

**Université de Toulouse**

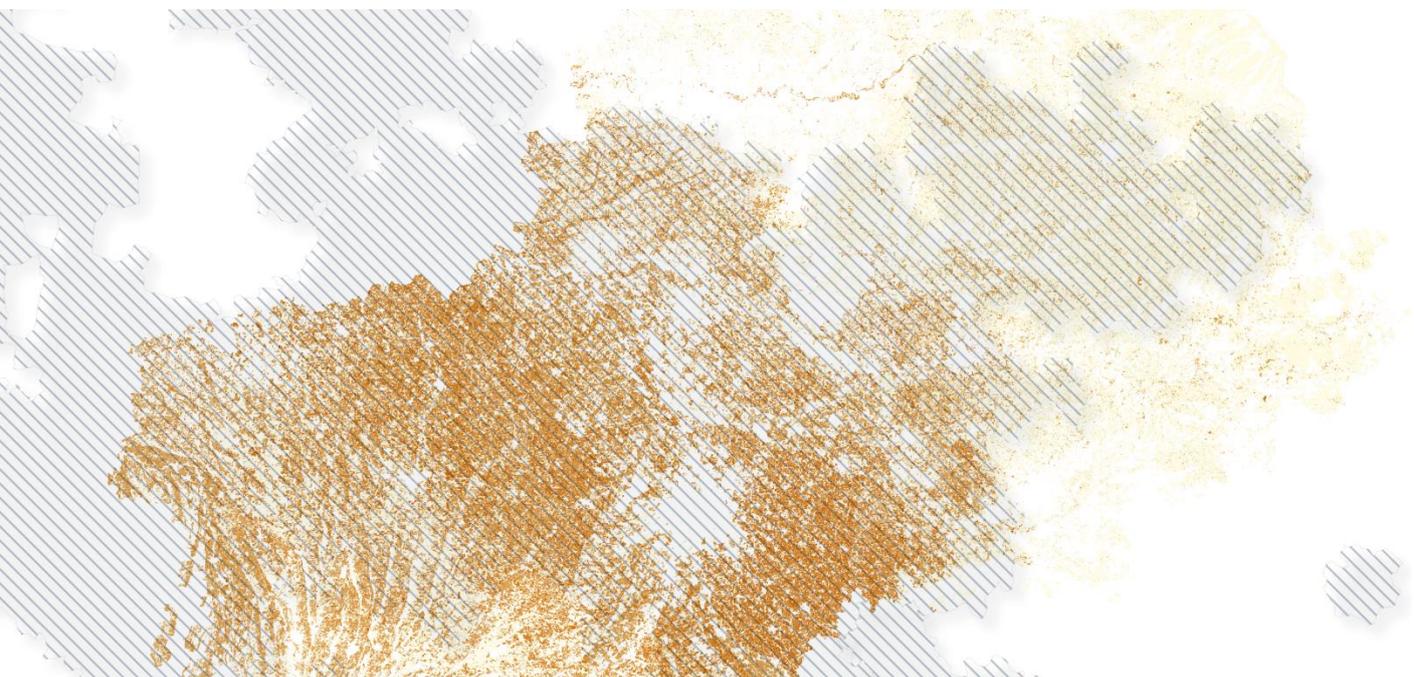
**MASTER 1 GEOMATIQUE**

« **ScIences Géomatiques en environneMent et AménageMent** » (**SIGMA**)

<http://sigma.univ-toulouse.fr>

**RAPPORT DE STAGE**

**Spatialisation des résultats d'un modèle d'offre agricole (AROPAj) sur les parcelles agricoles (Registre Parcellaire Graphique) pour l'étude des impacts environnementaux de l'agriculture française**



**BERMUDES Caroline**

**Encadrants de stage : CANTELAUBE Pierre, LUNGARSKA Anna**  
**Enseignant-référent : SHEEREN David**

## Remerciements

Je remercie tout d'abord Thomas Poméon, directeur de l'unité de service Observatoire du Développement Rural (ODR), pour m'avoir accueillie au sein de l'unité.

Je tiens à remercier mes encadrants de stage, Pierre Cantelaube et Anna Lungarska, pour toute la disponibilité qu'ils m'ont accordée, leur bienveillance et leurs conseils. J'espère que Pierre pourra considérer l'idée d'installer Rstudio un jour.

Je remercie également toute l'équipe de l'ODR pour leur accueil chaleureux et leur gentillesse.

# Table des matières

RESUME.....	3
INTRODUCTION .....	4
TABLES DES ABREVIATIONS ET DES ILLUSTRATIONS .....	6
1. L'ODR ET LES DONNEES .....	7
<b>1.1. La structure d'accueil.....</b>	<b>7</b>
<b>1.2. Des données .....</b>	<b>7</b>
1.2.1. Le Registre Parcellaire Graphique (RPG) .....	7
1.2.2. Le Réseau d'Information Comptable Agricole (RICA) .....	8
<b>2.3. Les précédentes spatialisations du modèle AROPAj.....</b>	<b>9</b>
2.3.1. Le modèle AROPAj.....	9
2.3.2. Le projet .....	10
2. LA SPATIALISATION VIA LE RPG .....	12
<b>2.1. La localisation des groupes types.....</b>	<b>12</b>
2.1.1. Etape 1 : enrichissement des données parcellaires RPG.....	12
2.1.2. Etape 2 et 3 : les probabilités selon deux variables : l'altitude et l'irrigation .....	15
<b>2.2. La simplification des parcelles du RPG .....</b>	<b>18</b>
2.2.1. Les éléments de sélection des parcelles.....	20
2.2.2. Modeleur graphique et ModelBuilder : avantages et inconvénients.....	22
2.2.3. PyQGIS : méthode d'automatisation.....	25
3. VISUALISATION DES RESULTATS ET PERSPECTIVES .....	28
<b>3.1 Quels groupes types sur quelles parcelles ? .....</b>	<b>28</b>
3.1.1. Les biais derrière l'association d'une parcelle à un groupe type .....	28
3.1.2. Des résultats plus justes avec le RPG .....	31
<b>3.2. Les sorties du modèle à l'échelle de la parcelle .....</b>	<b>32</b>
3.2.1. La diversité des sorties du modèle AROPAj.....	32
3.2.2. L'occupation du sol affectée à la parcelle : l'avantage du RPG .....	34
3.3.3. Quels liens sont possibles avec d'autres types de données ? .....	36
CONCLUSION .....	39
BIBLIOGRAPHIE .....	42
ANNEXES .....	43

## Résumé

Le modèle économique européen d'offre agricole AROPAj permet d'étudier les conséquences (économiques et environnementales) de réformes des politiques agricoles (Politique Agricole Commune, directives environnementales, changement climatique...). Il fonctionne sur des « agents économiques » qui sont des groupes d'exploitations agricoles similaires, et qui proviennent des données du Réseau d'Information Comptable Agricole (RICA). Les sorties du modèle sont initialement rattachées spatialement aux régions, et ont été précédemment désagrégées à une échelle plus fine : celle d'une carte des sols européenne construite sur la base un ensemble de couches de données géographiques sur l'occupation du sol.

Le Registre Parcellaire Graphique (RPG) est une base de données représentant une grande partie des parcelles agricoles françaises, concernées par les aides de la Politique Agricole Commune. Dans cette étude, le but est d'utiliser le RPG pour remplacer cette carte des sols (sur la France), avec tous les avantages qu'il peut apporter (précision thématique et géographique). L'idée est ainsi d'associer aux parcelles du RPG les probabilités d'accueillir les différents groupes d'exploitations agricoles du RICA qui font fonctionner le modèle AROPAj.

S'il arrive que ce rapport se présente parfois comme une note méthodologique, il expose également l'ensemble de tentatives pour trouver les meilleures manières de traiter ces données s'associant pour une première fois et expliquer les démarches pour arriver à cette association. Ce qui a été développé dans ce rapport est donc ouvert à la modification, face à ce que l'on cherche à réaliser avec ces données.

Mots clés : Agriculture, environnement, Registre Parcellaire Graphique (RPG), désagrégation spatiale, spatialisation, occupation du sol, modèle économique agricole.

## Abstract

The AROPAj model is an economic model dedicated to the simulation of European Union (EU) farming systems' behaviour facing change in economics (Common Agricultural Policy, European Directives regarding environment, climate change). AROPAj works on farm-groups provided by the Farm Accountancy Data Network (FADN), and the outputs can be located, initially, only at the regional level. A first, downscaling of these outputs has been done using a land cover map built for this purpose.

The aim of this study is to replace this land cover map by using the "Registre Parcellaire Graphique" (RPG), Land Parcel Information System (LPIS) in France. The RPG is indeed a geographical database mapping most of the agricultural parcels, and provides a very good thematic and geographical accuracy. The idea is to associate to each parcel probabilities of occurrence of AROPAj farm-groups.

This report is a methodological report as well as a description of attempts in data processing to achieve this aim.

Key words : Agriculture, environment, « Registre Parcellaire Graphique » (RPG), spatial disaggregation, spatialization, land cover, agro-economic model.

## Introduction

Principalement pour des raisons d'anonymat, les informations notamment économiques sur les exploitants agricoles et leurs pratiques ne sont souvent disponibles qu'à des échelles ne descendant pas, pour la plupart, en-deçà du niveau régional. Or, alors que nous considérons comme importants les effets des pratiques agricoles sur l'environnement, entre autres pollutions azotées et épandages de pesticides, il peut être considéré comme parfois difficile, à cette échelle, de les visualiser. Ces pratiques, pour l'Europe, sont étroitement liées à la Politique Agricole Commune (PAC), une des plus anciennes politiques européennes communautaires, et ses réformes. Mais elles sont aussi liées à des données physiques, environnementales ; l'altitude, la pente, le climat, le type de sol, mais aussi évidemment les types de cultures.

Il est complexe d'étudier les interactions entre des données économiques, (difficilement) accessibles à un niveau agrégé, régional voire national ou européen, et des données physiques, plus localisées et accessibles avec plus de facilité. L'étude de ces interactions est pourtant nécessaire pour analyser les comportements et les choix des exploitants. L'une des manières d'observer ces interactions passe par le biais de la désagrégation spatiale des informations. La désagrégation spatiale peut être définie comme le processus de séparation d'une entité en différents éléments par le biais de relations spatiales entre ces éléments. En outre, Vidal et al. (2001) cités dans l'article de Cantelaube et al. (2012) remarquent qu'il est beaucoup plus aisé de changer l'unité spatiale servant à la distribution des données plutôt que de changer l'échelle de production des données, ou de proposer un nouveau modèle travaillant sur cette unité spatiale. Sur ce point, nous pourrions également questionner l'intérêt d'un modèle à échelle fine mais dont les données d'entrées n'ont pas la même finesse.

Le modèle économique d'offre agricole AROPAj (Jayet et al., 2015) travaille sur des regroupements d'exploitations agricoles du RICA nommés « groupes types » (ce sont les agents économiques) dont la localisation connue la plus précise est la région.

Ce stage vise à la spatialisation des sorties d'AROPAJ, fonctionnant initialement donc à l'échelle infrarégionale sur les parcelles du Registre Parcellaire Graphique (RPG). Le RPG contient, depuis 2015, des informations très précises sur l'occupation du sol agricole français. Le modèle AROPAj, quant à lui, se nourrit notamment de données du Réseau d'Information Comptable Agricole ; données sur les exploitants européens, et de réformes de la PAC. Les objectifs du stage font partie de cette volonté d'observer, sur une nouvelle unité spatiale à très grande échelle, les interactions entre données économiques et données physiques, leurs résultantes sur les pratiques agricoles et les impacts environnementaux de ces dernières. Néanmoins, pour ce rapport, nous allons surtout nous concentrer sur la question de la désagrégation spatiale des sorties du modèle, et notamment voir si le RPG est idéal pour observer ces interactions entre données.

Ainsi, la problématique est la suivante : *En quoi le RPG représente-t-il un grand potentiel dans la spatialisation du modèle AROPAj ?*

Le modèle AROPAj a déjà été, auparavant, spatialisé au travers d'une « carte des sols » construite à cet effet, avec de nombreuses bases de données aidant à la localisation des groupes types agricoles AROPAj. La même méthodologie de localisation de ces groupes types a été appliquée en remplaçant cette carte des sols par le RPG agrémenté de quelques données. Les groupes types AROPAj étant à échelle infrarégionale, il était intéressant de

développer une méthode afin de regrouper les parcelles du RPG et les simplifier pour moins de lourdeur lors de traitements. Une fois les groupes types localisés sur les parcelles, plusieurs types de représentations et d'analyses des sorties peuvent être réalisées.

