

MARS 2023

SANTÉ
ENVIRONNEMENT

MÉTHODE ET REPÈRES

CONSTRUCTION D'UNE
CARTOGRAPHIE À ÉCHELLE FINE
DES CULTURES AGRICOLES
POUR LE TERRITOIRE FRANÇAIS
MÉTROPOLITAIN
PRINCIPES GÉNÉRAUX ET MÉTHODE

Abréviations

Anses	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail
ASP	Agence de services et de paiement
BNVD	Banque nationale des ventes de produits phytopharmaceutiques par les distributeurs agréés
CASD	Centre d'accès sécurisé aux données
CLC	Corine Land Cover (base de données européenne d'occupation biophysique des sols)
CVI	Casier viticole informatisé
DGDDI	Direction générale des douanes et des droits indirects
DGFIP	Direction générale des finances publiques
Draaf	Direction régionale de l'alimentation, de l'agriculture et de la forêt
Dreal	Direction régionale de l'environnement, de l'aménagement et du logement
DROM	Départements et régions d'outre-mer
Epicea	Épidémiologie des cancers de l'enfant et de l'adolescent (Inserm)
IGN	Institut national de l'information géographique et forestière
Insee	Institut national de la statistique et des études économiques
Inserm	Institut national de la santé et de la recherche médicale
IRIS	Ilot regroupé pour l'information statistique
MCE	Matrice Culture Exposition
OCS	Occupation du sol
OCS GE	Occupation du sol à grande échelle
PAC	Politique agricole commune (de l'Union européenne)
Pestiriv	Étude d'exposition aux pesticides chez les riverains de zones viticoles et non-viticoles
RGA ou RA	Recensement (général) agricole
RGE	Référentiel à grande échelle
RPG	Registre parcellaire graphique
SIG	Système d'information géographique
UE	Union européenne
UMR	Unité mixte de recherche
VPN	Valeur prédictive négative
VPP	Valeur prédictive positive
VV	Information sur la présence de vigne/vergers d'après le thème végétation de la BD TOPO® de l'IGN

Résumé

Construction d'une cartographie à échelle fine des cultures agricoles pour le territoire français métropolitain. Principes généraux et méthode

Afin de répondre aux besoins des études épidémiologiques portant sur les liens entre l'exposition environnementale aux pesticides et la survenue de certaines pathologies chroniques, et en l'absence de données quantitatives localisées finement sur les usages de produits phytopharmaceutiques agricoles en France, Santé publique France a développé une méthode de construction d'une cartographie fine des cultures. Cette dernière est homogène sur l'ensemble du territoire métropolitain, évolutive, utilisable à une échelle infra-communale, et elle permet, dès lors que les données commencent à être disponibles, la prise en compte de la latence de survenue de la plupart des pathologies pour lesquelles un lien avec l'exposition aux produits phytopharmaceutiques est suspecté ou documenté.

L'objet de ce rapport est de :

- Décrire précisément l'ensemble des données identifiées et disponibles pour répondre aux objectifs d'une telle cartographie ainsi que les données finalement retenues pour la construire ;
- Détailler la méthodologie de construction de la cartographie (enchaînement des différents traitements de combinaison, nettoyage et raffinement des données) ;
- Exposer l'utilisation de cette cartographie dans le cadre de l'étude épidémiologique Géocap Agri (étude nationale menée par Santé publique France en partenariat avec le Registre national des cancers d'enfant et l'Anses) ;
- Discuter des atouts et limites du recours à cette cartographie de la présence de cultures agricoles, de son apport potentiel aux études épidémiologiques et des évolutions à prévoir dans la perspective d'améliorer la description des expositions environnementales aux produits phytopharmaceutiques utilisés en agriculture.

MOTS-CLEFS : EXPOSITION, PESTICIDES, CULTURES AGRICOLES, SIG, GÉOGRAPHIE, CARTOGRAPHIE

Citation suggérée : Construction d'une cartographie à échelle fine des cultures agricoles pour le territoire français métropolitain. Principes généraux et méthode. Saint-Maurice : Santé publique France, 2022. 70 p. Disponible à partir de l'URL : <https://www.santepubliquefrance.fr>
http://portaildocumentaire.santepubliquefrance.fr/exl-php/vue-consult/spf_internet_recherche/SPF00004114

ISSN : 2647-4816 - ISBN-NET : 979-10-289-0801-0 - RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE - DÉPÔT LÉGAL : MARS 2023

Abstract

Construction of a fine-scale cartography of agricultural crops for metropolitan France. General principles and method

In order to help pesticide exposure assessment in epidemiological studies about health impact of pesticides, and without existing large-scaled located quantitative data on agricultural pesticides uses in France, Santé publique France developed a method to build a precise, homogeneous, scalable and usable at under-municipality level map of the geographic distribution of crops on France metropolitan territory. This map allows, as soon as data is available, to take into account the latency of most of pathologies which are in a suspicious or documented link with pesticide exposure.

The aim of this report is to:

- Precisely describe all of identified and available data susceptible to help to reach these objectives, and the data finally kept to build this map;
- Detail the methodology of map building (processing chain to combine data, cleaning and refinement of the map);
- Present the use of this map in the Geocap AGRI study;
- Discuss about assets and limits of the use of this map of crops presence, about the input of this map for epidemiological needs, and about evolutions to consider in order to better describe agricultural pesticides exposure.

KEY WORDS: EXPOSURE, PESTICIDES, AGRICULTURAL CROPS, GIS, GEOGRAPHY, MAPPING

Collaboration

Santé publique France et Inserm UMR 1153 Epicea (Épidémiologie des cancers de l'enfant et de l'adolescent)

Auteurs

Perrine de Crouy-Chanel, Laurence Guldner (Santé publique France)

Stéphanie Goujon, Matthieu Mancini (Inserm-Epicea)

Remerciements

Joëlle Le Moal, Denis Hemon, Jacqueline Clavel, Marine Sabastia

Relecteurs

Jebraiel Ben Raies, Frédéric Moisan, Laetitia Perrin, Johan Spinosi, Morgane Stempfelet

Date de réalisation : Juillet 2022

Sommaire

Abréviations	2
Résumé	3
Abstract	4
Collaboration.....	5
Auteurs	5
Remerciements.....	5
Relecteurs	5
1. Introduction	8
1.1 Contexte	8
1.2 Objectifs.....	8
2. Sources de données	10
2.1 Présentation des données disponibles	10
2.1.1 Le Recensement agricole	11
2.1.2 Corine Land Cover	13
2.1.3 Le Registre parcellaire graphique	15
2.1.4 Le Casier viticole informatisé (CVI).....	17
2.1.5 Les surfaces en vignes et vergers de la BD TOPO® de l'IGN	18
2.1.6 Les occupations du sol grande échelle	20
3. Sélection des couches à intégrer au SIG.....	23
3.1 Critères de sélection	23
3.2 Application dans le cadre de l'étude Geocap Agri	24
4. Combinaison des sources de données entre elles et construction de la cartographie .25	
4.1 Mise en cohérence des informations sur les types de cultures par le recours à des regroupements de cultures	25
4.2 Combinaison des sources de données dans une couche unique	27
4.3 Superposition avec le maillage administratif.....	30
4.4 Raffinement	30
4.5 Traitement des superpositions, dédoublement	33
4.6 Sélection d'une culture par polygone	35
5. Caractérisation de l'environnement agricole autour des sujets	36
6. Avantages, limites et perspectives	40
6.1 Avantages de la méthode	40
6.1.1 Couverture métropolitaine et méthodologie homogène.....	40
6.1.2 Optimisation de la qualité des sources de données	40
6.1.3 Possibilité de prise en compte de données météo et topographiques.....	40
6.2 Limites	40
6.2.1 Limites et obstacles techniques.....	40
6.2.2 Années d'étude et disponibilité des données	41
6.2.3 Incertitude quant à la qualité de l'information : les limites intrinsèques de chacune des sources de données influent nécessairement sur la cartographie	41

6.2.4 Non disponibilité pour l'outre-mer	42
6.2.5 Évolution des cultures au fil du temps.....	42
6.2.6 Une cartographie qui fournit une caractérisation de l'environnement agricole d'un sujet, et non son exposition environnementale aux pesticides agricoles.....	42
6.2.7 Prise en compte des incertitudes liées à l'imprécision de la géolocalisation des adresses des sujets.....	42
6.2.8 Absence de consensus sur les distances à prendre en compte.....	43
6.3 Perspectives	43
6.3.1 Amélioration de la prise en compte des surfaces non cultivées dans le raffinement.....	43
6.3.2 Utilisation des données d'occupation du sol grande échelle	43
6.3.3 Utilisation des matrices cultures exposition.....	43
6.3.4 Optimisation technique de la construction de la cartographie et du calcul des indicateurs de proximité.....	44
6.3.5 Passage de la cartographie sur une maille décimétrique (50 m).....	44
7. Conclusion.....	45
Annexes	46
Annexe 1. La France, grande puissance agricole, parmi les premiers utilisateurs européens de produits phytosanitaires	46
Annexe 2. Apports des différentes sources de données à la cartographie de présence des cultures : Comparaisons des surfaces agricoles disponibles avec les surfaces de référence du RGA.....	48
Annexe 3. Apport de l'information « Vigne » de la couche Végétation de la BD TOPO® ...	51
Annexe 4. Détail de la sélection des sources dans le cadre de l'étude Geocap Agri.....	55
Annexe 5. Apports et limites du casier viticole informatisé (CVI).....	60
Annexe 6. Traitement des superpositions dans l'étude Géocap Agri.....	66
Annexe 7. Choix de la source principale et donc de la culture principale	67
Annexe 8. Méthode de raffinement des données pour limiter la surestimation des surfaces agricoles	68
Bibliographie	70

1. INTRODUCTION

1.1 Contexte

L'évaluation de l'exposition aux pesticides dans des études épidémiologiques (Dereumeaux *et al.* 2020) pour la recherche de potentiels liens avec des effets sanitaires est un enjeu majeur et l'objet de fortes préoccupations en France, où l'activité agricole et les usages de pesticides restent parmi les plus importants en Europe (Annexe 1). La construction d'un système d'information géographique permettant d'évaluer la présence de certaines cultures en tout point du territoire français vise à répondre à ce besoin épidémiologique, tout en apportant des informations importantes sur l'exposition environnementale de la population générale aux pesticides (Booth *et al.* 2015) ; (Rappazzo *et al.* 2016; Jones *et al.* 2014) ; (Xiang, Nuckols, and Stallones 2000) ; (Ward *et al.* 2000) ; (Kamińska, Ołdak, and Turski 2004).

En effet, des travaux antérieurs de la Direction Santé Travail (DST) de Santé publique France sur les matrices cultures-expositions ont montré qu'il existe une forte corrélation entre l'utilisation de certaines familles de pesticides et certains types de cultures (expertise collective Inserm « [Pesticides et effets sur la santé : Nouvelles données](#) », 2021). L'estimation de la proximité des adresses de lieux de vie aux cultures agricoles, représente ainsi un moyen d'approcher l'exposition environnementale aux pesticides utilisés en agriculture (Teyssiere *et al.* 2020), (Gunier *et al.* 2011), (Ward *et al.* 2006 ; Deziel *et al.* 2015), (Felsot *et al.* 2011).

Plusieurs études écologiques ou cas-témoins avec géolocalisation des sujets reposant sur la caractérisation de l'activité agricole au voisinage des adresses de résidences ont rapporté un lien entre l'exposition des riverains des terres agricoles ou la proximité résidentielle à des zones d'épandages de pesticides (rayon < 1,5 km) et plusieurs pathologies (maladie de Parkinson et comportement évocateur des troubles du spectre autistique chez l'enfant)¹ (Garcia-Perez *et al.* 2016), (Patel *et al.* 2020).

1.2 Objectifs

L'objectif général est de construire une cartographie de la présence d'activités agricoles en tout point du territoire, qui permette :

- De caractériser la présence de cultures sur un maillage du territoire donné. Il s'agit d'une **approche agrégée/écologique**. L'échelle d'étude peut varier en fonction notamment de la prévalence du phénomène de santé étudié. Le niveau d'étude pourra varier de la commune, au canton, voire à des niveaux plus agrégés (département, etc.). Cette cartographie peut être utilisée, par exemple, pour étudier les variations d'incidence de certaines pathologies en fonction de la présence de ces cultures dans différentes zones géographiques.
- De caractériser l'environnement agricole en tout point du territoire en s'appuyant sur des indicateurs basés sur la géolocalisation précise des lieux de vie des sujets. Cette caractérisation est utilisée comme proxy **de l'exposition au lieu de vie de l'individu** : par exemple dans le cadre d'études cas-témoin, de cohorte, de biosurveillance, etc. pour lesquelles on disposerait d'une géolocalisation précise des lieux de résidence des sujets.

Cette méthode a été développée pour répondre aux besoins de l'étude Geocap Agri (Coste *et al.* 2020) sur les cancers de l'enfant et couvre donc la période d'étude de ce projet (2006-

¹ Inserm. Pesticides et effets sur la santé : Nouvelles données. Collection Expertise collective. Montrouge : EDP Sciences, 2021. <https://www.inserm.fr/information-en-sante/expertises-collectives/pesticides-et-sante-nouvelles-donnees-2021>

2013). Elle a néanmoins été construite pour être évolutive et pouvoir répondre aux besoins de futures études en s'adaptant aux données disponibles, en fonction de la période à couvrir dans la mesure où l'on dispose des données et des outils adaptés. Les objectifs opérationnels de cette méthode sont :

- De localiser les terrains (ou groupes de terrains) mis en culture, au sein d'un système d'information géographique, à l'échelle la plus fine possible ;
- D'en estimer les contours et la superficie ;
- Pour chaque point géolocalisé (correspondant au positionnement géographique d'un sujet) encerclé d'une zone tampon ou « buffer » (surface circulaire centrée sur le sujet) de rayon défini, d'affecter à ce « buffer » (et donc au sujet) un indicateur reflétant l'importance de la présence de chaque culture ;
- Pour chaque point géolocalisé, si besoin, d'évaluer la distance à la surface cultivée (quelle qu'elle soit ou d'une culture particulière) la plus proche, en récupérant le cas échéant le type de cultures présent dans la parcelle considérée.

Cette méthode doit :

- Être standardisée, pour permettre une caractérisation identique en tout point du territoire ;
- Être évolutive, car les sources de données disponibles évoluent dans le temps, et que la perspective dans un futur proche de la disponibilité de nouvelles sources doit permettre d'en améliorer la qualité ;
- Permettre de tenir compte de l'évolution historique des paysages et de l'environnement agricoles des territoires, notamment en utilisant des données de différents millésimes, afin de permettre une reconstitution chronologique de la présence d'activités agricoles, par exemple dans le cadre d'une étude épidémiologique exploitant des données sanitaires recueillies sur une période longue, et/ou portant sur une pathologie dont la latence entre l'exposition et la déclaration de la maladie est longue.

2. SOURCES DE DONNÉES

Les spécificités spatio-temporelles de chaque étude imposent une analyse précise des données existantes et de leur disponibilité sur la durée de la période d'étude, afin de sélectionner celles qui seront les plus adaptées. Ces dernières seront le socle de l'indicateur de présence/absence d'activités agricoles et la réflexion méthodologique est logiquement construite autour des possibilités et des limites intrinsèques de ces données.

Il demeure que les bases de données actuelles pour décrire les activités agricoles sont peu nombreuses, et ne couvrent pas toujours l'ensemble du territoire français, qui plus est sur une période longue. Dans un premier temps, il s'agit de présenter ces données en décrivant l'objet initial de leur création, leur temporalité, le territoire couvert et les éventuelles spécificités intrinsèques à chacune de ces sources de données. L'analyse détaillée des atouts et des limites de ces données au regard des objectifs et de la période de l'étude est également réalisée.

Les caractéristiques des données disponibles seront décrites en recourant aux critères suivants :

- **Précision** : finesse de localisation géographique des cultures (parcelle ou îlot de cultures) ;
- **Exhaustivité** : capacité de la source de donnée à décrire de façon exhaustive la présence de cultures en général, et/ou la présence d'une culture en particulier, en recensant l'intégralité des zones de cultures effectivement présentes sur le terrain.
- **Valeur prédictive positive (VPP)** : capacité à prédire la présence certaine de terrains agricoles ou d'un type de culture en un point donné ou dans une zone/surface/parcelle à caractériser (vrais positifs). Cette surface à caractériser peut avoir des contours administratifs (ex : cantons), des contours propres (surface cultivée délimitée géographiquement par une base de données géographiques) ou des contours tracés autour d'un point géolocalisé (zone-tampon, ou « buffer »).
- **Valeur prédictive négative (VPN)** : capacité à prédire la vraie absence de terrains agricoles ou d'un type de culture (vrais négatifs).

Les valeurs prédictives positive et négative sont évaluées de manière qualitative, en argumentant au vu des propriétés des sources de données utilisées. En effet aucune des sources de données permettant de localiser précisément les cultures ne peut constituer un « gold standard ».

2.1 Présentation des données disponibles

Afin d'établir une cartographie des cultures en tout point du territoire, les données doivent pouvoir être disponibles et construites de façon cohérente et homogène sur l'ensemble du territoire métropolitain. L'inventaire des sources de données disponibles a été réalisé dans le contexte de l'étude Geocap Agri et pour couvrir la période d'étude spécifique de ce projet.

2.1.1 Le Recensement agricole

Le Recensement agricole (RA) ou Recensement général agricole (RGA) est une donnée non géographique (mise à disposition uniquement dans un format tabulaire, et non disponibles sous forme de couche de données pour un système d'information géographique - SIG), qui recense de façon exhaustive les activités de toutes les exploitations agricoles du territoire métropolitain, des départements d'outre-mer et des deux collectivités d'outre-mer Saint-Barthélemy et Saint-Martin. Ce recensement est opéré tous les dix ans environ². Il est produit par le ministère chargé de l'agriculture et mis à disposition en ligne en tenant compte du secret statistique, ou après obtention d'une habilitation par le Comité du secret statistique, dans le système sécurisé « CASD³ ». Pour le territoire métropolitain, le recensement a été effectué en 1970, 1979, 1988, 2000, 2010 et 2020 (résultats non disponibles à ce jour). Les résultats du RA 2010 portent sur la campagne de recensement 2009-2010 (cultures et produits de l'exploitation entre octobre 2009 et juillet 2010). Le RA s'inscrit dans le cadre de la réglementation statistique de l'Union européenne. Il décrit non seulement les surfaces de la totalité des cultures pour l'année en cours (plus de 400 variétés), mais également l'orientation technico-économique et la gestion des exploitations.

Les informations du RA sont disponibles à l'échelle communale. De ce fait, cette donnée n'est pas optimale dans un objectif d'utilisation géographique à une échelle très fine. Par ailleurs, la déclaration de toutes les surfaces exploitées à la commune du siège de l'exploitation confère à l'information du RA, si on l'utilise à l'échelon communal, une certaine incertitude quant à la localisation précise des surfaces agricoles déclarées. En effet, le RA lie à la commune du siège de l'exploitation, ou du bâtiment principal, ou de la parcelle agricole la plus importante de l'exploitation⁴, l'ensemble des surfaces déclarées par l'exploitant ou son représentant, que ces dernières soient ou non localisées sur le territoire de cette commune. Un exploitant qui aurait le siège de son exploitation sur la commune A mais qui disposerait de parcelles de cultures sur les communes A, B et C, verra la somme des surfaces de cultures regroupée uniquement sur la commune A. Cela limite donc l'utilisation du RA pour décrire le voisinage proche des lieux de résidence. Cependant, son caractère exhaustif et systématique en fait une donnée de référence à une échelle plus large et notamment celle de la France métropolitaine. En effet, le RA permet de comparer la somme des surfaces agricoles utiles décrites à partir du SIG à une information de référence, en vue de valider celui-ci.

Afin d'utiliser exhaustivement les données de cette base à l'échelle communale, la levée du secret statistique a été obtenue après passage en commission du secret en 2017 et un accès au système de données sécurisées CASD a été possible afin d'extraire des données agrégées susceptibles d'être utiles à la construction et à la validation de la cartographie des surfaces agricoles.

Les données du RA 2010 sur les surfaces cultivées des exploitations fournissent une information précise sur un nombre élevé de cultures distinctes si on se reporte au catalogue des données du RA 2010 pour les déclarations de surfaces cultivées (Tableau 1).

² <https://agreste.agriculture.gouv.fr/agreste-web/methodon/S-RA%202020/methodon/>

³ CASD - Centre d'accès sécurisé aux données <https://www.casd.eu/source/recensement-general-de-lagriculture/>

⁴ L'exploitation agricole est définie, au sens de la statistique agricole, comme une unité économique et de production répondant simultanément à trois conditions. Elle a une activité agricole. Elle atteint ou dépasse une certaine dimension (superficie, nombre d'animaux, production). Elle est soumise à une gestion courante indépendante. Elle est localisée en un lieu unique et tous les résultats la concernant sont rattachés à la commune du siège de l'exploitation agricole. Ce qui peut être source d'un certain décalage entre les données chiffrées issues des recensements agricoles et la localisation réelle des activités agricoles.(source : <https://www.observatoire-des-territoires.gouv.fr/nombre-dexploitations>)

Tableau 1. Détail des cultures dans la base des surfaces cultivées du RA 2010

Variables 2010 (par exploitation/commune/canton)		
Superficie agricole utilisée (1)	Pois protéagineux	Tubercules (hors pommes de terre)
Superficie agricole utilisée hors arbres de Noël	Fève, féverole	_ Igname
Céréales	Lupin doux	_ Madère, dachine
Blé tendre	Plantes à fibres	_ Manioc
_ Blé tendre d'hiver et épeautre	Lin textile	_ Patate douce
_ Blé tendre de printemps	Chanvre y compris papier	_ Autres tubercules
Blé dur	Autres plantes textiles	Légumes secs
_ Blé dur d'hiver	Oléagineux Protéagineux non précisé (DOM)	_ Lentilles, pois chiches, fèves
_ Blé dur de printemps	Cultures industrielles (non compris racines d'endives en 2000)	_ Légumes secs hors lentilles, pois chiches, fèves
Orge et escourgeon	Betterave industrielle	Autres terres arables
_ Orge d'hiver et escourgeon	Houblon	Légumes frais, fraises, melons
_ Orge de printemps	Tabac	Fleurs et plantes ornementales
Avoine	Plantes médicinales, à parfum, aromatiques et condimentaires	Vignes
_ Avoine d'hiver	Semences grainières	Vignes à raisin de cuve
_ Avoine de printemps	Canne à sucre	Vignes à raisin de table
Triticale	Autres cultures industrielles (non compris racines d'endives en 2000)	Pépinières viticoles
Seigle	_ Chicorée à café	Vignes-mères de porte-greffe
Maïs-grain et maïs-semence	_ Racine d'endive	Cultures permanentes entretenues
Sorgho-Grain	_ Autres cult. indust. sauf chicorée ou endive	Vergers 9 espèces
Riz	Fourrages et superficies toujours en herbe	Fruits à noyaux
_ Riz Indica	Maïs fourrage et ensilage	Fruits à pépins
_ Autres riz	Plantes sarclées fourragères	Fruits à coque
Mélanges et autres céréales	Autres fourrages annuels	Agrumes
_ Mélanges d'hiver	_ Légumineuses annuelles	Petits fruits
_ Mélanges et autres cultures de printemps	_ Autres fourrages annuels (hors légumineuses)	Fruits tropicaux
Céréales non précisé (DOM)	Prairies artificielles	Cultures fruitières permanentes (DOM - 2000)
Oléagineux, protéagineux, plantes à fibres (Total)	Prairies temporaires	Pépinières ornementales, fruitières et forestières
Oléagineux	Superficie toujours en herbe (STH)	Cultures à vocation énergétique
Colza	Superficie toujours en herbe productive	Arbres de Noël-2010
_ Colza grain de printemps et navette	Superficie toujours en herbe peu productive	Autres cultures permanentes entretenues
_ Colza grain d'hiver	Pommes de terre et tubercules	Jachères
Tournesol	Pommes de terre primeurs ou nouvelles	Jachères sous contrat (floristique, pollinique et faunistique)
Soja	Pommes de terre demi-saison et conservation	Jachères autres
Lin oléagineux	Plants de pomme de terre	Jardins et vergers familiaux
Autres oléagineux (y compris lin en 2000)	Féculerie	
Protéagineux		

En résumé, on peut caractériser les données du RA en ces termes : exhaustivité et bonne VPP, relativement bonne VPN, mais ce n'est pas une couche de données géographiques (les contours géolocalisés des surfaces en cultures ne sont pas collectés dans le RA) utilisable pour réaliser une cartographie fine, et les informations de localisation disponibles peuvent être imprécises, dans la mesure où des surfaces cultivées seront systématiquement rattachées à la commune/au canton du siège de l'exploitation, alors qu'elles sont potentiellement localisées en réalité dans des communes voisines/cantons voisins.

