcelles susceptibles d'accueillir les surfaces agricoles hors RPG. Version 3. Note Méthodologie US ODR*, INRAE, septembre 2022

(1^{er} version : février 2021) (2^{ième} version : Octobre 2021)

Version 3, juillet 2022.

<https://odr.inrae.fr/>

Les travaux présentés dans ce document ont été réalisés (entre autres) sous financements du projet OFB dans le cadre de la convention OFB-INRAE (2020-2021). Les travaux précédents ont pu être menés à bien grâce aux financements dans le cadre du Projet 11 du projet spatialisation BNVD- plan Ecophyto (2012-2016 puis 2017-2020)



Avant-propos et résumé

La cartographie d'occupation du sol agricole a été élaborée à l'INRAE, unité ODR, dans le cadre du programme EcoPhyto. Le but est de compléter le RPG pour disposer, à l'échelle parcellaire, d'une représentation géographique exhaustive de la surface agricole française (métropolitaine). Il s'agit donc de créer une carte annuelle parcellaire, nommée RPG complété dont l'objectif peut se résumer grâce à la formule suivante : $RPG + RPG\ complété = couverture\ exhaustive\ de\ la\ surface\ agricole$ (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 « parcelles ODR ». Se décomposant en 2 étapes, le travail consiste (1) à repérer ces parcelles susceptibles d'accueillir ces 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 donc « RPG complété ») ont été établies pour compléter les campagnes 2012 à 2014 du RPG (ODR-Ecophyto, 2015, Cantelaube et al., 2016, Séard, 2016) suivant une première méthodologie (RPG complété version v1) depuis revue et améliorée.

Initialement, dans cette première version du RPG complété, 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 parlons alors d'« îlot ODR » plutôt que de « parcelles ODR ») de façon à ce que la somme des surfaces par culture ainsi obtenue, sur un département, corresponde exactement à la différence SAA - RPG. Les résultats avaient du sens uniquement 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 ayant leur siège social dans le département vs surfaces physiquement dans le département, différences temporelles, de nomenclatures) ainsi que le niveau géographique des entrées du modèle probabiliste (principalement l'échelle de la carte CORINE land COVER, CLC, 2009) rendaient ces résultats peu satisfaisants et peu utilisables à une échelle fine.

Dans la version 2 du RPG complété présentée ici, cette volonté de viser absolument la référence fournie par la SAA est abandonnée. 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). Cette information auxiliaire, issue de la télédétection, de qualité (thématiquement, géographiquement et temporellement) permet en outre d'attribuer une seule culture par parcelle et est devenue le support principal du RPG complété. Après avoir évalué l'utilisation d'OSO dans notre contexte et préparé notre méthode de construction à partir des cartes OSO pré-2018 (OSO en 17 classes, dont 2 agricoles + 1 classe dédiée aux prairies), une méthodologie plus « définitive » a été établie pour l'intégration de la carte OSO en 23 classes (dont 9 agricoles + 1 classe dédiée aux prairies).

Cette méthodologie est présentée dans le présent document.

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 (IGN, 2020b, 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). Une méthodologie a été développée par l'unité INRAE US ODR pour identifier et délimiter des parcelles dites « parcelles ODR » susceptibles d'être ces parcelles agricoles absentes du RPG¹.

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 (RPG complété version v1). La méthodologie a été revue pour produire le RPG complété version v2 pour les campagnes 2014 (refaite donc) et suivantes.

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 RPG. **Cette étape s'effectue via une succession de traitements géomatiques grâce à l'utilisation de nombreuses couches d'informations géographiques, notamment issues de l'IGN : Bd Parcellaire® (Cadastre), Bd Topo®.** La 2^{nde} phase attribue à ces parcelles identifiées une catégorie d'occupation du sol (culturales ou non agricole). **Cette dernière étape utilise de nombreuses informations auxiliaires ; et si la version v1 se basait essentiellement sur la statistique agricole annuelle (SAA, Agreste) et la couche CORINE Land Cover (CLC), la version v2 utilise désormais la carte d'occupation du sol issue de la télédétection ; la carte OSO développée par le CESBIO-theia².**

Cette note méthodologique décrit la phase 2 de l'élaboration du RPG complété, la méthodologie - basée sur l'utilisation de la carte OSO (Cesbio, 2018) - développée pour attribuer une classe d'occupation du sol (un type de culture) aux parcelles issues du fond parcellaire agricole hors RPG produit lors de la phase 1 (Lardot et al., 2022)

¹ Ce travail a été réalisé 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).

² <https://www.theia-land.fr/ceslist/ces-occupation-des-sols/>

2. Parcellaire agricole et carte d'occupation de sols OSO

A ce stade nous disposons d'un ensemble de parcelles susceptibles d'accueillir les surfaces agricoles hors RPG. Telles qu'elles ont été construites (voir Lardot et al., 2022), ces parcelles représentent encore une surface bien plus grande que celle occupée par ces cultures hors RPG. Certaines d'entre elles seront écartées car elles n'accueillent pas une surface dédiée à une activité agricole (zones artificialisées, bois, surfaces en herbe non agricoles, etc., éléments non détectables lors de 1^{ère} phase), les autres se verront attribuer une catégorie d'occupation du sol agricole de la nomenclature du RPG 28 classes, quelque peu élargie parfois à une sous-catégorie de la nomenclature étendue du RPG (soja, betteraves, pommes de terre, etc.).

Pour attribuer une catégorie d'occupation du sol (agricole ou pas, donc) aux parcelles, l'information principale provient de la couche OSO³ (Cesbio, 2018, Inglada et al., 2017), carte d'occupation du sol annuelle issue de la télédétection. Ce document décrit la méthode utilisée pour transférer l'information des pixels de cette carte OSO vers les unités spatiales définies par les parcelles ODR, grâce à l'aide d'information géographique et thématique complémentaire, via l'utilisation de modèles statistiques. Le produit final sera ainsi une couche d'information géographique à l'échelle de la parcelle, complémentaire au RPG, afin de disposer d'une couche parcellaire exhaustive de la surface agricole utile (SAU) française (métropole).

2.1 Le fond parcellaire agricole hors RPG

Construite lors de la 1^{ère} étape du projet RPG complété à partir la BD-Parcellaire[®] de l'IGN et d'autres sources d'information géographique dont la Bd-Topo[®] de l'IGN (voir *Note méthodologique*, partie 1), cette couche d'occupation du sol fournit donc un ensemble de parcelles qu'il faut maintenant qualifier thématiquement afin d'écarter les parcelles non agricoles encore présentes et de distinguer les types de cultures présentes sur celles dédiées à une activité agricole.

La Figure 1 ci-dessous montre des exemples de parcelles obtenues à ce stade. Les surfaces agricoles sont bien délimitées, mais la couche contient encore des parcelles clairement non agricoles (voir par exemple à gauche de la figure du milieu, la figure du bas en zone urbaine). Celles-ci devront être écartées de la couche géographique finale, c'est-à-dire l'ensemble des parcelles agricole hors-RPG : le RPG complété.

Ces parcelles qui deviendront les entités géographiques du RPG complété peuvent être vues comme des parcelles cadastrales retravaillées ; l'utilisation des parcelles cadastrales et leur découpage (éventuel) par des éléments artificialisés (chemins, routes, etc.) et des éléments paysagers (bois, forêt, haies, etc.) permet d'approcher correctement - d'après les observations - les délimitations des parcelles culturales 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 dans l'ensemble satisfaisant.

Il faut noter cependant la présence de parcelles culturales trop découpées : la parcelle culturale réelle (telle qu'elle est observée) est représentée par plusieurs de nos parcelles ODR, contiguës (par exemple sur la Figure 1, image du haut, parcelles en haut à gauche). Ce cas ne pose pas vraiment de problème si les parcelles ODR qui constituent la parcelle culturale sont qualifiées par une même culture.

L'inverse - c'est-à-dire la parcelle ODR contient plusieurs parcelles culturales « réelles » et donc la parcelle ODR devrait accueillir plusieurs types de cultures⁴ - est heureusement plutôt rare : lors des observations et contrôles effectués, la fréquence d'apparition de ce cas est inférieure à 1%.

³ La carte d'occupation du sol OSO : le Centre d'Expertise Scientifique OSO (Occupation du Sol) - par l'intermédiaire du CESBIO - produit cette carte d'occupation du sol annuelle à l'échelle du territoire métropolitain français et corse.
<http://osr-cesbio.ups-tlse.fr/~oso/>

⁴ Ce qui est impossible, à moins de redécouper la parcelle ODR. Seule la couche OSO pourrait permettre cela mais ce travail reste très délicat et incertain, comme le montrera la suite de ce document.



Figure 1 - illustrations de parcelles obtenues avant attribution des cultures, fond de carte à la fin de l'étape 1 de construction du RPG complété (de haut en bas : département 50, 51, 78).

2.2 La carte d'occupation des sols OSO

La carte d'occupation du sol OSO annuelle est issue de la télédétection (résultat de traitements automatiques massifs de séries temporelles d'images satellites optiques Sentinel-2, voir Cesbio 2018, Inglada et al. 2017b) et affiche une résolution de 10m (l'unité minimale de collecte est donc de 0.01 ha). A partir de la campagne 2018, cette carte (nommée « OSO » ci-après dans ce document) décrit l'occupation du sol grâce à 23 classes dont 8 catégories culturales (y compris riz), plus une catégorie

« prairies », 2 dédiées aux cultures pérennes (« vignes » et « vergers ») et également 2 dernières classes potentiellement d'intérêts ici (« landes ligneuses » et « pelouses »).

Nomenclature OSO (2018)	
1 Urbain dense	5 Colza
2 Urbain diffus	6 Céréales à paille
3 Zones indus. et comm.	7 Protéagineux
4 Routes	8 Soja
16 Forêts de feuillus	9 Tournesol
17 Forêts de conifères	10 Maïs
20 Surfaces minérales	11 Riz
21 Plages et dunes	12 Tubercules/Racines
22 Glaciers et neiges éternelles	13 Prairies
23 Eau	14 Vergers
	15 Vignes
	18 Pelouses
	19 Landes ligneuses

Tableau 1 :
Nomenclature OSO (2018) 23 catégories, Code couleurs (proposées par le producteur de la couche) reprises dans le reste de ce document.

Cette information, pour autant qu'elle soit de qualité (thématiquement et géographiquement), ne suffit pas à qualifier nos parcelles dans des classes agricoles similaires à celles du RPG (dans leur nomenclature 28 classes, élargie dans la mesure du possible). Par exemple, les *céréales à paille* d'OSO peuvent correspondre aux classes RPG *Blé, Orge, Autres céréales*. Les légumes et cultures industrielles peuvent se confondre dans OSO. Les prairies et les prairies permanentes ne sont pas distinguées. Les fourrages peuvent être classés dans différentes catégories OSO (Prairies, tubercules, voire céréales). Comment retrouver les jachères ou encore les surfaces temporairement non exploitées ?, etc.

De plus, les parcelles sont constituées d'un ensemble de pixel OSO et ces pixels ne sont pas toujours de la même classe d'occupation du sol. Le but est pourtant d'attribuer une culture unique à chaque parcelle.

Des informations auxiliaires seront donc nécessaires pour attribuer aux parcelles une catégorie d'occupation du sol unique. Ces informations seront de 3 types : (1) relatives à la parcelle (taille, forme, ensemble des pixels OSO présents), (2) complémentaires à OSO sur l'occupation du sol et (3) de contexte sur l'environnement « paysage agricole » dans lequel se trouve la parcelle.

Préalablement, il est nécessaire d'étudier les relations entre les parcelles agricoles et les pixels de la couche OSO. C'est l'objet des paragraphes suivants : pour les parcelles RPG (§2.3) et les « parcelles ODR », futures parcelles du RPG complété (§2.4).

2.3 Parcelles agricoles et catégories OSO

2.3.1 les catégories OSO confrontées aux cultures RPG

Les parcelles RPG (comme les parcelles ODR) sont constitués d'un ensemble de pixels OSO de taille 10m par 10m (seuls les pixels qui s'inscrivent entièrement dans la parcelle sont pris en considération).