## Tables des abréviations et des illustrations

### Table des abréviations :

RPG	Registre Parcellaire Graphique
ODR	Observatoire du Développement Rural
IGN	Institut national de l'information géographique
RICA	Réseau d'Information Comptable Agricole
PAC	Politique Agricole Commune
CLC	Corine Land Cover
MNT	Modèle Numérique de Terrain
BNVD	Banque Nationale des Ventes de produits phytopharmaceutiques par les Distributeurs agréés
SANDRE	Service d'Administration Nationale des Données et Référentiels sur l'Eau
OTEX	Orientation Technico-Economique des Exploitations
GDAL	Geospatial Data Abstraction Library
GRASS	Geographic Resources Analysis Support System
ASP	Agence de Services et de Paiement

### Table des illustrations :

Figure 1 : Organigramme de l'US-ODR.....	7
Figure 2 : Codes des régions RICA pour EU-28 ; source : FADN, European Commission .....	8
Figure 3 : Création de la carte des sols utilisée pour les spatialisations précédentes du modèle AROPAj .....	11
Figure 4 : Comparaison entre la classe 2 de Corine Land Cover (territoires agricoles) (à gauche) et le RPG (à droite, où l'on peut voir en bleu les estives) .....	11
Figure 5 : Visualisation des groupements parcellaires sur la commune de Champeau-en-Morvan (21139) dans le département de la Côte-d'Or.....	19
Figure 6 : Version finale du modeleur graphique .....	19
Figure 7 : Tri des polygones par groupements parcellaires.....	20
Figure 8 : Comparaison des fonctions GDAL (à gauche) et GRASS (à droite) .....	21
Figure 9 : Comparaison GDAL / GRASS .....	21
Figure 10 : Troisième version du modeleur, la plus aboutie avant la dernière .....	23
Figure 11 : Différences de sélection entre les communes de Champeau-en-Morvan et Ménessaire (Côte d'Or) .....	23
Figure 12 : Exemples de cartographies de l'atlas centré sur l'attribution des groupes types majoritaires par région.....	29
Figure 13 : Zoom sur les groupes types .....	30
Figure 14 : Les groupes types d'Aquitaine .....	30
Figure 15 : Comparaison de la spatialisation des groupes types entre le RPG (à gauche) et la carte des sols (à droite)..	31
Figure 16 : Orientations technico-économiques des exploitations (OTEX) des régions Champagne-Ardenne, Lorraine et Alsace d'après le Recensement agricole de 2010.....	32
Figure 17 : Schéma du calcul matriciel effectué pour le groupement parcellaire 5 de la région 42 (Alsace) .....	33
Figure 18 : Représentations de plusieurs sorties AROPAj sur la région Champagne-Ardenne .....	34
Figure 19 : La consommation d'engrais azotés à l'hectare confrontée aux zones vulnérables sur trois régions (Midi-Pyrénées, Alsace, Rhone-Alpes).....	37
Figure 20 : Diagramme de Gantt : durées estimées et réelles des tâches réalisées.....	40
Tableau 1 : Correspondances entre régions RICA et anciennes régions françaises	8
Tableau 2 : Pourcentages d'enlèvement des parcelles.....	26
Tableau 3 : Les informations conservées.....	27
Tableau 4 : Des probabilités de présences égales .....	29
Tableau 5 : Des sorties d'AROPAJ .....	33
Tableau 6 : Types de cultures (codes groupe du RPG) présents dans la classe de valeur la plus haute (classes d'effectifs égaux) pour les régions (de haut en bas) : 24 (Centre), 53 (Bretagne), 73 (Midi-Pyrénées) .....	35
Tableau 7 : Définition des besoins azotés pour la production de quelques plantes cultivées. Sources : Instituts techniques Arvalis et Terres Inovia, 2016, utilisé ici : <a href="https://www.encyclopedie-environnement.org/vivant/les-nitrates-dans-lenvironnement/">https://www.encyclopedie-environnement.org/vivant/les-nitrates-dans-lenvironnement/</a> .....	35
Schéma 1 : Données utilisées pour établir les probabilités de localisation des groupes types	15
Schéma 2 : Algorithme sur l'attribution des probabilités selon l'altitude et l'irrigation .....	17

# 1. L'ODR et les données

## 1.1. La structure d'accueil

Ce stage a été réalisé au sein de l'Unité de Service Observatoire du Développement Rural (US-ODR) du centre INRAE (Institut National de la Recherche en Agronomie et Environnement) de Auzeville-Tolosane, au sud-est de Toulouse. L'US-ODR, rattachée au département EcoSocio, a été créé en 2009 par l'INRA, le ministère de l'Agriculture (MAAF) et l'Agence de Services et Paiements (ASP).

Les champs de travail de l'ODR sont multiples : politiques de développement rural et agro-environnementales, activités et emplois agricoles, signes de qualité dans l'agriculture en France. Ce travail de spatialisation s'inscrit dans le réseau « Systèmes agricoles » de l'ODR, où sont réalisés des travaux en lien avec le RPG, notamment la création de la version « complétée » du RPG avec des « parcelles ODR » issues des données OSO et représentant les surfaces agricoles ne bénéficiant pas d'aides surfaciques de la PAC.

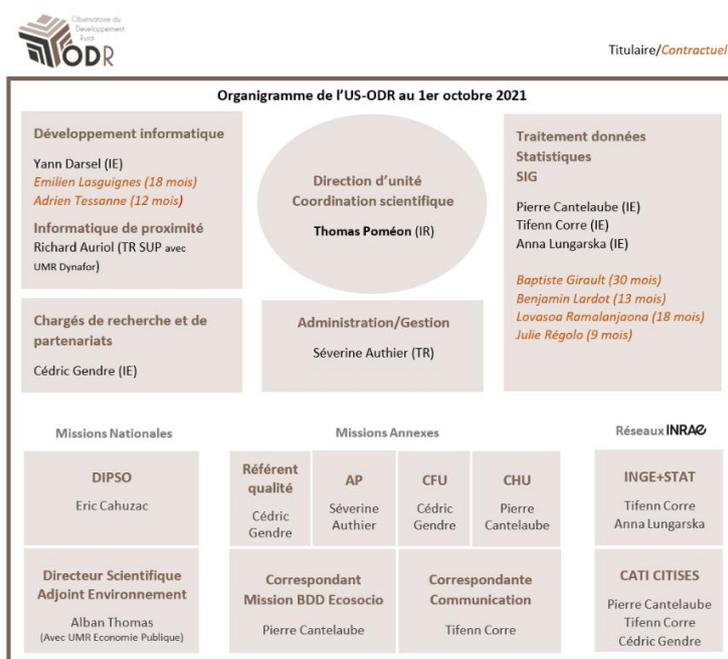


Figure 1 : Organigramme de l'US-ODR

## 1.2. Des données

### 1.2.1. Le Registre Parcellaire Graphique (RPG)

Le Registre Parcellaire Graphique est une base de données produite par l'Agence de Service et de Paiement (ASP), rassemblant des données géographiques et informations sur l'usage et l'occupation des sols agricoles. Elle est alimentée par les déclarations des exploitants en vue d'« instruire et enregistrer les dossiers de déclaration pour les aides dans le cadre de la politique agricole commune (PAC) » (IGN,2021) (Cantelaube et Carles, 2014) et de versements d'aides. A l'aide de photos aériennes, les exploitants dessinent les

contours de leurs parcelles de cultures, indiquant la nature des cultures et les surfaces des parcelles exploitées. Depuis 2015 sont dénotés 24 grands groupes types cultureux (« code groupe ») et 350 cultures principales (« code culture »). Les cultures sont donc très précisément distinguées ; les différents types de blé, de maïs, les prairies temporaires et permanentes.

### 1.2.2. Le Réseau d'Information Comptable Agricole (RICA)

Le RICA (ou FADN en version anglaise pour Farm Accountancy Data Network) est une base de données établie en 1965, fournissant des données sur le suivi des revenus et activités des exploitations agricoles européennes. Ces données sont collectées par le biais d'enquêtes réalisées pour chaque pays Membre de l'Union Européenne, et produites par le SSP, Bureau des statistiques sur les productions et comptabilités agricoles, dans le cas de la France.

La base est constituée par un échantillon d'individus, changeant chaque année, sur lesquels sont renseignées des informations sur la taille économique, l'orientation technico-économique de l'exploitation (OTEX\*) et la région dite région RICA. L'Union Européenne est découpée en région (découpage pouvant évoluer) chaque région ayant un identifiant unique dans la base de données. Pour la France, les données sont regroupées à l'échelle des anciennes régions administratives (avant la loi NOTRe de 2015).

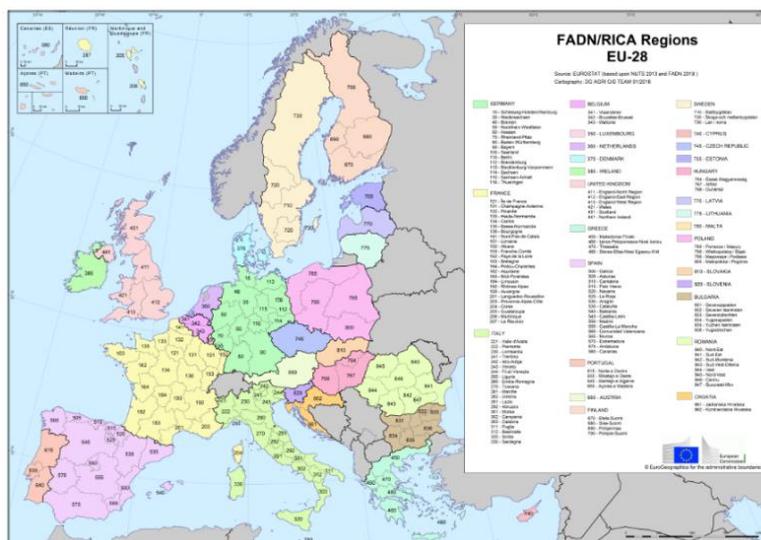


Figure 2 : Codes des régions RICA pour EU-28 ; source : FADN, European Commission

REGIONS RICA	ANCIENNES RÉGIONS FRANÇAISES
121	11
131	21
132	22
133	23
134	24
135	25
136	26
141	31
151	41
152	42
153	43
162	52
163	53
164	54
182	72
183	73
184	74
192	82
193	83
201	91
203	93

Tableau 1 : Correspondances entre régions RICA et anciennes régions françaises

Chaque individu de l'échantillon possède un poids correspondant au nombre d'exploitations agricoles qu'il représente réellement. Les exploitations prises en compte sont plutôt de tailles moyenne à grande, dans le but de prendre en considération une grande proportion des surfaces agricoles : cet échantillon n'est donc pas, malgré la pondération, très représentatif ; il ne « couvre que les exploitations agricoles de l'UE qui, en raison de leur

taille, peuvent être considérées comme professionnelles »<sup>1</sup>. En 2018, l'échantillon se fondait sur 80 400 « holdings » ou exploitations soit 3.9 fermes sur presque 10.3 millions<sup>2</sup>.

## \*LES OTEX

L'orientation technico-économique est une **classification des exploitations agricoles européennes selon leur spécialisation**, établie au départ sur la classification MBS (Marge brute standard) devenue obsolète après la réforme de la PAC de 2003, au profit des coefficients de **PBS (Production brute standard)**. Elle s'applique notamment sur les résultats du Recensement Agricole, ou du RICA, avec une nomenclature agrégée pour des raisons de confidentialité et/ou de représentativité.

Lorsque la PBS dépasse les deux tiers du total pour l'une des productions d'une exploitation, celle-ci devient la spécialisation de cette dernière. Si ce n'est pas le cas, la spécialisation de l'exploitation sera de la polyculture et/ou du polyélevage (même s'il ne s'agit que de deux types de productions importantes).

### POUR EN SAVOIR PLUS :

[HTTPS://AGRESTE.AGRICULTURE.GOUV.FR/AGRESTE-WEB/METHODON/N.1/!SEARCHURL/LISTETYPEMETHODON/](https://agreste.agriculture.gouv.fr/agreste-web/methodon/n.1/!searchurl/listetypemethodon/)

+

[HTTPS://ODR.INRA.FR/INTRANET/CARTO/CARTOWIKI/INDEX.PHP/OTEX\\_ET\\_ORIENTATION\\_AGRICOLE\\_DES\\_TERRITOIRES](https://odr.inra.fr/intranet/carto/cartowiki/index.php/otex_et_orientation_agricole_des_territoires)

### 2.3. Les précédentes spatialisations du modèle AROPAj

#### 2.3.1. Le modèle AROPAj

Notamment développé par le directeur de recherche en économie agricole Pierre-Alain Jayet au sein de l'Unité Economie Publique d'INRAE à partir des années 90, le modèle AROPAj est un modèle d'offre agricole européenne, mathématique, microéconomique, « fondé sur la représentation du comportement d'agents économiques individualisés, et de nature technico-économique dans le sens où les techniques de productions sont explicitement présentes » (Cantelaube et al., 2008).