2.1.2 Corine Land Cover

Le projet *Corine Land Cover (CLC)* est piloté par l'Agence européenne pour l'Environnement dans le cadre du programme européen de surveillance des terres de Copernicus⁵. Le projet vise à la production régulière d'une base de données géographiques d'occupation des terres homogènes sur l'ensemble du territoire européen ; elle est disponible en *open-data*⁶ pour le territoire métropolitain pour les années 1990, 2000, 2006, 2012 et 2018. Les dates de disponibilité varient pour les départements et territoires d'outre-mer, et pour les millésimes les plus anciens, CLC n'est disponible qu'au format raster (un format image) pour l'outre-mer, et non au format vectoriel, ce qui en limite les possibilités d'usage dans un SIG.

Cette donnée est produite par télédétection, c'est-à-dire par interprétation visuelle d'images satellitaires, « avec des données complémentaires en appui »⁷. Elle décrit des zones d'occupation du sol homogène d'au moins 25 hectares (soit l'équivalent d'un carré de 500 m de côté), et est produite à l'échelle du 1/100 000^{ème} (l'échelle de la construction de la base géographique est telle qu'1 cm sur la carte représente 1 km de la réalité du terrain), ce qui n'en fait pas une base de données utilisable seule à une échelle très fine. Par ailleurs, les comparaisons des superficies agricoles de CLC découpées par canton avec le RA à l'échelle du canton montrent une relative adéquation entre les surfaces des deux référentiels : s'il n'y a en effet pas de contradiction majeure entre les deux bases (les cantons identifiés comme agricoles dans CLC ont bien aussi une proportion de surfaces agricoles élevée selon le RA), par contre, elles font apparaître que CLC surestime globalement assez largement la superficie des terres agricoles (cf. Annexe 2). Ce phénomène est dû au manque de finesse de l'échelle de définition de CLC, qui, de ce fait, agrège aux terres agricoles des surfaces non agricoles (notamment des zones de végétation naturelle). CLC ayant tendance à largement surestimer les territoires agricoles, elle ne peut suffire à elle seule pour caractériser ces espaces.

CLC ne décrit pas seulement les territoires agricoles, mais caractérise également les territoires artificialisés, les forêts et milieux naturels, les zones humides et les surfaces en eau. Sa nomenclature décrit 44 postes d'occupation du sol en métropole dont 11 pour les territoires agricoles. Sa description des territoires agricoles (Tableau 2) reste assez généraliste, regroupant par exemple l'ensemble des terres arables dans une même catégorie. En revanche, la vigne, l'arboriculture, les prairies et les rizières disposent chacune d'un poste bien identifié (dont deux pour l'arboriculture).

⁵ « [...] programme d'observation de la Terre de l'Union européenne. Il s'intéresse à notre planète et à son environnement pour le bénéfice de tous les citoyens européens. Il offre des services d'information basés sur l'observation de la Terre par satellite et des données in situ (non spatiales). », <https://www.copernicus.eu/>

⁶ Le terme Open Data désigne des données auxquelles n'importe qui peut accéder, que tout le monde peut utiliser ou partager. Les critères essentiels de l'Open Data sont la disponibilité, la réutilisation et la distribution, et la participation universelle. (Open Knowledge Foundation, 2005)

⁷ <https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>

Tableau 2. Territoires agricoles décrits dans la nomenclature Corine Land Cover

Code occupation du sol	Descriptif
211	Terres arables hors périmètres d'irrigation
212	Périmètres irrigués en permanence
213	Rizières
221	Vignobles
222	Vergers et petits fruits
223	Oliveraies
231	Prairies
241	Cultures annuelles associées aux cultures permanentes
242	Systèmes culturaux et parcellaires complexes
243	Surfaces essentiellement agricoles, interrompues par des espaces naturels importants
244	Territoires agro-forestiers

Telle quelle, CLC ne saurait prétendre à caractériser à elle seule la présence d'activités agricoles de façon très fine en un point donné du territoire. Elle constitue néanmoins une base de données d'occupation du sol incontournable du fait de sa disponibilité, pour plusieurs millésimes, sur l'intégralité du territoire métropolitain. Mais une réflexion s'impose pour pouvoir l'utiliser de façon appropriée dans la mise en place de l'indicateur de proximité aux surfaces agricoles.

En résumé, il est possible de caractériser les données de Corine Land Cover en ces termes : VPP moyenne (la donnée a tendance à surestimer la surface agricole et il est par ailleurs impossible de distinguer les cultures en dehors de quelques cultures facilement repérables par télédétection) : la capacité de CLC à délimiter le fait agricole est correcte, mais elle n'est pas adaptée pour repérer des cultures précises. La VPN est assez bonne. La précision est modérée du fait à la fois de la résolution géographique de la base et du manque de détail dans la caractérisation des surfaces agricoles. L'exhaustivité n'est pas réellement qualifiable : le manque de précision sur les cultures et la résolution géographique assez grossière qui peut entraîner la non détection par CLC de très petites parcelles limitent nécessairement l'exhaustivité de cette source de données. Il y a un risque non négligeable d'erreur de classement sur de très petites parcelles agricoles dont la surface serait inférieure au seuil de 25 ha. Il y a également un risque d'erreur de classement sur des surfaces non agricoles imbriquées dans des zones agricoles, qui, du fait de cette résolution, se retrouvent « agrégées » à des espaces agricoles plus grands.

2.1.3 Le Registre parcellaire graphique

Le *Registre parcellaire graphique* (RPG) est une donnée géographique établie sur l'ensemble de la métropole française, les départements d'outre-mer et les deux collectivités d'outre-mer Saint-Barthélemy et Saint-Martin. Le RPG est la base de données de référence dans l'instruction des aides de la Politique agricole commune (PAC)⁸ créée par l'UE en 1958⁹. À travers un formulaire « TéléPAC », les exploitants déclarent leurs surfaces agricoles et indiquent leur localisation précise sur le territoire en en délimitant les contours graphiquement, ainsi que la culture pratiquée. Ces données sont collectées par l'Agence de services et de paiements (ASP) et disponibles dans une base de données géographiques depuis 2007.

Avant 2010, les exploitants n'étaient tenus de déclarer que les cultures faisant l'objet d'une demande de prime. À partir de 2010, l'ensemble des surfaces cultivées par l'exploitant doivent être déclarées si l'une des cultures fait l'objet d'une demande de prime. La complétude des données RPG est donc meilleure à partir de 2010. Cependant, malgré cette amélioration, le RPG demeure non exhaustif dans la mesure où les exploitations en monoculture pour des cultures non assujetties aux aides de la PAC (en particulier la vigne et l'arboriculture) ne sont pas tenues de déclarer leurs surfaces cultivées via TéléPAC.

Ces bases sont disponibles à l'îlot de cultures pour les millésimes antérieurs à 2015, et à la parcelle à partir de 2015 ; elles sont mises à jour annuellement dans la mesure où les aides de la PAC sont reversées chaque année aux exploitants. Avant 2015, seule la culture majoritaire d'un groupe de parcelles est assignée à un îlot de cultures (plusieurs parcelles, mais une description uniquement de la culture majoritaire). Le renseignement des îlots de cultures se fait au 1/5 000^{ème} : l'exploitant trace les contours de ses parcelles ou îlots de cultures en s'appuyant sur des orthophotographies¹⁰ mises à sa disposition dans l'outil TéléPAC¹¹.

À partir de 2015, l'information géolocalisée est disponible à la parcelle, ce qui évite le biais qui existait jusque-là avec la déclaration de la culture majoritaire de l'îlot assignée à toutes les parcelles qu'il regroupe. Le RPG est donc moins précis avant 2015 qu'à partir de ce millésime et pour les suivants. Il existe en complément des données du RPG avant 2015 une base non géographique (fichier RPG classe 4 anonymisé) qui décrit la proportion de chaque culture disponible pour un îlot de parcelles identifié. Par exemple, si un îlot de cultures composé de plusieurs parcelles simples est décrit comme étant un îlot à majorité de vigne, ce fichier complémentaire précisera qu'il y a, en réalité, 70% de vigne, mais aussi 30% de maïs. Il est en revanche impossible de connaître, à partir de ce fichier complémentaire, la localisation précise de ces deux cultures sur l'îlot en question. Par ailleurs cette base complémentaire n'est pas disponible simplement auprès de l'IGN, l'accès à ces données nécessite une demande spécifique auprès de l'Agence de services et de paiement (ASP).

La nomenclature du RPG avant 2015 décrit 28 classes de cultures (Tableau 3).

⁸ https://opera-connaissances.chambres-agriculture.fr/doc_num.php?explnum_id=147633

⁹ <http://www.europarl.europa.eu/factsheets/fr/sheet/103/la-politique-agricole-commune-pac-et-le-traite>

¹⁰ Les orthophotographies sont des images aériennes ou satellitaires de la surface terrestre rectifiées géométriquement et égalisées radiométriquement. Ces images se présentant sous forme de dalles couvrant une zone de la Terre peuvent être géoréférencées dans n'importe quel système de coordonnées.

¹¹ <https://www.casd.eu/source/registre-parcellaire-graphique/>

Tableau 3. Cultures majoritaires décrites dans la nomenclature du RPG entre 2007 et 2014

Code	CULTURE MAJORITAIRE	Code	CULTURE MAJORITAIRE	Code	CULTURE MAJORITAIRE	Code	CULTURE MAJORITAIRE
0	ABSENCE D'INFORMATION						
1	BLÉ TENDRE	8	PROTÉAGINEUX	15	LÉGUMINEUSES À GRAINS	22	FRUITS À COQUE
2	MAÏS GRAIN ET ENSILAGE	9	PLANTES A FIBRES	16	FOURRAGE	23	OLIVIERS
3	ORGE	10	SEMENCES	17	ESTIVES LANDES	24	AUTRES CULTURES INDUSTRIELLES
4	AUTRES CÉRÉALES	11	GEL (SURFACES GELÉES SANS PRODUCTION)	18	PRAIRIES PERMANENTES	25	LÉGUMES-FLEURS
5	COLZA	12	GEL INDUSTRIEL	19	PRAIRIES TEMPORAIRES	26	CANNE A SUCRE
6	TOURNESOL	13	AUTRES GELS	20	VERGERS	27	ARBORICULTURE
7	AUTRES OLÉAGINEUX	14	RIZ	21	VIGNES	28	DIVERS

À partir de 2015 la nomenclature évolue en plus de 250 variétés, la nomenclature en 28 groupes est conservée en parallèle.

La précision du RPG est proche de celle du Référentiel à grande échelle¹² sur lequel il se base et elle est proche également de celle de la BD TOPO de l'IGN. Cette précision en fait un atout indéniable pour l'utilisation à fine échelle. Cependant, jusqu'à 2015, il décrit les surfaces agricoles de façon imparfaite puisque seule la culture majoritaire de l'îlot est renseignée ; de plus, les cultures non assujetties aux aides de la PAC ne sont pas toutes déclarées dans le RPG, ce qui entraîne un défaut d'exhaustivité de cette source d'information, notamment pour les vignes et les vergers qui ne font pas l'objet d'aides de l'Union européenne (cf. Annexe 2).

En résumé, il est possible de caractériser le RPG par une bonne VPP quant à la délimitation des zones cultivées (même si le système de recueil des déclarations peut entraîner d'éventuelles sur-déclarations des surfaces cultivées par certains exploitants). Cette VPP est moins bonne quand on s'intéresse aux cultures individuellement : certes, le niveau de détail des cultures/groupes de cultures est fin, cependant, jusqu'en 2015, le regroupement des parcelles en îlots de cultures sur lesquels la culture majoritaire est seule à être déclarée entraîne une certaine possibilité d'erreur dans la caractérisation des cultures. Le RPG est doté d'une très bonne VPN, qui devient moins bonne si on considère les cultures non assujetties aux aides telles que vignes et vergers en particulier, une exhaustivité très correcte (sauf pour les cultures non assujetties aux aides puisque les exploitations en monoculture vignes et vergers n'apparaissent pas dans le RPG).

¹² L'État a confié à l'IGN, par décret, le 22 novembre 2004, le développement et la mise à jour du Référentiel à grande échelle (RGE). Le RGE regroupe plusieurs bases de données géographiques de précision métrique, qui relèvent de quatre grandes composantes (1. topographique – BD TOPO, 2. adresse – BD Adresse, 3. parcellaire – BD Parcellaire, et 4. orthophotographique – BD ORTHO). Les données du RGE sont accessibles sur la plateforme de téléchargement de l'IGN : <https://geoservices.ign.fr/>

2.1.4 Le Casier viticole informatisé (CVI)

Le Casier viticole informatisé (CVI) est une base de données détenue par la DGDDI (Direction générale des douanes et droits indirects), qui participe du corpus d'outils des États membres de l'UE pour contrôler la production viticole. Il couvre l'ensemble du territoire métropolitain pour la vigne. Basé sur les déclarations de productions des viticulteurs produisant du raisin de cuve à destination de vinification, il renseigne pour chaque parcelle rattachée à une exploitation viticole la présence ou l'absence de vigne, la variété et le label auquel il est affilié. Il contient également les informations nécessaires au traçage des mesures d'arrachage et aux replantations, opérations très contrôlées soumises à un suivi par les autorités. Censée garantir la traçabilité du vin, cette base se veut très précise mais n'est pas sans défaut. Ainsi, les données sur la production de vins de Champagne ne sont pas exhaustives dans cette base, et le CVI n'inclut pas les surfaces de vignes destinées aux raisins de table (0,78% des surfaces de vignes en France, d'après le RGA - Agreste 2010). L'accès à cette base est conditionné à l'obtention d'un accord du Comité du secret statistique dans la mesure où le cadastre constitue une information indirectement nominative.

Ces données sont disponibles sous forme de table, et non directement dans une base géographique. Afin d'obtenir une information géographique à partir du CVI, il est nécessaire de réaliser une jointure entre les tables du casier viticole informatisé et les parcelles du cadastre géoréférencé, par le biais des identifiants cadastraux des parcelles renseignés dans le CVI. Ces traitements sont rendus complexes par l'évolution continue du cadastre. Avant la mise à disposition des données du parcellaire sous forme de cartographie vectorielle utilisable dans un SIG, il est difficile à utiliser. La BD Parcellaire de l'IGN (base de données vectorielles du parcellaire fournie par l'IGN) n'a été mise à disposition dans le cadre de l'ouverture des données du RGE qu'à partir de 2011 pour les organismes publics, et l'ensemble des départements n'est alors pas encore couvert par le format vectoriel. Il faut attendre la mise à disposition du plan cadastral informatisé (à partir de 2018), construit par l'IGN à partir des données du plan cadastral actualisées trimestriellement par la DGFIP (Direction générale des finances publiques), pour disposer d'une information vectorielle sur l'ensemble du territoire et pleinement compatible avec l'appariement au CVI. Le CVI est par ailleurs remis en question actuellement par l'Union européenne dans sa forme actuelle. Néanmoins, dans l'idée de construire une cartographie évolutive, qui s'adapte à l'évolution des données disponibles pour décrire la présence de cultures en tout point du territoire, il convient de le retenir comme une source de données essentielle pour des cartographies devant répondre à des problématiques récentes.

Une étude de sensibilité a été réalisée avec une extraction du CVI sur le vignoble du Beaujolais, afin d'évaluer la part des surfaces viticoles décrites dans le CVI, et non renseignées par le RPG ou *Corine Land Cover* (cf. 2.2.1) afin d'approfondir les apports potentiels de cette base de données au regard des autres données géographiques disponibles (cf. Annexe 5). Elle montre que sur ce territoire, les bases de données disponibles sur la période pour laquelle la cartographie a été construite, détectent bien les parcelles en vigne, les surfaces que le CVI est le seul à identifier étant très réduites. La date de disponibilité de cette base de données, très postérieure aux dates pour lesquelles la cartographie est construite, remet en cause son utilisation ici.

En résumé, le casier viticole informatisé constitue une base de données partielle (elle ne concerne que la production de raisin de cuve), dont la VPP est très bonne, du fait de sa précision géographique et de son caractère déclaratif obligatoire. Sa VPN est également très bonne (sauf quand il s'agit des vins de Champagne ou du raisin de table qui ne sont pas complètement renseignés dans la base). Son niveau d'exhaustivité est bon si on met à part les vins de Champagne. La précision géographique est très bonne, le niveau de détail correspondant au découpage cadastral étant, par nature, très fin.

2.1.5 Les surfaces en vignes et vergers de la BD TOPO® de l'IGN

La BD TOPO® de l'IGN est la composante topographique du RGE. Elle est construite à partir de l'interprétation de photos aériennes ortho-rectifiées (redressées et projetées pour pouvoir être assemblées et lues dans un SIG), complétée par des observations sur le terrain. La BD TOPO® comporte de très nombreuses couches aux thématiques d'occupation du sol variées (bâtiments, infrastructures routières, surfaces en eau, végétation, etc.). Malheureusement, cette base de données ne comporte pas de thématique « agriculture », ce qui ne permet pas de l'utiliser seule pour la construction d'un indicateur géographique de présence d'activités agricoles.

L'information sur la présence de vignes ou de vergers est mentionnée dans la couche décrivant la végétation générale (forêts, bois, haies, vigne, vergers, landes ligneuses et peupleraies) en France métropolitaine, ainsi que dans les départements et régions d'outre-mer. À partir de cette base de végétation, les vignes et les vergers peuvent être extraits.

Cette information sur la présence de vignes et de vergers a été progressivement ajoutée à la couche « végétation » entre 2005 (pour les tout premiers départements) et 2015, sur la base d'interprétation de photos aériennes complémentaires, du CVI, du RPG et d'observations sur le terrain, sans réactualisation depuis. Ces informations sont ensuite mises en cohérence avec le reste de la végétation décrite dans la donnée zone de végétation (les bois, les haies, les forêts, etc.)¹³. Les unités géographiques correspondant aux vignes figurant dans la base ont une superficie minimum de 2 000 m² (soit 0,2 hectare). La méthodologie de création de la donnée « vergers » n'est pas décrite avec précision par l'IGN. Elle comprend les pépinières non forestières, les vergers à fruits y compris les vergers bas (petits fruits), les oliveraies, les châtaigniers à fruits et les noyeraies à fruits. Les roseraies sont elles aussi intégrées à cette source de données. Les unités géographiques correspondant aux vergers ont une superficie minimale de 5 000 m² (soit 0,5 hectares)¹⁴. L'inclusion des roseraies et des pépinières non forestières invite à une certaine prudence vis-à-vis de cette information sur les vergers, issue de la BD TOPO®.

Les données concernant la vigne et les vergers de la couche « végétation » de la BD TOPO® ont été progressivement mises à disposition à partir de 2007 pour les premiers départements, mais n'ont pas concerné immédiatement la totalité du territoire, et n'ont pas été actualisées depuis leur intégration. Avant 2017, les données sont parcellaires et ne recouvrent pas l'ensemble du territoire. Seules les données à partir de 2017 sont exhaustives. La date de collecte des données est connue pour chaque département, mais il n'existe pas un millésime unique pour ces informations. La BD TOPO® de l'IGN est en effet mise à jour au fil de l'eau et ces précisions y sont également ajoutées au fur et à mesure. Il en résulte que les données de 2017 ne représentent pas de manière certaine l'occupation du territoire cette année-là, mais une accumulation de données collectées sur plusieurs années. Les cultures de vigne et d'arbres fruitiers sont cependant des cultures dites pérennes et les variations d'une année à l'autre sont assez faibles. Mais si une vigne a été repérée et intégrée à la base en 2013, puis arrachée en 2014, il est possible qu'elle apparaisse encore dans la base en 2017. Il est toutefois probable que les surfaces concernées par ces changements soient relativement faibles.

Par ailleurs, les données fournies par la BD TOPO® étant très fines, elles ne sont pas mises à jour à une fréquence très élevée. Mais elles le sont par contre en continu : il n'existe donc pas de couche BD TOPO® de la végétation qui puisse être considérée comme certaine à une date précise sur l'ensemble du territoire français. L'intégration entre 2007 et 2017 de ces informations géographiques sur la présence de vignes et de vergers dans la BD TOPO® en

¹³https://geoservices.ign.fr/ressources_documentaires/Espace_documentaire/BASES_VECTORIELLES/BDTOPO/DC_BDTOPO_3-0.pdf

¹⁴https://geoservices.ign.fr/ressources_documentaires/Espace_documentaire/BASES_VECTORIELLES/BDTOPO/DC_BDTOPO_3-0.pdf

font une donnée intéressante pour la fin de la période d'étude de Geocap Agri, dans la mesure où elle apparaît comme plus complète que la même information issue du RPG. Les vignes et les vergers étant des cultures dites « pérennes », dont la mise en place prend plusieurs années, on peut avoir recours à cette information pour la fin de la période d'étude, tout en gardant à l'esprit les précautions d'usage liées à cette incertitude sur la date de l'information. Une étude détaillée sur la faisabilité de l'utilisation de la BD TOPO® « vignes et vergers » sur cette période est disponible en Annexe 3.

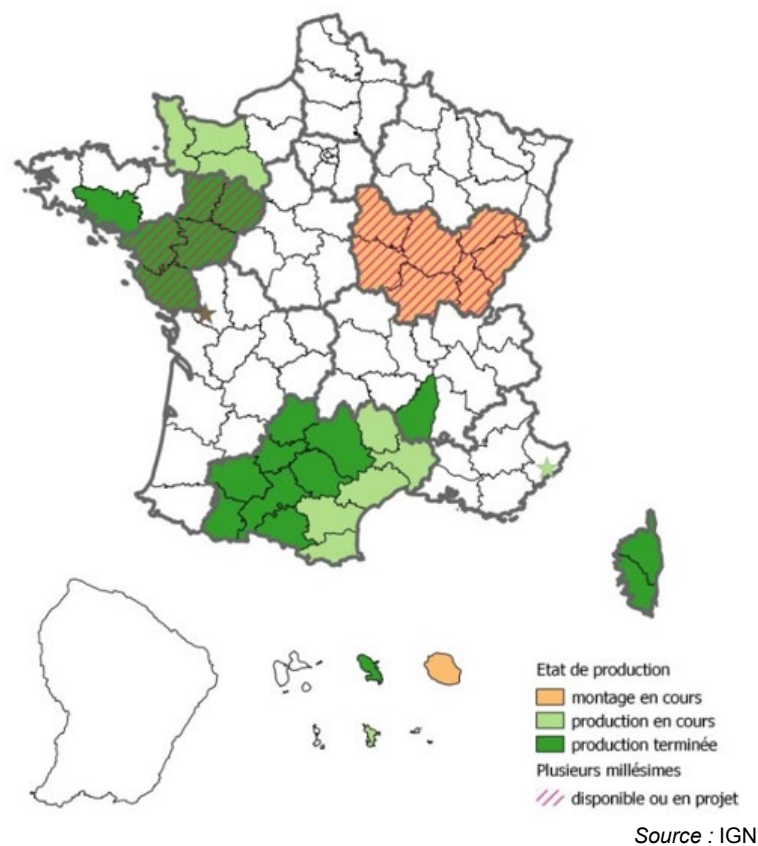
En résumé, la BD TOPO® est une base géographique précise par construction. Sa relative exhaustivité pour les informations concernant la vigne et les vergers en font une source d'information intéressante et sensible pour la fin de la période d'étude : la VPP de la BD TOPO® Vignes et Vergers est bonne pour ces deux groupes de cultures. Elle ne permet cependant pas d'identifier de façon spécifique les autres cultures et l'inclusion dans la catégorie « vergers » de surfaces exploitées en pépinières non forestières ou en plantations d'ornement invite à une relative prudence. La VPN de la BD TOPO® Vignes et Vergers est assez bonne, même si la catégorie « Vergers » recouvre une assez grande variété de surfaces plantées.

2.1.6 Les occupations du sol grande échelle

Afin de répondre à des problématiques locales, les collectivités territoriales ont, depuis une quinzaine d'années, œuvré régionalement pour la production de couches d'information géographique d'occupation du sol à très fine échelle (haute précision). Produites initialement par les collectivités elles-mêmes, ou sous-traitées à des opérateurs privés selon des cahiers des charges et des spécifications assez hétérogènes, ces données régionales ou départementales d'occupation du sol à grande échelle, également dénommées OCS GE, sont depuis quelques années produites par l'IGN, après un travail important de synthèse et d'harmonisation des spécifications et des nomenclatures.

Cependant, encore aujourd'hui, ces données ne couvrent pas l'ensemble de la France. La Figure 1, extraite de la documentation d'accompagnement des données de l'OCS GE, illustre les données disponibles au premier semestre 2019. À ce jour, la région Nouvelle-Aquitaine est désormais couverte.

Figure 1. État d'avancement OCS GE en France au premier semestre 2019



De plus, les années de production ne sont pas les mêmes d'un territoire à l'autre. En effet, les millésimes s'étalent de 2002 à 2016. Les OCS GE sont construites à partir de plusieurs bases issues de l'IGN dont la BD TOPO® pour les éléments dits « structurants » (bâti, réseaux routiers, ferrés et hydrographiques), la BD forêt (issue des données géographiques de l'Inventaire forestier national) pour les zones arborées et le RPG pour les cultures¹⁵. Les informations sont ensuite validées et complétées à partir de photo-interprétations des ortho-photographies de l'IGN. Ces données distinguent 14 postes pour la couverture du sol et 17 pour l'usage du sol. C'est à l'intérieur de cette classification que le poste « agriculture » est disponible.