En prenant en considération uniquement les parcelles RPG composées de pixels OSO d'une seule catégorie, les tableaux ci-dessous (Tableau 2, tables de contingence) comparent la culture déclarée (colonnes) et cette catégorie OSO (lignes). La culture du RPG est exprimée dans sa nomenclature en 28 postes⁵. La colonne « surf » indique la surface représentée par les parcelles RPG ainsi retenues (en hectare). Ces résultats sont présentés par région et se reportent à la campagne 2018 du RPG (comparaison avec OSO 2018).

Ainsi, 99,54% des parcelles contenant uniquement des pixels de la catégorie OSO « 6.Céréales à paille » dans la région Haut-de-France sont effectivement dans les groupes de culture *gc1,gc3,gc4* (respectivement blé tendre, orge, autres céréales) du RPG.

⁵ *Rappel : groupes de culture du RPG : 1-Blé tendre, 2-Maïs, 3-Orge, 4-Autres céréales, 5-Colza, 6-Tournesol, 7-Autres oléagineux, 8-Protéagineux, 9-Plantes à fibres, 11-Gel/Jachère, 15-Légumineuses à grains, 16-Fourrage, 17-Estives/Landes, 18 – Prairies permanentes, 19-Prairies temporaires, 24-Autres cultures industrielles, 25-Légumes/Fleurs, 28-Divers*

Hauts de France (nouvelle région)

	1,3,4	2	5	6	7	8	9	11	16	17	18	19	15	25	24	28	surf
5:colza	0.1	0.03	99.51			0.06		0.03	0.07		0	0.02	0.02	0.02	0.03	0.09	58 026
6:céréales	99.54	0.03	0.01			0.01	0.15	0.03	0.03		0		0	0.02	0.03	0.11	424 520
7:protéas	0.72		0			41.37	20.22	0.03	0.56			0.01	0.3	36.25	0.4	0	6 844
8 :soja																100	1
9 :tourn									0.35		0.03			98.19	0.53	0.91	34
10 :maïs	0.15	98.94					0.08	0.03	0.04		0.01	0.09	0	0.5	0	0.16	39 033
12 :tuber	0.12	0.06	0.01		0.01	0.05	0.09	0.03	0.53		0	0.02	0	34.17	64.77	0.12	8 5374
13 :prair	0.1	0.09				0.01		0.66	4.40	0.01	90.07	4.14	0	0	0	0.37	56 865
19 :lande	0.14							3.33	0.29		86.89	0.14		0.29	0.07	8.83	14
2:batidif	0.77	1.49				0.18	0.08	10	2.67	0.07	19.96	5.58		6.72	1.5	47.63	98
3:zonind	1.57	0.52	0.16			4.88		19.31	1.73		3.72	6.92		26.55		27.54	442

Auvergne (ancienne région)

	1,3,4	2	5	6	7	8	9	11	16	17	18	19	15	25	24	28	surf
5:colza	0.32	0.01	98.09	0.01	0.07	1.15		0.09	0.01			0.02				0.2	5 447
6:céréales	99.38	0.05	0.01	0.02	0.04	0.02		0.07	0.06			0.06	0.02			0.24	45 005
7:protéas	1.08	0.12			33.43	57.52					0.16	0.12	7.06	0.12		0.38	57
8 :soja					97.91			1.54	1.21			0.01				0.56	21.49
9 :tourn	0.06	0.82	0.02	95.98		0.01		0.07	1.21			0.01				0.42	1 688
10 :maïs	0.26	98.64	0.01	0.06	0.06	0.01		0.08	0.08			0.13		0.17		0.40	18 794
12 :tuber		0.32		0.1				0.1	0.33			0.02		0.16	98.65	0.3	2 695
13 :prair	0.15							0.13	4.01		81.44	13.71		0.01		0.12	323 155
18 :pelou	17.31	0.53	0.13			0.05	0.63	0.51	9.7	5.35	26.39	15.95		1.33	0.12	3.58	943
19 :lande								35.88		17.04	17.52		15.57			24.90	14
2: batidif	0.03							5.34		0.03	13.85	6.43	3.15		1.47	25.41	77
3:zonind	1.82	0.78						13.14	1.27		5.82			22.98	0.4	21.11	60

Midi-Pyrénées (ancienne région)

	1,3,4	2	5	6	7	8	9	11	16	17	18	19	15	25	24	28	surf
5:colza	0.13	0.04	98.54		0.07	0.35		0.08	0.18		0.11	0.11		0.15		0.34	14 194
6:céréales	98.9	0.01	0.02	0.06	0.07	0.06		0.13	0.2		0.03	0.1		0.01		0.40	166 888
7:protéas	1.26	0		0.01	1.79	88.36		0.91	3.34		0.01	0.21	3.16	0.43		0.36	1 557
8 :soja	0.38	0.51		0.26	97.86			0.06	0.12			0.03		0.07	0.04	0.40	6 280
9 :tourn	1.77	0.43	0.04	91.14	0.41	0.01	0.03	0.63	0.49		0.01	0.17	1.72	2.13	0.29	0.58	90 310
10 :maïs	1.04	97.5		0.1	0.19	0.09		0.21	0.17		0.03	0.09		0.1	0.01	0.45	55 603
12 :tuber									8.19				20.56	0	69.52	1.11	26
13 :prair	0.33	0.09		0.01		0.03		1.01	16.51	0.31	59.59	21.23		0.01		0.34	204 521
18 :pelou	3.1	0.02			0.02	0.06		1.49	9.67	51.44	25.83	6.65	0.01	0.09	0.03	1.21	17 025
19 :lande									0.09	89.51	8.91	0.06		0.03		1.17	320
2:batidif	0.18	0.12		0.05		0.01		8.65	0.32	1	13.21	3.87	0.01	10.05	0.51	38.0	116
3:zonind	0.16	0.02						9.39	0.85	0.52	4.13	1.51		42.5	0.78	32.95	113

Rhône-Alpes (ancienne région)

	1,3,4	2	5	6	7	8	9	11	16	17	18	19	15	25	24	28	surf
5:colza	0.36	0.10	98.96			0.11		0.04			0.03	0.01	0.01	0.09	0.02	0.19	6 975
6:céréales	99.07	0.11	0.02			0.14		0.05	0.17			0.1		0.01	0.05	0.24	31 693
7:protéas	9.06					74.54								15.23	0.11	1.01	37
8 :soja	0.01	0.15		0.4	98.52			0.03				0.15	0.01	0.21	0.07	0.42	1 587
9 :tourn	1.7	0.86		84.81	0.39			0.23	0.14				0.32	5.28	4.61	0.98	4 499
10 :maïs	0.84	97.91	0.02	0.03	0.22	0.01		0.09	0.08			0.12	0.02	0.08		0.52	27 990
12 :tuber					6.41				6.28					44.38	41.25	1.48	24
13 :prair	0.25	0.13		0.02				0.41	6.19	0.38	76.47	15.62		0.02		0.23	116 467
18 :pelou	1.94	0.03				0.03		1.3	4.63	34.60	46.75	2.3	0.01	0.19	1.91	2.16	20 038
19 :lande	0.02								1.31	64.40	9.35	0.31		0.16	14.64	8.72	118
2:batidif	0.89	0.44		0.11	0.01	0.01		3.32	0.79	0.11	12.28	3.18	0.01	4.80	5.33	20.05	408
3:zonind	1.29			0.09	0.02	1.22		7.84	0.37	0.01	1.48	1.62		56	0.41	24.72	82

Tableau 2 – tables de contingence entre les parcelles RPG d'une seule catégorie OSO et cette catégorie OSO.

Ces tableaux montrent bien la qualité d'OSO pour les grandes cultures annuelles les plus fréquentes dans le paysage agricole français (Maïs, Blé et plus largement l'ensemble des céréales, Colza, Soja ; correspondance exacte à plus de 97%). Pour les cultures moins présentes de manière homogène sur l'ensemble du territoire, par exemple le tournesol, l'information OSO se révèle un peu moins juste et surtout très hétérogène selon les régions. Par exemple, pour le tournesol : OSO correct à 91.14% en Midi-Pyrénées où cette culture est très présente, correct à 95.98% en Auvergne où elle est peu présente et (presque) systématiquement confondu avec des cultures maraichères dans les Hauts-de-France où les surfaces en tournesol sont très rares ou absentes.

Logiquement les parcelles en prairie selon OSO se répartissent bien entre prairies permanentes et temporaires du RPG, incluant parfois également les prairies artificielles (fourrages gc16). Les prairies du RPG peuvent également être vues comme des pelouses par OSO. D'un autre côté ces pelouses OSO peuvent correspondre parfois à des parcelles RPG déclarées en céréales à paille (par exemple en Auvergne, voir Tableau 2).

Des confusions peuvent apparaître pour des cultures plus « confidentielles » : par exemple la classe OSO protéagineux (fréquemment confondus avec des légumes). Les « tubercules OSO » correspondent (là aussi logiquement) avec des légumes (gc25 du RPG) ou des cultures industrielles (gc24), typiquement pommes de terre et betteraves dans les Hauts-de-France.

Enfin, des parcelles RPG peuvent entièrement être composées de pixels OSO de catégories « bâti » ou « zone industrielle » : parcelles en jachère, classées divers (particulièrement *surface temporairement non exploitée*), mais aussi en culture maraichère (ce qui peut s'expliquer notamment pour les serres agricoles ou les cultures sur bâches/toiles de paillage) et, plus surprenant, en prairie⁶.

2.3.2 Les cultures RPG confrontées aux catégories OSO.

Inversement, confrontons les déclarations RPG avec les catégories OSO. Les tableaux ci-dessous (Tableau 3) détaillent les catégories OSO majoritairement rencontrées dans les parcelles du RPG par type de culture déclarée.

Ainsi, 98.55% des parcelles en *Blé tendre* (gc1) sont composées majoritairement de pixels OSO 6-*Céréales à paille* dans les Hauts-de-France (94.76% en Midi-Pyrénées). En revanche, seulement 70.65% (resp. 72.%) des parcelles en « autres céréales » (gc4) sont composées majoritairement de pixels OSO catégorie 6 ; et les autres catégories OSO présentes dans de telles parcelles sont plutôt hétérogènes d'une région à l'autre (12-*Tubercules/racines* et 2,3,4-*Artificielisés* dans les Hauts-de-France ; 9-*Tournesol*, 10-*Maïs* et 13-*Prairie* en Midi-Pyrénées). Ce groupe de culture « autres céréales » est toutefois peu présent dans les Hauts-de-France.

Dans cette région, lorsque la parcelle est composée majoritairement de pixels OSO 12-*Tubercules/racines*, celle-ci est bien déclarée en cultures industrielles (à 97.46%) ou en légumes (un peu moins : 69.32%), cela est dû aux surfaces particulièrement importantes en betteraves (gc24) et pommes de terre (gc25) dans ce territoire. En revanche, les cultures maraichères en Midi-Pyrénées sont confondues par OSO avec du tournesol (OSO-9) et des surfaces artificialisées (OSO-2à4) ; et la catégorie majoritaire OSO est la catégorie 12-*Tubercules/racines* pour moins de 1% d'entre elles.

⁶ Des observations indiquent que certaines de ces parcelles RPG en prairie peuvent correspondre à des surfaces dédiées à l'élevage équin, mais cela ne constitue pas l'unique confusion.