Il s'agit d'un « *modèle opérationnel permettant de transformer des scénarios économiques en scénarios de changement d'usage des sols agricoles. On entend par « scénario économique » un scénario relevant de la Politique Agricole Commune, d'une politique de promotion des cultures énergétiques, ou de politiques environnementales* » (Cantelaube et al, 2008). En entrée du modèle sont intégrées des données économiques (les données RICA), les règles de la PAC, ses réformes et les restrictions ou opportunités qui en découlent, pouvant influencer sur la manière dont les exploitants gèrent et s'occupent de leurs terres. Par exemple, avec la réforme de la PAC de 2003 (les accords de Luxembourg) ; le découplage des aides a été ajouté comme scénario. Le modèle sert à évaluer les effets de

<sup>1</sup> [https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/farms-farming-and-innovation/structures-and-economics/economics/fadn\\_fr](https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/farms-farming-and-innovation/structures-and-economics/economics/fadn_fr)

<sup>2</sup> EUROPEAN COMMISSION (DIRECTORATE-GENERAL FOR AGRICULTURE AND RURAL DEVELOPMENT), Eu Farm Economics Overview FADN 2018. 2021, Brussels.

ces réformes successives ; mais aussi à analyser les relations entre l'agriculture et l'environnement, notamment sur l'utilisation de pesticides, la pollution liée aux engrais azoté et aux effluents d'élevage ; autant de données liées par ailleurs à des informations économiques.

Le modèle en est aujourd'hui à sa cinquième version, les versions ayant évolué à travers les années pour prendre en compte les pays membres de l'Union Européenne mais aussi pour s'adapter aux données européennes RICA. En effet, le modèle est en grande partie alimenté par des données du RICA, et ceci pour de nombreuses raisons exposées dans « A Handbook on the use of FADN Database in Programming Models » (Neuenfeldt and Gocht, 2014). Pour citer quelques exemples, la collecte de donnée est harmonisée sur toute l'Union Européenne et existe depuis déjà un certain nombre d'années. Ensuite, les données du RICA représentent statistiquement l'hétérogénéité des systèmes agricoles européens.

Les agents économiques du modèle AROPAj sont des regroupements d'au moins 15 individus observées, « *la mise à disposition du RICA interdisant toute utilisation directe, et a fortiori toute localisation d'information à caractère individuel* » (Jayet et al., 2009). Ces regroupements sont faits à partir d'informations portant sur l'altitude des exploitations, l'OTEX (ou « type agricole » du RICA), la dimension économique (classification « size » du RICA, en 9 classes) et l'irrigation. Ces variables, une fois rassemblées, forment des groupes présentant des caractéristiques communes ; que nous appellerons « groupes types » ou « GT » dans la suite de ce rapport. Pour représenter l'agriculture française sont dénombrés environ 160 groupes selon les années et versions d'AROPAj : un nombre élevé, indiquant la capacité du modèle à prendre en compte la grande diversité des systèmes de production agricoles. A ce stade, ces groupes types sont localisés dans les anciennes régions sans plus de précisions.

### 2.3.2. Le projet

La spatialisation de ces groupes types et, par extension, des « sorties » du modèle a déjà été réalisée, à petite échelle (sur le territoire européen), sur la base d'une carte des sols créée spécifiquement pour cette démarche (Cantelaube et al, 2012).

Cette dernière a été construite (voir Figure 3) à partir d'un modèle basé sur les polygones de CORINE Land Cover (CLC) classés arables ou prairie, enrichis avec des données topographiques de pente et d'altitude (MNT, Modèle Numérique de Terrain), de sol (EUROSOIL), météorologiques (les précipitations et températures fournies par la base de données MARS, Monitoring Agricultural ResourceS), cellules à l'échelle desquelles étaient localisés les groupes types. Le modèle est calibré sur les points d'observations de l'enquête LUCAS<sup>3</sup>. La construction de cette carte se fait selon une méthode proposée par Chakir (2009).

---

<sup>3</sup> LUCAS : enquête sur l'utilisation et l'occupation du sol. Voir : <https://ec.europa.eu/eurostat/fr/web/lucas/data/database> ou <https://artificialisation.developpement-durable.gouv.fr/bases-donnees/teruti-lucas>

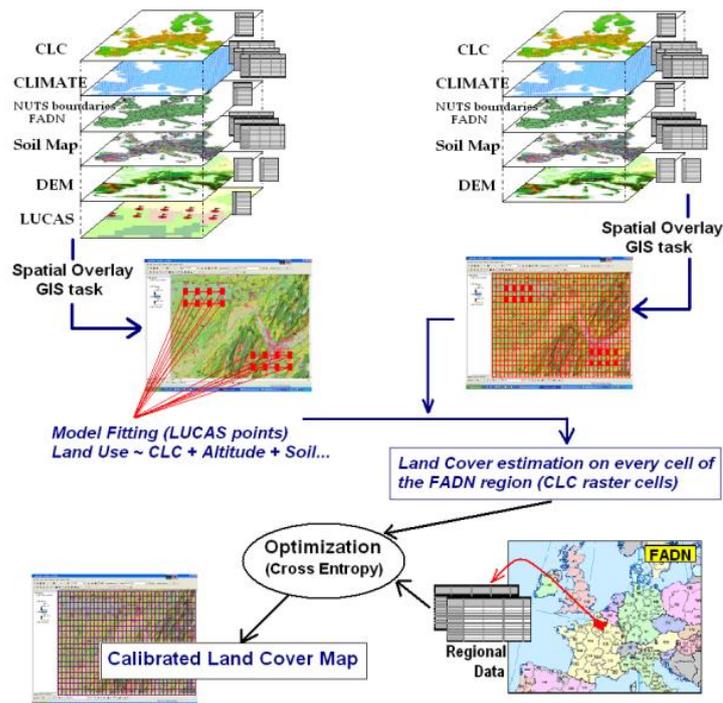


Figure 3 : Création de la carte des sols utilisée pour les spatialisations précédentes du modèle AROPAj

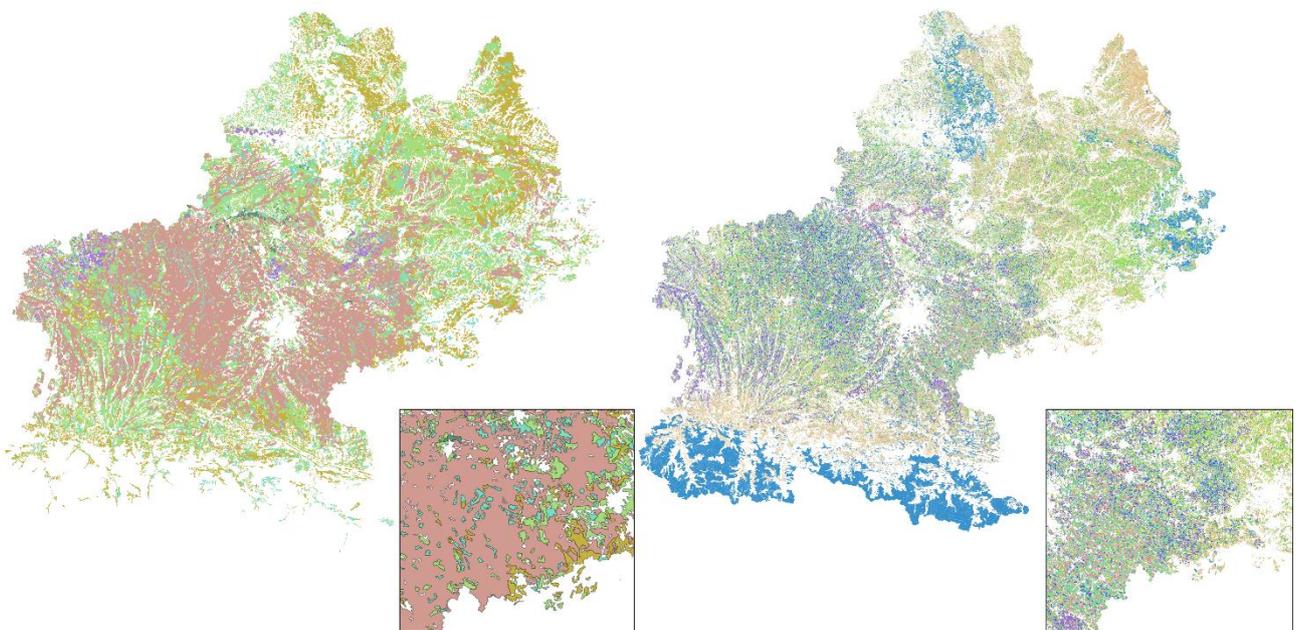


Figure 4 : Comparaison entre la classe 2 de Corine Land Cover (territoires agricoles) (à gauche) et le RPG (à droite, où l'on peut voir en bleu les estives)

Une première étape consiste à estimer la probabilité de présence des différents types de cultures pour chaque cellule, puis d'attribuer ensuite, à partir de ces couvertures du sol déduites, la probabilité de présence des groupes types AROPAj.

En utilisant les parcelles du RPG pour remplacer cette carte des sols, sur la France, nous savons avec précision où se trouvent telles ou telles cultures dans les régions (voir Figure 4). Les probabilités d'accueil des différents groupes types sur chaque région seront alors plus simples à attribuer, et la carte des sols n'a plus à être construite. Cependant, il reste néanmoins crucial de connaître deux informations en plus sur les parcelles afin de

faire coïncider les groupes types AROPAj et le RPG : l'irrigation et l'altitude, qui sont deux caractéristiques spécifiques aux groupes types.

Les précédentes spatialisations étaient réalisées à l'échelle européenne. Ici, nous nous intéressons seulement à la France, puisque nous avons facilement à disposition le RPG. Néanmoins, le but serait, avec le temps, de réaliser cette spatialisation à fine échelle pour d'autres pays européens, également concernés par le modèle AROPAj.

## 2. La spatialisation via le RPG

### 2.1. La localisation des groupes types

L'objectif, comme pour les spatialisations faites précédemment, est d'abord d'estimer la localisation des groupes types sur les régions françaises. Tout ce que nous décrivons ci-dessous a été réalisé grâce au langage de programmation R et au logiciel RStudio.

Le travail s'effectue en 3 étapes : d'abord les parcelles du RPG sont enrichies avec les informations nécessaires (altitude et irrigation), puis les surfaces par groupe type et catégorie AROPAj sont calculées, et enfin les probabilités de groupe type sont associées aux parcelles RPG.

#### 2.1.1. Etape 1 : enrichissement des données parcellaires RPG

##### *Les données utilisées*

Les données du RPG avec lesquelles nous travaillons datent de 2016 et sont fournies par département au format ESRI Shapefile. Cette année a été choisie car nous ne pouvons pas utiliser la dernière version de la base de données ; les dernières « sorties » du modèle AROPAj datent de 2018 (nous avons donc choisi, évidemment, les « sorties » de 2016 de la version 5 comprenant 167 groupes types). En outre, pour le RPG, les types de cultures sont associées à la parcelle depuis 2015, ils étaient avant associés à l'îlot.

Avant de lancer le processus de spatialisation, il est nécessaire de rajouter trois variables indispensables aux parcelles RPG : (1) la classification de l'altitude, (2) la présence ou non d'irrigation, et (3) les catégories de cultures auxquelles appartiennent les parcelles. Ces trois variables, également renseignées pour les surfaces AROPAj, vont permettre de calculer les probabilités de localisation.

##### *(1) Altitude :*

Grâce à une jointure spatiale (outil de statiques zonales sur QGIS) avec les valeurs d'un MNT à 25 mètres, l'ODR a associé une donnée d'altitude à chaque parcelle du RPG.

► Ces données RPG enrichies de l'altitude sont directement utilisées.

##### *(2) Irrigation :*

Les données sur l'irrigation sont tirées d'un jeu de données LUCAS (Land Use/Land Cover Area Frame Survey) : il s'agit de points d'observation du sol étalés selon une grille. Ces points nous permettent ici d'avoir des informations sur l'irrigation à travers une classification de 1 à 8 (les classes 1 (irrigation), 2 (potential irrigation), et 4 (irrigation and drainage) nous intéressent). A partir de cette information, une liste des communes irriguées a été établie : une commune est dite irriguée lorsqu'elle accueille au moins un point LUCAS irrigué.

► J'ai préparé la couche des communes françaises irriguées sur QGIS, bien que cela aurait tout aussi pu se faire avec R. Les limites administratives des communes françaises sont tirées de la couche IGN GEOFLA (2016), dont les géométries sont très simplifiées puisque tirées de la BDCARTO (Annexe 1). Bien que ces limites soient très approximatives et puissent donc engendrer des erreurs d'attribution de la donnée d'irrigation aux communes, elles permettent d'obtenir des temps de traitements SIG plus rapides. Il fallait tout d'abord ne garder dans les données ponctuelles que les points concernés par les trois classes relevées, puis réaliser une jointure spatiale entre ces derniers et les communes françaises afin de ne garder que les communes dans lesquelles étaient présents un ou plusieurs de ces points. J'aurais pu tenter de garder les communes à proximité de celles qui étaient concernées, car n'ayant pas elles-mêmes un point d'irrigation nous aurions pu considérer que l'irrigation ne s'arrête pas à la commune même. Ces pistes sont à explorer dans le futur mais ne résoudreont certainement pas le problème lié au manque d'information précise sur l'irrigation, chose d'autant plus étonnante que la gestion de l'eau d'irrigation est un des défis majeurs de l'agriculture dans le contexte du changement climatique amorcé.