¹⁵ <https://artificialisation.biodiversitetousvivants.fr/bases-donnees/ocs-ge>

Certaines OCS comme l'OCS régionale des espaces naturels agricoles forestiers et urbains (NAFU) de la plateforme d'échange de données en Nouvelle-Aquitaine (PIGMA), sont plus précises que d'autres (Tableau 4), et ont été produites pour davantage de millésimes que d'autres. Cette OCS présente une nomenclature en 60 postes basée sur les mêmes catégorisations que Corine Land Cover, dont 8 pour les territoires agricoles, avec une représentation isolée des vignobles, des vergers, des cultures annuelles, des cultures florales et légumières, des prairies, des terres arables irriguées, des surfaces en herbes non agricoles et des espaces agricoles en friches¹⁶. La couverture complète de la région Nouvelle-Aquitaine par cette base est effective depuis 2020.

Tableau 4. Millésimes des OCS à grande échelle disponibles pour la région Nouvelle-Aquitaine

Départements concernés	Millésimes OCS Nouvelle-Aquitaine (PIGMA)
24-Dordogne	2000, 2009, 2015, 2020
33-Gironde	2000, 2009, 2015, 2020
40-Landes	2000, 2009, 2015, 2020
47-Lot-et-Garonne	2000, 2009, 2015, 2020
64-Pyrénées-Atlantiques	2000, 2009, 2015, 2020
16-Charente	2009, 2015, 2020, 2020
17-Charente-Maritime	2009, 2015, 2020
19-Corrèze	2009, 2015, 2020
23-Creuse	2009, 2015, 2020
79-Deux-Sèvres	2009, 2015, 2020
86-Vienne	2009, 2015, 2020
87-Haute-Vienne	2009, 2015, 2020

Certaines données peuvent également exister à des échelles plus locales et sur des territoires plus restreints, telle qu'une commune ou une zone de la commune, un parc naturel régional, etc. Ces données produites par les Dreal (directions régionales de l'environnement, de l'aménagement et du logement), les chambres d'agriculture, divers syndicats ou collectivités, etc., sont multiples dans leur construction, leurs échelles, leurs précisions et leur fiabilité.

La complexité de l'accès à ces données d'occupation du sol aux cahiers des charges spécifiques est aussi hétérogène que la qualité de ces données.

L'avenir promet d'autres possibilités en matière d'occupation du sol à échelle fine, puisqu'aujourd'hui, l'IGN, certains instituts de recherche voire des entreprises privées développent des outils pour analyser de façon automatisée les images issues de prises de vues aériennes ou satellites afin d'en extraire des données géolocalisées sur l'occupation du sol. Un des objectifs de cette automatisation serait de pouvoir reconstituer des informations d'occupation du sol à échelle fine sur des périodes anciennes, à partir des photos aériennes disponibles. Effectivement, en l'état actuel des connaissances, il n'existe pas de données d'occupation du sol géolocalisées à une échelle plus fine que la commune, antérieure à 1990. Or, les problématiques des études de santé publique en particulier nécessitent souvent de pouvoir remonter sur des périodes antérieures à cette date.

¹⁶https://www.pigma.org/geonetwork/srv/api/records/8e078aae-0bf4-49fb-a49a-b48b41b7a859/attachments/Nomenclature_ocs_indicateurs_nafu.pdf

En résumé, les OCS GE constituent une donnée d'avenir en matière de description de la présence d'activités agricoles en tout point du territoire et à une échelle suffisamment fine pour décrire cette présence autour des lieux de vie des sujets des études épidémiologiques. Cependant, la grande hétérogénéité de ces données (date, spécification, construction) et leur indisponibilité, pour le moment, à l'échelle de la totalité du territoire national en font une donnée inutilisable à ce stade pour répondre aux objectifs des études nationales. Par ailleurs, l'absence de détail sur les types de cultures (regroupement de toutes les cultures annuelles en terres arables par exemple) hormis vignes et vergers, en limite également la précision.

3. SÉLECTION DES COUCHES À INTÉGRER AU SIG

3.1 Critères de sélection

À partir des données présentées précédemment, il est possible de déterminer, pour une étude donnée, quelles sont celles à privilégier pour la construction du SIG.

Certains critères de sélection sont imposés par le design et les caractéristiques de l'étude :

- L'emprise géographique : l'étude Geocap Agri est basée sur des données de cancers pédiatriques issues d'un registre couvrant le territoire de la France métropolitaine. Il est donc impératif de travailler sur toute la métropole pour pouvoir conserver la puissance statistique du jeu de données sanitaires lors de l'analyse des résultats. Cela impose de ne conserver que les données sources dont l'emprise géographique est cohérente avec cette condition ;
- L'amplitude temporelle : les données sanitaires étudiées résultent d'un recueil d'informations sur les cas de cancers pédiatriques, et leurs témoins sélectionnés par l'Insee, s'étalant de 2002 à 2013 ; toutes les sources de données existantes ne seront pas adaptées pour travailler sur cette temporalité ;
- La précision attendue : cette étude vise à connaître l'environnement agricole à proximité des lieux de résidence des sujets de l'étude Geocap Agri, et l'analyse de la présence des cultures est réalisée à une échelle assez fine, en cohérence avec les hypothèses qui peuvent exister sur l'exposition aux pesticides des riverains des cultures. Le SIG doit donc pouvoir être mobilisé à un niveau de précision élevé, or certaines données existantes ne disposent pas d'une échelle de précision suffisante pour être exploitables dans ce cadre.

Ces trois critères déterminent le choix des bases de données utilisées pour la construction de la cartographie des surfaces agricoles, et orientent la méthode mise en œuvre pour les combiner dans le SIG afin d'obtenir une information exploitable pour caractériser la présence de cultures à proximité des lieux de vie des sujets.

3.2 Application dans le cadre de l'étude Geocap Agri

Pour répondre à ces différents critères d'homogénéité géographique et de contrainte temporelle, les données retenues pour contribuer à la construction de la cartographie des cultures pour l'étude Geocap Agri sont :

- Le Recensement général agricole
- Le Registre parcellaire graphique
- Corine Land Cover
- La sélection des classes vignes et vergers (VV) et la BD TOPO® « Végétation » pour les dernières années de la période d'étude.

La source OCS GE n'a pas été retenue pour cette étude du fait de sa couverture partielle du territoire et sur des années différentes. De même, la source CVI n'a pas été retenue, car trop récente par rapport à la période d'étude. Enfin, le RA est utilisé à des fins de comparaison et de validation de la cartographie des cultures issue de la combinaison des données géographiques, à l'échelle communale et à l'échelle cantonale. Dans la mesure où il n'est pas une donnée géographique, il n'est pas possible de l'inclure au processus de combinaison géométrique utilisé dans le SIG pour construire la cartographie des cultures.

Au vu de la disponibilité des données identifiées pour la construction de l'indicateur de proximité aux cultures (Tableau 5), deux « sous-périodes » ont été définies pour l'étude Geocap Agri et deux cartographies ont été construites afin de combiner entre elles des données dont les dates de production soient les plus proches possibles :

- La période 2006-2009, qui s'appuie sur une cartographie combinant les sources RPG 2007 et CLC 2006 – on parlera de cartographie 2007, et cette cartographie est considérée comme représentative de la période 2006-2009 ;
- La période 2010-2013 et une cartographie basée sur les sources RPG 2012 et CLC 2012 et BD TOPO® « Végétation » - on parlera de cartographie 2012, et cette cartographie est considérée comme représentative de la période 2010-2013.

Les années 2002-2005, pour lesquelles le RPG n'est pas disponible, n'ont finalement pas été prises en compte. Pour la construction d'une cartographie à échelle fine, la disponibilité du RPG est un minimum incontournable pour pouvoir développer des indicateurs de proximité aux cultures suffisamment précis, étant donné la faible résolution de CLC et la non disponibilité à ce jour des OCS à grande échelle sur l'ensemble du territoire français.

Tableau 5. Dates de disponibilité des différentes bases de données mobilisables sur la période d'étude de Geocap Agri

Source de donnée	Dates de mise à jour des données
Corine Land Cover	1990 ; 2000 ; 2006 ; 2012 ; 2018
Registre parcellaire graphique	à partir de 2007, tous les ans
Vignes et vergers et la BD TOPO®	entre 2010 et 2017, pas avant ; pas de mise à jour
Recensement général agricole	1970, 1979, 1988, 2000 et 2010

4. COMBINAISON DES SOURCES DE DONNÉES ENTRE ELLES ET CONSTRUCTION DE LA CARTOGRAPHIE

Une fois sélectionnées et traitées pour les doublons potentiels propres à chaque source de données, l'ensemble des sources de données sont ensuite combinées dans un système d'information géographique (ArcGIS Pro 2.7.2 – ESRI Inc), afin de construire les cartographies 2007 et 2012 de présence des cultures sur l'ensemble du territoire métropolitain.

4.1 Mise en cohérence des informations sur les types de cultures par le recours à des regroupements de cultures

Chacune des sources de données identifiées pour contribuer à la construction du SIG informe, de manière variable et plus ou moins précise, sur la nature des cultures identifiées.

Un regroupement des cultures a été opéré de manière à simplifier la lecture et à faciliter l'exploitation des informations d'exposition quand elles seront intégrées dans les études épidémiologiques. Ces regroupements ont été faits en fonction du niveau de détail de caractérisation des cultures dans les bases de données sources, et :

- En cohérence avec le travail de construction des matrices culture-exposition qui détaillent, en regard de chaque culture, les types de produits phytosanitaires utilisés et les pratiques d'épandages recueillis sur la base de dires d'experts. Treize regroupements de cultures ont pu être construits. C'est notamment le cas des céréales à paille, du maïs et de la vigne. La matrice arboriculture n'est pas produite pour le moment.
- Soit parce qu'ils correspondent à des groupes plutôt classiques en terme de statistique agricole. Il existe une matrice pour les pommes de terre, et un projet de matrice pour la betterave, mais ces deux cultures sont dans les faits souvent regroupées dans « cultures industrielles ». Pour les oléagineux, le regroupement est arbitraire : concernant les traitements phytosanitaires, il y a des écarts entre colza et tournesol par exemple. Mais il n'est pas prévu de réaliser à court terme des matrices sur ces différentes cultures, elles ont donc été regroupées, d'autant que le RPG ne les distingue pas.
- Soit enfin, sur la base de dires d'experts (ingénieurs agronomes), dans l'optique de ne pas garder un trop grand nombre de cultures différentes. Par exemple, la faisabilité de la construction d'une matrice n'est pas très convaincante sur les cultures maraîchères, notamment pour les données anciennes, néanmoins il est assez cohérent de les regrouper.

Ainsi, dans la base VV, seules deux cultures (la vigne et les vergers) peuvent être distinguées. Parallèlement, la base CLC permet aussi de distinguer ces deux cultures auxquelles s'ajoutent distinctement les prairies et les rizières, les autres types de cultures n'étant pas suffisamment distinguées, elles ont été regroupées sous l'item « cultures indifférenciées ». Enfin, la base RPG est celle qui comporte le plus de détail sur les cultures agricoles. La vigne, les vergers, les prairies et les rizières y sont retrouvées. Les autres cultures ont été regroupées selon les règles citées précédemment. On retrouve les céréales à paille et le maïs qui font l'objet de matrices culture-expositions, les oléagineux (qui regroupe le tournesol, le colza et d'autres cultures indéterminées), les cultures industrielles sans distinction (dans lesquelles peuvent être retrouvées les pommes de terre et les betteraves faisant l'objet de matrices), les cultures maraîchères, les jachères, les semences et les cultures agricoles hors étude.

Le Tableau 6 détaille, pour chaque regroupement de cultures, les classes de cultures identifiées par les différentes sources d'information qui contribuent à la construction de la cartographie.

Tableau 6. Groupes de cultures et correspondance avec les modalités des nomenclatures par source de données

Groupes de cultures	Types de cultures identifiées par les bases de données		
	RPG	CLC	VV
Viticulture	Vignes	Vignobles	Vignes
Arboriculture	Vergers ; Fruits à coque ; Oliviers	Vergers et petits fruits ; Oliveraies	Vergers
Céréales à paille	Blé tendre ; Orge ; Autres céréales		
Maïs	Maïs grain et ensilage		
Oléagineux	Colza ; Tournesol ; Autres oléagineux		
Cultures industrielles	Autres cultures industrielles		
Cultures maraîchères	Protéagineux ; Légumineuses à grains ; Légumes à fleurs		
Prairies	Prairies permanentes ; prairies temporaires	Prairies	
Jachères	Gel industriel ; Gel (surfaces gelées sans production) ; Autres gels		
Semences	Semences		
Rizières	Rizières	Rizières	
Cultures indifférenciées		Terres arables hors périmètres d'irrigation ; périmètres irrigués en permanence ; Cultures annuelles associées aux cultures permanentes ; Systèmes culturaux et parcellaires complexes ; Surfaces essentiellement agricoles, interrompues par des espaces naturels importants ; territoires agroforestiers	
Cultures agricoles hors étude	Plantes à fibres ; Fourrage ; Estives-landes ; Canne à sucre ; Divers		



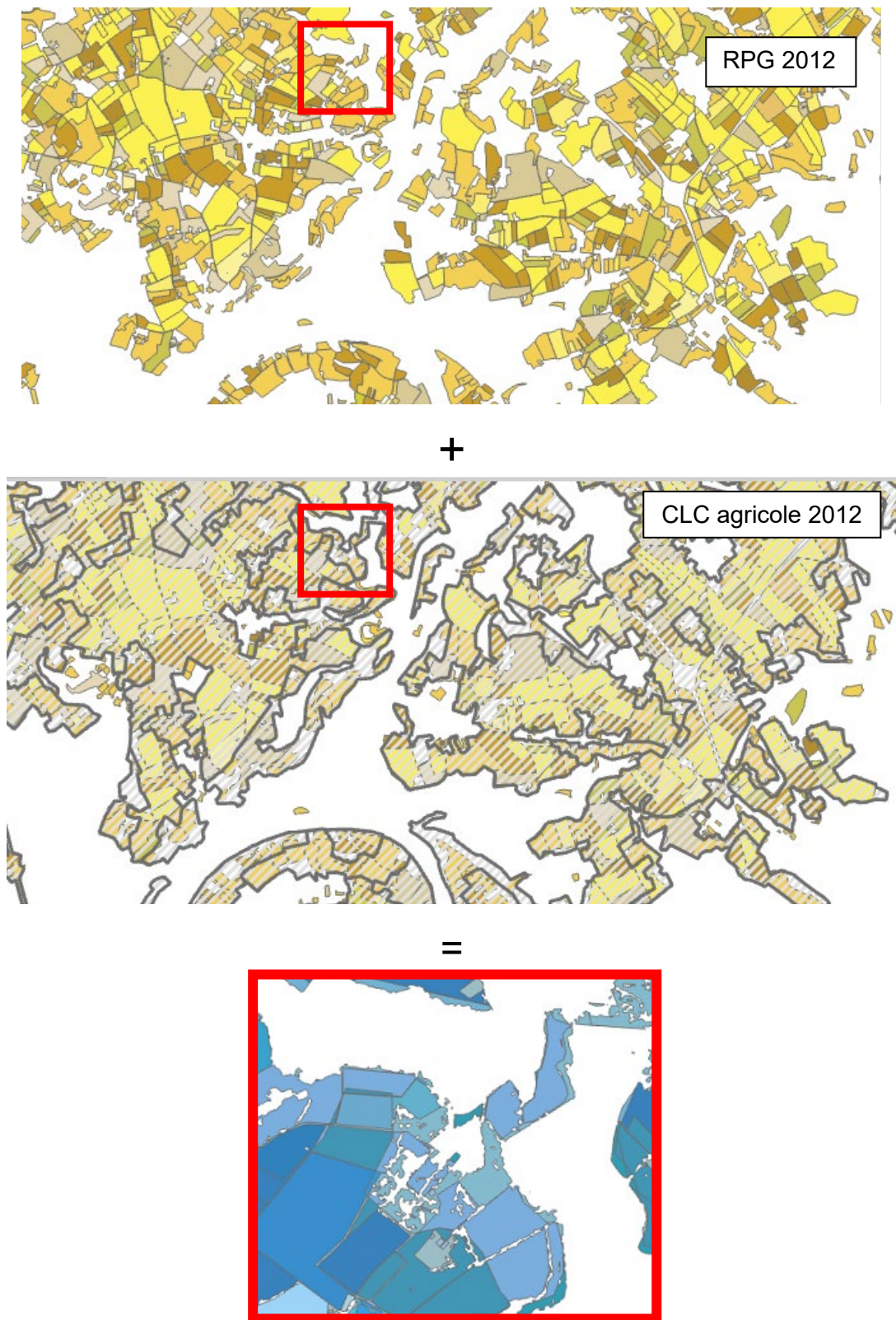
Information non disponible dans la base de données

4.2 Combinaison des sources de données dans une couche unique

Après vérification des géométries des polygones de cultures sur chacune des couches de données en entrée pour éliminer les auto-intersections (anomalies de construction des polygones entraînant des recouvrements entre les contours d'un même objet) et les objets de géométrie nulle, les couches sont successivement combinées entre elles deux à deux en utilisant une fonctionnalité d'union, qui, lors du croisement, conservent à la fois les géométries en superposition et les géométries propres aux couches d'entrée.

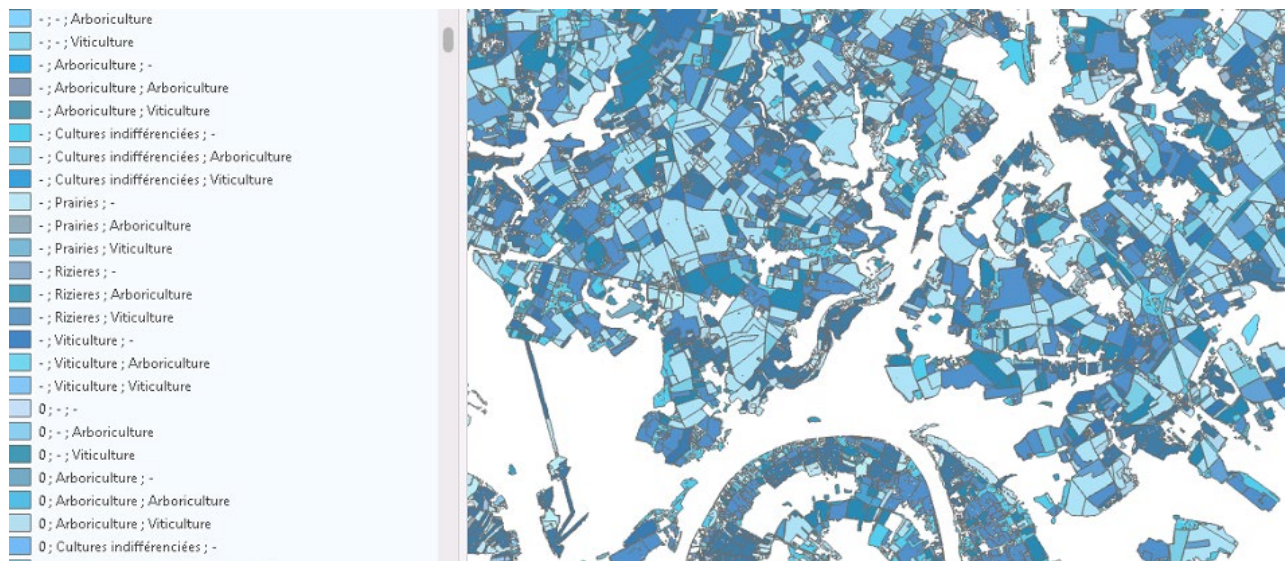
Une fois l'ensemble des données d'entrée pour une période donnée combinées entre elles, on obtient une couche géographique qui peut être implémentée dans un SIG, dont la table attributaire est ensuite nettoyée afin d'éliminer les champs surnuméraires et inutiles, et de rendre lisible les informations d'intérêt : type de culture renseigné sur chaque polygone, et par quelle source de donnée. Les superpositions de polygones présents dans plusieurs couches de données en entrée sont ainsi identifiables. À l'issue de cette combinaison, on obtient une couche cartographique de présence des cultures (Figure 2).

Figure 2. Illustration de la combinaison de deux sources de données pour la construction de la cartographie des surfaces cultivées



Les limites des surfaces agricoles selon CLC ne coïncident pas avec les limites des îlots de cultures du RPG comme le montre ce « zoom » sur la superposition du RPG et de CLC 2012 à l'est de la Boucle de Brotonne (Seine-Maritime). Il n'y a pas non plus de concordance entre les limites des zones couvertes par la vigne et les vergers dans VV de la BD TOPO® et les limites des îlots de cultures du RPG ou les surfaces CLC. Une fois les trois couches combinées entre elles dans le SIG, les délimitations des surfaces cultivées, dans la couche cartographique résultat de la combinaison, ne correspondent donc plus à des limites d'îlots de culture ou de regroupements de parcelles, mais à la délimitation des surfaces cultivées identifiées par les données d'entrée :

Figure 3. Extrait de la cartographie des cultures issues de la combinaison des sources de données au Nord de la Boucle de Brotonne (Seine-Maritime)



(la couleur varie selon le type de cultures renseigné par les trois sources combinées pour construire la cartographie des cultures 2012)

Les zones ainsi délimitées par la cartographie finale ne sont plus des parcelles ou des îlots de cultures, mais des surfaces cultivées, issues d'une, ou plusieurs sources de données, et pour lesquelles la culture principale est déterminée en fonction :

- Des cultures renseignées par les différentes sources, puisque selon les configurations, les sources de données concordent, ou divergent, quant à la culture détectée ;
- Des sources de données elles-mêmes puisqu'elles n'ont pas toutes la même fiabilité.

Pour obtenir des parcelles il faudrait recroiser cette cartographie avec une couche géographique de parcellaire cadastral.

Le traitement qui vise à déterminer la culture à retenir pour chaque polygone de la cartographie est décrit en détail dans la partie 4.6.

4.3 Superposition avec le maillage administratif

Renseigner l'appartenance administrative de chaque polygone cartographié dans la couche résultat de la combinaison des données permet d'utiliser cette cartographie de présence des cultures à l'échelle d'une maille administrative donnée. Ce type d'utilisation peut être envisagé dans des études écologiques, ou pour des croisements avec des données sanitaires pour lesquelles une géolocalisation par géocodage d'adresse au point précis n'est pas envisageable, ce qui est de plus en plus fréquent dans les études épidémiologiques qui s'appuient de plus en plus souvent sur les données des bases médico administratives.

L'IRIS¹⁷, maille utilisée par l'Insee pour le recueil et la diffusion des informations du recensement de la population, est retenu comme « brique de base » du redécoupage de la cartographie. La finesse du découpage géographique à l'IRIS (découpage infra-communal) est cohérente avec la granulométrie des sources de données entrant dans la construction de la cartographie de la présence de cultures. Ce redécoupage permet également de reconstituer l'information à l'échelle de la commune, du canton, du département et de la région. De cette manière, il est possible de caractériser de manière plus ou moins fine le territoire par rapport aux activités agricoles. Il est alors également possible de comparer les informations apportées par la cartographie aux données de référence du recensement agricole qui sont agrégées à la maille de la commune du siège des exploitations.

4.4 Raffinement

La tendance de CLC à surestimer la superficie agricole de façon générale sur le territoire métropolitain a été évoquée précédemment. La nature même de la base de données, ses caractéristiques de construction et de précision, la granulométrie des informations qui y sont répertoriées induisent logiquement cette tendance. Pour remédier à cela, une méthode de raffinement a été développée afin d'extraire de la couche CLC les surfaces qui sont, de manière certaine, non agricoles.

Après combinaison de l'ensemble des données sources, il apparaît que la couche résultat surestime elle aussi la surface agricole, ce qui est cohérent. En effet CLC, qui entre dans la composition de la cartographie, surestime la surface agricole. Par ailleurs dans le RPG, du fait notamment de la saisie des parcelles cultivées sur un fond d'orthophotos par les déclarants eux-mêmes, certains îlots de culture peuvent englober des surfaces non agricoles (chemins d'exploitations, lisières de forêt, haies, etc.), ce qui entraîne aussi, ponctuellement, une surestimation des superficies consacrées à l'agriculture.

Cette méthode de raffinement est décrite plus précisément en Annexe 7. Le raffinement a été appliqué sur la couche de la cartographie, résultat de la combinaison des différentes sources d'informations disponibles. Le choix méthodologique a été d'identifier tous les éléments de la BD TOPO[®] qui ont une emprise surfacique au sol, et qui ne sont pas des surfaces agricoles, et ce, de façon certaine (le bâti, certaines constructions, des espaces anthropiques artificialisés, de la végétation). En superposant les îlots de culture du RPG, il est possible de constater que l'emprise des bâtiments, stades, aérodromes, parkings, etc. est parfois englobée dans les surfaces agricoles que délimite le RPG, comme elle l'est également dans les classes agricoles d'occupation du sol de CLC, notamment la vigne et les vergers.

¹⁷ L'IRIS (îlots regroupés pour l'information statistique) est le découpage du territoire en maille de taille homogène en matière de population (environ 2 000 habitants) sur lequel repose la diffusion des données infra-communales. Les communes de plus de 10 000 habitants et une part importante des communes dont la population est comprise entre 5 000 et 10 000 habitants sont subdivisées en IRIS. Le dernier découpage en IRIS du territoire français date de 2008.