Région Hauts-de-France, parcelles RPG dans le gc [ligne] composées en majorité de pixels OSO [colonne]

	2, 3 ou 4	5	6	7	8	9	10	12	13	14 ou 15	18 ou 19	Nb parcelles
gc1	0.3	0.16	98.55	0.28	0	0.01	0.12	0.35	0.16	0	0.01	86 784
gc2	1.89	0.46	2.49	0.18	0	0.11	90.58	3.36	0.57	0.07	0	29 496
gc3	0.97	0.77	93.16	2.77	0	0.09	0.22	1.4	0.45	0.03	0.01	20 365
gc4	4.13	1.71	70.65	10.51	0.03	0.3	2.09	7.84	2.04	0.06	0.08	3 636
gc5	0.3	98.13	0.9	0.05	0	0.01	0.07	0.31	0.17	0	0	13 403
gc6	6.67	0	20	6.67	0	20	6.67	30	0	0	0	30
gc7	9.64	0	13.25	46.99	1.2	2.41	6.02	18.07	1.2	0	0	83
gc8	2.63	3.93	4.51	79.2	0	0.45	1.34	7.41	0.27	0	0	2 240
gc9	1.38	0.2	24.25	61.84	0	0.48	0.38	11.26	0.08	0.04	0	5 018
gc11	29.42	3.59	14.37	0.87	0.01	0.14	2.89	6.09	34.65	0.19	0.82	16 698
gc15	3.08	3.08	10.77	50.77	0	1.54	1.54	29.23	0	0	0	65
gc16	7.95	2.18	6.29	2.56	0.01	0.17	4.85	22.75	51.48	0.11	0.17	7 111
gc17	7.19	0	0.72	0	0	0	0	0.72	38.85	0	1.44	139
gc18	6.02	0.14	0.64	0.02	0	0	0.46	0.27	90.3	0.02	0.13	47 572
gc19	14.16	2.44	6.28	0.66	0.03	0.15	6.56	6.52	59.88	0.43	0.29	6 814
gc24	0.8	0.16	0.64	0.31	0	0.04	0.34	97.46	0.16	0	0.01	22 504
gc25	6.73	0.13	1.04	6.43	0.22	0.18	4	79.32	0.31	0.09	0.05	27 591
gc28	29.31	3.73	28.02	1.24	0.01	0.18	5.25	8.59	15.57	0.35	0.36	44 583

Région Midi-Pyrénées, parcelles RPG dans le gc [ligne] composées en majorité de pixels OSO [colonne]

	2, 3 ou 4	5	6	7	8	9	10	12	13	14 ou 15	18 ou 19	Nb parcelles
gc1	0.33	0.18	94.76	1.77	0.04	0.47	0.21	0	1.47	0.09	0.54	45 663
gc2	0.51	0.11	0.69	0.15	1.67	4.9	89.76	0.01	1.58	0.13	0.18	34 799
gc3	0.44	0.49	89.82	2.03	0.14	0.92	0.45	0.01	3	0.26	2.16	20 649
gc4	0.71	0.64	72.6	3.06	2.55	7.04	6.26	0.02	4.64	0.3	1.74	32 692
gc5	0.27	95.06	1.64	0.93	0.06	0.8	0.34	0.03	0.67	0.03	0.1	7 119
gc6	0.42	0.11	0.55	0.24	0.4	96.57	1.15	0.01	0.29	0.09	0.05	27 473
gc7	0.68	0.37	2.44	2.76	74.83	11.96	6.14	0.03	0.46	0.15	0.05	9 217
gc8	1.38	5.28	13.13	67.91	0.48	3.69	1.4	0.22	4.73	0.46	0.92	4 144
gc9	3.12	0	3.12	3.12	9.38	75	3.12	0	3.12	0	0	32
gc11	12.16	1.55	10.98	2.09	0.99	11.2	4.99	0.04	40.84	4.56	5.86	54 910
gc15	0.82	0.22	0.39	25.09	0.39	55.68	0.74	15.16	0.3	0.56	0.43	2 308
gc16	2.22	1.17	4.98	2.35	0.45	6.25	3.1	0.26	70.45	1.67	6.35	49 932
gc17	1.25	0.04	0.51	0.02	0.01	0.12	0.11	0	20.47	0.17	38	36 245
gc18	2.2	0.15	0.83	0.1	0.02	0.42	0.3	0	82.72	0.39	7.56	215 671
gc19	2.72	1.45	4.42	1.23	0.16	1.85	1.82	0.03	77.86	0.83	6.49	56 100
gc24	9.6	2.89	1.29	3.42	2.4	37.12	4.25	26.22	7.94	0.74	3.05	1 083
gc25	21.44	1.67	3.75	4.53	1.65	39.77	6.53	0.84	10.98	3.47	2.78	8 260
gc28	15.35	2.1	18.52	0.84	1.95	14.44	8.29	0.05	21.87	2.92	4.15	95 318

Tableau 3 - tables de contingence entre les parcelles RPG et les catégories OSO (somme = 100% en ligne)

Sans détailler toutes les informations présentes dans ces tableaux, ainsi que dans les tableaux du paragraphe précédent, des conclusions s'imposent rapidement :

- De fortes hétérogénéités entre les régions/territoires.
- Plus de difficultés pour OSO à décrire les cultures peu présentes sur un territoire donné.

2.3.3 Parcelles agricoles et pixels OSO

Observer la présence et l'arrangement spatial des catégories OSO dans une parcelle (RPG ou ODR) amène à distinguer 3 cas : (1) la parcelle contient une seule catégorie OSO, (2) la parcelle contient plusieurs catégories OSO mais l'une d'entre elles est prépondérante (couvre une grande partie de la parcelle, plus de x % de sa surface, x dépendant de la taille de la parcelle) et enfin (3) la parcelle contient de nombreuses catégories OSO chacune occupant une partie conséquente de la parcelle.

Le cas (1) est facile à traiter, le report de la catégorie OSO sur la parcelle est soit immédiat (si équivalence de la classe OSO avec un groupe de culture RPG), soit nécessite un traitement de répartition (OSO « Céréales » vers gc1, gc3 ou gc4 du RPG, voir plus bas) ou une « correction » thématique (OSO « Prairie » vers Prairie temporaire, vers fourrages, etc. ; selon le Tableau 2).

Le cas (2) concerne généralement les décalages spatiaux autour des limites de la parcelle due à la résolution OSO ou bien à la difficulté pour OSO à qualifier des espaces en bordures de champs ou d'éléments non agricoles (éléments paysagers de type haies ou bois, chemins, etc.) autour des parcelles⁷. Plus la parcelle est de grande taille, moins ce cas est difficile à traiter.

Le cas (3) est généralement dû soit à des difficultés pour OSO pour définir clairement l'occupation du sol observée (notamment pour le maraichage ou certaines cultures industrielles), soit à une parcelle qui présente réellement un couvert hétérogène (par exemple pour des parcelles temporairement non exploitées ou en jachère).

Les figures ci-dessous (Figure 2 et Figure 3) illustrent ces cas sur des exemples de parcelles du RPG, pour lesquelles donc la culture est connue. Sur chaque figure, à droite l'image satellite (attention : référence temporelle incertaine !) et à gauche l'occupation du sol vue par OSO (2018). Les traits bleus correspondent aux limites des parcelles RPG (2018).

La Figure 2 proposent des exemples de parcelles RPG composées de pixels OSO de plusieurs catégories. Les derniers exemples (Figure 3) montrent la difficulté majeure pour la production du RPG complété à partir de la couche OSO : l'erreur thématique sur des cultures maraichères (et certaines cultures industrielles) où des pixels OSO sont qualifiés comme classes artificialisées (généralement « zone industrielle et commerciale »).

Remarquons que sur ce type de parcelles agricoles de nombreuses serres sont généralement présentes. L'intérêt pour le RPG complété est de conserver et recenser de telles parcelles. Nous verrons ultérieurement que cela est réalisé grâce à la couche 'Bati' de la Bd-Topo[®] de l'IGN qui distingue les bâtiments de type *serres agricoles*.

Notons également que les vergers et les vignes sont globalement mal rapportées par OSO (La qualité est hétérogène selon les territoires (Les vergers sont par exemple bien localisés dans le sud-ouest, moins bien dans le nord).

Les erreurs thématiques de la carte OSO restent peu nombreuses, cette carte d'occupation du sol présente une grande qualité, notamment pour les cultures annuelles (voir les *matrices de confusions*, Cesbio⁸, 2018). Les grandes cultures sont bien rapportées (blé, maïs, colza, etc.), les prairies également. Cependant, les cas extrêmes présentés ci-dessus où soit sont notées des confusions thématiques, soit la distribution de nombreuses catégories OSO dans la parcelle est complexe, montrent que le lien entre la culture présente sur une parcelle et les pixels qui composent cette parcelle n'est pas systématiquement évident et immédiat. D'autant plus que les légumes, les vergers et les vignes sont les cultures les plus absentes du RPG et sont d'un grand intérêt pour des utilisations prévues du RPG complété (en lien avec les produits phytosanitaires par exemple).

⁷ Il existe des indices rendant compte de la structure et l'organisation spatiale de différentes classes d'une unité source (le pixel par exemple) rencontrées dans une unité géographique cible plus grande. Ces indices permettent de distinguer des cas où le nombre de classe et leurs surfaces respectives dans une unité cibles sont identiques mais leurs distributions différentes. C'est l'exemple classique de la répartition de 2 classes « noir » et « blanc » sur une grille de 8x8, 64 cellules : ces 2 couleurs peuvent être distribuées telles que les cases noires et blanches d'un échiquier ou bien en 2 groupes bien distincts les blancs à droite et les noirs à gauche par exemple. Voir par exemple Eiden et al. (2002)

Nous n'avons pas utilisé de tels indices dans ce travail. Nous estimons qu'ils apporteraient une amélioration significative (probablement) uniquement dans peu de cas (au vue des observations réalisées), et surtout cela nécessiterait un « travail poussé » que nous n'avons pas eu le temps de réaliser. Cela reste une piste d'amélioration. Ces indices permettraient éventuellement de traiter les cas de parcelles « odr » regroupant 2 parcelles culturales « réelles » accueillant des cultures différentes (cas évoqué ci-dessus §2.1)

⁸ <http://osr-cesbio.ups-tlse.fr/~oso/>



Figure 2 – Exemples de distribution des pixels OSO dans des parcelles du RPG.

Sources : IGN, ASP pour le RPG, ESRI World Imagery Map (ESRI, Digital Globe, Earthstars geographic and CNES/Airbus) pour l'image satellite.



Figure 3 – (haut) Exemples de confusion thématique sur des pixels OSO entre « artificiel » et culture maraîchères, vues sur des parcelles du RPG (département d62)
(bas) Exemples de confusions thématiques sur des pixels OSO et de difficultés pour qualifier les parcelles dans des « paysages agricoles complexes » (département 13)

2.3.4 parcelles ODR décrites avec une seule catégorie OSO

Environ 1/3 des parcelles ODR accueillent une seule catégorie OSO. Ce résultat est évidemment dépendant de la taille des parcelles ; ainsi en terme de surface cela correspond à environ 20% de la surface totale. Mais ces pourcentages varient selon les régions. Nous illustrons ce propos sur deux régions⁹ aux paysages agricoles différents et assez représentatives de l'ensemble des régions françaises métropolitaines : l'une avec de nombreuses grandes cultures, légumes et culture industrielles (Hauts de France), l'autre où les prairies prédominent (Auvergne).

⁹ Une « ancienne » région, une « nouvelle » région, de tailles relativement similaires (mais les Hauts de France contiennent cependant à ce stade plus de parcelles en zones urbaines, parcelles qui seront écartées du RPG complété par la suite).

- Cas global, ensemble des catégories OSO

Les tableaux ci-dessous dénombre les parcelles ODR dans lesquelles une seule catégorie OSO est rencontrée. Les tableaux distinguent la taille des parcelles.

Une première remarque est qu'environ 30% des parcelles accueillent une seule catégorie OSO, cependant la surface que cela représentent varie selon les régions : 18% dans les Hauts de France, 27% en Auvergne où pourtant globalement les parcelles sont plus grandes. Ceci s'explique car ces parcelles « mono-OSO » sont majoritairement des prairies. En Auvergne, ce sont même presque exclusivement des prairies, du moins pour les parcelles de taille supérieure au ½ hectare.

Ces tableaux montrent en outre que la surface occupée par les parcelles de taille inférieure à un demi hectare constituent un peu plus de la moitié de la surface totale étudiée¹⁰, celles de taille supérieure à 1 hectare un peu plus d'un tiers (colonne « surface totale » dans les tableaux ci-dessous).