### *(3) Les catégories de cultures du modèle AROPAj*

Pour connaître et associer les catégories de cultures AROPAj aux parcelles, un tableau de correspondance est construit. Tous les types de cultures renseignés dans le RPG correspondent à une classe de culture AROPAj ; la classe 0 correspondant aux types de culture qui ne sont pas traités par le modèle, dont par exemple la viticulture, l'arboriculture et l'horticulture. Chaque parcelle étant liée à une classe, nous pouvons retirer du RPG les parcelles classées 0.

► Pour chaque parcelle du RPG, j'ai donc remplacé la culture déclarée initialement par une catégorie de culture AROPAj.

### *Le script pour préparer les données*

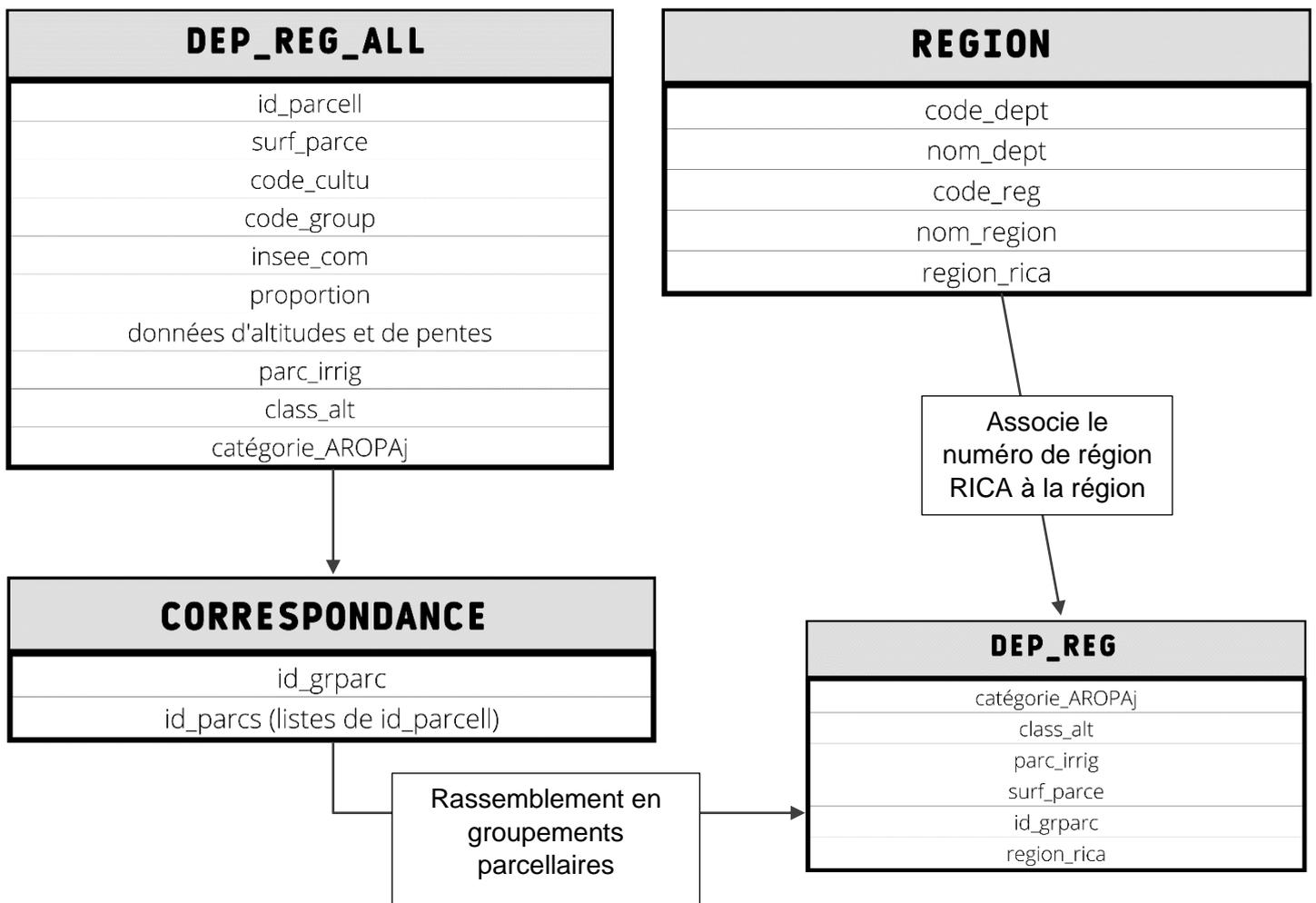
#### *(1) Travail par région*

Dans le script, je rassemble les départements en région (anciennes régions françaises), puisque les données que nous cherchons à lier sur ces parcelles sont à cette échelle. J'appelle donc les départements par région dans notre dossier data, qui sont ensuite regroupés pour obtenir un unique fichier pour la région. J'importais d'abord les données en .shp ; mais à la vue des premiers résultats, cherchant à faire plusieurs tests sur différentes régions et modifiant régulièrement les variables en sortie, j'ai préféré le chargement des tables attributaires seulement via les fichiers .dbf moins lourds. Ainsi, tout le script se lançait beaucoup plus rapidement.

(2) Regroupement des parcelles selon les 3 critères (altitude, irrigation, groupe de culture)

Dans un souci d'allégement des traitements et de la cartographie j'ai rassemblé les parcelles selon les paramètres que sont la catégorie de culture, l'altitude et l'irrigation ; celles-ci font l'objet d'un *group\_by* via le package Dplyr. Nous nous retrouvons à la suite de cette opération avec un fichier RPG où chaque groupe de parcelles ayant des caractéristiques communes (même catégorie, altitude et état d'irrigation) est lié à un numéro d'identification : le fichier est nommé « [dep\\_reg](#) ». Pour retrouver les identifiants exacts des parcelles liées à ces groupements, nous créons un [tableau de correspondance](#).

Nous importons ensuite un fichier « [class\\_regs](#) », contenant les catégories AROPAj et leur équivalence en codes RICA. Ces codes, nous les retrouvons dans un [fichier contenant des surfaces de cultures AROPAj](#), complétés par des informations « AA » ou « IRAA » pour indiquer une potentielle irrigation des surfaces. De ce fichier des surfaces, nous extrayons celles qui correspondent à la région RICA avec laquelle nous travaillons, et nous les plaçons dans un tableau nommé « [surf\\_reg](#) ». Nous faisons de même pour le fichier « [class\\_reg](#) » pour ne récupérer que les codes RICA de la région.



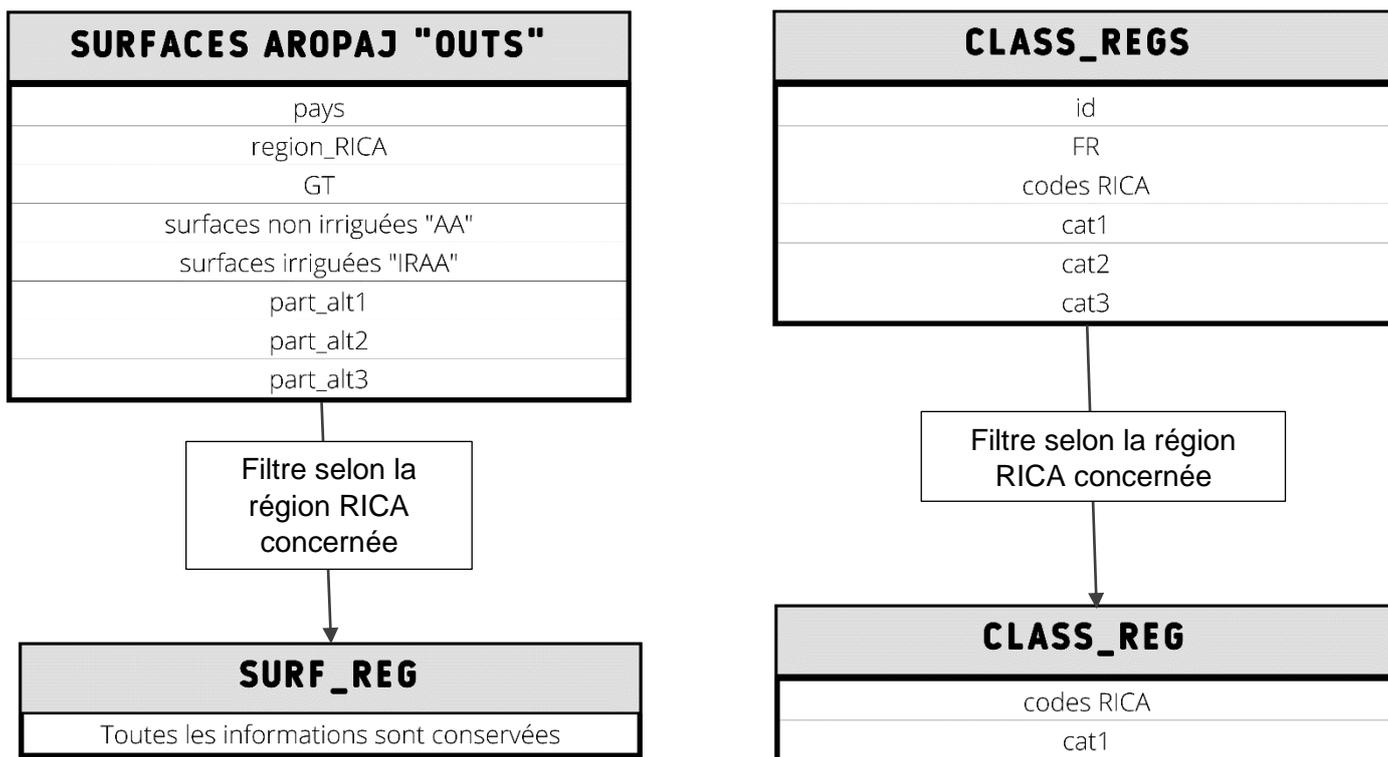


Schéma 1 : Données utilisées pour établir les probabilités de localisation des groupes types

Concernant les deux prochaines étapes, la difficulté résidait plus dans le fait de comprendre les calculs, ce qui était fait pour obtenir les résultats et le passage d'équations rédigées sur une méthodologie à un code fonctionnel ; le code ayant déjà été réalisé et ne nécessitant que peu de modifications. Il est question d'éclairer leur fonctionnement et leur but et les adapter à ces nouvelles données et nouveaux objectifs.

#### 2.1.2. Etape 2 et 3 : les probabilités selon deux variables : l'altitude et l'irrigation

##### (1) Etape 2

Ces étapes traduisent la méthodologie d'obtentions des probabilités expliquée dans le manuel du modèle AROPAj, dans la partie 8.2.2, « Estimation of probabilities for farm types ». Ainsi, comme indiqué dans la méthodologie, nous devons connaître, à l'échelle régionale :

- la surface totale des parcelles, dénotée  $M$  : `M <- sum(dep_reg$surf_parce)`
- la surface totale des parcelles irriguées, dénotée  $I$  : `I <- sum(dep_reg$surf_parce[which(dep_reg$parc_irrig == 1)])`
- et les sommes des surfaces pour les trois classes d'élévation, dénotées  $A1$ ,  $A2$  et  $A3$  : `A1 <- sum(dep_reg$surf_parce[which(dep_reg$class_alt == 1)])`

La méthodologie indique également qu'il est nécessaire de connaître :

- la somme des surfaces RICA non irriguées « sum\_sjk » : `sum_sjk = sum(surf_reg[, grep("[0-9]AA", colnames(surf_reg))])`
- et irriguées : « sum\_sjk\_i » : `sum_sjk_i = sum(surf_reg[, grep("IRAA", colnames(surf_reg))])`

Des ratios de représentativité  $\rho$  (rho) sont calculés pour les surfaces totales :  $\rho = \text{sum\_sjk}/M$  ; ainsi que des  $\delta$  (delta) de distorsions pour les surfaces concernant l'irrigation et l'altitude :  $\text{delta\_irr} = ((I*\rho)/\text{sum\_sjk\_i}) - 1$ . D'abord sont calculées les sommes des surfaces de chaque classe d'altitude du RICA : `sum_U1 = sum(surf_reg$part_alt1)` ; la même chose est faite pour `sum_U2` et `sum_U3` avec les classes correspondantes. Puis les deltas de distorsions :  $\text{delta\_a1} = ((A1*\rho)/(\text{sum\_U1}*\text{sum\_sjk})) - 1$  ; la même chose étant réalisée pour `delta_a2` et `delta_a3`.