La méthode de raffinement de la couche de présence d'agriculture combinant les diverses sources de données disponibles a permis d'extraire les surfaces suivantes :

- Le bâti « indifférencié »
- Le bâti industriel
- Le bâti remarquable
- Les cimetières
- Les constructions légères
- Les constructions surfaciques
- Les pistes aérodromes
- Les réservoirs
- Les réservoirs d'eau
- Les surfaces en eau (lacs, rivières et fleuves, estuaires, etc.)
- Les surfaces de route (échangeurs, parkings)
- Les terrains de sport
- Les espaces couverts par la végétation (naturelle ou surfaces boisées d'origine anthropique, hors cultures viticoles et vergers)

En 2008, la surface totale des éléments retenus pour caractériser de façon certaine l'absence de cultures en France, à partir de la BD TOPO[®], couvre 17,8 millions d'hectares, parmi lesquels 1,3 millions d'hectares de surfaces artificialisées (hors végétation). Dans cette surface totale non cultivée de 17,8 millions d'hectares, 4,78 millions d'hectares se trouvent en intersection avec des surfaces identifiées comme cultivées par la cartographie de 2007 : ce sont près de 27% des surfaces caractérisées comme non agricoles de façon certaine par la BD TOPO[®] qui sont effectivement en intersection avec des surfaces agricoles d'après les bases de données utilisées pour la cartographie de la présence de cultures (RPG et CLC).

L'ensemble des surfaces non agricoles caractérisées par ces données permet d'informer **sur la vraisemblance élevée d'absence d'agriculture**. Ces surfaces, issues de la BD TOPO[®] et utilisées pour le raffinement, sont regroupées dans une couche d'information séparée, identifiée comme couche de données « No culture », négatif certain de la présence de cultures.

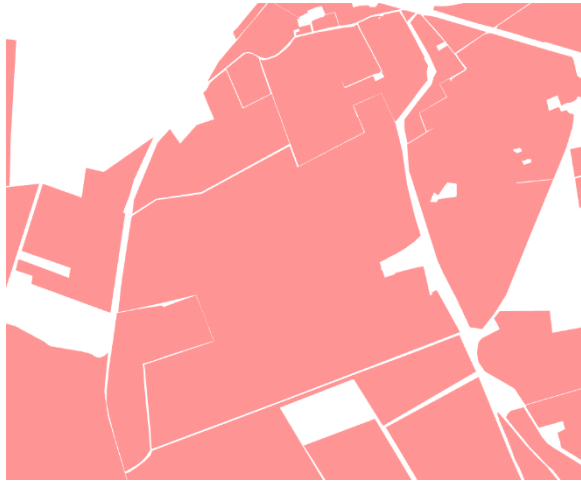
La 4 ci-dessous illustre ce processus de raffinement sur la couche de la cartographie des cultures, avec « effacement » des surfaces végétales (haies) et des surfaces en eau.

Figure 4 :

4.a Extrait de la cartographie des cultures avant raffinement

4.b Superposition de diverses données de la BD TOPO® « No Culture » (haies, surface en eau)

4.c Extrait de la cartographie après raffinement affinées par la BD TOPO® « No Culture »



4.a



4.b



4.c

Cette méthode de raffinement de l'information agricole permet de limiter l'erreur de classement, en évitant de classer comme agricole une surface qui ne l'est pas de façon certaine, ou quasiment.

4.5 Traitement des superpositions, dédoublonnage

Une des informations utilisées pour caractériser l'environnement agricole des sujets dans le cadre d'une étude épidémiologique est la surface de cultures présentes dans un rayon donné autour d'un sujet localisé par géocodage. La quantification de ces surfaces cultivées et l'utilisation de cet indicateur impose de veiller à ne pas conserver de surfaces en superposition qui seraient comptées plusieurs fois, et pourraient entraîner une surestimation de ces surfaces agricoles.

Le fait de combiner l'ensemble des couches sources de données et de conserver, pour chaque polygone, le résultat des intersections entre les couches sources et les informations issues de sa couche d'origine (caractérisation des cultures, et source(s) de données qui la renseigne(nt)), permet a priori de ne pas avoir de problème de superpositions entre les sources de données (c'est le principe du découpage géométrique mis en œuvre dans le SIG).

Cependant, après les premiers traitements réalisés sur les couches de données retenues pour construire la cartographie des cultures, que ce soit la cartographie 2007, ou la cartographie 2012, il apparaît que des superpositions sont bien présentes dans les couches résultat. Elles ne sont pas liées à des problèmes de croisement géométrique dans le SIG (anomalies éventuelles au moment des traitements d'intersection), mais à la présence de superpositions d'objets dans les différentes couches de données sources. Il apparaît en effet qu'en dehors de CLC qui présente une topologie propre, les deux autres sources de données (RPG et BD TOPO[®]) comportent des entités en superposition, de par leurs spécificités de construction. Ainsi, dans le RPG, il peut y avoir une superposition partielle de deux îlots de culture, avec mention de cultures différentes ; dans la BD TOPO[®], ces superpositions entre objets sont également présentes.

Si on conserve ces superpositions lors de la combinaison des sources de données dans la cartographie, on obtient, lors des intersections entre les couches, des objets en doublons, voire en triplettes ou quadruplets parfois, créés sur ces superpositions. Ces surfaces se retrouvent comptées en double (voire davantage) dans la cartographie. Ces superpositions, si elles sont présentes dans la zone de recherche de l'exposition autour des sujets, peuvent être à l'origine d'une surestimation de la surface en culture.

Une méthode, détaillée en Annexe 4 a été spécifiquement développée par l'équipe Epicea afin de traiter ces situations de doublon, à l'échelle de la cartographie, en tenant compte à la fois de la source de données qui renseigne la présence de culture sur ces surfaces en doublons, et de la culture ou des cultures mentionnées pour ces surfaces.

La cartographie, pour répondre aux objectifs de l'étude Geocap Agri, est construite pour être représentative de deux sous-périodes (2006-2009 et 2010-2013). En termes de volumétrie, il est à noter que la cartographie (avant traitement des superpositions), contient, pour la période 2006-2009, 10 138 320 entités, et pour la période 2010-2013, compte 1 315 661 entités. Après traitement de ces superpositions, la cartographie 2006-2009 compte 10 048 581 entités et la cartographie 2010-2013, 1 310 647 (Tableau 7).

Tableau 7. Comparaison des couches pour 2007 et 2012

Couches	2007	2012
Cartographie des cultures	10 048 581 objets	1 310 647 objets
Surface agricole totale	30 824 397 ha	30 020 683,6 ha
RPG	5 897 853 objets	6 132 686 objets
CLC	134 964 objets	136 500 objets
VV BD TOPO®	-	571 263 objets
No Culture	46 881 004 objets	22 796 848 objets

L'évolution de la surface totale considérée comme agricole entre les deux millésimes est assez cohérente (la diminution de la SAU est un phénomène attendu sur cette cartographie car bien documenté au fil des années).

Néanmoins, le nombre d'objets constitutifs de la couche « No Culture », qui sert à « raffiner » la cartographie, est notoirement différent entre les deux éditions de la cartographie des cultures. Cette différence notable entre les deux couches « No Culture » a pu, pour la version 2007, entraîner une fragmentation plus importante lors du raffinement.

Il n'est pas exclu par ailleurs que cette édition 2007, qui implémente la première édition du RPG numérisé, comporte davantage de superpositions d'objets que la version 2012 (saisie moyennement propre par les déclarants) : cela n'entraîne pas d'écart sensible sur le nombre d'objets de chaque couche RPG, mais s'il y a plus de chevauchements, lors de la combinaison, tout est redécoupé : en 2007, le nombre d'objets en superposition liés à ces chevauchements d'entités des couches sources correspond à 89 739 objets, soit environ 0,88% des objets de la cartographie avant traitement des superpositions, alors qu'en 2012, le nombre d'objets liés à ces chevauchements d'entités dans les couches sources correspond à 5 014 objets, soit seulement 0,38% des objets de la cartographie avant traitement de ces superpositions. Sur la cartographie globale, ces impacts des superpositions d'entités dans les couches sources sont faibles en termes de surfaces puisque cela va souvent concerner des artefacts (chevauchements très peu étendus en superficie) et donc des surfaces très réduites. Cependant, il est nécessaire de les traiter, car lors de l'utilisation de la cartographie en intersection avec des buffers de rayons d'échelles hectométriques autour des sujets, ces surfaces superposées peuvent entraîner des aberrations et générer des doubles-comptes de superficie agricole, conduisant à la surestimation de l'exposition des sujets.

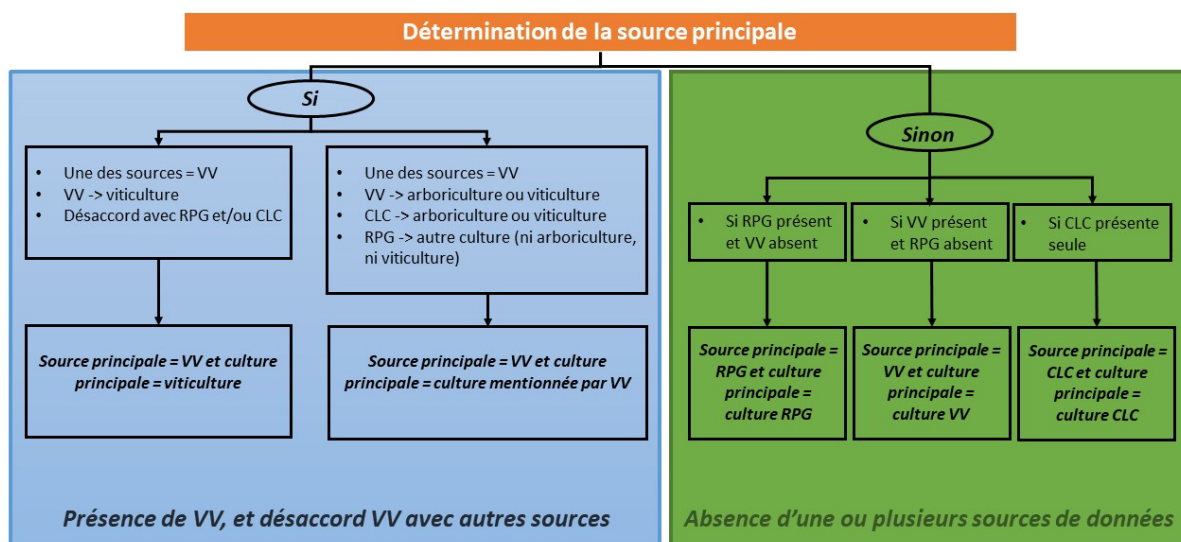
4.6 Sélection d'une culture par polygone

Les données combinées dans les cartographies des activités agricoles sont de nature inégale, de précision variable, ce qui induit une hétérogénéité de la qualité de l'information construite par combinaison de ces données. En fonction des informations présentes pour un même polygone, des types de cultures possiblement présents sur celui-ci, et de la fiabilité des différentes sources de données, un algorithme de sélection des cultures peut être appliqué.

Chaque polygone (dans un buffer ou sur la cartographie complète) est associé à une ou plusieurs sources d'information et un ou plusieurs types de cultures, qui peuvent ne pas être concordants.

L'arbre de décision ci-dessous (Figure 5) décrit le processus guidant la sélection de la source à retenir, en fonction de la culture, quand les sources ne concordent pas. L'algorithme utilisé dans le cadre du projet Geocap Agri par l'équipe Inserm Epicea pour déterminer, pour chaque polygone, la source d'information principale (et la culture associée) est présenté en Annexe 5.

Figure 5. Détermination de la source principale et de la culture principale en fonction des sources de données présentes et des cultures détectées par chaque source de données



Pour les années pour lesquelles VV ne peut être mobilisé la situation est simple :

- Si RPG présent, alors source principale = RPG et culture principale = celle du RPG ;
- Si CLC seule, alors source principale = CLC et culture principale = celle de CLC

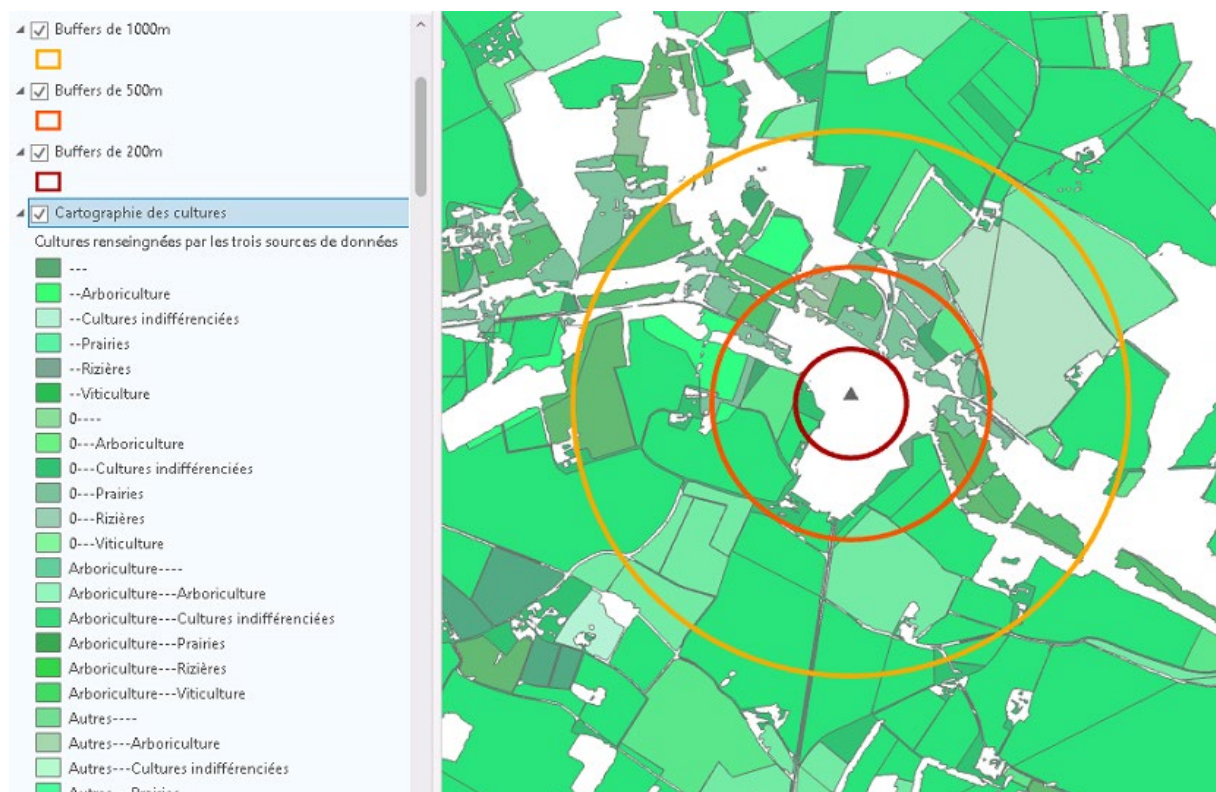
5. CARACTÉRISATION DE L'ENVIRONNEMENT AGRICOLE AUTOUR DES SUJETS

Pour certaines études épidémiologiques pour lesquelles un recueil des adresses des sujets est réalisé, un géocodage est possible, permettant une localisation précise des sujets dans le territoire. Une analyse assez fine de la présence de surfaces agricoles dans leur environnement immédiat est donc envisageable.

La littérature n'est pas consensuelle quant à la distance maximale d'exposition aux produits phytosanitaires agricoles à prendre en considération entre la géolocalisation du sujet et le champ, considéré comme la source d'exposition ; cette distance peut être très différente selon les substances, la formule des produits épanchés, le matériel utilisé, la hauteur des cultures, etc. Elle évoque des problématiques de dispersion dans l'air, de dépôt au sol et de remise en suspension de particules susceptibles d'entrer dans l'habitat, par l'air ou par transport sous les chaussures, ainsi que des problématiques de dérive (une partie des produits épanchés sont transportés par le vent avant de parvenir au sol et peut atteindre les riverains). Ces différents potentiels d'exposition conduisent à analyser la présence des cultures dans des rayons de distance assez faible autour des lieux de vie des sujets. Pour le projet Geocap Agri par exemple, quatre rayons ont été définis pour la caractérisation de l'environnement des sujets : 200, 500, 1000 et 1 250 m (seuil haut de distance documentée pour la dérive). Seuls les trois rayons 200, 500 et 1 000 m ont été retenus dans l'analyse.

Dans un SIG, la couche globale de la cartographie des cultures agricoles est découpée selon l'emprise respective des différents buffers délimitant les zones d'exposition (Figure 6). Les surfaces agricoles sont quantifiées à l'intérieur de chaque disque ou « buffer », chacun étant par ailleurs propre à lieu de vie d'un sujet de l'étude identifié par un code identifiant d'anonymisation qui permet de relier, pour l'analyse statistique, les informations d'exposition aux informations sanitaires recueillies par le registre.

Figure 6. Illustration de la méthode d'un découpage des cultures autour d'un point cas/témoin



Plusieurs indicateurs peuvent être construits à partir des résultats de ces traitements d'intersection avec les buffers (Jones *et al.* 2014) ; (Rappazzo *et al.* 2016) ; (Xiang, Nuckols, and Stallones 2000) ; (Booth *et al.* 2015) ; (Ward *et al.* 2006).

La présence de culture à proximité de l'adresse géolocalisée (i.e. dans un buffer donné), constitue un premier indicateur possible.

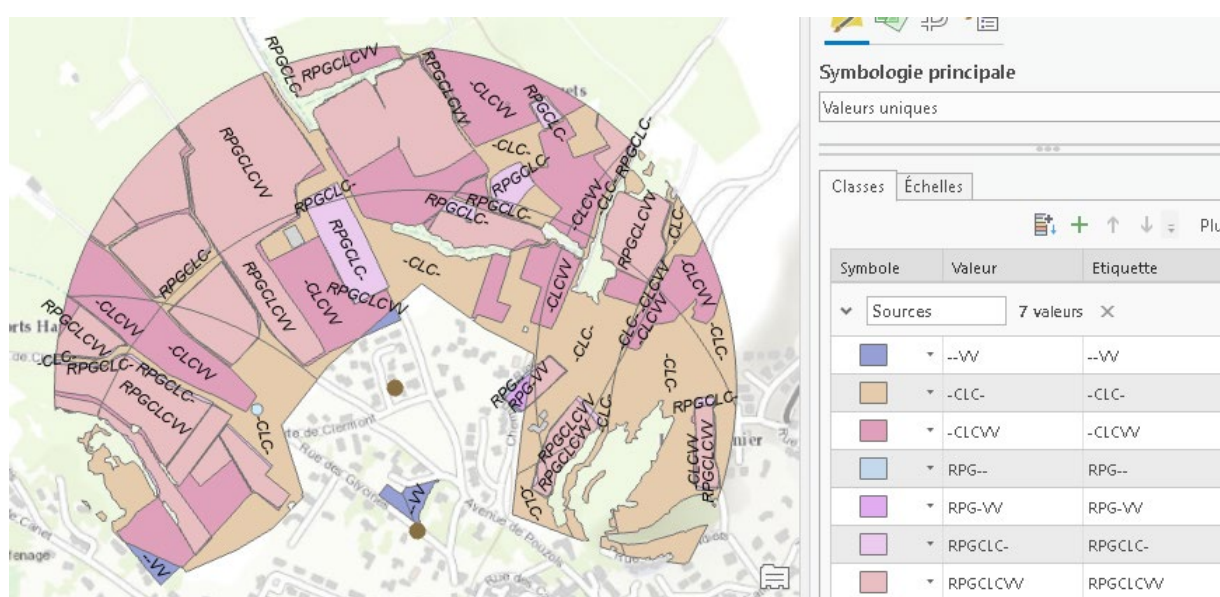
La superficie des cultures (par type de cultures, le niveau de détail à retenir dans la caractérisation des cultures étant à préciser) présentes dans un rayon donné constitue un autre exemple d'indicateur, partant du principe qu'il est vraisemblable d'imaginer qu'un individu résidant à proximité d'une toute petite parcelle a un risque moindre d'être exposé aux pesticides agricoles qu'un individu autour de l'adresse duquel la superficie agricole est importante, dans un rayon donné.

Des indicateurs de distance peuvent aussi être calculés, en fonction des besoins, par exemple :

- Soit en calculant les coordonnées des centroïdes des objets géographiques (polygones) correspondant aux surfaces cultivées dans un buffer donné, puis en calculant la distance euclidienne entre l'adresse du sujet et chacun des centroïdes de ces zones cultivées. L'indicateur retenu pourra alors être le minimum, ou la moyenne des distances calculées entre le point adresse du sujet et l'ensemble des surfaces cultivées présentes dans un buffer donné, ou encore le minimum ou la moyenne des distances à une sélection de surfaces cultivées pour une culture spécifique.
- Soit en calculant (par une méthode de jointure spatiale dans le SIG), la distance entre chaque sujet et l'enveloppe (la limite) de la zone cultivée la plus proche.

La caractérisation de l'environnement agricole des sujets est réalisée par l'intersection des buffers (cf. Figure 7) avec la couche globale de la cartographie, et avec la couche « No Culture », décrivant les espaces assurément non agricoles. Ces informations (Tableau 8 et Tableau 9) sont livrées pour l'analyse dans différents tableaux (un pour chaque rayon de buffer). Ce sont les mêmes informations qui sont présentes dans la couche de la cartographie globale.

Figure 7. Résultat d'intersection entre un buffer de 500 m et la cartographie 2012



Les tables attributaires des couches SIG résultats d'intersections entre les buffers et la cartographie de présence de cultures contiennent de nombreuses informations attributaires, pour chacune des entités qui résultent de l'intersection, à savoir chaque polygone résultant du découpage de la cartographie par les buffers.