Région Hauts de France

Taille parcelle	Nb total de parcelles	Nb parcelles avec 1 catégorie OSO	soit en %	Surface totale	Surf. parcelles avec 1 catégorie OSO	soit en %
>500	383 409	108 657	28.3%	104 507	18 749 ⁽¹⁾	17.9%
> ½ ha	45 597	6 739	14.8%	54 909	7 161 ⁽²⁾	13%
> 1 ha	17 075	2 277	13.3%	35 175	4 092 ⁽³⁾	11.6%

(1) Prairies 9 080 ha (48%), Artificialisé 6 017 ha (38%), Agricole 2 759 ha (15%), Forêts 460 ha (2%)

(2) Prairies 4 627 ha, (65%), Agricole 2 037 ha (28%), Artificialisé 307 ha (4%), Forêt 113 ha (1%)

(3) Prairies 2 392 ha (58%), Agricole 1 474 ha (36%), Forêt 161 ha (4%), Artificialisé 52 ha (1%)

Région Auvergne

Taille parcelle	Nb total de parcelles	Nb parcelles avec 1 catégorie OSO	soit en %	Surface totale	Surf. parcelles avec 1 catégorie OSO	soit en %
>500	316 064	103 534	32.8%	102 613	27 864 ⁽¹⁾	27.1%
> ½ ha	48 160	12 791	26.6%	56 930	13 555 ⁽²⁾	23.8%
> 1 ha	17 559	4 136	23.5%	35 844	7 612 ⁽³⁾	21.2%

(1) Prairies 25 386 ha (91%), Artificialisé 1 351 ha (5%), Agricole 412 ha (1.5%), Forêt 233 ha (1%)

(2) Prairies 13 229 ha (97.5%), Agricole 121 ha (1%), Forêt 45 ha (0.3%), Artificialisé 12 ha (0.1%)

(3) Prairies 7 512 ha (98.5%), Agricole 74 ha (1%), Forêt 8 ha (0.1%), Artificialisé 3 ha (0.03%)

La Figure 4 ci-dessous présente les surfaces (en cumul) en fonction du nombre de catégorie OSO dans les parcelles. La courbe noire concerne l'ensemble des parcelles, la rouge (respectivement verte) uniquement les parcelles de taille supérieure à un 1/2ha (resp. à 1ha). Cette figure montre que les parcelles avec 4 catégories OSO ou moins couvrent environ 85% - 90% de la surface totale étudiée. En considérant les parcelles avec jusqu'à 3 catégories OSO, le cumul entre les deux régions est significativement différent : il atteint 70% du total pour Hauts-de-France, 80% pour Auvergne (courbe noire)

L'écart entre les courbes vertes, rouges et noires est différent sur les 2 figures, pour les parcelles avec 2,3 ou 4 catégories OSO. Cela peut s'expliquer par une plus grande « hétérogénéité OSO » dans les régions de grandes cultures¹¹.

¹⁰ « Surface étudiée », dans le sens : surface occupée par les *parcelles hors RPG éventuellement agricole* issue de la première étape de notre méthode, avant attribution d'une catégorie de culture ou avant d'être écartées du RPG complété.

¹¹ La présence des parcelles « 100% urbaine » d'après OSO avec plusieurs catégories (urbain dense, diffus, zones industrielle et commerciales, routes) est assez similaire dans les deux régions, surtout pour les parcelles de taille supérieure à un ½ hectare. Il y a sans surprise plus de parcelles de ce type dans les Hauts de France qu'en Auvergne, mais celles-ci sont globalement de petites tailles.

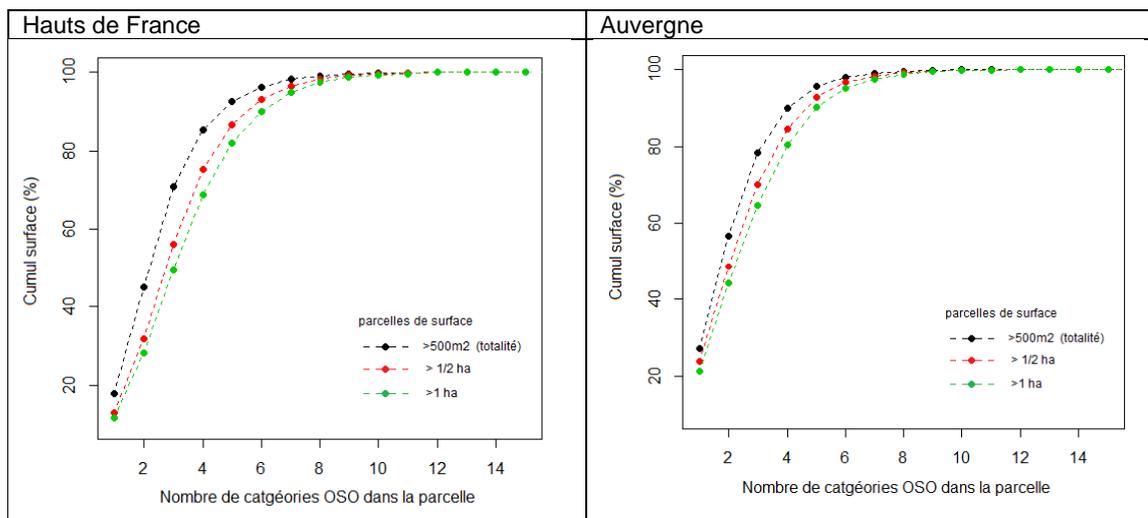


Figure 4 – Cumul de la surface occupée par les parcelles selon le nombre de catégories OSO différentes dans la parcelle (parcelle du RPG complétée avant attribution d’une classe de culture)

- Cas des catégories agricoles de la nomenclature OSO (hors prairies)

Les tableaux suivants concernent ainsi les parcelles que la couche OSO qualifie à 100% d’agricole (c’est à dire l’ensemble des catégories OSO présentes sur la parcelle sont des catégories agricoles, hors prairie).

Région Hauts de France (parcelles 100% agricole d’après OSO)

Taille parcelle	Nb total de parcelles	Nb parcelles avec 1 catégorie OSO	soit en %	Surface totale	Surf parcelles avec 1 catégorie OSO	soit en %
>500	7 219	5 329	73.8%	4 241	2 615 ⁽¹⁾	61.6%
> ½ ha	2 183	1 412	64.7%	3 300	1 904 ⁽²⁾	57.7%
> 1 ha	1 083	644	59.4%	2 526	1 367 ⁽³⁾	54.1%

(1) Céréales : 1 716 ha (66%), Tubercule/racines 463 ha (17%), Mais 242 ha (10%), Colza 157 ha (6%)

(2) Céréales : 1 291 ha (68%), Tubercule/racines 345 ha (18%), Mais 142 ha (7%), Colza 112 ha (6%)

(3) Céréales : 929 ha (68%), Tubercule/racines 244 ha (18%), Colza 90 ha (7%), Mais 78 ha (6%)

Région Auvergne (parcelles 100% agricole d’après OSO)

Taille parcelle	Nb total de parcelles	Nb parcelles avec 1 catégorie OSO	soit en %	Surface totale	Surf parcelles avec 1 catégorie OSO	soit en %
>500	1 175	927	78.9%	362	249 ⁽¹⁾	68.8%
> ½ ha	158	107	67.7%	199	120 ⁽²⁾	60%
> 1 ha	63	37	58.7%	135	74 ⁽³⁾	54.8%

(1) Céréales : 181 ha (73%), Mais 43 ha (17%), Colza 17 ha (8%), Protéagineux 4 ha (2%)

(2) Céréales : 86 ha (72%), Mais 21 ha (17.5%), Colza 10 ha (8%), Protéagineux 2 ha (1.5%)

(3) Céréales : 49 ha (66%), Mais 13 ha (17.5%), Colza 9 ha (12%), Protéagineux 2 ha (3%)

La part des parcelles « mono-OSO » est plus élevée que dans le cas global vu ci-dessus. D’un côté ce résultat n’est pas surprenant (par sélection les parcelles contiennent au plus 8 catégories), mais ces taux supérieurs à 60% pour l’ensemble des parcelles et supérieur à 50% pour les parcelles les plus grandes sont encourageant.

Les catégories les plus présentes dans ces parcelles « mono-OSO 100% Agricole » correspondent sans surprise aux cultures les plus rencontrées sur le terrain, avec quelques notes moins attendues : inversion entre le maïs et du colza sur les grandes parcelles dans les Hauts-de-France, apparition des protéagineux en Auvergne (alors que d’autres cultures sont plus présentes, le tournesol par exemple).

La catégorie OSO « tubercules/racines » correspondra à la fois aux légumes et aux cultures industrielles (en l'occurrence les betteraves) dans les Hauts-de-France.

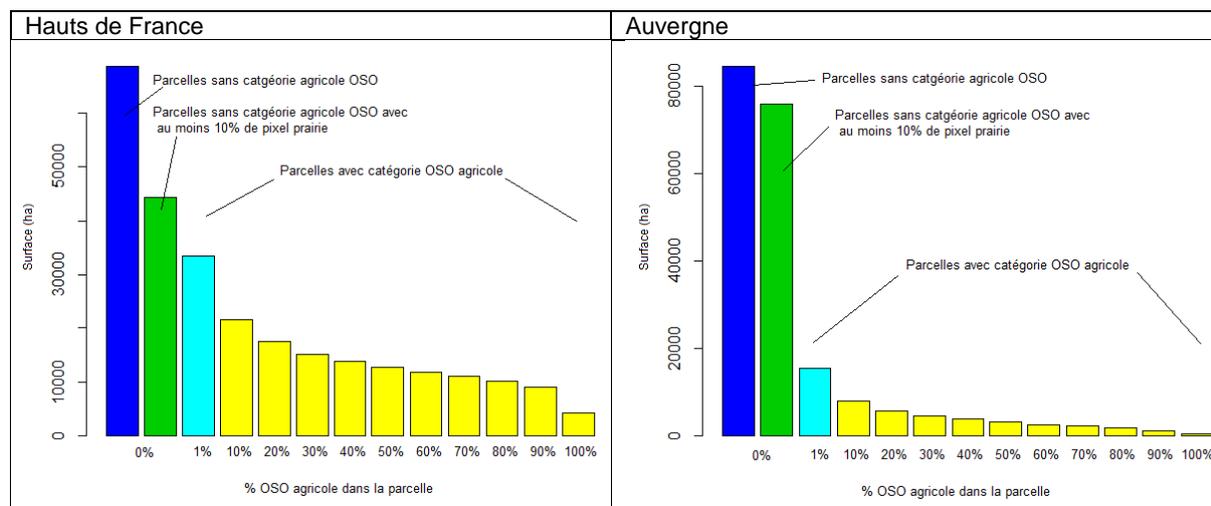


Figure 5 – Surfaces occupées par les parcelles du RPG complété (avant l'attribution d'une classe) selon la présence de catégories OSO agricole (jaune) ou absence de catégories OSO agricole (bleu :total ou vert : avec présence de prairie). (En abscisse : 10% signifie partie agricole \geq 10% et < 20%, etc.)

La Figure 5 résumé la composition en termes de catégories OSO des parcelles ODR avant l'attribution d'une classe (culture ou non agricole). Parmi les parcelles qui ne contiennent aucune catégorie agricole (prairies non comprises), la barre verte (contient au moins 10% de surface en herbe) est incluse dans la barre bleu foncé. La différence de hauteur entre ces 2 barres correspond aux parcelles entièrement artificialisées encore présentes à ce stade. Celles-ci seront écartées (sauf si elles étaient précédemment présentes dans le RPG d'une ou plusieurs campagnes précédentes. Elles seront alors qualifiées comme « abandonnée », voir §4.4)

Cette figure montre que les parcelles ODR contenant uniquement des catégories OSO agricoles sont peu nombreuses (Elles sont toutefois plus fréquentes dans certains départements de grandes cultures comme les céréales ou le maïs). Ceci peut s'expliquer entre autre par la présence de parcelles qui sortent du RPG tout en restant des parcelles agricoles mais potentiellement non exploitées temporairement¹², des parcelles en « phase d'abandon »¹³, de nombreuses parcelles en maraichage absentes dans le RPG (et ces cultures sont difficiles à repérer pour OSO). Ces observations sont à nuancer cependant par des « erreurs » sur les bordures de certaines parcelles dues à la résolution de la carte OSO. Enfin, gardons à l'esprit que les parcelles ODR sont des *parcelles cadastrales* et non - au sens strict - des *parcelles culturales*.

2.4 Conclusions sur la confrontation parcelle agricole et OSO

Ces analyses ont permis de révéler des hétérogénéités régionales ; en terme de composition des parcelles ODR et en terme de qualité de qualification OSO.