Une fois ces variables calculées, une boucle crée un fichier nommé en sortie « `gt_surfs` », rassemblant les sommes RICA des surfaces non irriguées et irriguées, par groupes types et par catégories de culture AROPAj.

Cette boucle tourne donc sur tous les groupes types présents dans le fichier « [surf\\_reg](#) », fichier où l'on retrouve, pour la région RICA donnée, toutes les surfaces des types de cultures non irriguées pour chaque groupe type présent dans la région. Pour chacun de ces groupes types, une seconde boucle « imbriquée » dans la première est faite pour aller observer toutes les classes présentes dans le fichier « [class\\_reg](#) » contenant toutes les catégories AROPAj et leur correspondance RICA, et additionner dans une variable les surfaces des cultures RICA par catégories lorsque la correspondance RICA est présente dans le fichier « [surf\\_reg](#) ». La même chose est réalisée pour les surfaces des types de cultures irriguées.

La seconde boucle se sert des calculs de ratios et de distorsions réalisés précédemment ainsi que du fichier « `gt_surfs` », nous l'expliquons ci-dessous. Des premiers résultats ont été obtenus avant sa modification, sous la forme d'un fichier DBF avec des groupements parcellaires (par caractéristiques communes) et des probabilités par groupe type. Ces probabilités, après vérification, ne paraissaient pas être correctes. En effet, les probabilités étaient bien trop basses pour chacun des groupes types (bien que certains se démarquaient malgré tout par rapport à d'autres). Cependant, ces premières « sorties » nous ont malgré tout donné l'opportunité de commencer à travailler sur le RPG en lui-même ([simplification parcellaire](#)) en parallèle.

## (2) Etape 3

Nous commençons par retirer du fichier « `gt_surfs` » les surfaces égales à 0 irriguées et non irriguées. Nous créons le tableau qui va accueillir les probabilités :

```
prob_gt = data.frame(matrix(vector(), 0, (1 + length(surf_reg$GT)),
dimnames=list(c(), c("gid", paste0("GT_", surf_reg$GT))), stringsAsFactors=F).
```

Ce tableau s'adapte en nombre de colonnes selon le nombre de groupes types considérés dans la région. Cette deuxième boucle du code va générer les probabilités ; et cette fois-ci encore, la méthodologie indiquée dans le rapport AROPAj est suivie, dans laquelle sont exposés deux cas de figure. Afin de rendre les explications plus lisibles pour ce rapport, j'ai traduit ces cas de figure sous la forme d'une sorte d'algorithme :

## L'ATTRIBUTION DES PROBABILITÉS SELON L'ALTITUDE ET L'IRRIGATION

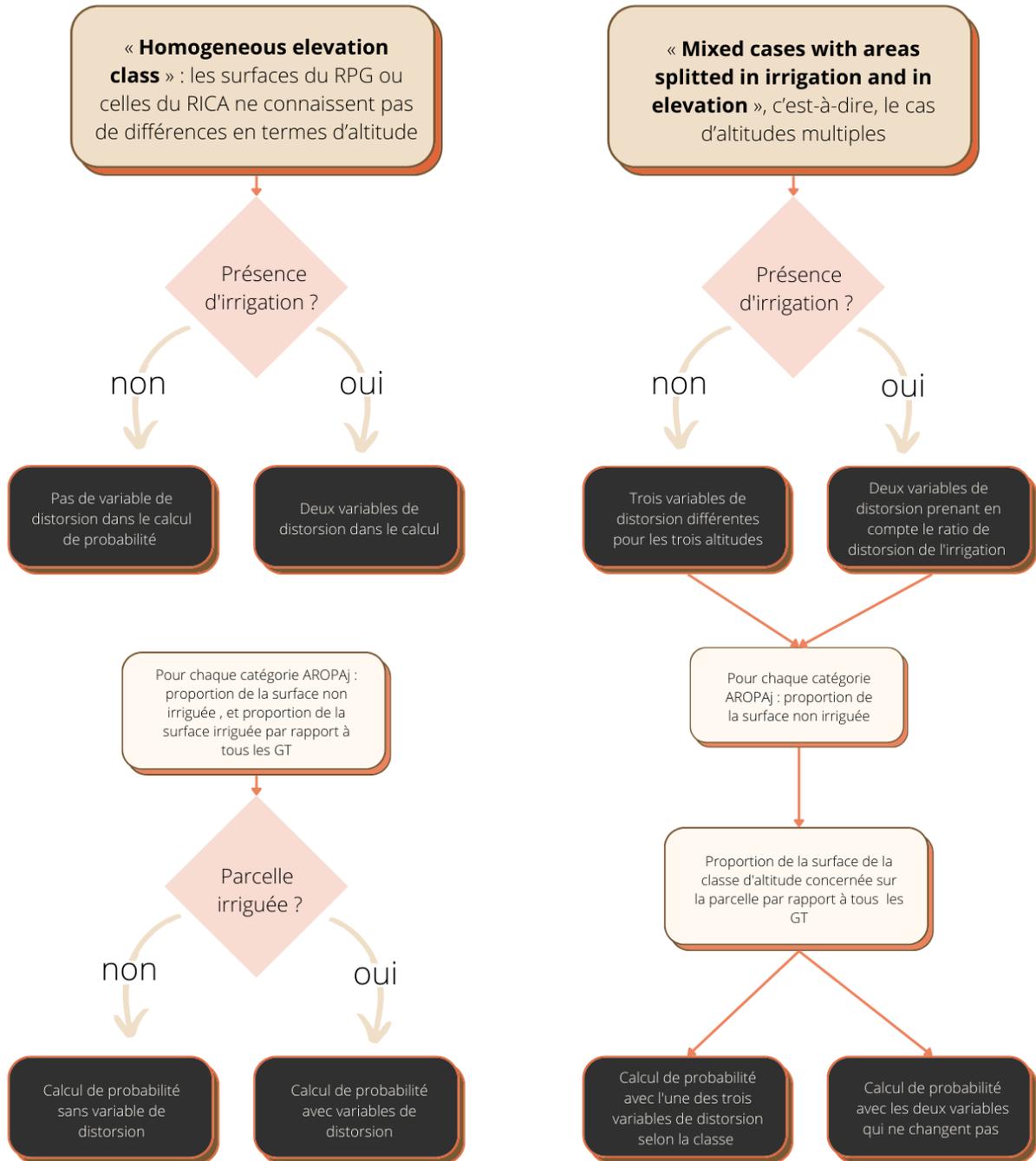


Schéma 2 : Algorithme sur l'attribution des probabilités selon l'altitude et l'irrigation

Nous pouvons donc voir au travers de cet algorithme l'importance de ces deux variables que sont l'irrigation et l'altitude lors de l'attribution des probabilités. Pour chacun de ces cas nous avons donc une équation déterminant les probabilités (Annexe 2) traduite dans le code R. La modification majeure apportée à cette boucle calculant les probabilités était de ne plus considérer que les types de culture restaient à être définis pour les parcelles. En effet, avec le RPG, nous sommes certains de l'occupation du sol des parcelles, ce qui n'était pas le cas lorsque la spatialisation était réalisée avec la carte des sols.

### Conclusion de la partie 2.1,

La localisation des groupes types :

L'enrichissement du RPG par le rajout de données d'altitude et d'irrigation était l'étape la plus importante, car l'attribution des probabilités de localisation des groupes types s'établit essentiellement sur ces variables. Il est donc important de garder à l'esprit que ces variables peuvent évoluer selon les données utilisées, leurs sources et leur précision. De même, réaliser des groupements parcellaires selon ces deux caractéristiques ainsi que la typologie de cultures AROPAj est nécessaire, afin d'obtenir une unité spatiale correspondant mieux au modèle qui détermine ainsi ses groupes types.

Ces groupements vont pouvoir nous servir à réaliser une simplification et une réduction du nombre de parcelles dans le RPG.

## 2.2. La simplification des parcelles du RPG

En observant des premiers résultats à l'échelle communale, nous pouvons nous rendre compte que l'information est d'une grande précision, et que si nous souhaitons observer, plus tard, les résultats sur une échelle plus petite, telle que celle de la région, il nous faudrait simplifier les géométries, ainsi que supprimer les parcelles en dessous d'une certaine surface. Cela nous permettra d'y voir plus clair sur les représentations que nous pourrons faire par la suite ; et un aussi grand niveau de précision n'a pas forcément un fort intérêt, puisque nous aurons du mal à voir des grandes tendances apparaître si nous observons de trop près les résultats.

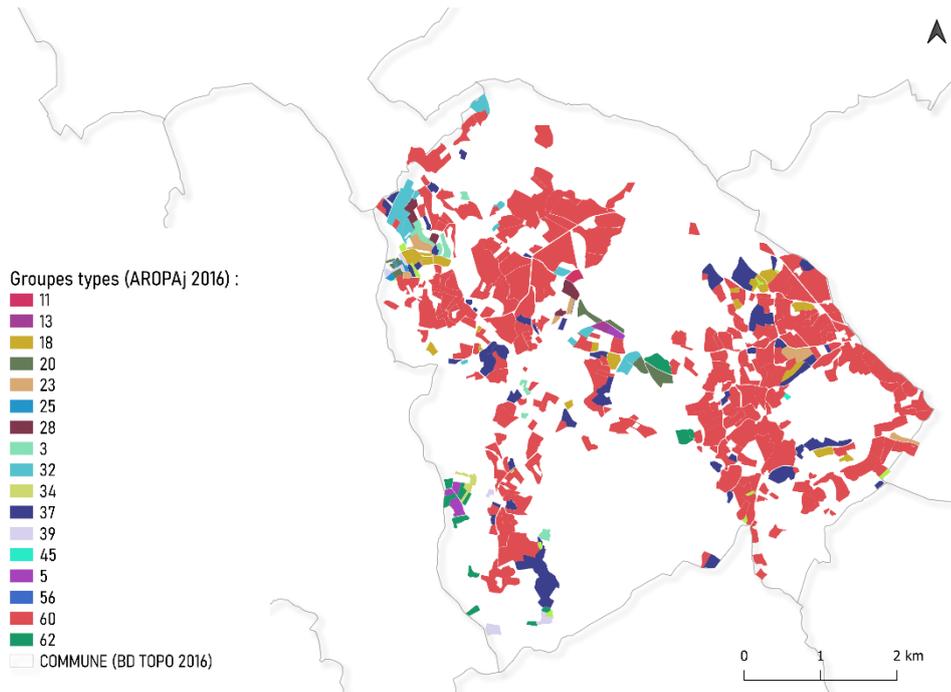


Figure 5 : Visualisation des groupements parcellaires sur la commune de Champeau-en-Morvan (21139) dans le département de la Côte-d'Or

J'ai donc choisi de réaliser un modèleur graphique sur le logiciel de SIG QGIS, dont voici la version finale. Afin d'arriver à ce modèleur final et son utilisation, plusieurs étapes ont été réalisées.

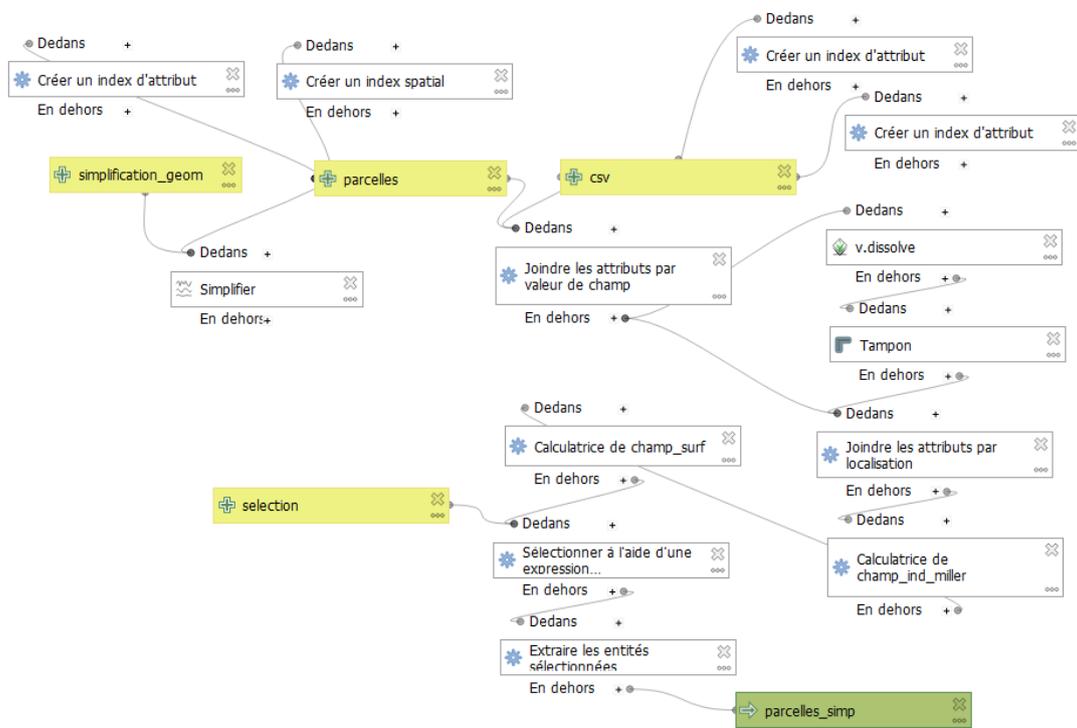


Figure 6 : Version finale du modèleur graphique

### 2.2.1. Les éléments de sélection des parcelles

Le plus gros des traitements nécessite de connaître, pour chaque parcelle, le groupement parcellaire auquel elle appartient afin de pouvoir faire des premiers regroupements spatiaux. Pour avancer plus rapidement, j'ai fait des tests sur deux communes uniquement, dans le département de la Côte-d'Or dans l'actuelle Bourgogne-Franche Comté : Champeau-en-Morvan et Ménessaire. J'ai choisi la région Bourgogne car il s'agit de la première région représentée avec des résultats de localisation de groupes types dans la méthodologie AROPAj. L'une des communes, Ménessaire, a été choisie par rapport à la forme de ses parcelles, correspondant à des prairies, formes intéressantes pour observer la justesse de notre sélection.