Tableau 8. Variables de la table attributaire des résultats des intersections des buffers avec la cartographie (une table pour chaque rayon de buffer)

Identifiant sujet
Code IRIS
Code commune
Code canton
Code département
Code région
Source RPG : si la valeur est « RPG » cela signifiera que le RPG couvre ce polygone sinon c'est que le RPG n'est pas une des sources de données présente à cet endroit
Source CLC : si la valeur est « CLC » cela signifiera que le CLC couvre ce polygone sinon c'est que CLC n'est pas une des sources de données présente à cet endroit
Source VV : si la valeur est « VV » cela signifiera que VV couvre ce polygone sinon c'est que VV n'est pas une des sources de données présente à cet endroit
Source GLOBALE : cette valeur sera une concaténation des valeurs des colonnes <i>Source RPG</i> , <i>Source CLC</i> et <i>Source VV</i> , on retrouvera par exemple « RPGCLCVV » lorsque les trois sources de données seront présentes
Code Culture RPG : c'est la valeur du code de la culture proposée par le RPG, par exemple la valeur « 21 » (qui correspondra à la vigne) ; il se peut que la valeur soit vide si le RPG n'est pas présent
Code culture CLC : c'est la valeur du code de la culture proposée par CLC, par exemple la valeur « 221 » (qui correspondra aux vignobles) ; il se peut que la valeur soit vide si CLC n'est pas présent
Code culture VV : c'est la valeur du code de la culture proposée par VV, par exemple la valeur « VV21 » (qui correspondra à la vigne) ; il se peut que la valeur soit vide si VV n'est pas présent
Code culture GLOBAL : c'est la concaténation des trois codes cultures, par exemple « 21221VV21 »
Type Culture RPG : le type de culture est la dénomination du <i>Code culture du RPG</i> , par exemple si <i>Code culture RPG</i> est « 21 » alors le <i>Type de culture RPG</i> sera « Vigne »
Type culture CLC : le type de culture est la dénomination du <i>Code culture de CLC</i> , par exemple si <i>Code culture CLC</i> est « 221 » alors le <i>Type de culture CLC</i> sera « Vignobles »
Type culture VV : le type de culture est la dénomination du <i>Code culture de VV</i> , par exemple si <i>Code culture VV</i> est « VV21 » alors le <i>Type de culture VV</i> sera « Vigne »
Type culture GLOBAL : c'est la concaténation des trois types de cultures, par exemple « Vigne ; Vignobles ; Vignes »
Groupe de culture RPG : il s'agit d'un des 13 groupes de cultures décrits plus haut, par exemple si le <i>Type de culture RPG</i> est « vigne » alors le <i>Groupe de culture RPG</i> sera « Viticulture »
Groupe de culture CLC : il s'agit d'un des 13 groupes de cultures décrits plus haut, par exemple si le <i>Type de culture CLC</i> est « vignoble » alors le <i>Groupe de culture CLC</i> sera « Viticulture »
Groupe de culture VV : il s'agit d'un des 13 groupes de cultures décrits plus haut, par exemple si le <i>Type de culture VV</i> est « vigne » alors le <i>Groupe de culture VV</i> sera « Viticulture »
Groupe culture GLOBAL : c'est la concaténation des groupes de cultures, par exemple « Viticulture ; Viticulture ; Viticulture »
Source principale : source de donnée la plus fiable parmi les sources de données concourant à la présence de l'information de présence de cultures – source de données qui détermine la culture principale
Culture principale : culture identifiée comme la plus certaine d'après les différentes informations des sources de données - après passage de l'algorithme de priorisation des cultures
Surface de la zone de culture en intersection avec le buffer (en m²) : surface géométrique des polygones de la couche de la cartographie en intersection avec le buffer
Rayon du buffer : la valeur ne pourra être différente de « 200 », « 500 », « 1 000 », ou « 1 250 »
Surface du buffer (en m²) : la valeur ne pourra être différente de « 125 663 », « 785 398 », « 3 141 592 », « 4 908 738 », ou « 19 634 954 »

Si on examine la morphologie de la table, celle-ci se déploie de la façon suivante :

Tableau 9. Aperçu de la table attributaire en sortie des intersections des buffers avec la cartographie des cultures (extrait pour la cartographie 2012 – une ligne = un polygone en intersection avec un buffer)

Identifiant du cas	Code IRIS	Code commune	Code canton	Code département	Code région	Source RPG	Source CLC	Source VV	Source GLOBALE	Code Culture RPG	Code culture CLC	Code culture VV	Code culture GLOBAL	Type Culture RPG	Type culture CLC	Type culture VV	Type culture GLOBAL	Groupe de culture RPG	Groupe de culture CLC	Groupe de culture VV	Groupe culture GLOBAL	Surface de la culture (en ha)	Rayon du buffer
1	73300000	07330	0715	07	84	RPG	CLC	/	RPGCLC	21	223	/	21223	Vignes	Oliveraies	/	Vignes ; Oliveraies ;	Viticulture	Arboriculture	/	Viticulture ; Arboriculture ;	12	500
1	73300000	07330	0715	07	84	RPG	/	VV	RPGVV	20	/	VV20	20VV20	Vergers	/	Vergers	Vergers ; ; Vergers	Arboriculture	/	Arboriculture	Arboriculture ; ; Arboriculture	10	500
1	73300000	07330	0715	07	84	RPG	/	/	RPG	20	/	/	20	Vergers	/	/	Vergers ; ;	Arboriculture	/	/	Arboriculture ; ;	8	500
1	73300000	07330	0715	07	84	RPG	/	/	RPG	3	/	/	3	Orge	/	/	Orge ; ;	Céréales à pailles	/	/	Céréales à pailles ; ;	13	500
1	73300000	07330	0715	07	84	/	/	VV	VV	/	/	VV20	VV20	/	/	Vergers	; ; Vergers	/	/	Arboriculture	; ; Arboriculture	2	500
2	471570106	47157	4798	47	75	RPG	/	/	RPG	1	/	/	1	Blé tendre	/	/	Blé tendre ; ;	Céréales à pailles	/	/	Céréales à pailles ; ;	60	1250
2	471570106	47157	4798	47	75	RPG	CLC	VV	RPGCLCVV	21	221	VV21	21221VV21	Vignes	Vignobles	Vignes	Vignes ; Vignobles ; Vignes	Viticulture	Viticulture	Viticulture	Viticulture ; Viticulture ; Viticulture	20	1250
2	471570106	47157	4798	47	75	RPG	CLC	/	RPGCLC	22	222	/	22222	Fruits à coques	Vergers et petits fruits	/	Fruits à coques ; Vergers petits fruits ;	Arboriculture	Arboriculture	/	Arboriculture ; Arboriculture ;	15	1250
2	471570106	47157	4798	47	75	RPG	CLC	/	RPGCLC	22	223	/	22223	Fruits à coques	Oliveraies	/	Fruits à coques ; Oliveraies ;	Arboriculture	Arboriculture	/	Arboriculture ; Arboriculture ;	15	1250
2	471570106	47157	4798	47	75	RPG	/	VV	RPGVV	23	/	VV20	23VV20	Oliviers	/	Vergers	Oliviers ; ; Vergers	Arboriculture	/	Arboriculture	Arboriculture ; ; Arboriculture	30	1250
2	471570106	47157	4798	47	75	RPG	/	/	RPG	23	/	/	23	Oliviers	/	/	Oliviers ; ;	Arboriculture	/	/	Arboriculture ; ;	40	1250
2	471570106	47157	4798	47	75	RPG	/	/	RPG	21	/	/	21	Vignes	/	/	Vignes ; ;	Viticulture	/	/	Viticulture ; ;	35	1250
2	471570106	47157	4798	47	75	/	CLC	/	CLC	/	223	/	223	/	Oliveraies	/	; Oliveraies ;	/	Arboriculture	/	; Arboriculture ;	15	1250
2	471570106	47157	4798	47	75	RPG	/	/	RPG	23	/	/	23	Oliviers	/	/	Oliviers ; ;	Arboriculture	/	/	Arboriculture ; ;	20	1250

6. AVANTAGES, LIMITES ET PERSPECTIVES

6.1 Avantages de la méthode

La cartographie construite pour différents millésimes (2007 et 2012 dans le contexte de Geocap Agri) permet d'envisager une utilisation dans les études épidémiologiques sans recueil supplémentaire de données, en dehors des adresses pour la géolocalisation des sujets pour lesquels on considère que la proximité des cultures agricoles constitue un assez bon proxy de l'exposition environnementale aux produits phytopharmaceutiques utilisés en agriculture.

6.1.1 Couverture métropolitaine et méthodologie homogène

Les données utilisées pour la construction de la cartographie sont disponibles sur l'ensemble du territoire métropolitain, ce qui permet d'envisager une évaluation de l'exposition homogène en tout point du territoire lors du croisement avec des données sanitaires.

6.1.2 Optimisation de la qualité des sources de données

Dans la cartographie, le fait de combiner plusieurs sources de données différentes, construites à des fins différentes, permet d'améliorer la qualité de la caractérisation de l'exposition aux cultures ainsi obtenue : chaque source de données (CLC, RPG, BD Topo) vient en effet renforcer l'information des autres quand, en un même point, les différentes sources de données documentent la présence d'une même culture, et chaque source vient compléter les autres, quand en un point, l'une d'elle documente la présence d'une culture que les autres sources ne détectent pas. En tenant compte des caractéristiques propres à chaque source de données, de la fiabilité de l'information apportée par chacune en fonction du type de cultures, il est ainsi possible d'obtenir une cartographie qui localise assez bien les informations exhaustives du RA, qui ne peuvent être localisées avec une très grande précision.

6.1.3 Possibilité de prise en compte de données météo et topographiques

La construction d'une cartographie à échelle fine permet également la prise en compte des données complémentaires sur les vents dominants (données météo) ou d'éventuels obstacles de terrain (données géographiques de topographie fine ou de rugosité) pouvant influencer sur la diffusion des produits phytopharmaceutiques, afin de préciser l'exposition des sujets d'une étude épidémiologique.

6.2 Limites

6.2.1 Limites et obstacles techniques

La combinaison de sources de données à grande échelle sur l'ensemble du territoire métropolitain entraîne la génération de couches de données géographiques de volumétrie importante. Lors du croisement avec les buffers générés autour des lieux de vie géocodés des sujets des études épidémiologiques, les fichiers d'intersection obtenus sont susceptibles de comporter plusieurs millions d'entités. La question de la volumétrie des données traitées doit nécessairement être prise en compte dans les moyens requis (ressources machines et logicielles mobilisées pour la réalisation des calculs, disponibilité d'espace de stockage sur les serveurs). Le traitement des superpositions dans les données sources engendre, lors de la réalisation des intersections entre sources de données, l'apparition de surfaces en doublons. Cette possibilité de traitement est un atout qui nécessite toutefois des fonctionnalités de gestion de la topologie.

6.2.2 Années d'étude et disponibilité des données

Une des limites de ce travail réside dans la difficulté à couvrir la période d'étude de Geocap Agri. En effet, il n'a pas été possible d'acquérir des données sur des périodes permettant de décrire les activités agricoles pour les années les plus anciennes et notamment entre 2002 et 2005. L'absence de disponibilité des données les plus précises (le RPG) sur la période antérieure à 2007 a conduit à ne pas retenir pour l'analyse la période courante, dans l'étude Geocap Agri, de 2002 à 2005. Cela n'est pas sans conséquence sur l'analyse et limite la puissance statistique.

Par ailleurs, le RPG, utilisé comme source de données essentielle dans la construction de la cartographie, présente une relative stabilité d'une année sur l'autre pour les cultures pérennes, qui évoluent peu, et plus lentement ; mais, pour les cultures annuelles, il est très dépendant des variations liées aux rotations des cultures : une partie importante des surfaces cultivées en cultures annuelles est en effet concernée par les rotations agraires¹⁸, qui affaiblissent la représentativité du RPG d'un millésime donné pour une période qui s'étend sur plusieurs années. Les données géographiques produites annuellement comme le RPG sont très tributaires de ces pratiques de rotations culturales, puisqu'une même parcelle allouée par un exploitant à une culture annuelle, sera plantée différemment d'une année sur l'autre. Pour ces cultures, le choix d'un millésime du RPG pour la construction de la cartographie aura donc des répercussions, de même que pour la caractérisation, à partir de cette cartographie, de l'environnement agricole d'un sujet à proximité duquel se trouvent des parcelles concernées par ces rotations.

Le fait de choisir un millésime du RPG plutôt qu'un autre a donc forcément des conséquences sur la prise en compte des cultures annuelles dans la cartographie. Pour une cartographie considérée comme représentative d'une période s'étalant sur plusieurs années, des précautions sont nécessaires quant à la prise en compte des surfaces en cultures annuelles, probablement en les considérant comme un regroupement de cultures « cultures annuelles » et non en tenant compte du détail précis de chaque culture.

6.2.3 Incertitude quant à la qualité de l'information : les limites intrinsèques de chacune des sources de données influent nécessairement sur la cartographie

L'autre limite de ce SIG se traduit par la qualité des données disponibles pour l'objectif recherché, très directement liée aux modalités de recueil et aux caractéristiques de construction intrinsèque à chacune des sources de données mobilisées. Ainsi, pour rappel, dans le RPG pour les années antérieures à 2015, la description à l'ilot constitue un frein à la description précise des cultures. Le recueil des données du RPG, destiné au calcul du montant des versements des aides PAC, entraîne une sous-déclaration de certaines cultures non assujetties aux aides de l'Union européenne et donc à une relative incomplétude de la base sur certaines cultures. Quant à CLC, elle n'est pas assez précise sur la définition des territoires agricoles, certains territoires urbains, industriels ou autre étant intégrés aux territoires agricoles ou inversement ; par ailleurs, la caractérisation des cultures y est peu détaillée. Tout cela doit être pris en considération lors de l'utilisation de la cartographie pour la caractérisation de l'environnement agricole des sujets. Enfin, l'utilisation du thème « Végétation » de la BD Topo, par une sélection des classes « Vignes » et « Vergers », n'est pas sans limite non

¹⁸ La technique des rotations, que les archéologues ont permis d'identifier dès l'Antiquité, consiste à pratiquer, sur un terrain donné, une culture différente d'une année sur l'autre, selon un cycle régulier sur deux, trois ou quatre ans - on parle aussi d'assolement biennal, triennal, ou quadriennal, même si l'assolement aujourd'hui désigne plutôt la répartition des terres d'une même exploitation agricole en « soles », consacrées chacune à une culture donnée pendant une saison culturale. L'assolement et les rotations culturales sont des notions complémentaires : en fonction du climat et de la nature du sol, un exploitant va effectuer des rotations annuelles entre les parcelles de son exploitation (une parcelle en lin une année, en maïs la suivante, en blé d'hiver la troisième par exemple). La rotation inclut aussi la mise en jachère qui permet la mise au repos des sols.

plus : limite liée à la temporalité de la construction de cette information, recueillie et complétée progressivement entre 2012 et 2015, et non disponible pour les années antérieures, et non revue depuis, limite liée au chevauchement de certaines surfaces cartographiées (la BD Topo n'est pas une couche d'occupation du sol), notamment.

6.2.4 Non disponibilité pour l'outre-mer

Actuellement, la cartographie couvre l'ensemble du territoire métropolitain. Les régions ultramarines pourraient être couvertes dans un second temps, la disponibilité variable des données en entrée pour certains territoires constituant à ce stade un frein pour le développement pour les DROM.

6.2.5 Évolution des cultures au fil du temps

Le paysage agricole français est en constante évolution, d'une part à cause des rotations agricoles déjà évoquées précédemment, qui entraînent des mutations du paysage d'une année sur l'autre pour les surfaces allouées aux cultures annuelles ; mais aussi parce que les surfaces attribuées à l'agriculture sur notre territoire diminuent au fil des années, de façon assez conséquente : « Depuis 1982, en France métropolitaine, les sols artificialisés augmentent en moyenne chaque année de 57 600 hectares (ha), les sols naturels de 8 300 ha au détriment des terres agricoles qui perdent 65 900 ha »¹⁹. Cette constante évolution du paysage agricole incite à tenir compte de façon approfondie de la période d'étude et des mutations agricoles qui ont pu marquer cette période dans le choix des données à inclure à la cartographie des cultures qui sera considérée comme représentative de cette même période.

6.2.6 Une cartographie qui fournit une caractérisation de l'environnement agricole d'un sujet, et non son exposition environnementale aux pesticides agricoles

La cartographie fournit des informations détaillées et fiables sur la présence de cultures agricoles en tout point du territoire métropolitain. Mais même si elle est utilisée comme proxy de l'exposition aux produits phytosanitaires, cette cartographie ne peut pas servir à estimer les expositions aux pesticides à proprement parler. Un enrichissement de l'information sur les surfaces cultivées qu'elle contient par des informations sur les substances (en fonction des années), à partir des matrices cultures-exposition ou à partir des données de la BNVD, permettrait de mieux approcher l'estimation de l'exposition aux produits phytosanitaires à proprement parler.

6.2.7 Prise en compte des incertitudes liées à l'imprécision de la géolocalisation des adresses des sujets

Si les algorithmes de géocodage, la qualité des bases de données d'adresses qui sont mobilisées dans ces algorithmes, comme la qualité du recueil des adresses dans les bases de données sanitaires, ont eu tendance à s'améliorer au fil des années, il n'en demeure pas moins qu'aujourd'hui encore, subsistent des incertitudes quant à la qualité de la géolocalisation précise des sujets, notamment en zone rurale. La plupart des algorithmes de géocodage fournissent des indicateurs de précision de la géolocalisation en sortie de traitement, et quels que soient les outils, le géocodage des adresses en milieu rural est toujours plus complexe : absence de noms de rue ou de voie, absence de numéro, adresse se limitant à un lieu-dit parfois assez étendu, etc. De ce fait, lorsqu'un buffer de rayon très fin est défini autour d'un sujet géolocalisé, si la précision de la géolocalisation est médiocre, le risque que le buffer encercle des surfaces agricoles qui, dans la réalité, ne se trouvent pas à

¹⁹ L'occupation du sol entre 1982 et 2018, les Dossiers d'Agreste, avril 2021.
https://www.terre-net.fr/ulf/data/DelphineJ/Dossiers2021-3_TERUTI-1.PDF

proximité immédiate du sujet, est élevé. Cela entraîne un risque important d'erreur de classement dans la caractérisation de l'environnement des sujets. Pour utiliser au mieux cette cartographie à l'échelle individuelle, une qualité optimale dans le recueil des adresses est à rechercher afin de limiter au maximum l'effet de l'imprécision du géocodage sur la caractérisation de l'environnement des sujets.

6.2.8 Absence de consensus sur les distances à prendre en compte

À ce stade, la littérature ne permet pas de dégager des distances optimales pour la caractérisation de l'environnement agricole des sujets, et l'état des connaissances actuelles implique d'analyser la présence de cultures sur différents rayons (buffers variables), et d'interroger différents indicateurs (superficie, distances, etc.) pour tenter d'identifier les métriques les plus adaptées. La revue de la littérature réalisée dans le cadre de l'expertise collective de l'Inserm Pesticides et Effets sur la santé de 2021 montre que les distances prises en compte dans les travaux portant sur la proximité aux cultures agricoles varient le plus souvent entre 100 m et 1500 m maximum. L'amélioration des connaissances permettra de mieux décrire les composantes de l'environnement résidentiel des sujets susceptibles d'avoir un impact sur leur exposition.

6.3 Perspectives

6.3.1 Amélioration de la prise en compte des surfaces non cultivées dans le raffinement

La couche « No Culture » comme les traitements de raffinement de la cartographie ne prennent pas en compte, à ce stade, les routes et les chemins de fer, dans la mesure où l'emprise surfacique de ces réseaux n'est pas disponible dans la BD TOPO® (sauf pour quelques gros échangeurs et grandes gares de triage). Ces réseaux sont inclus à la BD TOPO® dans des couches d'objets linéaires. L'information de largeur est disponible dans les tables attributaires de ces couches, mais elle est à considérer avec précaution (elle ne prend en compte en général que l'emprise précise des voies de circulation et non ce qui les entoure : talus, bordures, etc.). Ce n'est sans doute pas une des limites les plus importantes de cette méthode, mais ces infrastructures présentant une forte emprise territoriale, leur absence de prise en compte dans le traitement de raffinement est susceptible d'avoir un effet sur l'estimation des surfaces totales de cultures.

6.3.2 Utilisation des données d'occupation du sol grande échelle

Les données d'occupation du sol antérieures à 2007 ne sont pas disponibles à ce stade à des échelles fines pour l'intégralité du territoire métropolitain. Cependant, les données satellitaires ou les prises de vues aériennes anciennes font l'objet de nombreux projets dans le but d'en tirer des informations géolocalisées sur l'occupation du sol pour des périodes passées. Par ailleurs, la couverture progressive du territoire par les OCS grande échelle promet des avancées non négligeables dans la qualité et l'homogénéité des données sur la présence de cultures. Reste à voir le niveau de détail avec lequel les nomenclatures de ces occupations du sol seront construites, et donc la précision, en matière de cultures, qu'il sera permis d'atteindre. Des travaux sont en cours actuellement afin de comparer les OCS grande échelle, dans une région où elles sont disponibles, avec la cartographie des cultures.

6.3.3 Utilisation des matrices cultures exposition

La construction d'une cartographie permettant d'estimer de façon assez fiable les surfaces cultivées pour une période donnée permet d'envisager l'application des matrices cultures-exposition sur les surfaces cultivées de la cartographie. L'objectif est d'affiner l'exposition aux produits phytosanitaires employés en agriculture, en évaluant les probabilités, fréquence

(nombre de passages annuels), et, pour certaines matrices, l'intensité (quantités de substances diffusées à l'hectare) d'utilisation des substances sur ces cultures, en fonction des périodes²⁰.

Si les travaux comparant les OCS grande échelle permettent, dans les régions où elles sont disponibles de confirmer que la cartographie des cultures ainsi construite est concluante quant à la localisation effective du fait agricole et des cultures en présence, l'application, sur les surfaces cultivées, des informations issues des matrices cultures-exposition (MCE), permettrait de fournir des probabilités d'utilisation de substances appliquées (en fonction des cultures et des périodes). Elles permettraient ainsi de passer de la caractérisation de l'environnement agricole des sujets, à l'évaluation des probabilités d'exposition aux produits phytosanitaires utilisés en agriculture (pour les MCE disponibles à tout le moins). Un travail similaire, mettant en œuvre les MCE sur les surfaces agricoles déclarées dans le recensement de l'agriculture a été réalisé à l'échelle dans le cadre du projet PestNeuro (Rapport à paraître).

6.3.4 Optimisation technique de la construction de la cartographie et du calcul des indicateurs de proximité

Le traitement de gros volumes de données dans un SIG entraîne des temps de calcul importants et même aboutir à la saturation des capacités de traitement : travailler avec un système de gestion de bases de données type PostgreSQL/FME, ou externaliser les intersections géométriques dans R pourraient permettre d'optimiser les temps de calculs.

6.3.5 Passage de la cartographie sur une maille décimétrique (50 m)

La question de transposer la cartographie sur une maille carrée très fine est à l'étude. 50 m paraît être un bon compromis entre la précision du géocodage classiquement obtenue en zone rurale, les besoins des études épidémiologiques d'analyser des expositions dans des buffers de rayon variant de 100 à 1 500 m, et la granulométrie des bases de données sources. Cette solution permettrait :

- de s'affranchir des difficultés de traitement des artéfacts d'intersection. Lorsque les sources de données sont superposées dans le SIG, du fait de leur granulométrie et de leur mode de recueil différents, les limites d'une même parcelle cultivée (un seul contour dans la réalité terrain) peuvent ne pas coïncider : pour une source, le polygone décrivant un champ va englober la route qui le borde, pour une autre source, la limite sera bien en bordure de la route, pour une troisième, le contour pourra éventuellement mordre sur la forêt qui jouxte la route. La combinaison des trois sources génère des polygones parasites, de surface souvent faible, dont la caractérisation agricole est délicate.
- de combiner plusieurs millésimes du RPG et donc de mieux contrôler la variabilité des cultures pour les surfaces en cultures annuelles soumises aux rotations. Sur une maille carrée, il est possible d'implémenter autant de sources de données que l'on veut sans générer de trop nombreux artéfacts comme évoqués précédemment.

Cette réflexion doit être poursuivie, en tenant compte de la volumétrie de données à traiter que cette méthode est susceptible de générer.

²⁰ Cf. programme Matphyto de Santé publique France et rapports à paraître : 1/ Définition d'indicateurs d'exposition aux pesticides à partir des matrices culture-exposition Matphyto et du recensement agricole. Perrin L., Moisan F., Spinosi J., Chaperon L., Jezewski-Serra D. et Elbaz A. ; et 2/ Matphyto, des matrices cultures-expositions aux produits phytopharmaceutiques ; Principes et méthodologie. Johan Spinosi, Laura Chaperon, Laetitia Perrin, Mathias Lunghi, Katia Touzri, Clémence Fillol.

7. CONCLUSION

La méthodologie de construction d'une cartographie de la présence de cultures, développée par Santé publique France en combinant les informations des différentes bases de données disponibles, permet de répondre à l'objectif de description fine, en tout point du territoire métropolitain, de la présence des cultures agricoles sur une période prolongée. Les bases de données sources se complètent entre elles et les méthodes de raffinement mises en œuvre visent à une description la plus fine possible, ce qui limite au maximum des erreurs de classement. La méthodologie développée, déclinée pour deux dates distinctes dans le cadre du projet Geocap Agri, peut-être remobilisée sur d'autres périodes, en ajustant les données d'entrée afin de répondre au mieux aux besoins d'autres études. La cartographie obtenue permet de décrire finement l'environnement agricole de territoires dans le cadre d'une approche écologique, mais aussi, pour des études dans lesquelles un géocodage des sujets est envisageable, afin de caractériser l'environnement agricole des adresses géolocalisées.

Cette caractérisation de l'environnement agricole reste toutefois un proxy de l'exposition aux produits phytosanitaires agricoles. La poursuite de ce travail, via la spatialisation de données de ventes de produits phytosanitaires ou via la spatialisation fine des MCE, permettrait d'aboutir à une meilleure évaluation de l'exposition aux substances dans les études épidémiologiques. Une réflexion est en cours, afin de préciser les probabilités d'usage de produits phytopharmaceutiques sur les surfaces agricoles détectées par la cartographie.

ANNEXES

Annexe 1. La France, grande puissance agricole, parmi les premiers utilisateurs européens de produits phytosanitaires

La construction d'une cartographie des cultures en tout point du territoire français vise à contribuer à répondre au besoin épidémiologique de connaître l'exposition environnementale de la population générale aux pesticides.

L'estimation de l'exposition aux cultures, estimée géographiquement par la proximité des adresses de lieux de vie à celles-ci, représente donc un moyen d'approcher l'exposition environnementale aux pesticides.

Le terme d'exposition environnementale ne préjuge d'aucune voie d'exposition privilégiée et recouvre toutes les situations potentiellement exposantes liées à la proximité de cultures et à la contamination des milieux/médias par les pesticides : inhalation d'air et de poussières en milieu intérieur et extérieur, contact cutané avec des objets, des animaux, des végétaux ou des personnes, ingestion d'aliments, de poussières et d'eau contaminés localement, etc.

La France est, de loin, le premier pays producteur agricole européen²¹. En 2017, la France à elle seule représente 18% de la valeur des productions végétales européennes et 15% de la valeur des productions animales. La France est le premier pays producteur de céréales, de plantes fourragères, de vins, de pommes de terre et de bétail de l'Union européenne. La production agricole française représente 72,6 milliards d'euros, contre 53,4 pour l'Allemagne, second pays producteur agricole de l'UE.

Avec plus de 50% de son territoire consacré à la production agricole, un potentiel agronomique élevé, et grâce à des savoir-faire qui se maintiennent, la France conserve encore aujourd'hui sa place prépondérante. L'agriculture est le troisième secteur en excédent commercial de l'économie française, derrière le secteur aéronautique et spatial et le secteur de la chimie. La surface du territoire français couverte par l'agriculture correspond à 16% des surfaces agricoles européennes²².

Premier pays agricole de l'UE, la France est le deuxième pays consommateur de produits phytosanitaires (en volumes épandus, 75 287 tonnes en 2014) pour l'agriculture, juste derrière l'Espagne, après avoir longtemps occupé la première place. Rapporté à la surface agricole utile (SAU), cela correspond à une moyenne de 2,7 kg par hectare et par an.