Ainsi, nous proposons d'utiliser la modélisation statistique pour attribuer une classe de culture aux parcelles de la couche créée à partir du parcellaire cadastral et contenant ces portions du territoire susceptibles d'accueillir les aires agricoles absentes du RPG. Cette modélisation statistique devra également écartier les parcelles *non agricoles* encore présentes dans notre couche.

¹² Dans ces parcelles, des pixels classés dans des catégories agricoles par OSO sont présents.

¹³ D'anciennes parcelles agricoles en phase de transition vers de l'artificialisation ; « le temps des travaux » l'occupation du sol peut être très particulière (sol nu, zone de stockage de matériaux, sable, etc.) et évoluer rapidement ce qui complique le classement dans les catégories OSO des pixels qui composent ces parcelles.

3. Attribution d'une catégorie de culture aux parcelles ODR, modélisation statistique

Ce travail est traité comme un problème de changement d'échelle, de l'unité *source* le pixel OSO vers l'unité *cible* la parcelle. Lorsqu'une seule catégorie OSO est présente sur la parcelle, l'attribution est immédiate, excepté pour les cas présentés ci-dessus de confusions thématiques ou bien pour des cultures qui n'ont pas une correspondance immédiate dans la nomenclature OSO.

Ce changement d'échelle étant réalisé en utilisant la modélisation statistique, nous présentons successivement la variable à expliquer, les données de calibrage du modèle, les variables explicatives et le modèle statistique utilisé.

3.1 Variable réponse, la variable à expliquer

Les catégories d'occupation du sol ciblées pour qualifier nos parcelles correspondent aux catégories de la nomenclature en 28 postes du RPG, élargie d'une classe « non agricole » (comme dit précédemment il reste encore beaucoup de parcelles situées sur des surfaces artificialisées, ainsi que quelques parcelles sur des zones forestières). Les parcelles qualifiées dans cette dernière classe seront écartées du produit final RPG complété. Dans un premier temps, les catégories *Blé*, *Orge* et *Autres céréales* (gc1, gc3 et gc4) sont regroupées, suivant la nomenclature OSO (catégorie « Céréales à paille »). Dans la catégorie RPG gc7, *Autres oléagineux*, le détail de la nomenclature OSO permet de distinguer le Soja. Dans la catégorie gc28 *divers*, nous retrouverons de nombreuses parcelles classées SNE (surface temporairement non utilisée).

3.2 Données d'acquisition/données de calibrage du modèle

Ces données sont de 3 sortes : des déductions (parcelles qualifiées directement par une catégorie OSO), des parcelles issues d'un échantillon du RPG et des observations.

3.2.1 Les déductions : parcelles décrites avec une seule catégorie OSO

Il s'agit des parcelles ODR décrites dans le paragraphe 2.3.3 ci-dessus. Des parcelles composées d'une seule catégorie OSO sont dites « parcelles déduites » et deviennent des données de calibrage du modèle. Ces déductions suivent les « liens » régionaux entre OSO et le RPG présentés précédemment sous forme de tables de contingence (paragraphe 2.3.1 et 2.3.2).

Par exemple : suivant le Tableau 2 page 8 :

- dans les Hauts-de-France, une parcelle composée de pixels OSO « tournesol » sera considérée comme étant une parcelle en *légume gc25*.
- en Midi-Pyrénées ou en Auvergne, une telle parcelle est considérée comme étant bien du *tournesol gc6*.
- en Rhône-Alpes, une parcelle composée de pixels OSO tubercule sera assimilée à une parcelle soit *légume* soit *culture industrielle*, avant d'être répartie ultérieurement dans l'une de ces 2 classes (voir §4.2 ci-dessous)
- l'ensemble des parcelles composées uniquement - selon OSO - de Céréales, Maïs, Colza, Soja ou Prairie sont affectées dans cette classe.
- les parcelles en prairies d'après OSO sont effectivement affectées à cette catégorie. La distinction entre prairies permanentes et prairies temporaires se fait dans un 2nd temps (voir §4.3 ci-dessous). En revanche, cette façon de procéder ne permet pas de distinguer les prairies artificielles (fourrages herbagers), bien présentes parmi les parcelles en prairie selon OSO (voir les différentes tables de contingence du Tableau 2, par exemple celle pour l'ancienne région Midi-Pyrénées).

Ces déductions sont en outre très importantes pour écarter les parcelles *non agricoles* encore présentes dans notre fond de carte : sols artificialisées (parcelles composées de pixel OSO des catégories 1 à 4), zones arborées *non* qualifiées de verger par la Bd-Topo® de l'IGN (catégorie OSO 16 ou 17) ou sol clairement non agricole (surface minérales, eau, etc. : catégories OSO 20 à 23).

3.2.2 Un échantillon de parcelles du RPG

Afin de mieux caractériser le lien entre, d'un côté, un ensemble de catégories OSO et leurs répartitions (quantitative) dans la parcelle et, de l'autre côté, la culture réellement présente, des parcelles du RPG sont introduites parmi les données de calibrage. Ces parcelles RPG sont indispensables pour repérer et distinguer les types de culture non directement présentes dans la nomenclature OSO. Par exemple les jachères ou les surfaces temporairement non exploitées, mais aussi donc (comme vu plus haut) le maraichage ou les cultures fourragères (hors prairies artificielles), ou d'autres cultures « marginales » (autres oléagineux, plantes à fibres, etc.).

Elles sont également indispensables pour qualifier les parcelles contenant plusieurs catégories OSO.

Un échantillon aléatoire de parcelles du RPG est donc constitué. Le tirage au sort est effectué par Petite Région Agricole (voir ci-dessous §3.3.3), sous contrainte, pour garantir la présence de cultures d'intérêts (jachères, fourrages, légumes, etc.) selon les régions traitées.

3.2.3 Des observations

L'ajout d'observations améliore significativement la performance prédictive du modèle, notamment sur des cultures peu nombreuses dans le RPG mais d'intérêt pour le RPG complété. Ces observations sont donc importantes, mais elles nécessitent des sources d'informations... Consulter des images satellites en est une, cependant les décalages temporaires entre les images disponibles et la campagne désirée du RPG complété, ainsi que l'incertitude entre l'interprétation « visuelle » et les cultures réellement présentes sur des parcelles, limitent l'information utilisable. C'est pourquoi ces observations seront essentiellement des cultures pérennes (vergers, vignes) ou assimilées comme étant pérennes sur période courte. Cela inclut les cultures maraichères, industrielles, prairies permanentes, landes ou des parcelles artificialisées.

En revanche, ces observations permettent de repérer des parcelles qui ne sont pas agricoles : situées sur des terres artificialisées, boisées, des friches, etc. C'est particulièrement intéressant pour les parcelles « mixtes OSO » (composées à la fois de pixels OSO de catégories artificialisées et de catégories agricoles ou prairies) et notamment des bordures de routes ou voies ferrées.

3.3 Les variables explicatives

Notre méthodologie est donc basée sur l'information fournie par la carte OSO, cette information est utilisée en entrée du modèle via un ensemble de variables explicatives. Celle-ci sont complétées par des variables décrivant la parcelle elle-même (taille, forme), des variables relatives à l'occupation du sol à plus large échelle pour décrire l'environnement paysager et territorial de la parcelle (échelle de la Petite Région Agricole) et enfin, dans le même esprit, des variables relatives au contexte agricole mais plus locales, afin de détecter des zones agricoles plus particulières et atypiques.

3.3.1 Variables explicatives OSO

Afin de rapporter sur la parcelle l'information fournie par l'ensemble des pixels et de leur catégorie, les variables suivantes sont utilisées :

- **Variété** : nombre de classe OSO rencontrées dans la parcelle (variable continue)
- **Catégorie majoritaire** : classe la plus fréquente dans la parcelle (variable discrète)
- **Part de des différentes catégories OSO** : *part_urbain* (catégories OSO n.1 à 4), *part_forets* (OSO n.16 et 17), *part_prairies* (OSO n.13) , *part_pelouses* (OSO n.18) et *part_landes* (OSO n.19) et *part_divers* (OSO n.20 à 23), *part_vergers* (OSO n.14), *part_vignes* (OSO n.15)
Les catégories agricoles OSO sont incluses individuellement dans le modèle : *part_colza* (OSO 5), *part_cereales* (OSO 6), ..., *part_mais* (OSO 10) *part_tubercules* (OSO 12).

- **Variété agricole** : nombre de catégorie OSO agricole présentent dans la parcelle (donc parmi les catégories OSO 5 à 12)

3.3.2 Variables explicatives décrivant la topologie de la parcelle

Pour caractériser la « topologie intrinsèque » de la parcelle, la taille et forme de la parcelle sont incluses dans le modèle : **surface** (m²) et **indice de circularité de Miller** (Miller, 1953, Piombini et al. 2014).

3.3.3 Variables explicatives décrivant le « contexte de production agricole »

Les Petites Régions Agricoles (PRA, Agreste¹⁴) discriminent les territoires selon leur vocation agricole dominante pour former des zones homogènes. Rattacher des parcelles à une même PRA permet donc de les rassembler géographiquement suivant une logique de production agricole et de les discriminer des parcelles localisées dans d'autre PRA, dédiées à d'autres activités agricoles. Cependant, le zonage en lui-même n'apporte pas d'information supplémentaire sur ces activités, et cela limite les apports au modèle. Par conséquent, les PRA sont utilisées en entrée du modèle à travers un ensemble de variables construites à cette échelle.

La première variable concerne l'**orientation** technico-économique de la PRA, et celle-ci est estimée à travers les surfaces occupées dans le RPG par les différents types de cultures (SCOP, autres cultures annuelles, prairie, cultures pérennes, voir en annexe). Cette orientation est une variable discrète prenant les valeurs *grande cultures strictes*, *grandes cultures larges*, *mixte*, *élevage*, *cultures pérennes* ou *autre* (voir définition détaillée en annexe)

La deuxième variable est la **part de la SAU** (Surface Agricole Utile) dans la superficie totale de la PRA.

Les autres variables sont les parts occupées dans la SAU par différents grands groupes de cultures d'intérêts pour notre problème : **part des SCOP** (surface céréales oléo-protéagineux), **part des fourrages**, **part des légumes** et **part des cultures industrielles**. Ces taux sont calculés via le RPG et la SAU est la donc « SAU RPG » des exploitations PAC (calcul sur l'ensemble des groupes de culture excepté le gc28 divers).

Dans certaines régions, la part des SCOP est trop corrélée avec l'orientation estimée de la PRA ou avec la part de la SAU. Dans ce cas, elle est retirée du modèle. A contrario, dans d'autres régions cette variable permet d'améliorer sensiblement le modèle.

Les PRA peuvent représenter des zones assez grandes, il n'est pas rare que les départements soient subdivisés uniquement en 5, 6 ou 7 entités seulement.

Décrire, à travers les PRA, le contexte territorial agricole dans lequel se situe la parcelle est instructif mais insuffisant notamment pour détecter des particularismes très locaux et particulièrement les surfaces dédiées au maraichage. C'est pourquoi un second type d'indicateur est utilisé.

3.3.4 Variables explicatives décrivant le « contexte agricole local »

L'une des cultures les plus absentes du RPG est celle des légumes. Or, la couche OSO peine parfois à les repérer (notamment par confusion avec des surfaces artificialisées) et/ou ne permet pas les distinguer d'autres types de cultures, particulièrement des cultures industrielles (voir exemple ci-dessus). Pomme de terre et betteraves se retrouvent notamment (dans les régions septentrionales particulièrement) dans la même classe OSO (tubercules/racines) sur des parcelles qui se ressemblent beaucoup (taille, forme, etc.).

Pour cela, il est nécessaire de connaître le type de zone agricole dans lequel se trouvent les parcelles plus précisément que ce que permet l'approche via les PRA décrites ci-dessus.