Il nous faut donc d'abord faire une jointure des attributs par valeur de champ entre les polygones du RPG et les fichiers DBF des probabilités. Indexer les champs de jointure entre les deux permet de rendre les traitements plus rapides : sont donc indexés, pour le RPG, le champ « id\_parcell », et pour le fichier des probabilités, les champs « id\_parcell » et « id\_grparc ». Cette jointure se fait grâce aux identifiants des parcelles, rajoutés au tableau des probabilités sur R grâce à la table de correspondance que nous avons gardée entre les groupements parcellaires et l'intégralité des identifiants parcellaires.

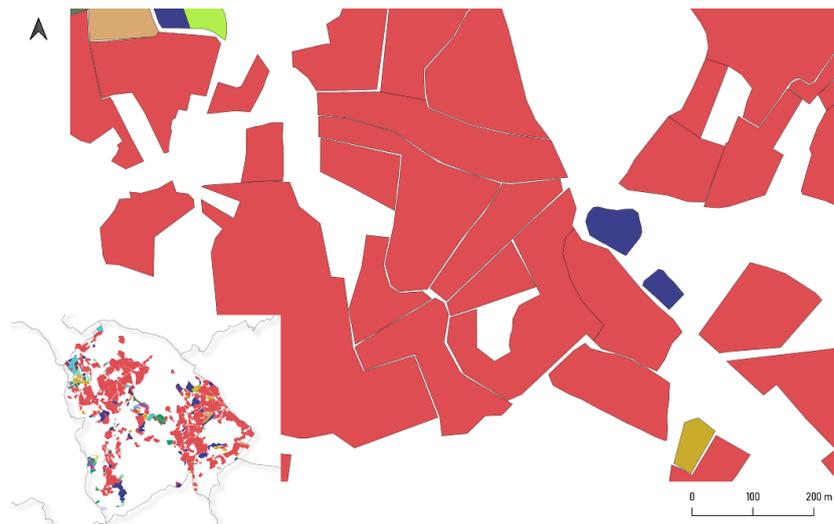


Figure 7 : Tri des polygones par groupements parcellaires

Comme nous avons pu le constater sur la figure 4, nombre de parcelles appartiennent au groupement parcellaire numéro 60. Et, nous pouvons voir, sur la capture d'écran ci-dessus, des polygones appartenant à ce groupement, (en rouge) qui semblent être très proches les uns des autres, voire qui semblent se toucher. Nous envisageons donc de les regrouper, ou pour être plus claire de les fusionner pour qu'ils ne forment plus qu'un seul polygone. Pour cela, il existe sur Qgis deux traitements possibles grâce aux bibliothèques GDAL et GRASS : la fonction « regrouper » de GDAL, qui permet de regrouper les polygones proches et selon un champ d'attribut, et de récupérer en sortie le reste des attributs présents dans la couche en entrée ; la fonction « v.dissolve » de GRASS, qui, quant à elle, ne permet pas cette dernière option, mais nous pouvons cependant choisir le seuil de regroupement géométrique. Ce seuil, bien qu'intéressant, a ses limites : nous ne sommes pas certains de l'unité utilisée ; tout ce que nous sommes en mesure de faire, c'est d'effectuer plusieurs tests selon plusieurs seuils.

La fonction GDAL ne nous accorde pas le choix de la tolérance d'agrégat ; les résultats semblent donc satisfaisants mais parfois questionnables lorsqu'on les compare à des premiers résultats de la fonction GRASS (seuil réglé sur 1).

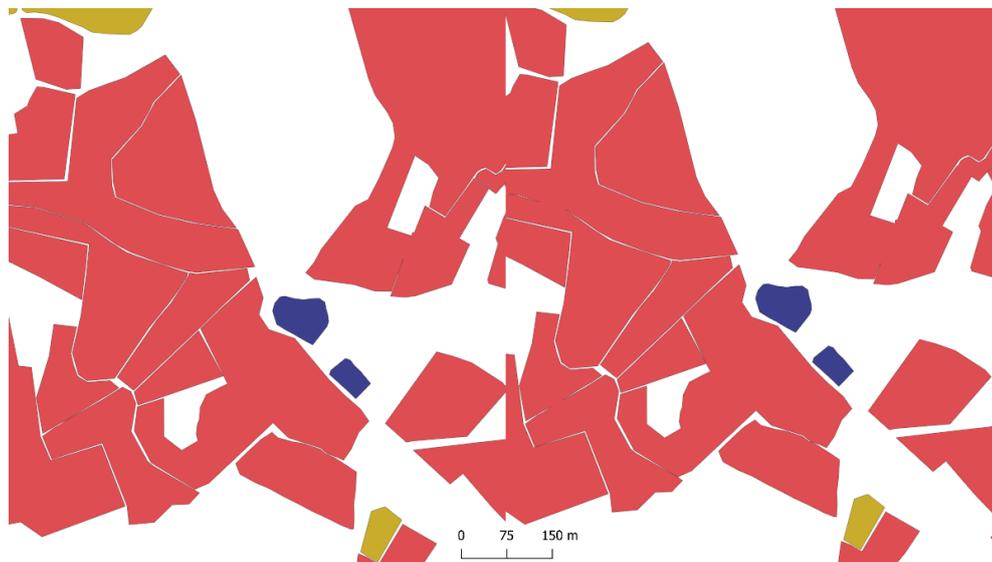


Figure 8 : Comparaison des fonctions GDAL (à gauche) et GRASS (à droite)

Si nous regardons les polygones ci-dessous (à gauche), la fonction GDAL ne les fusionne pas alors qu'ils sont pourtant très proches. La fonction GRASS les fusionne. En se rapprochant au maximum, aucun espace n'apparaît entre les deux polygones : GRASS semblerait avoir juste.

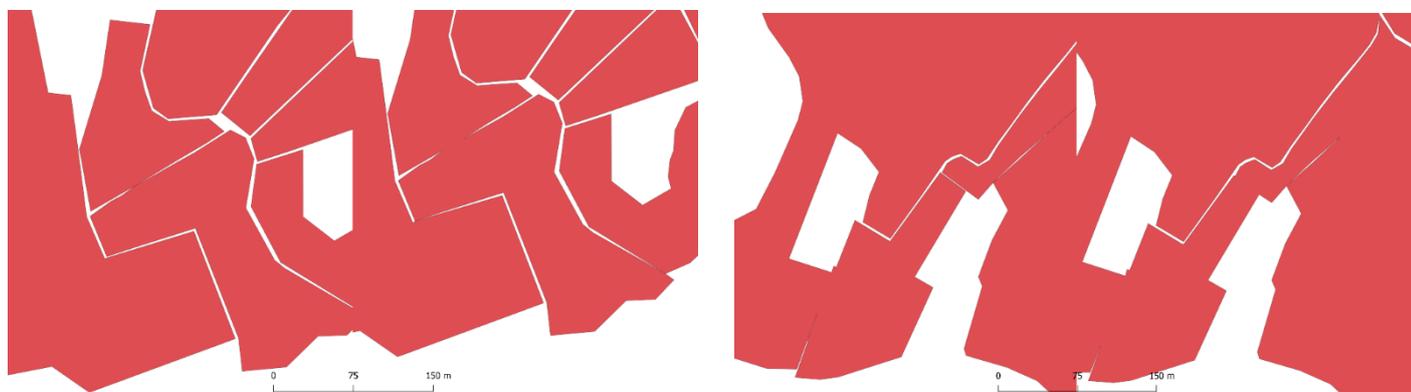


Figure 9 : Comparaison GDAL / GRASS

Néanmoins, GRASS est ici à un seuil de 1, et si l'on observe les différents résultats obtenus avec ce seuil parmi toutes les parcelles, nous nous rendons compte que certains polygones sont regroupés alors qu'ils ne se touchent pas (à droite).

Il faut donc faire attention avec cette tolérance, la fonction semble rajouter des surfaces aux polygones selon le réglage de celle-ci ; cela ne paraît pas énorme à échelle parcellaire, mais pourrait grandement changer les résultats sur une échelle départementale ou régionale. Un seuil pouvant être contrôlé a donc des avantages et, après avoir réalisé plusieurs tests avec différents seuils (allant de 1 à 0,001), nous nous rendons compte que le plus efficace serait celui fixé à 0,01.

Utiliser la fonction GRASS ne nous permet pas cependant de récupérer les attributs, il faudra donc passer par une jointure des attributs par localisation pour les récupérer. C'était principalement pour cette raison que nous avons, au départ, préféré la fonction GDAL, dont les résultats se rapprochaient de ceux de GRASS lorsque ce dernier avait un seuil réglé sur 0,01. Cependant, comme nous le verrons, nous avons mis en place un modeleur graphique, dans lequel la fonction GDAL ne paraissait pas fonctionner correctement. Nous choisissons donc GRASS avec un seuil de 0.01.

Une fois cela fait, il nous faut trier les polygones restants : les plus fins et longilignes qui n'ont pas pu se fusionner avec d'autres ainsi que ceux connaissant une surface plus petite. Pour les polygones longilignes ou d'une forme particulière, nous avons choisi d'appliquer l'indice de circularité de Miller, également utilisé pour trier les polygones dans la méthode du RPG complétée de l'ODR.

L'indice de Miller s'écrit ainsi : «  $(4 * \pi * \text{AREA}) / \text{PERIMETRE}^2$  ». Sur QGIS, nous l'appliquons dans une nouvelle variable (ind\_miller) : «  $(4 * \pi * \$\text{AREA}) / (\$PERIMETER * \$PERIMETER)$  »

Il s'agit du rapport de la superficie d'une entité, ici, un polygone, à un cercle de même périmètre. Cet indice nous donne des résultats de 0 à 1, 0 indiquant une ligne et 1, un cercle. Nous avons choisi, après plusieurs tests effectués, d'éliminer les polygones en dessous d'un seuil de « 0.12 ». Les polygones les plus concernés sont les plus longilignes et ceux n'ayant pas fusionné avec d'autres polygones après le traitement de regroupement géométrique. Pour ensuite supprimer les polygones avec des petites surfaces, nous recalculons d'abord les surfaces en hectare, QGIS ne mettant pas automatiquement à jour les surfaces, modifiées par le traitement de regroupement des polygones («  $\$area/10000$  »).

Nous nous retrouvons avec cette formule de sélection : ("IND\_MILLER">'0.12' AND "SURF\_PARCE">'0.1' )

Nous aurions pu aussi nous pencher sur le sujet des jachères : faut-il les retirer également des parcelles que nous étudions ? Il peut arriver que des agriculteurs parfois déclarent des zones en jachère alors qu'elles n'en sont pas réellement : seulement des zones qui leur appartiennent mais qui ne sont pas mises en culture. Il aurait éventuellement fallu chercher à supprimer les polygones enregistrés en jachère lorsqu'ils ne dépassent pas un certain seuil de surface.

Il s'agit donc de sélectionner uniquement les polygones se situant au-dessus des seuils choisis, et de les enregistrer dans une nouvelle couche. Afin d'enchaîner tous ces traitements, nous créons un modèle grâce aux options de Modeleur graphique de QGIS, qui va nous permettre de les automatiser.