Cependant, à la différence de ses concurrents, la France voit sa production agricole stagner, alors que celle de ses concurrents augmente. En 1980, le secteur agricole représentait 12% de l'emploi total en France, contre 5,5% aujourd'hui. Ce recul de l'emploi agricole est amené à se poursuivre avec les nombreux départs à la retraite à venir dans les prochaines années. D'autre part, la surface consacrée à l'agriculture a chuté en France de 17% depuis les années 1960, soit une perte de près de 60 000 km² (6 millions d'hectares), ce qui équivaut à la surface totale de la région Grand Est. Cette chute des surfaces agricoles concerne toute l'Union européenne. La production française de viande bovine baisse depuis vingt ans, concomitamment à un recul de la consommation individuelle. La production nationale de lait est proche de celle qui prévalait au moment de la mise en place des quotas laitiers dans les

²¹ Rapport d'information du sénat – M. le sénateur Laurent Duplomb, au nom de la commission des affaires économiques, par le groupe d'études « Agriculture et alimentation », sur la place de l'agriculture française sur les marchés mondiaux. Session ordinaire du Sénat 2018-2019 – 28 mai 2019. 31 p. Consulté le 26/09/2022. <http://www.senat.fr/rap/r18-528/r18-5281.pdf>

²² Agriculture française, chiffres clés – données RGA 2010. APCA : Références et études économiques, 2015. 2 p. Consulté le 26/09/2022. https://chambres-agriculture.fr/fileadmin/user_upload/National/002_inst-site-chambres/pages/infos_eco/FicheAgri_Francais.pdf

années 1980, mais avec un cheptel de vaches laitières divisé par deux, ce qui traduit des gains très importants en productivité. Enfin, la production de céréales plafonne depuis de nombreuses années, en raison de la stabilité des surfaces qui lui sont consacrées, mais aussi de ses rendements.

Du fait de ces récentes évolutions, le paysage agricole est en constante évolution, évolution qui s'est accélérée depuis le début des années 2000, qui ont par exemple vu de grandes campagnes d'encouragement à l'arrachage dans le secteur viticole, afin d'accompagner une évolution à la baisse et vers davantage de qualité de la consommation en vin des Français. Par ailleurs, l'évolution de l'agriculture française au cours des dernières décennies est marquée également par une évolution rapide des produits et des familles de substances utilisés dans le traitement des cultures.

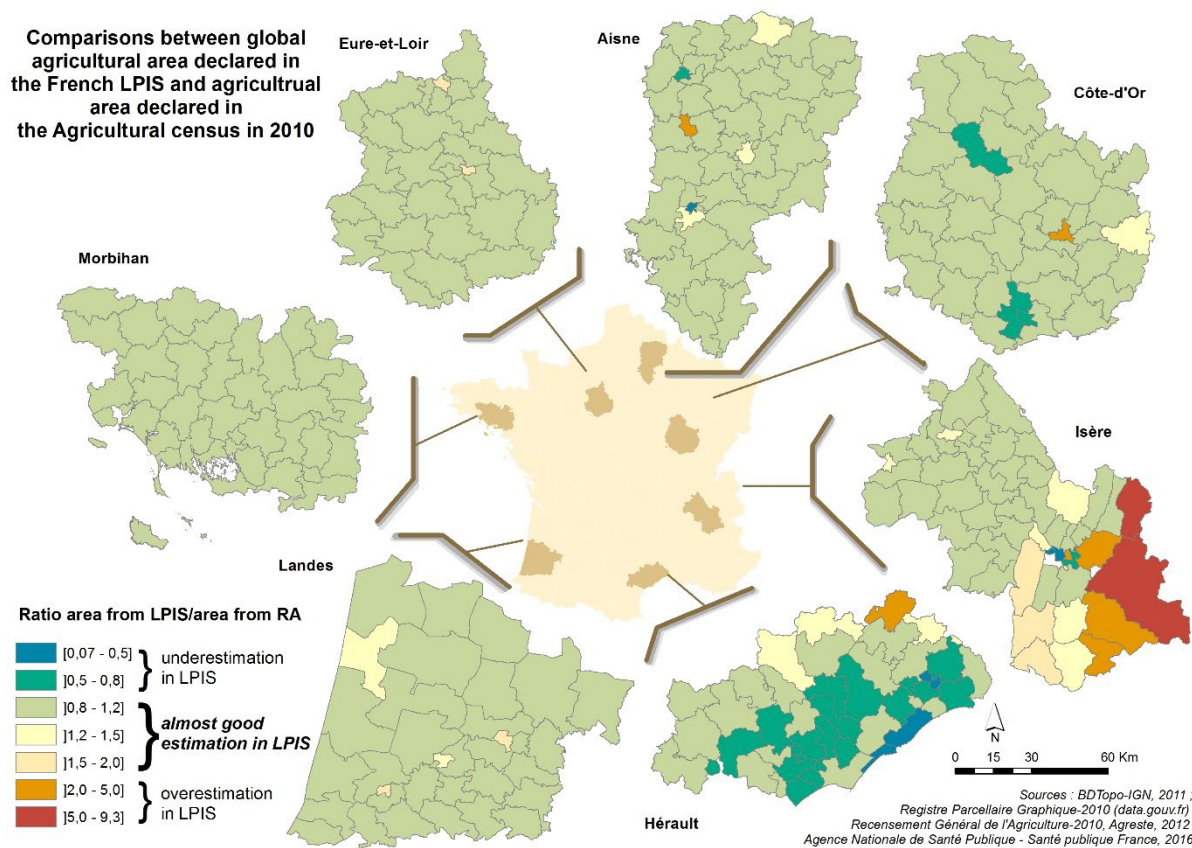
La construction d'indicateurs de présence d'activités agricoles à des fins de caractérisation des expositions dans l'environnement dans des études épidémiologiques doit donc tenir compte de ces évolutions continues survenues dans les deux ou trois dernières décennies. Pour une étude épidémiologique exploitant des données sanitaires recueillies sur une période longue, il convient de veiller à tenir compte de cette évolution des paysages et de l'environnement agricoles des territoires, notamment en utilisant des données de différents millésimes. En effet, l'utilisation de ces informations pour des études épidémiologiques prenant en compte un délai de latence long dans la survenue de certaines pathologies peut nécessiter de pouvoir chercher à reconstituer une exposition sur plusieurs années, voire plusieurs décennies, et, dans certains cas, à considérer un cumul d'exposition vie entière. Dans ces cas, la reconstitution de l'historique de l'exposition à travers l'exploitation de différents millésimes des sources de données existantes est à envisager.

Annexe 2. Apports des différentes sources de données à la cartographie de présence des cultures : Comparaisons des surfaces agricoles disponibles avec les surfaces de référence du RGA

Comparaisons du RPG au RGA

Les données du recensement agricole (RGA) sont exhaustives. Elles peuvent donc constituer une donnée de référence, pour une année de recensement donnée, pour la quantification des surfaces cultivées. En ceci le RGA constitue une donnée intéressante qui peut être comparée aux autres sources de données afin d'en évaluer la complétude et les limites (Figure 7).

Figure 8. Comparaison des surfaces agricoles (toutes cultures confondues) déclarées au RPG et de la SAU du RGA



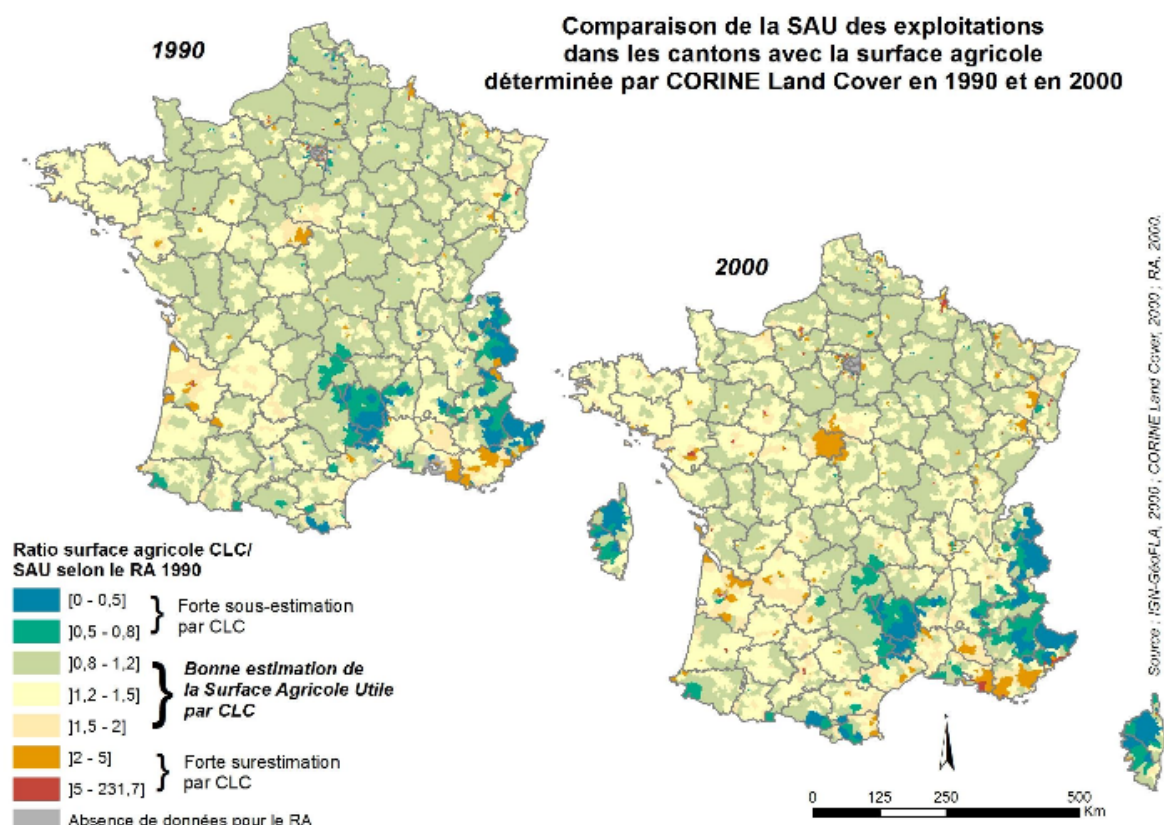
Des comparaisons entre les données du RPG et les données du RGA ont été conduites sur quelques départements métropolitains, retenus pour illustrer de façon représentative la diversité agricole française : l'Aisne, l'Eure-et-Loir, le Morbihan, les Landes, l'Hérault, l'Isère et la Côte-d'Or. Si on compare la totalité des surfaces agricoles déclarées au RPG de 2010 avec les données du RGA 2010, toutes cultures confondues, on observe dans les départements tests une relative concordance entre les données du RPG et les données du RGA (cantons en beige clair). Pour quelques cantons de l'Hérault et de la Côte-d'Or (en bleu sur les cartographies ci-dessous), on note une sous-estimation des surfaces cultivées par le RPG par rapport aux données de référence du RGA, mais cela s'explique assez aisément par la présence importante, dans ces départements, des cultures viticoles ou des vergers, qui ne sont pas assujetties aux aides de la PAC et donc assez nettement sous-déclarées dans le RPG.

(Pour l'Isère, la surestimation - cantons en orangé et en rouge - des surfaces agricoles calculées dans le SIG pour le RPG s'explique notamment par l'absence de prise en compte des pentes lors de la réalisation de l'analyse).

Comparaisons de Corine Land Cover au RGA

Afin d'évaluer les apports possibles de Corine Land Cover au SIG et d'en identifier au mieux les potentielles limites, une comparaison des surfaces agricoles délimitées par CLC a été réalisée avec les données de référence du RGA (Figure 8).

Figure 9. Comparaisons entre les surfaces agricoles détectées par CLC (toute culture confondue) et les surfaces « SAU » de référence du RGA pour les millésimes 1990 et 2000, par canton



Cette étude de sensibilité, réalisée en 2014, avant que ne soit mise à disposition CLC 2012 (disponible en 2015), compare deux éditions de la base de données CLC (1990 et 2000) avec les surfaces du RGA. Les classes d'occupation du sol de CLC ne permettent pas d'identifier avec précision une grande variété de cultures, puisque les seules cultures qui sont spécifiquement isolées par la classification sont les vignes, les vergers, les rizières, les autres regroupements évoquant plutôt des formes agraires que des cultures. Néanmoins, il est déjà intéressant de comparer la surface agricole totale (à l'échelle du canton dans les comparatifs ci-après), toutes cultures confondues, avec la SAU donnée par le RGA pour les mêmes années. En vert clair et en jaune pâle figurent les cantons pour lesquels l'approximation, par CLC, de la SAU, est assez bonne, même si la surestimation (rapport de superficie CLC/RGA autour de 1.2 à 1.5) est assez systématique. Pour les zones en bleu, et en turquoise, CLC sous-estime nettement les surfaces agricoles, tandis que pour les zones en orangé ou en rouge (assez rares), les surfaces agricoles délimitées par CLC sont très nettement surestimées. Ces erreurs d'estimation sont à relier à la méthode de construction de la base de données, qui peut sans doute conduire, notamment dans les zones de montagne, à mal

classer des espaces agricoles (zones de pâturages par exemple) en espaces naturels. En 2000, la surestimation assez marquée des surfaces agricoles est notable, par exemple, dans la région de la Sologne, là encore il faut probablement y voir une erreur de classification (zone naturelle classée à tort comme zone agricole). La surestimation relativement généralisée des surfaces agricoles par CLC s'explique sans difficulté par les spécificités de construction (échelle, unité minimum cartographiée) de Corine Land Cover.

D'autres analyses ont été conduites sur la base de données CLC 2012. La surface agricole totale de cette base a été évaluée à 32 699 875 ha. On évalue la surestimation à 5 730 681 ha des surfaces agricoles pour ce millésime par CLC puisque le RGA 2010 quantifie la surface agricole totale de la France métropolitaine à 26 969 194 ha, soit un ratio surface agricole CLC/SAU RGA de 1.2 en moyenne sur l'ensemble du territoire français. Cette surestimation, dans une perspective d'utilisation de CLC comme base géographique complémentaire aux autres données disponibles pour la construction du SIG, nécessite que soit envisagée la mise en œuvre d'une méthodologie de traitement des données en vue de limiter les erreurs liées à cette surestimation (cf. 4.4).

Une méthode permettant d'affiner l'information fournie par les sources de données géographiques afin de se prémunir autant que possible de la surestimation des surfaces agricole a donc été élaborée. Cette méthode a été élaborée puis testée dans un premier temps sur Corine Land Cover, dans la mesure où les surestimations par CLC des surfaces agricoles étaient particulièrement notoires.

Annexe 3. Apport de l'information « Vigne » de la couche Végétation de la BD TOPO®

La caractérisation des vergers et des vignes dans la couche Végétation de la BD TOPO® a été ajoutée au fil de l'eau par l'IGN aux catégories précédemment présentes dans la base de données entre les années 2007-2017 pour les vergers, et 2010-2017 pour les vignes. La BD TOPO® est mise à disposition en continu par l'IGN, et si la base de données à son acquisition comporte une date de livraison, celle-ci ne saurait toutefois être considérée comme un millésime, année de validité et de dernière mise à jour de l'intégralité des informations présentes dans la base. La première couche livrée mise à disposition par l'IGN à Santé publique France comportant une information homogène sur la caractérisation des vignes et des vergers est datée de 2017. Il est de ce fait nécessaire de s'interroger sur la pertinence de l'intégration de cette source de données pour des études antérieures à cette date.

La vigne et les vergers sont des cultures pérennes, dont la mise en place nécessite plusieurs années avant que les premières récoltes ne soient possibles. Les rotations et assolements n'existent pas pour ces cultures, et il est de ce fait plausible de considérer que les superficies sont relativement stables d'une année sur l'autre, même si cette hypothèse doit être validée dans la mesure où des campagnes d'arrachage ont été conduites à la fin du XX^e siècle et au début des années 2000 pour accompagner l'évolution de la consommation de vin par la population européenne.

Afin de vérifier que cette hypothèse sur la stabilité des surfaces viticoles permet d'envisager d'utiliser une base de donnée dont l'information a été achevée de collecter en 2017, donc très récemment, pour des études portant sur des périodes antérieures, une analyse portant plus spécifiquement sur l'évolution de la viticulture entre les années 2000 et 2018 a été réalisée.

La viticulture est une culture pérenne, dont le cycle de vie varie de vingt-cinq à quarante ans. Les coûts d'implantation comme d'arrachage sont très élevés. La plantation ou l'arrachage de plants de vigne induit donc des conséquences à long terme, et l'adaptation au marché est évidemment plus lente que dans d'autres secteurs de l'agriculture.

La consommation européenne de vins, depuis une quarantaine d'années, a beaucoup évolué. Le consommateur est passé d'un recours fréquent à des vins de consommation courante et de qualité moyenne, à une consommation moins fréquente et plus exigeante. Les quantités consommées ont fortement régressé, au profit d'une demande plus sensible à la qualité des produits. Cette évolution de la consommation a entraîné une surproduction de vins de consommation courante, une crise du secteur viticole, et nécessité assez rapidement une adaptation de la production à l'évolution du marché.

En 1976 est mise en place une interdiction de plantations nouvelles dans les exploitations viticoles, interdiction plusieurs fois prorogée jusque dans les années 2010. Les replantations ne sont autorisées que sous réserve de l'obtention d'une autorisation, suspendue à un arrachage d'une superficie équivalente à celle replantée. La production viticole étant alors largement excédentaire, une incitation à l'abandon définitif des surfaces plantées en vigne a même été conduite. Ces mesures d'incitation, dites aussi « politique d'arrachage », sont ouvertes pour quasiment toutes les superficies en vigne. L'exploitant qui en fait la demande, en fonction de critères d'éligibilité fixés par l'État membre (région, cépage, âge des plants de vigne) reçoit une prime (en moyenne 6 300 euros par hectare arraché), en contrepartie de laquelle il s'engage à renoncer au droit de replantation. En Europe, environ 500 000 hectares de vignes ont été arrachés entre 1992 et 2007, dont 21% pour le seul territoire français. Ces arrachages se sont déroulés selon un rythme soutenu les premières années (50 000 ha par an environ). Ces campagnes se sont ensuite nettement ralenties, pour ne plus concerner que la France ces dernières années, les autres États membres producteurs ayant fait exclure tout ou partie de leur territoire du régime d'arrachage.

En France, les surfaces en vignes ont fortement régressé au cours des années 1980-1990 afin de limiter les excédents. Entre 2000 et 2011, ce recul se poursuit : le vignoble de raisin de cuve français perd 13% de sa superficie. La crise qui touche le secteur viticole dans les années 2000 a entraîné la poursuite de ces arrachages à une période où ils deviennent moins fréquents dans d'autres pays producteurs de l'Union européenne. Depuis 2011, le recul des surfaces est moins important. Entre 2012 et 2017, en cinq ans, le vignoble français perd 10 000 ha, répartis de façon assez homogène sur l'ensemble des régions productrices, avec toutefois une régression un peu plus importante dans certains grands bassins viticoles : la Nouvelle-Aquitaine (vignobles de Bordeaux et de Charentes), l'Occitanie (vignobles du Languedoc), et la région Auvergne-Rhône-Alpes, alors que dans certaines régions, les surfaces en production progressent même légèrement (Grand Est et Bourgogne-Franche-Comté) (Figure 8, Figure 9 et Tableau 10). Le vignoble du Languedoc (Occitanie), traditionnellement largement orienté vers les vins de consommation courante, a été particulièrement touché par la crise de la reconversion de la production viticole, ce qui a entraîné, pour cette région plus que pour d'autres, un recours important à l'arrachage, qui explique que les superficies agricoles aient régressé de façon plus notable en Occitanie que dans d'autres régions productrices²³.

Ces informations permettent d'envisager d'utiliser, sans courir de trop grands risques de biais de classement, les superficies documentées comme plantées en vigne dans la BD TOPO® acquise en 2018 pour compléter l'identification des superficies viticoles repérées par le RPG 2012²⁴.

Ces différentes études menées sur les sources de données existantes permettent de mieux maîtriser leurs atouts et leurs limites et de mieux appréhender la contribution de chacune au SIG.

²³ Patricia Abrantes, Christophe Soulard, Françoise Jarrige et Lucette Laurens, « Dynamiques urbaines et mutations des espaces agricoles en Languedoc-Roussillon (France) », *Cybergeo : European Journal of Geography* [En ligne], Espace, Société, Territoire, document 485, mis en ligne le 13 janvier 2010, consulté le 26 septembre 2022. <http://journals.openedition.org/cybergeo/22869> ; DOI : 10.4000/cybergeo.22869. <http://journals.openedition.org/cybergeo/22869>

²⁴ - Rapport d'information n° 348 (2006-2007) de M. Gérard CÉSAR, fait au nom de la commission des affaires économiques, déposé le 22 juin 2007 <https://www.senat.fr/rap/r06-348/r06-348.html>
- Patricia Abrantes, Christophe Soulard, Françoise Jarrige et Lucette Laurens, « Dynamiques urbaines et mutations des espaces agricoles en Languedoc-Roussillon (France) », *Cybergeo : European Journal of Geography* [En ligne], Espace, Société, Territoire, document 485, mis en ligne le 13 janvier 2010, consulté le 26 septembre 2022. URL : <http://journals.openedition.org/cybergeo/22869> ; DOI : 10.4000/cybergeo.22869. <http://journals.openedition.org/cybergeo/22869>
- France Agri-Mer – évolution 2008-2018. <https://www.franceagrimer.fr/fam/content/download/62836/document/chiffres-fili%C3%A8re-viti-vinicole-2008-2018.pdf>

Figure 10. Évolution de la surface viticole en production en France, entre 1990 et 2018
 (source : DGDDI – FranceAgriMer, 2019)

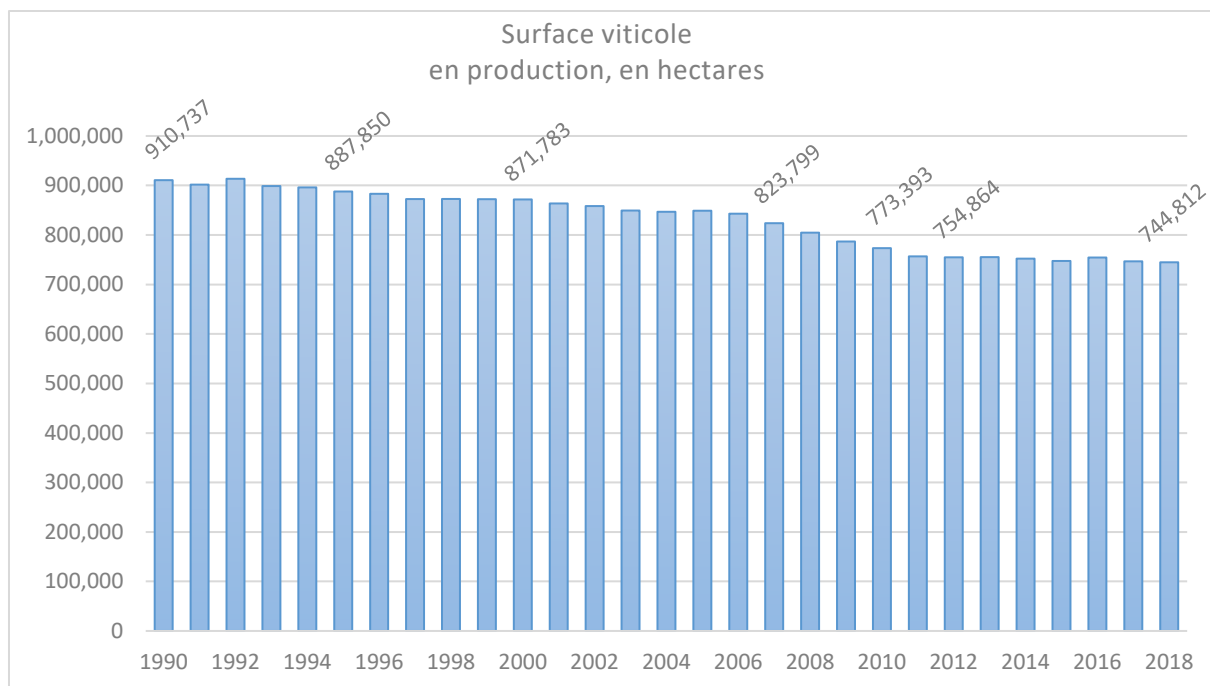


Figure 11. Évolution des superficies viticoles en production, par région, entre 2012 et 2018
 (source DGDDI – Direction Générale des Douanes et Droits indirects – FranceAgriMer, 2019)