¹⁴ PRA référentiel 2017. Voir : <https://agreste.agriculture.gouv.fr/agreste-web/methodon/Z.1/!searchurl/listeTypeMethodon/>

La carte européenne d'occupation des sols Corine Land Cover (CLC, 2009) permet de délimiter des espaces agricoles homogènes, particuliers et isolés dans un territoire (maraichage en zone péri-urbaine, dans de milieux hétérogène avec cultures permanentes...), à condition que ceux-ci dépassent 25ha (échelle minimale de CLC). Une unité cartographique CLC définit géographiquement une zone caractérisée par une même occupation du sol qui se distinguent de ce point de vue des territoires qui l'entourent (les unités CLC voisines). Elles sont de tailles très hétérogènes (minimum 25ha donc) et nombreuses : un département administratif compte en moyenne plus de 3000-4000 unités CLC.

A l'échelle de ses unités cartographiques, des indicateurs sont calculés pour rendre compte de l'activité agricole : **part de la SAU** dans la superficie, **part de légumes**, **part des cultures industrielles**, **part des fourrages**. Ces indicateurs, ressemblant à ceux calculés à l'échelle de la PRA présentés ci-dessus, sont centrées sur ces catégories particulièrement difficiles à repérer ou à distinguer entre elles. Ils sont indispensables pour distinguer 2 catégories pour lesquelles les autres variables explicatives (notamment celle relatives à OSO) sont similaires ou pour repérer les cultures légumières. Ainsi, afin de bien localiser ces zones de maraichages, la **part occupée par les serres agricoles** dans l'aire totale de l'unité CLC sera également une entrée du modèle (cette variable est normalisée). L'information sur les serres est fournie par la Bd-Topo®, couche BATI.

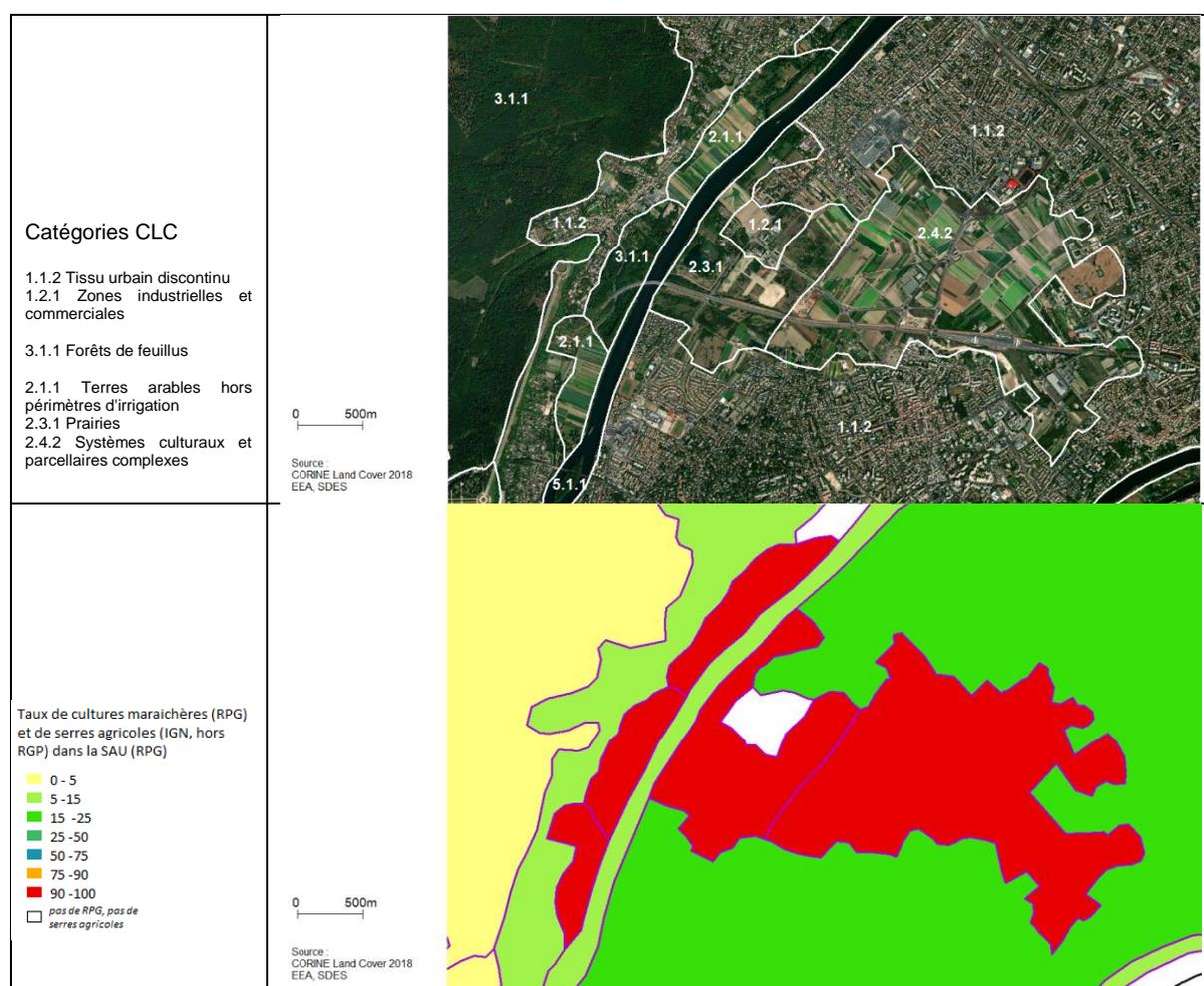


Figure 6 - exemples d'unités cartographiques CLC dans une zone urbaine (département 78) et (bas) indicateur de présence de cultures maraichères (RPG) et de serres agricoles (IGN)

L'occupation du sol dans la nomenclature CLC est détaillée pour les milieux semi-naturels et les zones agricoles hétérogènes (*Cultures annuelles associées aux cultures permanentes, surfaces essentiellement agricoles, interrompues par des espaces naturels importants, systèmes culturaux et parcellaires complexes, ...*). Ces zones sont distinguées des « grands territoires arables » ou des

prairies. La **catégorie CLC (code CLC)** de la parcelle est donc utilisée elle aussi comme variable explicative du modèle. Les modèles perdent en qualité si cette information est ôtée de l'ensemble des variables explicatives.

CLC est un inventaire biophysique de l'occupation des terres. La qualification de l'occupation du sol et les délimitations des unités cartographiques dérivent purement de l'observation satellitaire et sont indépendants de toutes considérations de systèmes de production (les PRA peuvent être vues comme dépendantes d'une fonction socio-économique). La version *CLC 2018* est utilisée.

Ces variables calculées à l'échelle de l'unité cartographique CLC permettent de repérer des zones agricoles plus atypiques dans un territoire voué globalement à d'autres types d'agricultures. Ou encore des zones de maraichages ou de cultures industrielles en zone péri-urbaine, voire entourées de zones urbaines.

3.4 Le modèle

Le but est de prédire une variable à expliquer (variable réponse, ici l'occupation du sol dans les parcelles) en fonction d'un ensemble de variables explicatives.

Les variables explicatives peuvent être quantitatives (numériques) ou qualitatives (ici polytomiques, ordonnables ou non ordonnables).

Le modèle utilisé est donc un modèle de régression logistique multinomiale (la variable à expliquer possède plus 2 modalités). Le principe de la méthode du Logit Multinomial est de relier la probabilité d'apparition des modalités de la variable à expliquer en fonction des variables explicatives.

Nous notons Y la variable réponse (catégorielle) et X le vecteur des variables explicatives. Le modèle Logit multinomial est construit suivant l'hypothèse suivante : la probabilité qu'un individu i (ici : *la parcelle*), compte tenu de ses caractéristiques X_j (les variables explicatives), prenne la catégorie j dépend d'une combinaison linéaire des X_j . Ainsi, cette probabilité d'apparition de la catégorie j (occupation du sol, J catégories possibles) dans la parcelle i s'écrit :

$$P_{i,j} = \frac{\exp(X'_{ij}\beta_{ij})}{\sum_{j=1}^J \exp(X'_{ij}\beta_{ij})} + \varepsilon_{ij}$$

Où les X'_{ij} sont les variables explicatives (associées à l'occupation du sol j dans la parcelle i), ε_{ij} le terme d'erreur et les β_{ij} sont les paramètres à estimer.

Finalement, pour une parcelle donnée, les probabilités $P_{.,j}$ de rencontrer l'occupation du sol j sont directement fournies par les paramètres estimés $\beta_{.,j}$ par le modèle. Ce modèle est calibré sous le logiciel R en utilisant fonction `multinom` issue du package `nnet`, fonction qui cale le modèle via la méthode des réseaux de neurones (Venables and Ripley, 2002).

Lorsqu'il est utilisé en mode prédictifs, c'est-à-dire sur de nouveaux individus (des parcelles qui n'ont pas servi à la calibration), l'intérêt d'utiliser le modèle Logit multinomial réside donc dans la facilité d'interprétation des sorties associées à ces nouveaux individus. Ainsi, pour chaque parcelle en entrée du modèle, le résultat en sortie est la probabilité associée à chaque catégorie d'occupation du sol.

Dans le RPG complété, chaque parcelle se voit attribuer la classe d'option du sol (catégorie culturale ou « non agricole ») qui présente la probabilité estimée la plus élevée, et cette probabilité est également reportée en donnée attributive. L'utilisateur du RPG complété peut ainsi se rendre de la robustesse de la catégorie prédite, éventuellement écarter des parcelles pour lesquelles le modèle n'a pas réussi à distinguer clairement une catégorie particulière (la probabilité maximale n'est alors pas élevée, par exemple inférieure à 0.5 ou 0.6).

Un modèle global ou des modèles territoriaux ?

Utiliser un modèle unique sur l'ensemble du territoire français n'a pas vraiment de sens au vu de la diversité et des spécificités régionales de l'agriculture française. Des unités géographiques cohérentes pour notre problème doivent présenter une cohérence entre une certaine homogénéité

du point de vue agricole et une représentation le plus exhaustive possible des diversités locales. Ces unités de simulations ne doivent pas être trop petites non plus car le cela rendrait l'échantillon de parcelles de calibrage trop restreint et donnerait trop de poids à certaines combinaisons des variables explicatives (en particulier des catégories OSO) tout en n'en omettant d'autres¹⁵.

Des essais ont été réalisés, pour travailler de l'échelle du département à celle de la (nouvelle) région administrative. L'échelle du département est clairement trop petites pour « capter » l'ensemble de la diversité des variables explicatives. La nouvelle région administrative peut être présenter une hétérogénéité agricole/agronomique trop grande qui « gomme » certains particularisme locaux (notamment Nouvelle Aquitaine, Auvergne-Rhône Alpes ou Occitanie). Ainsi nous utilisons un mix entre ancienne région (Midi-Pyrénées, Aquitaine, Auvergne, Rhône-Alpes) et nouvelle région (les autres : Hauts-de-France, Normandie, etc. et celles qui n'ont pas changé).

4. Finalisation, compléments

4.1 Répartition des céréales (à paille)

Dans la nomenclature OSO, les céréales à paille sont regroupées dans une classe unique. Le modèle MNL prédit cette catégorie de culture. Or, dans le RPG, ces céréales apparaissent dans 3 groupes : Blé tendre (gc1), Orge (gc3) et Autres céréales (gc4, Blé dur, avoine, triticale, sorgho, seigle, épeautre, etc.), et nous visons ces mêmes classes dans le RPG complété.

Dans le RPG complété il n'y a donc pas une catégorie « céréales », mais bien les mêmes catégories gc1, gc3 et gc4 du RPG. Les surfaces prédites en céréales sont réparties entre ces 3 groupes en suivant leur répartition relevée dans le RPG à l'échelle des unités CLC. Lorsqu'il n'y a pas assez de données relatives aux céréales à cette échelle, c'est la répartition à l'échelle de la PRA qui est utilisée (ce qui équivaut généralement à donner un poids très important au *Blé tendre*, excepté localement aux *Autres céréales* en région Paca ou Pays de la Loire)¹⁶. La probabilité associée à ces parcelles reste celle fournie par le modèle MNL, c'est à dire celle de la classe « céréales à paille », sans tenir compte de la répartition entre gc1, gc3 et gc4.