## 2.2.2. Modeleur graphique et ModelBuilder : avantages et inconvénients

Plusieurs modèles ont été réalisés et testés pour ces premières sélections. Nous rajoutons, à la suite de ces premiers test, l'outil de simplification des polygones, après avoir réalisé des tests à 10, 5, 3, et 2 mètres (Annexe 4)

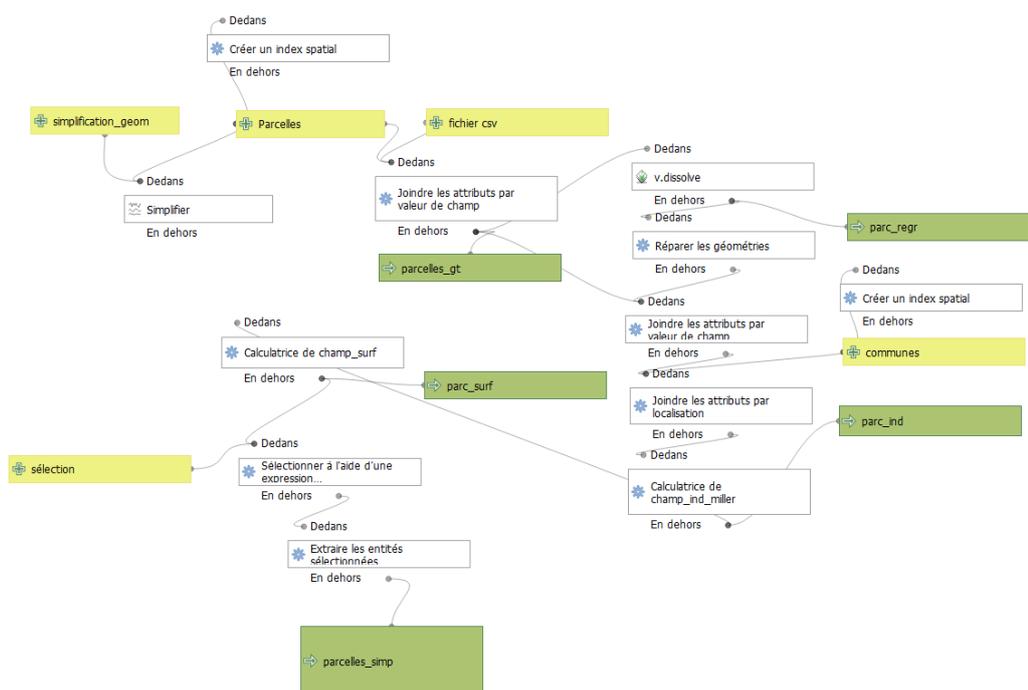


Figure 10 : Troisième version du modèleur, la plus aboutie avant la dernière

Ci-dessous les résultats après avoir lancé le modèle : nous avons à gauche des résultats cohérents et dans la lignée de ce que nous essayons de faire, cependant, à droite, les parcelles ayant des formes plus complexes ont été retirées à cause d'un indice de Miller faible. Même en changeant les paramètres de la fonction GRASS, certains polygones vont inévitablement n'en former plus qu'un seul. Il est nécessaire d'ajuster les paramètres de sélection ce modèleur, après les regroupements, afin que des polygones aussi importants que ceux-ci ne soient pas systématiquement retirés.

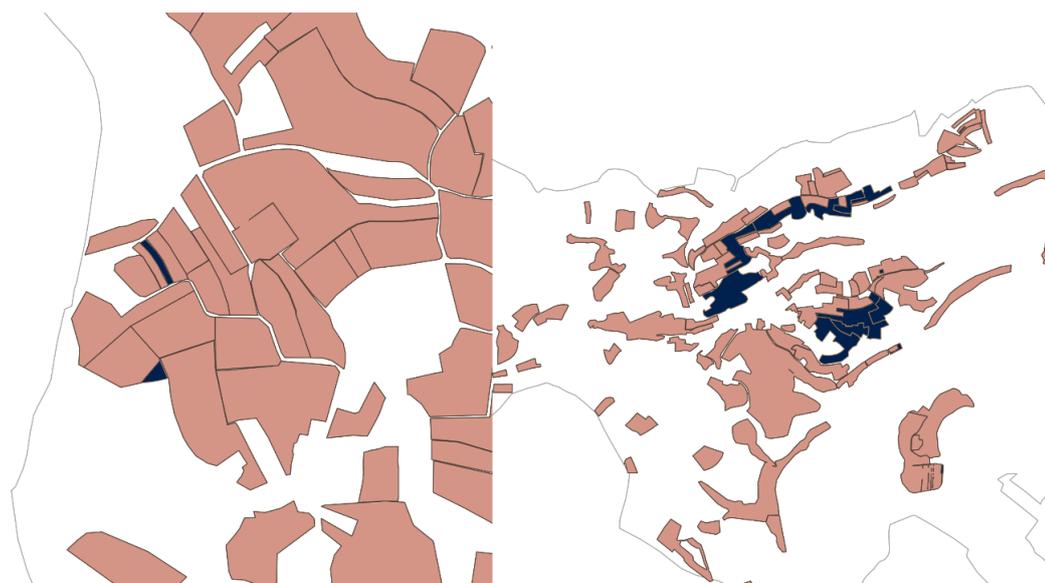


Figure 11 : Différences de sélection entre les communes de Champeau-en-Morvan et Ménessaire (Côte d'Or)

Nous décidons donc de changer la formule de sélection :

`("IND_MILLER">'0.12' AND "SURF_PARCE">'0.1' ) OR ("IND_MILLER"<'0.12' AND "SURF_PARCE" >'0.2')`

Le seuil choisi pour supprimer les polygones est celui de 0.1, lorsque l'indice de Miller est supérieur à 0.12, et de 0.2 lorsque l'indice est inférieur à 0.12, afin de tout de même garder certains polygones dont l'indice est faible mais dont la surface dépasse un certain seuil.

Une fois le modèle réajusté, il s'agissait ensuite de trouver des solutions afin que le modèle tourne sur tous les fichiers parcellaires départementaux à la suite. Il aurait été possible de réaliser un traitement par lot, mais il aurait fallu remplir pour chaque fichier RPG le fichier DBF correspondant manuellement, et cela aurait pu mener à d'éventuelles erreurs. Une autre solution était de se tourner vers la console Python de QGIS : PyQgis. Faire tourner le modèle une fois permet ensuite d'aller chercher dans l'historique des traitements la manière d'appeler et de paramétrer celui-ci dans la console (Annexe 5). Dans la console python, j'ai donc essayé de faire un test qui tournerait sur plusieurs fichiers RPG départementaux et un fichier DBF régional.

Néanmoins, la boucle s'arrête en indiquant une erreur au niveau du traitement de réparation des géométries, le traceback menant seulement vers la ligne où est écrit le nom du modèle « model :simplification\_parcelles » ; nous n'avons donc que peu d'informations sur l'origine réelle du problème. Après avoir essayé de régler le problème, il s'avère que ce traitement ne fonctionne pas sur la plupart des départements ; sans doute l'outil Dissolve crée des polygones avec des erreurs trop importantes ou complexes à régler. Le traitement est pourtant indispensable après celui de regroupement qui modifie une grande partie des polygones. Cependant, « réparer les géométries » ne constitue pas le seul outil qui peut être utilisé sur Qgis pour la correction des topologies. GRASS possède aussi l'outil « v.clean » et nous pouvons également tenter de réparer les géométries avec un simple tampon de 0 mètres : cette dernière option fonctionnant sans problème et ne nécessitant pas un temps de traitement très long, je l'ai choisie. La boucle dans PyQGIS peut retourner avec ce nouveau modèle.

En parallèle, j'ai essayé de reconstruire le modèle sur le ModelBuilder de ArcGIS, afin de comparer les outils disponibles, les différences de temps de traitements, les avantages et inconvénients donnés par chacun pour établir ce que nous voulons réaliser. J'ai commencé à considérer cette option lors de l'erreur survenue à cause de l'outil de réparation des géométries.

Tout d'abord, un grand avantage d'ArcGIS est la multitude d'options d'itérations possibles sur ModelBuilder : dans des dossiers, sur des fichiers de plusieurs types, ou encore uniquement sur des enregistrements ou champs de fichiers. Néanmoins, nous ne pouvons utiliser deux options d'itérations à la fois ; la solution aurait été de réaliser un premier modèle itérant sur les parcelles, puis de glisser ce modèle à l'intérieur d'un autre, ce dernier itérant sur les fichiers DBF.

L'outil de regroupement étant l'étape la plus importante de notre modèle, il nous faut retrouver une équivalence des fonctions de GRASS ou GDAL sur ArcGIS. Deux outils se distinguent parmi le catalogue du logiciel ESRI : l'outil « dissolve with boundaries », mais qui, tout comme GDAL, ne nous permet pas de régler l'intensité du regroupement ; ainsi que l'outil « agréger les polygones » qui, comme décrit dans la documentation « *est destiné à une agrégation et à une réduction d'échelle modérées lorsque les entités en entrée ne peuvent plus être représentées individuellement en raison de l'espace limité de la carte ou de la résolution requise pour les données* »<sup>4</sup>, correspondant donc très bien à ce que nous

---

<sup>4</sup> <https://pro.arcgis.com/fr/pro-app/latest/tool-reference/cartography/aggregate-polygons.htm>

cherchons à produire. De nombreux seuils de regroupement sont possibles et sont clairement définis : millimètres, kilomètres...

Comme pour le reste, nous nous heurtons à des « impasses » : ce modèle (Annexe 3), pour un département, met parfois une dizaine d'heures à s'effectuer ; notamment la partie sur l'agrégation des polygones. J'avais également songé à réaliser ces traitements sous R, les jointures et le traitement des données étant beaucoup plus rapides ; cependant l'agrégation spatiale avec R semble complexe, voire ne semble pas possible, en tout cas pas aussi facilement qu'avec des logiciels de SIG.

Tout en sachant qu'ArcGIS reste une option explorée et dont les résultats sont satisfaisants malgré un temps d'exécution beaucoup trop lent, nous revenons sur QGIS (dont le temps de traitement dépasse rarement les 30 minutes pour un département). L'avant-dernière version du modèle (Figure 10) ne récupérait que les codes INSEE des communes par jointure spatiale car nous pensions n'avoir besoin que de ces codes ; cependant, les identifiants parcellaires ainsi que les codes culture et les codes groupe de culture du RPG, bien qu'altérés par le regroupement, doivent être récupérés au mieux : ils constituent l'avantage du RPG pour cette spatialisation, celui de connaître précisément l'occupation du sol. Le modèle est donc modifié ; une jointure spatiale est directement faite pour récupérer toutes les informations du fichier RPG agrémenté des informations du fichier DBF, avec le paramètre « intersection ». Pour les polygones regroupés, le polygone final prend l'identifiant du polygone dont la surface est la plus importante au départ.

### 2.2.3. PyQGIS : méthode d'automatisation

Sur PyQgis, nous faisons donc une boucle qui récupère tous les shapefiles départementaux correspondant à une région et réalise les traitements du modèle « simplification\_parcelles » sur chacun d'entre eux (Annexe 6).

Pour ce faire, nous devons importer plusieurs bibliothèques. D'abord, la bibliothèque «processing » pour réaliser des algorithmes sur la console. Panda est importée afin de pouvoir lire et trier comme un data.frame les données de notre csv sur les régions et départements. Enfin, nous importons « glob », l'abréviation de « global », utilisée pour aller extraire tous les fichiers correspondant au pattern indiqué après la commande « glob.glob ». Avant d'utiliser glob, j'avais tenté d'utiliser le module « os », qui permet d'interagir avec les fonctionnalités du système d'exploitation, mais qui permet surtout de travailler avec les fichiers et les répertoires. J'indiquais simplement à Python que les éléments sur lesquels tourner étaient présents dans un chemin défini grâce à la commande `os.listdir(chemin)` qui devenait l'itérateur. Cependant, cette méthode posait des problèmes : un jour elle fonctionnait, le lendemain elle ne fonctionnait plus ; l'explication la plus plausible résidait dans le fait que PyQGIS ne reconnaissait peut-être plus les chemins d'accès permettant de récupérer les fichiers.

La boucle a d'abord été testée simplement en indiquant un numéro de région ; et une fois que nous savions que ce script fonctionnait après avoir fait plusieurs tests, les numéros des régions ont été intégrés à la boucle afin de pouvoir la lancer sans avoir à préciser chaque numéro de région. Nous comptons environ trois jours pour que la boucle passe par tous les départements et réalise les traitements du modèle (le traitement le plus long étant celui de la jointure par localisation).