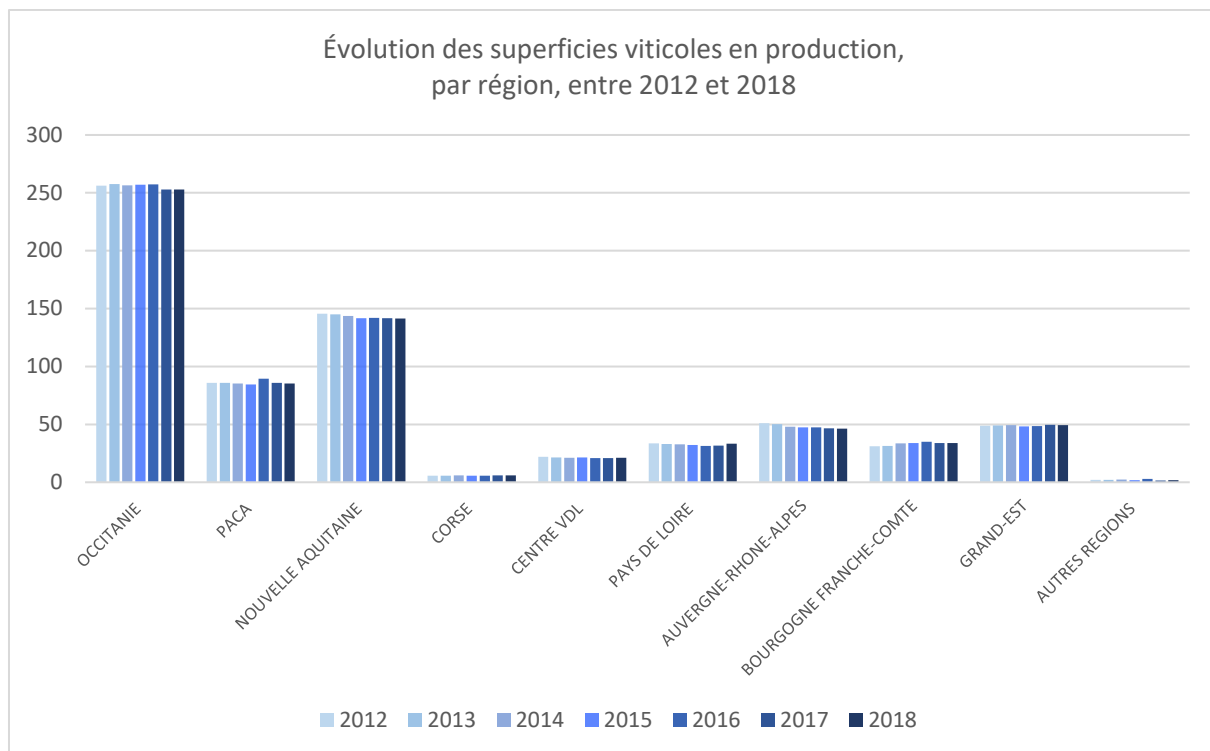


Tableau 10. Évolution de la surface viticole en production en France par région, entre 2012 et 2018

Régions	2012	2013	2014	2015	2016	2017	2018
Occitanie	256,3	257,5	256,5	257,1	257,2	252,9	252,8
PACA	86	86	85,4	84,6	89,6	86	85,2
Nouvelle-Aquitaine	145,6	145,1	143,6	141,6	142	141,8	141,4
Corse	5,8	5,8	5,9	5,8	5,8	5,9	5,9
Centre-Val de Loire	21,9	21,4	21,2	21,4	20,8	21	21,1
Pays de la Loire	33,5	33,1	32,7	32,3	31,5	31,6	33,3
Auvergne-Rhône-Alpes	51	50,2	48,1	47,5	47,5	46,5	46,3
Bourgogne-Franche-Comté	31,1	31,5	33,6	33,8	34,9	33,8	33,9
Grand Est	48,7	49,2	49,4	48,3	48,6	49,6	49,3
Autres régions	2	2,2	2,3	1,8	2,8	1,6	1,7

Annexe 4. Détail de la sélection des sources dans le cadre de l'étude Geocap Agri

Emprise géographique

Certaines données décrivent les activités agricoles sur des territoires plus ou moins vastes. Ainsi, les OCS GE par exemple ne couvrent que des départements ou des régions, d'autres données produites par ou pour divers organismes locaux couvrent même parfois des territoires plus réduits encore. L'exploitation de ce type de données est difficilement envisageable dans le cadre des objectifs du projet Geocap Agri. En effet, afin de limiter les biais et de faciliter leur qualification et quantification, il a paru plus pertinent de travailler sur des données disponibles sur l'échelle qui correspond à l'emprise géographique de l'étude, soit des sources de données qui couvrent l'ensemble de la France métropolitaine. Pour les données d'occupation du sol grande échelle, les justifications ont été données :

- D'une part, les données des OCS GE ne sont pas disponibles partout en France, seule une faible partie du territoire est finalement couverte par ces données-là. Cela est d'autant plus vrai pour les données « métiers » issues des Dreal, des chambres d'agriculture, etc.
- D'autre part, la construction de ces données par de multiples structures donne lieu à de multiples résultats, chacun comportant ses propres biais. La quantification de ces derniers paraît difficilement envisageable et donnerait lieu à une importante hétérogénéité de l'information à l'échelle de la France métropolitaine dans son ensemble, avec le risque d'avoir des territoires « mieux » couverts que d'autres.

Enfin, ces données ne sont pas générées aux mêmes dates et leur méthode de création n'est pas automatisée ou normée de manière à obtenir une donnée homogène.

Pour ces trois raisons, ces données locales ne peuvent pas être intégrées dans l'indicateur avec un objectif de caractérisation homogène de la présence des cultures en tout point du territoire. Il importe cependant de pouvoir les mobiliser dans le cadre d'études aux objectifs différents (études locales par exemple) ou éventuellement à des fins de validation locale d'un indicateur global.

Amplitude temporelle

Pour définir la contribution des données à la construction de l'information d'exposition à mobiliser pour l'étude Geocap Agri, il est nécessaire de tenir compte de la période d'étude sur laquelle porte le recueil des données sanitaires, soient les cancers pédiatriques diagnostiqués entre 2002 et 2013.

Ce paramètre est important puisque les sources d'information existantes pour localiser le fait agricole ne sont pas mises à jour annuellement. De plus, certaines de ces sources de données sont récentes et il est difficile d'envisager un historique long pour certaines sources. Corine Land Cover n'est actualisée que tous les six ou dix ans, et la première édition remonte à 1990. Le Registre parcellaire graphique qui est la base la plus exhaustive pour décrire les types de cultures est générée chaque année mais seulement à partir de 2007. Quant aux classes de végétation Vigne et Verger de la BD TOPO®, elles ont été construites sur une période courant entre 2012 et 2017 et n'a pas été mise à jour depuis. Le Tableau 11 résume l'étendue temporelle de ces sources de données.

Tableau 11. Dates de disponibilité des différentes bases de données mobilisables sur la période d'étude de Geocap Agri

Source de donnée	Dates de mise à jour des données
Corine Land Cover	1990 ; 2000 ; 2006 ; 2012 ; 2018
Registre parcellaire graphique	À partir de 2007, tous les ans
Vignes et vergers et la BD TOPO®	Entre 2010 et 2017, pas avant 2017 pour disposer de la base sur l'ensemble des départements
Recensement général agricole	1970, 1979, 1988, 2000 et 2010

Le Casier viticole informatisé, information précise et exhaustive sur les surfaces viticoles en production, n'est pas disponible avant 2016. Même en tablant sur une certaine stabilité des surfaces pour des cultures pérennes comme la vigne, il est néanmoins inenvisageable de retenir cette source de données pour une analyse dont la période d'étude s'arrête en 2013. De ce fait, le casier viticole informatisé n'a pas été implémenté pour l'étude Geocap Agri. Il convient néanmoins de retenir que cette base de données, disponible pour des millésimes plus récents, constitue une source d'information majeure, à mobiliser très certainement dans le cas de travaux portant sur des expositions plus actuelles.

Pour répondre à ces différents critères d'homogénéité géographique et de contrainte temporelle, les données retenues pour contribuer à la construction de l'indicateur pour l'étude Geocap Agri sont :

- Le Recensement Général Agricole ;
- Le Registre Parcellaire Graphique ;
- Corine Land Cover ;
- La sélection des classes Vignes et vergers et la BD TOPO® « Végétation » pour les dernières années de la période d'étude.

Au vu de la disponibilité des données identifiées pour la construction de l'indicateur de proximité aux cultures, il est possible d'identifier deux « sous-périodes » d'étude (Tableau 12) :

- Une première période, de 2006 à 2009, pour laquelle les sources de données permettront de construire une information sur la proximité aux cultures de qualité 'moyenne' : le RPG n'est disponible qu'à partir de 2007, il ne peut donc être mobilisé qu'à partir de cette date. La BD TOPO® informant sur la présence de vignes et de vergers ne saurait être utilisée avant 2010 dans la mesure où l'information spécifique sur la présence de vignes et de vergers n'a été ajoutée dans cette base qu'entre 2012 et 2015. Corine Land Cover (2006) est certes disponible, mais avec les limites qui lui sont propres. L'information de proximité aux cultures pour cette sous-période ne pourra donc qu'être de qualité moyenne, puisqu'il n'est pas possible d'envisager de descendre à une échelle fine avant la mise à disposition du premier millésime du RPG en 2007, ce qui implique de considérer que la répartition géographique fine des activités agricoles pour cette année-là soit représentative de l'ensemble de la période courant de 2002 à 2009, ce qui induit nécessairement des incertitudes difficiles à mesurer.
- Une seconde période, bien couverte, de 2010 à 2013 : des sources de données plus nombreuses sur cette période permettent la construction d'un indicateur d'exposition multi-source plus robuste ; le RGA de 2010 fournit une information de référence, et on peut utiliser conjointement CLC 2012 et le RPG 2012, pour affiner la localisation des cultures. Les classes vignes et vergers de la BD TOPO®, renseignées entre 2010 et 2017, viennent compléter l'information sur ces cultures pérennes en particulier, qui se trouvent être aussi particulièrement soumises aux traitements phytosanitaires.

Tableau 12. Couverture des périodes de l'étude en fonction des données disponibles

Qualité de la couverture par les données	Périodes	Couches présentes et dates
Correctement couverte	2010-2013	RPG ; CLC (2012 uniquement) ; VV ; RGA 2010
Moins bien couverte	2002-2009	CLC ; RGA (à la commune) ; RPG à partir de 2007 seulement (avec donc une période 2002-2006 assez mal couverte)

Ces deux périodes (2002-2009) et (2010-2013) font appel à des données différentes.

Pour la période 2002-2009, les données utilisées sont :

- RPG 2007 : le plus ancien millésime disponible pour cette source ; la plus précise des sources de données pour localiser le fait agricole ; la plus exhaustive des données en terme de typologie de cultures disponibles bien que la précision soit altérée par le mode de regroupement à l'îlot ; une donnée incomplète à cause des sous déclarations pour certaines cultures non assujetties aux aides PAC ;
- CLC 2006 : la plus corrélée temporellement au RPG 2007 ; moins précise dans sa description des cultures et du fait agricole, elle permet néanmoins de confirmer la présence de cultures renseignées par le RPG ;
- RGA 2000/2010 : ces valeurs bien que biaisées par le mode d'agrégation des surfaces à la commune du siège de l'exploitation et de trois à quatre ans postérieures aux autres données, pourront nous servir à estimer la justesse globale de notre indicateur.

À partir de ces données, combinées, est construite la cartographie 2007 considérée comme représentative des activités agricoles sur la période 2006-2009.

Pour la période 2010-2013, les données utilisées seront :

- RPG 2012 : la plus précise des sources de données pour localiser le fait agricole ; elle correspond le mieux à la période étudiée ; la plus exhaustive des données en matière de typologie de cultures disponibles bien que la précision soit altérée par le mode de regroupement à l'îlot ; toujours incomplète mais moins que pour la première sous-période d'étude puisqu'à partir de 2010 un agriculteur devra déclarer toutes ses parcelles agricoles dès lors qu'une de ses parcelles fera l'objet d'une demande de prime à la PAC, avant 2010 seules les parcelles pour lesquelles une aide était demandée devaient être déclarées.
- CLC 2012 : la plus corrélée au RPG2012 ; moins précise dans sa description des cultures et également dans sa description du fait agricole ; elle permettra d'augmenter la vraisemblance de présence du fait agricole.
- VV 2010-2017 : source de donnée qui a influencé le choix des millésimes pour le RPG et CLC pour la période B ; la plus précise et la plus exhaustive pour représenter la vigne et les vergers ; sa construction sur plusieurs années apporte un biais important mais la pérennité de ces cultures permet d'envisager de l'utiliser en confiance, et à la considérer comme source de donnée clé de l'indicateur pour la période B.
- RGA 2010 : malgré le mode d'agrégation des surfaces déclarées à la commune du siège de l'exploitation et de deux ans antérieures aux autres données, les données du RGA restent la source de données la plus exhaustive pour la description des surfaces cultivées et sont utilisées comme données de référence pour estimer la justesse globale de la cartographie.

À partir de ces données, combinées, est construite la cartographie 2012, considérée comme représentative des activités agricoles entre 2010 et 2013.

Précision

Un des objectifs de l'étude Geocap Agri est d'identifier les activités agricoles présentes à proximité des cas/témoins recensés par le registre des cancers de l'enfant. Les adresses des sujets ont été géocodées²⁵ pour obtenir leurs coordonnées x et y, ce qui permet d'exploiter leur géolocalisation précise dans un SIG et d'étudier la présence d'activités agricoles autour de chacun d'eux. Les opérations de géocodage ont fait l'objet de vérification et de contrôles qualité. La qualité de la localisation fournie est globalement élevée, ce qui implique, en regard :

- Que la définition des cultures identifiées et localisées autour de chaque sujet soit précise sur le plan sémantique,
- Et que la localisation comme la quantification des surfaces cultivées soient elles aussi précises, d'autant que les hypothèses sur les distances d'exposition aux produits phytosanitaires épandus sur les cultures imposent une recherche dans des rayons réduits (la distance d'exposition prise en compte est de 1250 m au maximum autour des sujets)

Ce niveau de détail nécessaire tant en terme de classification que de localisation et de quantification ouvre la réflexion sur la capacité réelle de l'indicateur à identifier précisément une culture (blé, maïs, etc.) que ce soit pour la **première ou la seconde « sous-période » de l'étude Geocap Agri**. Effectivement, seul le RPG distingue le blé, du colza, du tournesol, etc., cependant, sa précision n'est pas infaillible : le regroupement de parcelles à l'îlot de culture majoritaire jusqu'en 2014 limite sa précision sémantique, et sa non exhaustivité pour certaines cultures non assujetties aux aides en limite également la précision géographique. Cette précaution s'impose d'autant plus qu'il a pu être observé au fil des années que certains agriculteurs ne délimitaient pas forcément avec la plus haute précision les contours de leurs surfaces cultivées, englobant chemins d'exploitation, haies, etc., correspondant sur le terrain à des surfaces non cultivées. Par ailleurs, l'utilisation d'un millésime unique du RPG (2007 pour la première sous-période 2002-2009, 2012 pour la seconde sous-période 2010-2013) pour plusieurs années amène à considérer avec prudence, sur la totalité de la période, les surfaces cultivées en grandes cultures souvent intégrées à des processus de rotations des cultures : d'une année sur l'autre, les parcelles sont plantées différemment (par exemple : une année du blé, une autre du lin, une troisième une légumineuse ou du maïs fourrager). La prise en compte de regroupements ou de grandes catégories de cultures, dans la mesure où elle reste cohérente avec les usages de produits phytosanitaires, peut permettre de s'affranchir de ces imprécisions et de ces questions de rotation de culture. Des analyses de sensibilité évaluant les retombées du changement de millésime du RPG dans la construction de la cartographie seraient à conduire pour approfondir cette question.

L'information sur la présence de vignes ou de vergers présente dans la BD TOPO® après 2017 ne peut être mobilisée sur la première sous-période (2002-2009). Enfin, CLC ne permet de distinguer que ces deux dernières cultures en plus des rizières, des prairies et des jachères ; sinon, la description des territoires agricoles y est très généralisée et distingue plutôt des formes d'organisation agraire que des cultures à proprement parler. La localisation du fait agricole y est moyennement précise, avec une tendance à surestimer les surfaces agricoles,

²⁵ Le géocodage est un procédé qui permet de passer d'une adresse postale à un point géolocalisé dans une couche de données géographiques. La qualité du géocodage dépend de la qualité de l'adresse recueillie au départ, de la précision de la base adresse mobilisée pour attribuer aux adresses recueillies un couple de coordonnées dans un référentiel géographique donné, et des algorithmes de géocodage qui permettent d'interpoler la localisation d'une adresse recueillie entre deux points adresses référencés dans la base adresse utilisée comme « localisateur ». Le géocodage peut être réalisé dans tout SIG, il est parfois confié à des sociétés prestataires qui ont fait de la qualité des résultats de ces opérations et de l'optimisation de leurs algorithmes de géocodage un argument commercial majeur.

et, plus ponctuellement, également des sous-estimations liées à des erreurs d'interprétation des images satellites.

Les informations utilisées pour localiser la vigne et les vergers sont de qualité fiable pour la **seconde sous période** (à partir de 2010), avec la possibilité de combiner les informations du RPG et de la BD TOPO®. Vignes et vergers sont des cultures pérennes, cependant les surfaces en production évoluent, des mesures d'arrachage ont été prises et ont encouragé les producteurs à accompagner l'évolution des modes de consommation européens et même si la localisation géographique de ces cultures est susceptible de rester assez stable d'une année sur l'autre, il n'est cependant pas possible de considérer que les bases de données disponibles pour la fin de la période d'étude (2010-2013) puissent être représentatives du début (2002-2009).

Ces constats amènent à envisager que la méthode de construction de l'indicateur se doit d'être évolutive entre les deux sous-périodes qui subdivisent la période d'étude Geocap Agri, et à accepter de ne pouvoir disposer de la même information en termes de précision sémantique sur la présence des cultures entre ces deux sous-périodes. Des analyses de sensibilité sont également menées par l'équipe Epicea pour étudier l'apport réel de la couche Vigne et Verger dans la seconde période.

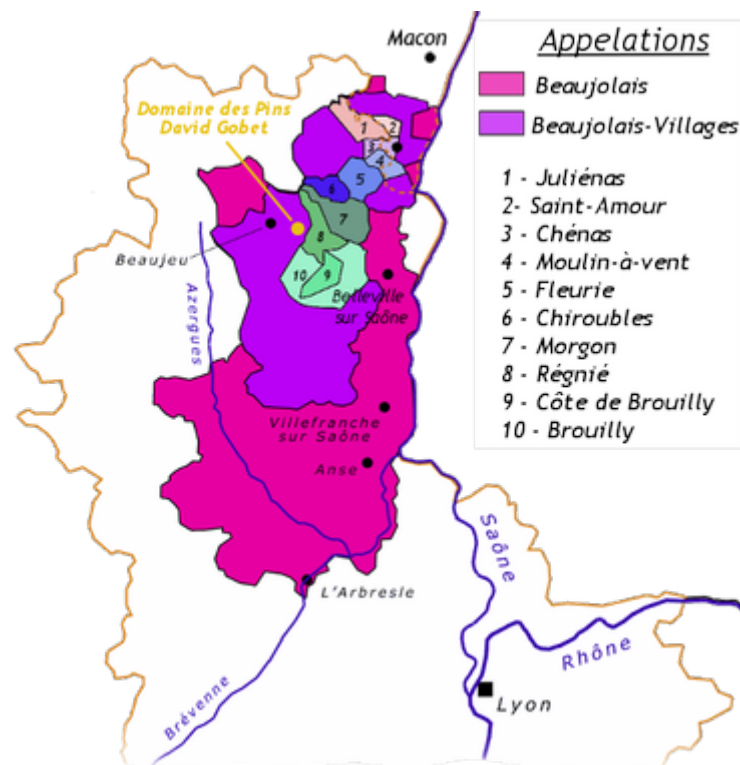
Annexe 5. Apports et limites du casier viticole informatisé (CVI)

Le CVI est une source de données dont l'utilisation est rendue complexe par la nécessité d'un appariement avec une base de données vectorielle du cadastre. Son obtention est également délicate comparée aux autres sources de données présentées ici et susceptibles d'entrer dans la construction de la cartographie des cultures. De ce fait, le choix de l'intégrer à la cartographie des cultures doit être interrogé à l'éclairage de la connaissance de la période d'étude de l'étude épidémiologique dans laquelle elle est employée comme proxy de l'exposition aux pesticides. Afin d'étudier les apports du CVI pour la localisation des surfaces viticoles par comparaison aux autres sources disponibles, une analyse a été conduite sur une zone viticole réduite, le vignoble du Beaujolais en région Auvergne-Rhône-Alpes, pour lequel les données étaient mises à disposition dans un jeu de données dans le cadre de l'open-data par la Draaf (Direction régionale de l'alimentation, de l'agriculture et de la forêt). Les opérations suivantes ont été effectuées :

- Récupération de la table des parcelles viticoles du CVI pour la région Rhône-Alpes (ancien découpage, sans l'Auvergne). Cette table est mise à disposition sur la plateforme de téléchargement de l'infrastructure régionale de données géographiques de la région par la Draaf, avec l'autorisation de la DGDDI (détentricrice de la donnée) et de FranceAgriMer qui exploite ces données dans le cadre de ses activités.
- Jointure nécessaire avec la couche des parcelles de la BD Parcellaire (composante du Référentiel grande échelle de l'IGN).
- Choix du département du Rhône (d'autres départements de la région sont incomplètement couverts par la version 2011 de la BD Parcellaire mise à disposition par l'IGN à Santé publique France). Dans la moitié nord du département du Rhône (69) se trouve notamment le vignoble du Beaujolais (Figure 10), à l'ouest de la Saône.

Dans le département du Rhône on détecte, grâce aux différentes sources de données disponibles, deux zones viticoles : le Beaujolais, qui s'étend assez largement dans la moitié nord du département, et une autre zone viticole moins étendue localisée au sud de Lyon.

Figure 12. Localisation géographique du vignoble du Beaujolais²⁶



Le Registre parcellaire graphique (RPG)

La superposition du RPG (violet foncé) aux parcelles viticoles du CVI sur le Rhône illustre parfaitement le manque d'exhaustivité de cette base par rapport à la réalité du terrain. Ce manque d'exhaustivité est une limite bien identifiée de cette base de données, en particulier sur des cultures comme la vigne ou l'arboriculture (Figure 11). Ce constat conforte la position selon laquelle le RPG ne pourrait être utilisé seul pour évaluer avec justesse la présence de ce type de cultures.

Corine Land Cover - vignes

La superposition de la classe des vignes par rapport aux parcelles viticoles du CVI donne d'assez bons résultats, notamment dans le Beaujolais, où les zones cultivées en vignes s'étendent sur des superficies assez importantes (et supérieures au seuil de représentation dans CLC). CLC détecte donc bien ces vignes-là et ne les surestime apparemment pas trop. Par contre, sur la zone viticole plus resserrée qui se trouve au sud du département, la couverture par CLC-Vignes est un peu moins bonne.

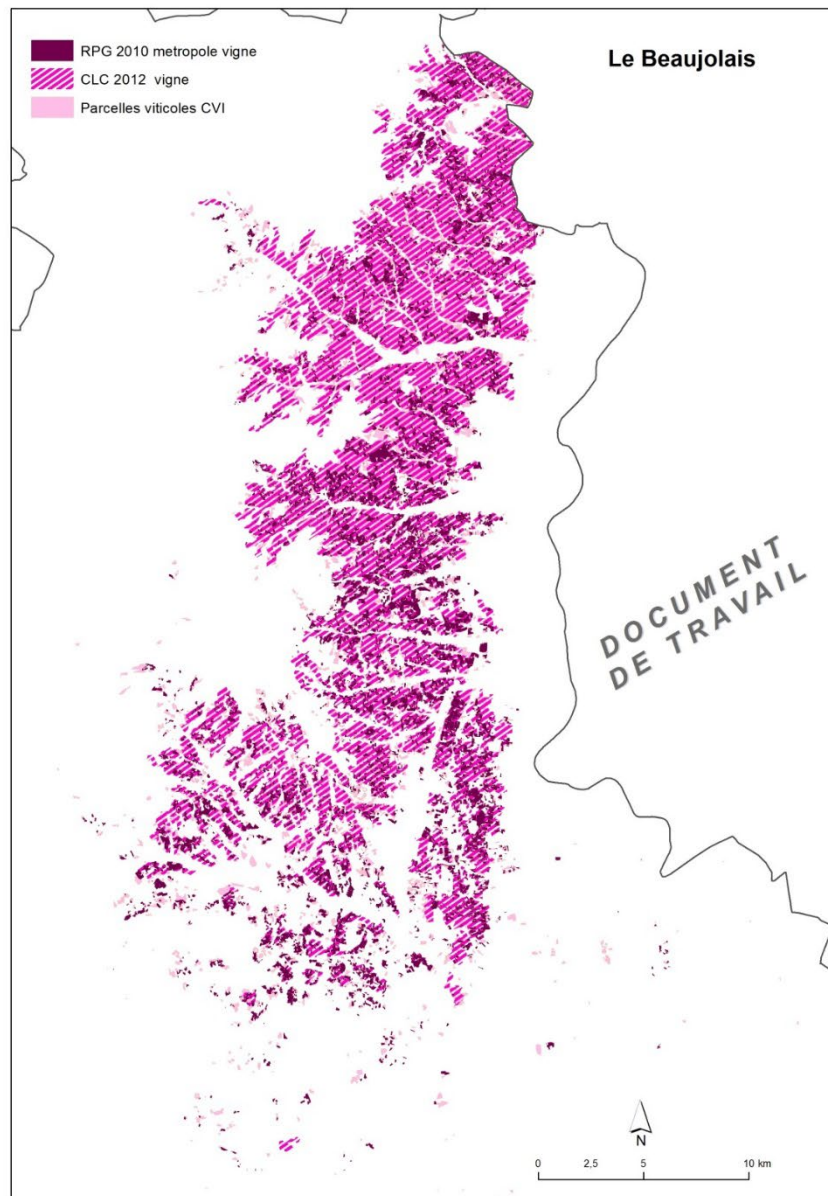
Corine Land Cover – localisation de la surface agricole

Afin de s'assurer que les sources de données étaient susceptibles de renseigner de façon assez juste la présence des vignes dès lors qu'on les utilise en les combinant, un test est également réalisé en cartographiant la surface agricole extraite des données de CLC puisque celle-ci est aussi utilisée pour délimiter les zones de présence vraisemblable des cultures dans l'indicateur (Figure 11).