La somme des surfaces RPG et des surfaces du RPG complété de ces 3 groupes semblent cohérentes avec les données de la SAA à l'échelle des départements (avec un bémol, toutefois, pour certains territoires de la région Pays de la Loire, où le *Blé tendre* est sans doute surestimé au détriment des *Autres céréales*). En d'autres termes, à cette échelle départementale, cette méthode de répartition permet globalement de retrouver les surfaces absentes du RPG dans les 2 groupes minoritaires gc3 et gc4. Au vue de la qualité d'OSO pour les céréales à paille, cette catégorie est l'une des mieux prédites dans le RPG complété.

4.2 Répartition entre légumes et cultures industrielles

Les légumes et cultures industrielles peuvent se retrouver sous la catégorie OSO n.12 tubercules/racines ; typiquement *pommes de terre* (gc25 du RPG) et *betteraves* (gc24). De même ces deux classes du RPG peuvent se confondre sous des parcelles similaires du point de vue OSO (diversité des catégories, catégories artificialisées). Ainsi, une répartition entre ces deux grands groupes de cultures est effectuée, répartition similaire à celle effectuée pour les céréales (voir paragraphe précédent) à l'échelle des unités CLC, à partir de la répartition observée dans le RPG. Ceci uniquement pour les parcelles de taille supérieur à 5000m² (les plus petites parcelles sont supposées plutôt dédiées au maraichages, mais cela reste une hypothèse à discuter).

¹⁵ Sans évoquer les problèmes de ressources informatiques nécessaires et temps de calculs pour la modélisation statistique, un modèle à trop grande échelle n'est pas vraiment envisageable.

¹⁶ A noter que le *Blé tendre* est largement majoritaire : à l'échelle de la France (métropole) il occupe 64% des surfaces en céréales à paille, l'orge 23% et les autres céréales 13% (selon Agreste, SAA 2018). Cette répartition reste très hétérogène sur le territoire, les *Autres céréales* notamment sont rares dans certains territoires.

Notons que repérer les cultures industrielles du type plantes aromatiques est assez difficile ; ces cultures n'apparaîtront dans le RPG complété presque uniquement lorsqu'elles sont majoritaires ou quasi exclusives parmi les cultures industrielles manquantes dans le département (par exemple dans les Alpes-Maritimes). De nombreuses parcelles de lavande sont repérées par leur présence dans les RPG précédents (2015-2017 pour le RPG complété 2018, ..., 2015-2019 pour le RPG complété 2020, ceci dans le sud-est).

En corolaire : légumes sous serres, jardins collectifs.

- L'information fournie par la Bd-Topo® de l'IGN, couche bâti, permet de repérer les **serres agricoles**, qui peuvent représenter des surfaces non négligeables dans certains territoires. Dans le RPG complété, ces parcelles « serres » (qui font l'objet d'un découpage ad-hoc lors de la phase 1, voir Lardot et al., 2022) sont qualifiées comme cultures légumières sous serre (*Légumes (Serre)*, *gc25*), sauf rares exceptions constatées après observations (serres dédiées à la culture de fleurs, pépinières).
- Des essais ont été réalisés (Hauts-de-France, Ile-de-France) pour tenter de distinguer les cultures de betteraves ou de pomme de terres d'autres types de (respectivement) cultures industrielles ou légumes. Ce type d'estimations restent à valider et vérifier, mais semblent prometteuses, et intéressante au vu de leur importance en terme de surface et compte tenu des pratiques culturelles qui leur sont associées.
- Enfin, des parcelles prédites en culture légumière peuvent correspondre en réalité à des jardins collectifs : jardins ouvriers, jardins partagés, jardin communautaire etc. Souvent rencontrées dans des zones urbaines ou péri-urbaines, ils restent difficiles à détecter, et les essais réalisés en ce sens restent peu fiables à ce stade. Par conséquent, il n'y a pas d'attribution automatique pour ce type de jardins collectifs. Toutefois, pour évaluer ces essais, des observations ont été faites et lorsque la parcelle correspond effectivement à un jardin collectif, cette information a été conservée dans le RPG complété (classe *gc28*, *jardin collectif*)¹⁷.

4.3 Détection du caractère temporaire des prairies

Une prairie est dite permanente s'il s'agit d'une prairie naturelle ou bien si elle est implantée (semée) depuis au moins 6 ans. Dans le cas contraire elle est dite temporaire (incluse dans une rotation culturale).

Les cartes OSO 2016 et 2017 ont été superposées avec le RPG complété 2018. Plus généralement, cartes OSO 2016 à $n-1$ sont superposées avec le RPG complété de l'année n . Ainsi, l'occupation du sol des parcelles lors des années précédentes est connue.

Pour les parcelles du RPG complété *campagne n* prédites en *prairie*, cette information permet donc de savoir si elles ont été cultivées dans les années précédentes. Si c'est le cas, c'est à dire si elles accueillent une des classes agricoles d'après les cartes OSO de 2016 à l'année $n-1$, alors elles sont modifiées et qualifiées de *prairie temporaire*.

Pour les parcelles présentes précédemment dans le RPG, cette information sur le passé des surfaces en prairie peut également être connue en observant la culture déclarée alors, comme l'indique le paragraphe suivant.

¹⁷ Il est important de noter que cette information dans le RPG complété n'est pas du tout exhaustive – outre le fait que nous savons que nous les détectons pas tous – car ces jardins collectifs apparaissent dans le BD-Topo® de l'IGN (couche *zone d'activité ou d'intérêt*) et sont alors écartés du fond parcellaire produit lors de l'étape 1 de production du RPG complété. Ceux présents dans le RPG complété n'apparaissent pas dans la Bd-Topo® (du moins dans la version/mise à jour utilisée).

4.4 Parcelles présentes dans les RPG des campagnes précédentes

Les parcelles présentes dans le RPG au moins une année depuis 2015 mais absentes pour la campagne n ($n = 2018$ à 2020) sont conservées dans le RPG complété *campagne n*, quel que soit la classe finalement prédite. Cette information est rapportée dans la table attributaire décrivant les parcelles. Ce sont les colonnes (pour le RPG complété 2018) : RPG_2017_groupe, RPG 2017_code, RPG_2016_groupe, RPG 2016_code et RPG_2015_groupe, RPG 2015_code.

Cette information sur les RPG précédents permet de modifier la catégorie attribuée par le modèle :

- Lorsque cette catégorie est un des groupes de cultures annuelles du RPG et que la parcelle a été prédite comme étant de la prairie, la parcelle est qualifiée de prairie temporaire (comme ci-dessus). En effet, ces surfaces sont clairement en herbe depuis moins de 5 ans, et peuvent donc être qualifiées de prairies temporaires.

- Lorsque la parcelle a été prédite comme *non agricole/artificialisé*, son absence en 2018 signifie qu'elle n'est plus une surface agricole, ce qu'elle était au moins 1 fois entre 2015 et 2016.

De telles parcelles absentes du RPG déjà en 2017 sont alors qualifiées comme « *Abandon 2017* », « *Abandon 2016* » pour celles présentes uniquement dans le RPG 2015. Pour ces 2 catégories, un nouveau groupe est ajouté à la nomenclature habituelle en 28 postes du RPG : le gc29. Pour les parcelles qui sortent du RPG 2017, le recul temporel n'est pas assez important et les qualifier de terres « agricoles abandonnées » est assez aléatoire. Ces parcelles sont alors qualifiées comme *surfaces temporairement non exploitées* (SNE, gc28)

L'idée est d'évaluer la possibilité d'utiliser ces informations sur les sorties du RPG pour estimer et quantifier l'abandon de terres agricoles, la déprise agricole aux profits de zones artificialisées. Cette idée pourra être toutefois mieux analysée et évaluée avec les productions des futures campagnes du RPG complété.

Plus généralement à propos des prairies, la difficulté principale (OSO montre de très bons résultats pour cette catégorie) reste dans la qualification de l'usage fait de ces surfaces en herbes : certaines peuvent ne pas être destinées à un usage agricole. La première étape de notre méthodologie permet d'écarter efficacement¹⁸ de telles surfaces à usage récréatif (parcs, jardins communaux, châteaux,...) ou sportifs (terrains de sport, parcours de golf, etc.) ou encore correspondantes à des aéroports/aérodromes. Il peut néanmoins subsister dans le RPG complété des surfaces en herbe pas réellement agricoles comme des jardins individuels ou des bordures de routes (plutôt de type autoroutes ou échangeurs) ou de voies ferrées.

5. Cas des vignes et vergers

5.1 Les vignes

Les parcelles ODR qui sont situées sur des *Vignes* de la couche Végétation de la Bd Topo® se voient attribuées directement cette classe de culture, et uniquement celles-ci. A ce stade, la couche OSO ne peut servir à localiser les vignes : trop d'absences et d'erreurs thématiques sur les pixels OSO sont observées. L'information fournie par OSO pourrait aider à « confirmer » la présence de vignes, mais cette information reste difficile à utiliser. Par conséquent, il est difficile de repérer des surfaces temporairement non exploitées (restructuration par exemple) ou abandonnées (à quelques années près, selon la période temporelle couverte par les données de la Bd Topo®).

Le cas des vignes dans le RPG complété sera revu dans les années à venir, afin d'améliorer sa précision temporaire notamment (analyse plus poussée de l'utilisation d'OSO conjointement à l'information de la Bd-Topo®, potentiellement en validant les prévisions grâce au casier viticole informatisé CVI).

Il faut noter que la part des surfaces en vigne absentes du RPG est très hétérogène selon les

¹⁸ Grâce à la Bd-Topo® de l'IGN, complété parfois par CLC (golfs et aérodromes particulièrement).

départements : par exemple (et « grossièrement »), le RPG est presque complet pour les vignes du Gers, recense la moitié des vignobles languedociens (Hérault, Gard), et pratiquement totalement absent sur le vignoble champenois (Marne).

5.2 Les vergers

Les parcelles ODR qui sont situées sur des *Vergers* de la couche *Végétation* de la Bd Topo® font l'objet d'une modélisation indépendante des autres parcelles. Il n'est pas possible de leur attribuer directement une catégorie (comme pour les vignes) car :

- les vergers dans cette couche *Végétation* sont surestimés : ceci est dû à des confusion avec des zones forestières (hors fruits), et aussi à la présence de prairies parsemées de quelques vergers seulement, etc.

- ces vergers peuvent correspondre à différentes catégories du RPG, différents types de cultures fruitières qu'il est nécessaire de distinguer : vergers, fruits à coque, oliviers et également pépinières, châtaigneraies (truffières dans le RPG), ou encore petit fruits rouge.

D'un autre côté, la carte OSO semble sous-estimer les vergers, particulièrement dans les régions du nord ou du nord-ouest. Par conséquent, la prédiction des vergers repose sur un modèle particulier et indépendant de celui dédié à l'ensemble des parcelles. Il s'appliquera sur l'ensemble des parcelles qui accueillent cette classe *Végétation* de la Bd-Topo® et uniquement sur celles-ci.

Un modèle Logit multinomial est à nouveau utilisé. Les variables explicatives sont identiques excepté pour les variables de « contexte agricole » (à l'échelle des PRA et unités cartographiques CLC) où les variables part de culture ou de légumes sont remplacés par les *part des vergers*, *part des fruits à coques*, *part des oliviers* et *part des truffières* (le taux « serre agricole » n'est pas utilisé).

Pour les données de calibrage (voir §3.2), les parcelles présentes précédemment dans le RPG et qualifiées de Verger par la BD-Topo® sont utilisées. L'échantillon RPG est prélevé également parmi les parcelles RPG « en Verger selon la Bd-Topo® ».

Ainsi, les classes ciblées en sortie des modèles suivent les déclarations RPG (présentes ou passées) rencontrées : gc20 (Vergers¹⁹), gc22 (Fruits à coques), gc23 (Oliviers²⁰), ainsi que les pépinières (PEP) et les truffières (TRU) du gc28 (divers) ; et également des catégories qui ne correspondent pas à des cultures fruitières ou arboricoles : prairie (gc18), landes (gc17) ou encore SNE (gc28)

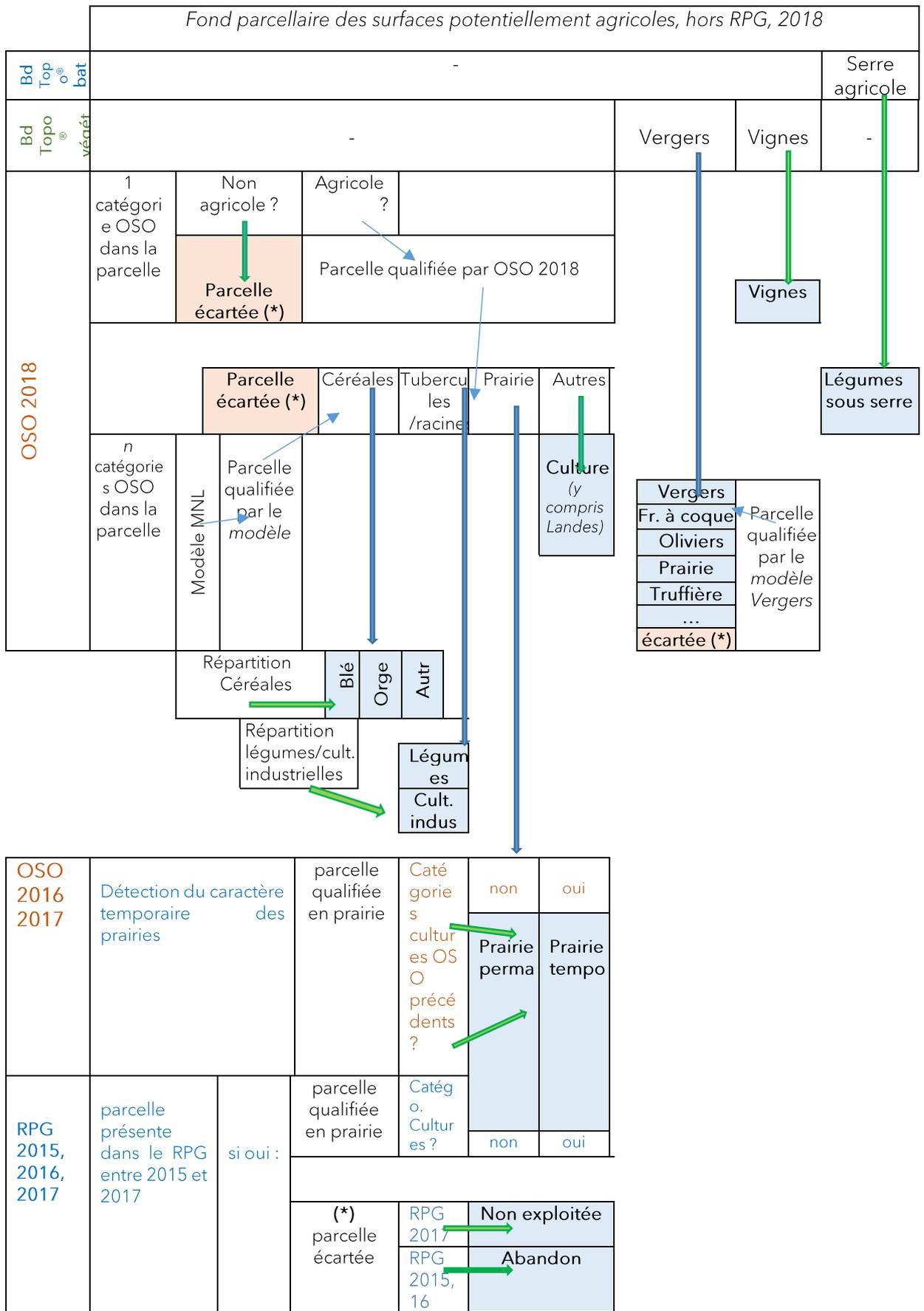
Cependant, au vue du peu de données dans ces catégories pour les 2 types de données de calibrage présentées ci-dessus, des parcelles dites « *déductions* » (voir §3.2.1) sont également utilisées : ce sont les parcelles ODR pour lesquelles OSO nous informe de la présence de vergers (classe OSO 14, même si la parcelle n'est pas entièrement occupée par cette classe). Ces parcelles sont qualifiées de verger (gc20). De plus, les parcelles qui d'après OSO sont entièrement couvertes par des forêts (classe OSO 16 et 17) sont également prises en compte comme « *déductions non agricoles* ».

Néanmoins, ajouter des observations reste le meilleur moyen d'accroître la performance prédictive du modèle. Cela nécessite néanmoins du temps... et surtout des sources d'informations !

¹⁹ La « sous-classe » du gc20 Agrumes est également présente en Corse.

²⁰ Uniquement dans le sud-est et sur le contour méditerranéen (PACA, Languedoc-Roussillon, Corse et département de la Drôme).

6. Résumé de la méthode



7. Produit final

Le RPG complété se présente sous la forme de fichiers SIG au format shapefile® ESRI. Un fichier par département.

Les attributs sont décrits dans la table ci-dessous (ici pour la campagne 2018) :

Champ	Contenu	Remarques
id_parc	Identifiant de la parcelle	
surf_ha	Surface en hectare	
surf_m2	Surface en m ²	
culture	Culture : nom	
code_culture	Culture : code du groupe dans la nomenclature RPG en 28 classes	28 classes + 1 nouvelle classe (n. 29) pour les parcelles « non agricole » (*)
proba	Probabilité (0-1) de la présence de la culture sur la parcelle	Sortie du modèle LM
com_insee	Code Insee de la commune	Référentiel Insee 2018
code_pra	Code de la PRA	
rpg17_gpe	Nom culture RPG 2017	(*) <i>uniquement si parcelle présente dans les campagnes RPG précédentes</i>
rpg17_cult	Code culture RPG 2017	
rpg16_gpe	Nom culture RPG 2016	
rpg16_cult	Code culture RG 2016	
rpg15_gpe	Nom culture RPG 2015	
rpg15_cult	Code culture RG 2015	

Le champ « culture » contient un nom simple et sans ambiguïté (Blé, Autres céréales, Maïs, Prairie, Soja, Betterave, etc.), sans accent. Ce champ fait également parfois référence au code culture du RPG en 3 lettres : SNE pour *surface temporairement non exploitée*, SPH pour *surface pastorale*, légumes CHU pour *choux*, etc.

« *Abandon (2016)* » signifie que la parcelle était présente en (ou jusqu'en) 2015 et disparaît du RPG en (à partir de) 2016 au profit de l'artificialisation. Cette culture est donc classée, dans la nomenclature en 28 postes du RPG, dans la nouvelle catégorie « 29 ».

Le RPG complété contient également des parcelles qui correspondent à des entités particulières comme les *jardins collectifs* (jardins partagés, jardins familiaux), *Bâtiment agricoles divers*, qui sont classées dans le groupe 28.

Les RPG complétés 2018, 2019 et 2020 sont d'ores et déjà mis à disposition et téléchargeables sur l'entrepôt de données (dataverse) de l'unité US ODR d'INRAE, hébergé sur l'entrepôt pluridisciplinaire Recherche Data Gouv.

https://entrepot.recherche.data.gouv.fr/dataverse/odr_sysagri

https://entrepot.recherche.data.gouv.fr/dataverse/rpg_complete_2020

https://entrepot.recherche.data.gouv.fr/dataverse/rpg_complete_2019

https://entrepot.recherche.data.gouv.fr/dataverse/rpg_complete_2018

8. Conclusion et perspectives

La méthodologie présentée dans ce document pour la production du RPG complété permet, avec l'aide précieuse de la carte d'occupation issue de la télédétection OSO (dans sa version en 23 catégories), d'attribuer une catégorie de culture aux parcelles identifiées comme potentiellement accueillant des surfaces agricoles hors RPG.

Le travail principal consiste à passer d'un ensemble de pixels OSO - présentant éventuellement des classes d'occupation de sol différentes - à une catégorie d'occupation du sol unique sur une entité assimilée à une parcelle culturale. La modélisation proposée montre une certaine qualité. Cette qualité dérive de celle de la carte OSO, et les cultures non directement identifiées dans la nomenclature OSO restent les moins bien représentées dans notre carte finale (Légumes, fourrages, ainsi que des cultures moins fréquentes telles que les plantes à fibres, etc.). Un autre enjeu est de proposer un modèle robuste pour repérer les surfaces temporairement non exploitées ou en jachère.

Le cas particulier des vignes pourra être revu pour inclure les informations fournies par OSO pour mieux suivre temporellement leur présence.

Le RPG complété tente également d'assurer un suivi des parcelles qui sortent du RPG, certaines d'entre elles étant liées à la déprise agricole.

Cette méthodologie pourra être appliquée aux futures campagnes du RPG pour proposer annuellement un RPG complété, suivant la mise à disposition du RPG et de la carte OSO par les producteurs de ces 2 couches d'information géographique annuelle.

La production des RPG complété 2018 à 2020 suit exactement cette méthodologie. Pour les campagnes suivantes, il est prévu d'effectuer une mise à jour des modèles, via de nouvelles calibrations notamment en revoyant le mode d'échantillonnage des données RPG (§3.2.2) et l'évaluation des bénéfices potentiels pour la qualité prédictive des modèles apporté par l'utilisation de nouvelles variables explicatives (altitude par exemple). Comme indiqué plus haut, l'estimation des vignes sera également revue. Enfin, suite à des retours d'utilisateurs, d'autres catégories de cultures seront ciblées et distinguées ; par exemple les betteraves et les pommes de terre.

Références

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

Inglada Jordi, Vincent Arthur and Thierion Vincent., 2017a. Theia OSO Land Cover Map 2106 [Data set, Zenodo].

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., & Rodes, I., 2017b. Operational high resolution land cover map production at the country scale using satellite image time series, *Remote Sensing*, 9(1), 95 (2017). <http://dx.doi.org/10.3390/rs9010095>

<http://doi.org/10.5281/zenodo.1048161>

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

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

Lardot Benjamin, Cantelaube Pierre, Carles Marie, Seard Claire et Truche Camille. Construction d'une base de données géographiques exhaustive à échelle fine sur l'occupation agricole du sol : le RPG complété : Partie 1 : Production de la couche géographique des parcelles susceptibles d'accueillir les surfaces agricoles hors RPG. 2022. hal-03647500. <https://hal.inrae.fr/hal-03647500v1>

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)

Piombini Arnaud, Leduc Thomas et Woloszyn Philippe, 2014. Usage de la morphométrie dans la révélation des préférences de mobilité. Application aux cheminements piétons. *Revue Internationale de Géomatique*, Lavoisier, 2014, 24 (1), pp.101-130. 10.3166/ri.24.101-130. hal-01345703.

Miller V.C., 1953. A quantitative geomorphic study of drainage basin characteristics in the Clinch mountain area Virginia and Tennessee. Project NR 389-042, Tech. Report 3. Columbia University, Department of Geology, New York.

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.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.

Annexe

Orientation des Petites Régions Agricoles.

Initialement proposées pour estimer l'orientation des exploitations agricoles du RPG, cette typologie est construite ici à partir des surfaces déclarées dans les PRA, selon leur part dans la SAU :

<i>Orientation</i>	<i>Définition</i>
Élevage	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) \geq 65% de la SAU
Mixte	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) entre 35 et 65% SAU
Cultures Pérennes	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) < 35% SAU <i>et</i> (Vignes+ Vergers+ Arboriculture+ Oliviers) > 25% SAU
Grandes Cultures « Strict »	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) < 10% SAU <i>et</i> (Vignes+ Vergers+ Arboriculture+ Oliviers) < 25% SAU <i>et</i> SCOP > 0
Grandes Cultures « Large »	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) entre 10% et 35% SAU <i>et</i> (Vignes+ Vergers+ Arboriculture+ Oliviers) < 25% SAU <i>et</i> SCOP > 0
Autres	(Fourrage + Prairies Temporaires + Prairies Permanentes + Estives/Landes) entre 10% et 35% SAU <i>et</i> (Vignes+ Vergers+ Arboriculture+ Oliviers) < 25% SAU <i>et</i> SCOP = 0



Pierre Cantelaube (pierre.cantelaube@inrae.fr)

INRAE, Centre Occitanie - Toulouse
Unité de Service de l'Observatoire du Développement Rural (0685)
24 chemin de Bordé Rouge, Auzeville - CS 52627
31326 Castanet-Tolosan Cedex, France

odr.inrae.fr