²⁶ <https://www.lepetitballon.com/blog/la-veritable-histoire-du-beaujolais-nouveau.html>

La surface agricole délimitée par CLC recouvre quasiment intégralement les superficies identifiées par le CVI (cf. Figure 8 : Localisation des zones viticoles identifiées par le CVI non détectées par CLC et le RPG). On peut en déduire qu'a priori, les sources de données se compléteront assez bien pour repérer la vigne. Seules quelques petites parcelles du CVI tombent à l'extérieur de la surface agricole de CLC.

Figure 13. Évaluation de CLC Vignes et du RPG pour la localisation de la présence des cultures viticoles par comparaison avec les parcelles du CVI

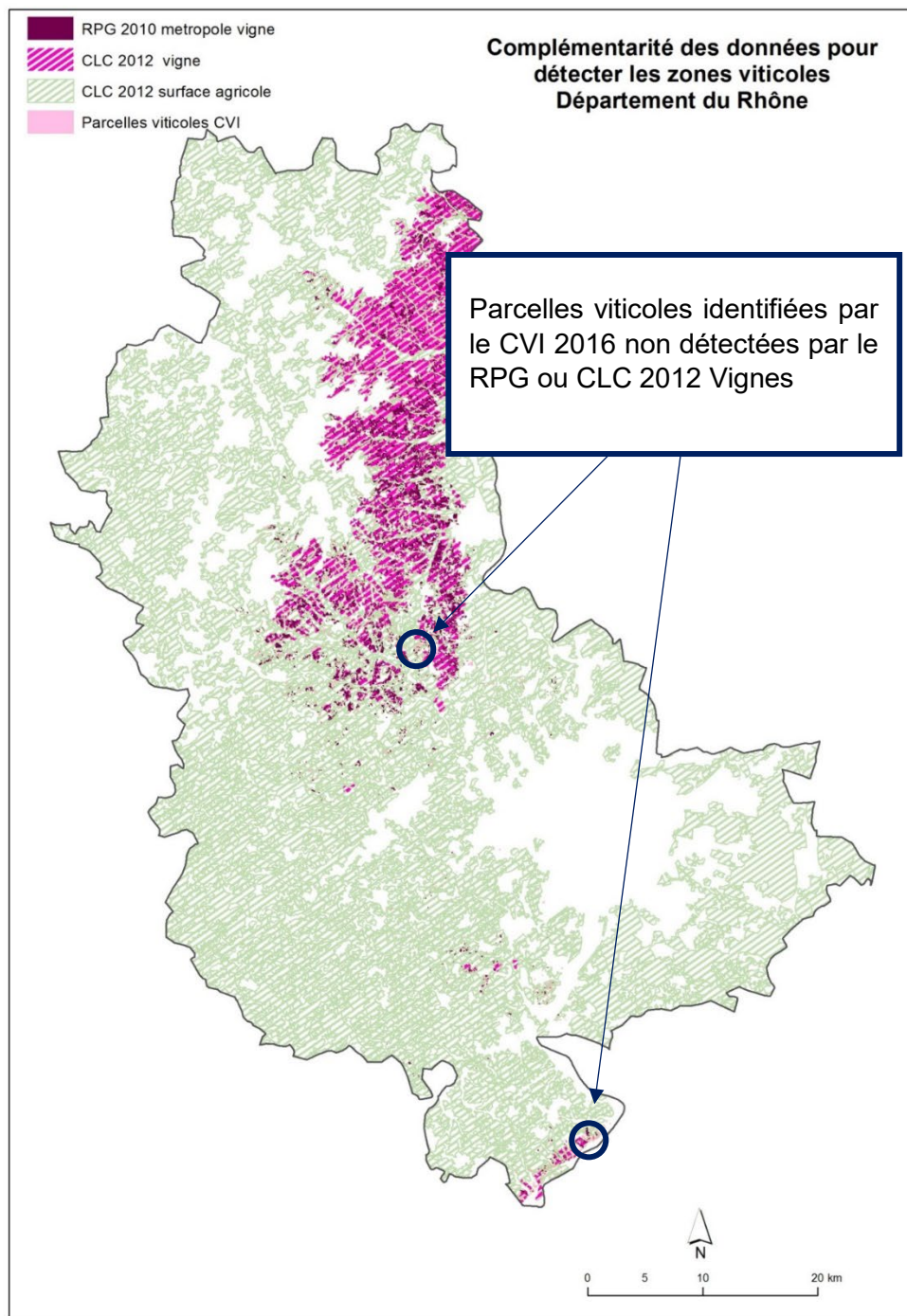


Il importe d'identifier les classes d'occupation du sol qui les englobent, afin de ne pas commettre d'erreur en utilisant les autres classes de CLC pour repérer les zones non exposées.

Identification de l'occupation du sol dans les zones où les parcelles viticoles ne sont pas détectées par les sources de données utilisées dans la construction de l'indicateur

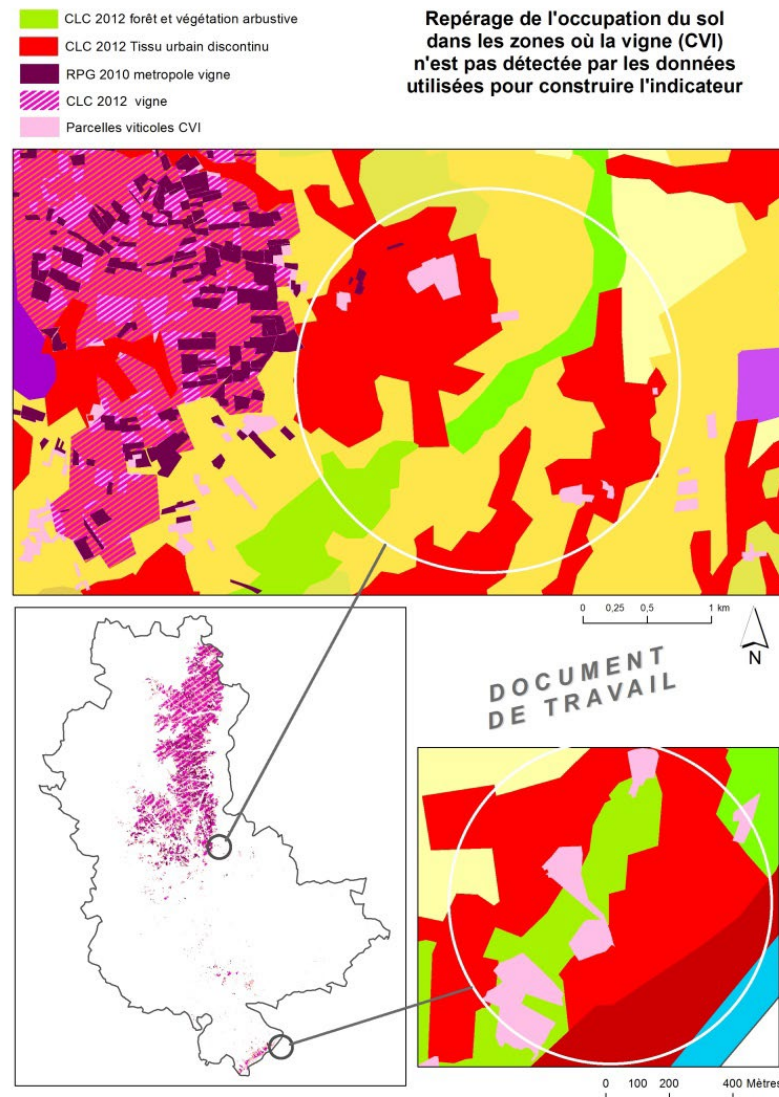
Visuellement deux zones assez peu étendues en superficie semblent n'être détectées par aucune des sources de données mobilisées dans la construction de l'indicateur. L'une de ces zones se trouve en bordure méridionale du Beaujolais et l'autre au sud du département (Figure 12)

Figure 14. Localisation des zones viticoles identifiées par le CVI non détectées par CLC et le RPG



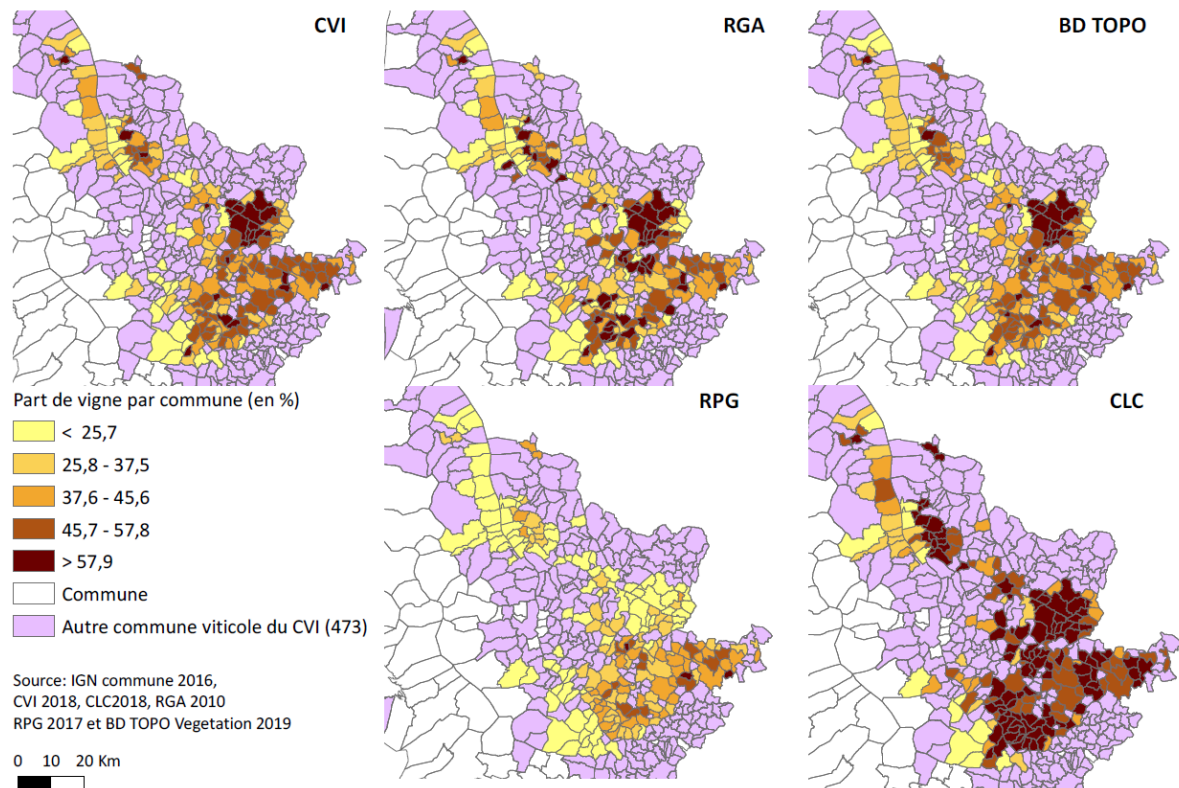
L'occupation du sol qui caractérise ces zones se partage entre tissu urbain discontinu (ce qui, pour le cas de la vigne, n'est pas une surprise, car l'intrication du bâti avec ce type de cultures est assez répandue), et de la végétation arbustive en mutation (avec laquelle des vignes ont pu être confondues éventuellement lors de la constitution de la base CLC). Les superficies concernées, qui échappent à l'identification par les données du RPG ou de CLC, sont assez faibles : autour de 9 ha pour la plus étendue au sud du Beaujolais, et environ 1 ha pour les parcelles du sud du département. Ces parcelles sont toutes plantées de vigne jeune (entre 0 et 10 ans) d'après le CVI.

Figure 15. Caractérisation par CLC des zones viticoles détectées uniquement par le CVI



Une autre analyse comparant les sources de données disponibles pour caractériser les surfaces viticoles sur quelques communes de Gironde (Figure 15) montre par ailleurs que la source de données Vignes et Vergers caractérisés par la BD TOPO® décrit assez bien la présence de vignes documentée par le CVI, ces deux sources de données fournissant des informations très proches. Les millésimes disponibles ne remontent pas à des dates antérieures à 2016, ce qui rend l'utilisation du CVI difficilement envisageable dans le contexte spécifique d'études épidémiologiques dont la période d'étude n'inclurait pas cette date (cf. 2.1.4 Le Casier viticole informatisé - CVI).

Figure 16. Comparaison de la part de vigne sur quelques communes de Gironde en fonction de différentes sources de données – Réalisation dans le cadre de l'étude Pestiriv (Santé publique France)



En résumé, le CVI constitue donc indéniablement une source de données incontournable pour l'étude de la localisation des cultures viticoles. Très précise du fait de sa construction sur la base du cadastre et des objectifs de suivi et de contrôle qu'elle sert, cette base n'est cependant pas construite pour l'usage géographique que nécessitent les objectifs des études épidémiologiques. Sa relative exhaustivité pour les informations concernant la vigne en font une source d'information intéressante et sensible, cependant, la complexité de sa mise en pratique en fait une donnée difficilement mobilisable pour des études couvrant des périodes plus anciennes que la mise à disposition d'un parcellaire cadastral vectoriel exhaustif sur l'ensemble du territoire national.

Annexe 6. Traitement des superpositions dans l'étude Géocap Agri

Dédoublonnage de la cartographie 2012 – sources RPG, CLC, VV (la plupart des duplications proviennent de la couche VV)

- Si polygone plus qu'en double (triples, quadruples, etc.)
Peu de cas et surfaces petites
 - On garde un polygone aléatoirement (pour garder une bonne estimation de la surface en culture totale)
- Si polygone en double
 - Si les deux polygones sont parfaitement identiques (mêmes source(s) et type de culture)
 - On garde un des deux polygones
 - Si la seule différence vient de la source VV (un polygone sans VV et un avec VV, le reste étant identique par ailleurs)
 - On garde le polygone avec l'information VV
 - Si la seule différence vient des sources RPG et/ou CLC (un polygone sans et un avec) avec VV présent
 - On garde le polygone le plus complet, i.e. avec l'information VV + RPG et/ou CLC
 - Si la seule différence vient du type de culture trouvé par VV (un polygone en viticulture et l'autre arboriculture)
 - On garde un des deux polygones et on renomme le type de culture en « Viticulture ou Arboriculture »
 - Tous les autres cas (5% des doublons)
 - On garde un polygone aléatoirement (pour garder une bonne estimation de la surface en culture totale)

Dédoublonnage de la cartographie 2007 – sources RPG, CLC (les duplications sont toutes dues à la couche RPG)

- Si polygone plus qu'en double (triples, quadruples, etc.)
Peu de cas et surfaces petites
 - On garde un polygone aléatoirement (pour garder une bonne estimation de la surface en culture totale)
- Si polygone en double
 - Si les deux polygones sont parfaitement identiques
 - On garde un des deux polygones
 - Si la seule différence vient du type de culture trouvé par RPG et qu'un des deux polygones a comme type « 0 » ou « »
 - On garde le polygone avec une véritable culture
 - Si la seule différence vient du type de culture trouvé par RPG et qu'un des deux polygones a comme type « Prairies »
 - On garde le polygone non Prairies (car « Prairies » n'est pas considéré dans l'étude, on privilégie la VPP à la VPN)
 - Tous les autres cas
 - On garde un polygone aléatoirement (pour garder une bonne estimation de la surface en culture totale)

Annexe 7. Choix de la source principale et donc de la culture principale

Chaque polygone (dans un buffer ou sur la cartographie complète) est associé à une ou plusieurs sources d'information et un ou plusieurs types de cultures, qui peuvent ne pas être concordants.

Dans le cadre du projet Geocap Agri, nous avons déterminé pour chaque polygone la source d'information principale (et la culture associée).

Pour chaque polygone, on suit l'arbre de décision suivant :

- Si une des sources est VV avec comme culture « Viticulture » et qu'il y a un désaccord sur le type de culture entre les sources RPG et CLC (on considère qu'il y a aussi désaccord si une des deux sources est absente)
 - o Source principale = VV et culture principale = Viticulture
- Sinon, si culture VV = « Arboriculture » ou « Viticulture ou Arboriculture » et culture CLC = « Arboriculture » ou « Viticulture » et culture RPG différente de « Viticulture » ou « Arboriculture »
 - o Source principale = VV et culture principale = celle de VV
- Sinon, si source RPG présente
 - o Source principale = RPG et culture principale = celle de RPG
- Sinon, si source VV présente
 - o Source principale = VV et culture principale = celle de VV
- Sinon
 - o Source principale = CLC et culture principale = celle de CLC

Si on raisonne uniquement sur les sources RPG et CLC, par exemple si l'on souhaite étudier une période commençant avant 2010, l'algorithme devient alors très simple :

- Si source RPG présente
 - o Source principale = RPG et culture principale = celle de RPG
- Sinon
 - o Source principale = CLC et culture principale = celle de CLC

Annexe 8. Méthode de raffinement des données pour limiter la surestimation des surfaces agricoles

Lors des travaux préliminaires de comparaison des sources de données disponibles, la surestimation des superficies allouées à l'agriculture, notamment par CLC, a été mise en évidence. Des traitements complémentaires de raffinement pour supprimer les emprises de surfaces non agricoles (bâtiments, végétation naturelle, surfaces en eau, etc.) caractérisées par d'autres bases de données géographiques d'occupation du sol d'échelle plus fine, ont été menés. Un premier test de raffinement de CLC 2012, en extrayant les superficies de végétation naturelle, de surfaces en eau et de grosses infrastructures routières décrites par la base de données BD TOPO® 2011 de l'IGN a permis de réduire cette surestimation de façon notable :

Sur 32 699 875 ha de surface décrite comme agricole par Corine Land Cover 2012, on évalue la surestimation à 5 730 681 ha (puisque le RGA 2010 quantifie la surface agricole totale de la France métropolitaine à 26 969 194 ha). Grâce au premier raffinement, 4 315 518 ha de surfaces non agricoles ont pu être extraites des surfaces décrites comme agricoles par Corine Land Cover, soit plus de 75% de l'excédent.

Ce constat a conduit à approfondir ce travail de raffinement en complétant l'ensemble des surfaces non agricoles par la suppression des bâtiments, et des surfaces des aéroports et aérodromes qui sont disponibles dans la BD TOPO®, et en utilisant cette base de données comme source d'information essentielle pour caractériser de façon certaine l'espace non agricole. Le Tableau 13 illustre les surfaces gagnées sur la surface agricole délimitée par Corine Land Cover en 2012, par raffinement de surfaces assurément non agricoles d'après la BD TOPO®.

Tableau 13. Évaluation des surfaces reprises sur la surface agricole CLC 2012 par raffinement des surfaces non agricoles de la BD TOPO®

Nature occupation du sol des surfaces que l'on souhaite supprimer de la couche « CLC agricole » 2012	Surface raffinée (supprimée de la couche CLC agricole)	Part de la surestimation de la SAU représentée par cette occupation du sol
Bâti indifférencié de la BD TOPO®	123 465 ha	2,15 %
Cimetières	6 276 ha	0,1 %
Bâti remarquable	1 750 ha	0,03 %
Bâti industriel	58 239 ha	1,01 %
Construction légère	9 605 ha	0,17 %
Construction surfacique	573 ha	0,01 %
Piste aérodrome	1 001 ha	0,02 %
Réservoirs et réservoirs d'eau	692 ha	0,01 %
Surfaces en eau	188 744 ha	3,3 %
Surface infrastructures routières	919 ha	0,01 %
Terrains de sport	11 202 ha	0,2 %
Total hors végétation	401 547 ha	7,00 %
Végétation (hors vignes et vergers)	4 138 677 ha	72,22 %
Total surfaces raffinables dans CLC	4 540 224 ha	79,22 %

Cette méthodologie destinée à affiner la précision des surfaces agricoles décrites par les différentes sources de données permet d'obtenir des résultats significatifs sur Corine Land Cover, comme le montrent les informations issues des analyses de sensibilité répertoriées au Tableau 13, mais également sur les autres bases de données plus précises utilisées pour la cartographie. Le raffinement est donc opéré sur la cartographie complète synthétisant les différentes sources d'information géographique.

BIBLIOGRAPHIE

- Booth, B. J., M. H. Ward, M. E. Turyk, and L. T. Stayner. 2015. 'Agricultural crop density and risk of childhood cancer in the midwestern United States: an ecologic study', *Environ Health*, 14: 82.
- Coste, A., S. Goujon, L. Faure, D. Hémon, and J. Clavel. 2020. 'Agricultural crop density in the municipalities of France and incidence of childhood leukemia: An ecological study', *Environ Res*, 187: 109517.
- Dereumeaux, C., C. Fillol, P. Quenel, and S. Denys. 2020. 'Pesticide exposures for residents living close to agricultural lands: A review', *Environ Int*, 134: 105210.
- Deziel, N. C., M. C. Friesen, J. A. Hoppin, C. J. Hines, K. Thomas, and L. E. Freeman. 2015. 'A review of nonoccupational pathways for pesticide exposure in women living in agricultural areas', *Environ Health Perspect*, 123: 515-24.
- Felsot, A. S., J. B. Unsworth, J. B. Linders, G. Roberts, D. Rautman, C. Harris, and E. Carazo. 2011. 'Agrochemical spray drift; assessment and mitigation--a review', *J Environ Sci Health B*, 46: 1-23.
- Garcia-Perez, J., A. Morales-Piga, J. Gomez, D. Gomez-Barroso, I. Tamayo-Uria, E. Pardo Romaguera, P. Fernandez-Navarro, G. Lopez-Abente, and R. Ramis. 2016. 'Association between residential proximity to environmental pollution sources and childhood renal tumors', *Environ Res*, 147: 405-14.
- Gunier, R. B., M. H. Ward, M. Airola, E. M. Bell, J. Colt, M. Nishioka, P. A. Buffler, P. Reynolds, R. P. Rull, A. Hertz, C. Metayer, and J. R. Nuckols. 2011. 'Determinants of agricultural pesticide concentrations in carpet dust', *Environ Health Perspect*, 119: 970-6.
- Jones, R. R., C. L. Yu, J. R. Nuckols, J. R. Cerhan, M. Airola, J. A. Ross, K. Robien, and M. H. Ward. 2014. 'Farm residence and lymphohematopoietic cancers in the Iowa Women's Health Study', *Environ Res*, 133: 353-61.
- Kamińska, I. A., A. Oldak, and W. A. Turski. 2004. 'Geographical Information System (GIS) as a tool for monitoring and analysing pesticide pollution and its impact on public health', *Ann Agric Environ Med*, 11: 181-4.
- Patel, D. M., S. Gyldenkaerne, R. R. Jones, S. F. Olsen, G. Tikellis, C. Granstrom, T. Dwyer, L. T. Stayner, and M. H. Ward. 2020. 'Residential proximity to agriculture and risk of childhood leukemia and central nervous system tumors in the Danish national birth cohort', *Environ Int*, 143: 105955.
- Rappazzo, K. M., J. L. Warren, R. E. Meyer, A. H. Herring, A. P. Sanders, N. C. Brownstein, and T. J. Luben. 2016. 'Maternal residential exposure to agricultural pesticides and birth defects in a 2003 to 2005 North Carolina birth cohort', *Birth Defects Res A Clin Mol Teratol*.
- Teyssere, R., G. Manangama, I. Baldi, C. Carles, P. Brochard, C. Bedos, and F. Delva. 2020. 'Assessment of residential exposures to agricultural pesticides: A scoping review', *PLoS One*, 15: e0232258.
- Ward, M. H., J. Lubin, J. Giglierano, J. S. Colt, C. Wolter, N. Bekiroglu, D. Camann, P. Hartge, and J. R. Nuckols. 2006. 'Proximity to crops and residential exposure to agricultural herbicides in Iowa', *Environ Health Perspect*, 114: 893-7.
- Ward, M. H., J. R. Nuckols, S. J. Weigel, S. K. Maxwell, K. P. Cantor, and R. S. Miller. 2000. 'Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System', *Environ Health Perspect*, 108: 5-12.
- Xiang, Huiyun, Jay R. Nuckols, and Lorann Stallones. 2000. 'A Geographic Information Assessment of Birth Weight and Crop Production Patterns around Mother's Residence', *Environ Res*, 82: 160-67.