En moyenne, 30,67 % des parcelles sont retirées du RPG, qu'elles soient véritablement retirées ou alors fusionnées avec d'autres. Cependant, nous pouvons remarquer que cela dépend parfois selon les régions. Voici le nombre de parcelles au départ de chaque région puis le nombre de parcelles à l'arrivée :

REGION	PARCELLES AU DÉPART	PARCELLES À L'ARRIVÉE	PARCELLES RETIRÉES	% RETIRÉES
11	119 621	95 812	23 809	19,90
21	307 961	228 891	79 070	25,68
22	289 209	226 964	62 245	21,52
23	161 326	127 117	34 209	21,20
24	536 080	403 243	132 837	24,78
25	352 050	275 218	76 832	21,82
26	423 604	297 821	125 783	29,69
31	263 486	194 966	68 520	26,01
41	295 118	208 095	87 023	29,49
42	243 612	158 194	85 418	35,06
43	181 666	140 400	41 266	22,72
52	691 524	482 823	208 701	30,18
53	780 344	584 044	196 300	25,16
54	599 497	438 266	161 231	26,89
72	592 330	384 106	208 224	35,15
73	1 032 425	718 452	313 973	30,41
74	346 275	256 967	89 308	25,79
82	748 463	497 687	250 776	33,51
83	555 423	403 963	151 460	27,27
91	491 605	190 725	300 880	61,20
93	282 288	142 245	140 043	49,61
94	35 643	11 967	23 676	66,43

Tableau 2 : Pourcentages d'enlèvement des parcelles

Sans surprise, les régions les plus concernées par l'enlèvement de leurs parcelles sont celles constituant aujourd'hui la région PACA : en effet, celle-ci se caractérise par une agriculture tournée principalement vers l'horticulture, qui n'est pas prise en compte par AROPAj. La région 42, l'Alsace, est concernée par de la viticulture, également non traitée par AROPAj.

Tous les fichiers traités étant sortis, nous pouvons considérer le nettoyage automatique des tables attributaires : nous avons depuis le début gardé toutes les informations stockées sur nos fichiers shapefile ; le détail du RPG, les altitudes, les pentes... Il est malgré tout important de garder une trace de toutes ces données, même si nous considérons qu'elles ne nous sont plus forcément utiles à ce stade de la spatialisation. Nous allons donc conserver ces fichiers shapefiles, mais en créer de nouveaux, plus allégés au niveau de la table attributaire. Nous conservons les champs :

CHAMPS	INFORMATIONS
Fid	Nouvel identifiant après les regroupements
Id_parcell	Identifiant à conserver afin de faire le lien avec d'autres bases de données (BNVd)
Surf_parce	Information à conserver
Code_culture	Information à conserver
Groupe culture	Information à conserver
Insee_com	Information à conserver
Proportion	Proportion de la parcelle à l'intérieur de la commune : Information à conserver
A_min	La variable nous ayant permis de faire les classes d'altitude : à conserver
Id_grparc	Les groupements parcellaires : à conserver
Les groupes types	Faire le lien avec les « sorties » du modèle AROPAj
Gr_parcom	Groupements parcellaires avec identifiant INSEE de la commune : peu mener à d'autres types de représentations
Ind_miller	Information à conserver si jamais nous souhaitons par la suite continuer à trier des parcelles selon leur forme

Tableau 3 : Les informations conservées

Avec R, nous pouvons facilement réaliser cette opération.

## Conclusion de la partie 2.2,

### La simplification des parcelles du RPG :

Avec un peu plus de 9 millions de parcelles pour la version 2016 du RPG, et alors que nous avons déterminé un « groupement parcellaire » correspondant aux caractéristiques de désagrégation spatiale pour chaque parcelle, il nous a semblé nécessaire de diminuer le nombre de polygones présents dans la base de données, afin d'« alléger » au mieux les représentations.

Finalement, nous ne remarquons pas beaucoup de changement visuellement, sauf si nous nous rapprochons à l'échelle de la parcelle. Cependant, le chargement des parcelles est moins lourd sur QGIS et des possibilités pour encore plus simplification s'ouvrent : nous avons tout de même conservé des seuils de regroupement et de simplification géométriques assez bas pour plus de sécurité face à ce que nous cherchons à faire par la suite. Presque 3 millions de parcelles ont été retirées, soit 30,67%.

Plusieurs tests ont été réalisés afin de trouver les meilleures méthodes, ou en tout cas, dans la mesure du possible, les plus rapides, mais tout cela reste à être amélioré.

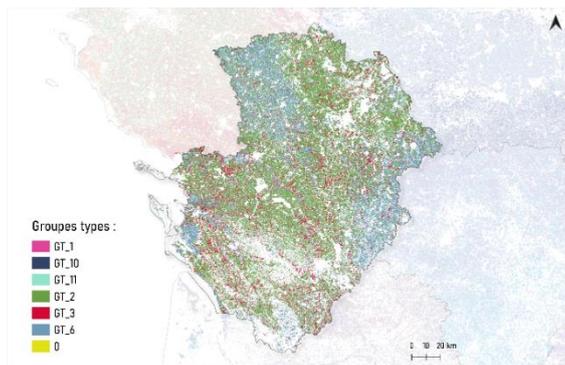
## 3. Visualisation des résultats et perspectives

### 3.1 Quels groupes types sur quelles parcelles ?

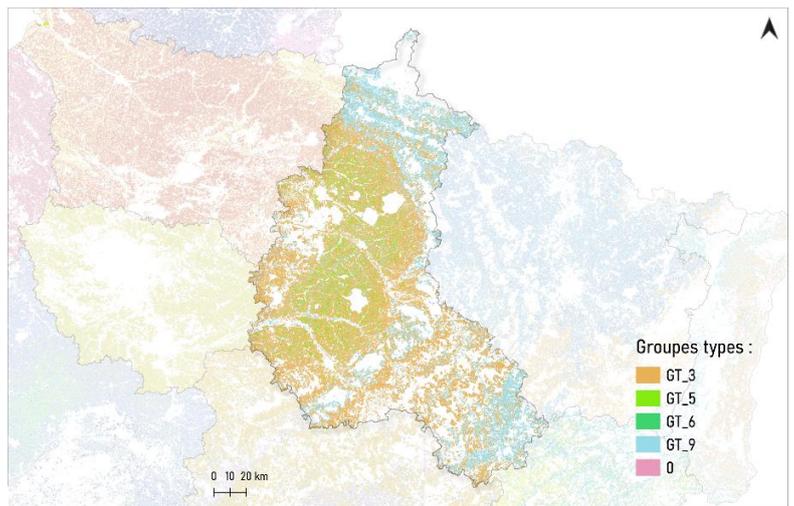
Nous avons à présent des fichiers de parcelles allégés en termes de géométrie et sur lesquels nous connaissons les estimations de groupes types. Cependant, avec la volonté d'observer et de cartographier les groupes types en une seule représentation, nous pouvons essayer d'obtenir la probabilité de groupe type la plus haute pour chaque parcelle.

#### 3.1.1. Les biais derrière l'association d'une parcelle à un groupe type

Avec R, nous pouvons modifier les DBF des Shapefiles (tant que nous n'enlevons pas d'enregistrements, cela pouvant altérer les géométries) pour rajouter une colonne contenant le groupe type avec la probabilité maximum sur chaque parcelle. Comme nous l'avons précédemment fait pour le script des probabilités, nous encapsulons ce script en tant que fonction, et nous créons une boucle pour que sur chaque fichier de parcelles soit rajoutée cette colonne. Sur QGIS, nous pouvons lancer un simple script python qui rassemble en région les fichiers départementaux, dans le but de faire un atlas pour cartographier et visualiser les groupes types de chaque région.



Les groupes types de la région Poitou-Charentes (54) en 2016



Les groupes types de la région Champagne-Ardenne (21) en 2016

Figure 12 : Exemples de cartographies de l'atlas centré sur l'attribution des groupes types majoritaires par région

Source : RFS (2016), AR6PA (2016), IGN, GEOFLA  
Réalisation : Bernadette C. 2022

L'atlas représentant l'ensemble des régions françaises métropolitaines est en annexe (Annexe 7). Pour les limites administratives des régions, nous avons utilisé les données GEOFLA de l'IGN.

D'abord, nous notons que malgré le travail de simplification parcellaire, nous nous retrouvons toujours devant des résultats cartographiques très fins. Il serait à l'avenir possible d'envisager d'augmenter les paramètres de simplification.

Dans la majorité des régions cartographiées, nous observons que certaines parcelles n'ont été rattachées à aucun GT : cela vient du fait que le fichier contenant les surfaces AROPAJ utilisé pour déterminer les probabilités ne contient pas toujours des surfaces pour chaque type de culture RICA selon les régions. Par exemple, les types de culture RICA 121, 125 ou encore 127, correspondant respectivement au blé dur, à des céréales mixtes, et au riz, n'ont que peu ou pas de surfaces renseignées dans le fichier : les probabilités d'appartenir à un groupe type AROPAj sont donc, logiquement, altérées voire impossibles.

Inversement, des GT comme par exemple les GT 1, 4, 7 ou encore 8 ne sont eux jamais représentés sur des parcelles. Ils sont pourtant bien présents dans nos fichiers de probabilités, mais ne sont jamais majoritaires. Il est nécessaire de souligner le biais qui réside dans le fait de choisir le GT majoritaire pour chaque parcelle : en effet, les autres GT peuvent ne pas être majoritaires sur les parcelles mais obtenir tout de même une probabilité de présence élevée. De plus, certains GT ont des probabilités de présences égales (nous pouvons citer les groupes types 19, 20, 21 et 22 de la région 23, c'est le groupe type 22 qui a été sélectionné).

id_gparc	GT_17	GT_18	GT_19	GT_20	GT_21	GT_22
16	0,167196121049992	0,164019394750042	0,167196121049992	0,167196121049992	0,167196121049992	0,167196121049992
3	0,167196121049992	0,164019394750042	0,167196121049992	0,167196121049992	0,167196121049992	0,167196121049992

Tableau 4 : Des probabilités de présences égales

Après avoir fait ces deux observations, nous remarquons d'autres choses. Premièrement, pour plusieurs régions (surtout des régions plus petites, avec deux départements seulement) un seul groupe type seulement apparaît ; c'est donc dans ces cas-là qu'il faut

faire attention avec ce choix de n'avoir, pour chaque parcelle, représenté que le groupe type principal « retenu ». J'avais la volonté, grâce à cette démarche, de simplifier la cartographie des sorties, et bien que ce ne fut pas forcément une bonne initiative ; nous avons pu avoir une première visualisation de la répartition des résultats, et du RPG en tant que nouvelle carte des sols.

Ensuite, même quand certaines régions, plus grandes, abritent plus de GT, deux voire trois paraissent souvent majoritaires à première vue sur la région. Néanmoins, si l'on regarde avec plus de précision, tous les groupes types présents dans la légende sont bien représentés. Ces cartes sont donc intéressantes car nous pouvons à la fois voir « de loin » des grandes tendances, mais lorsque l'on s'y intéresse de plus près nous voyons l'hétérogénéité des résultats et donc des types d'agriculture sur les parcelles.

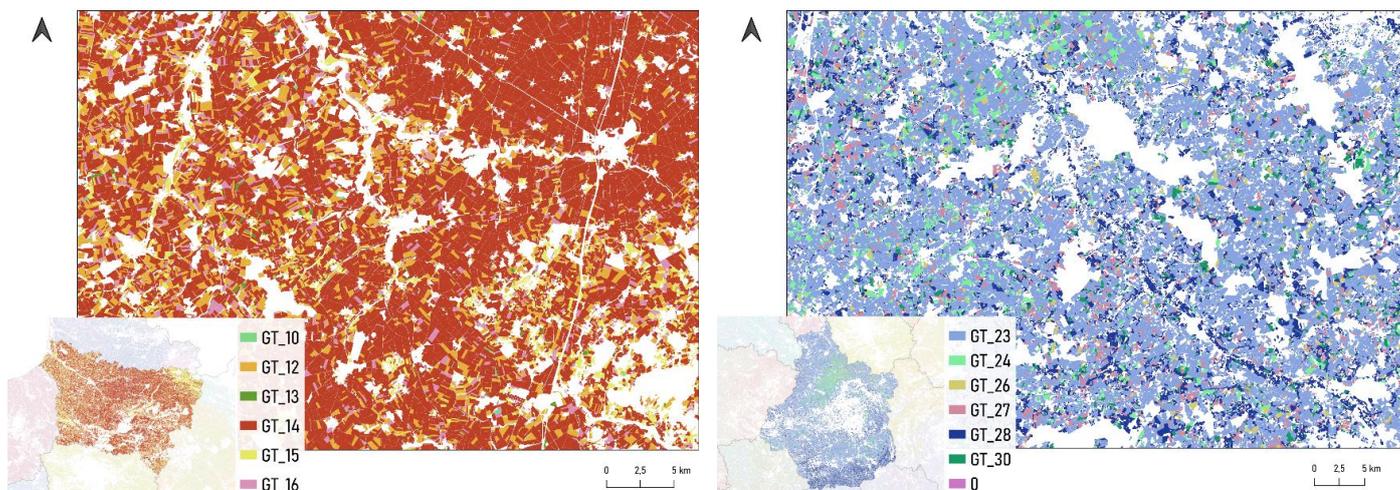
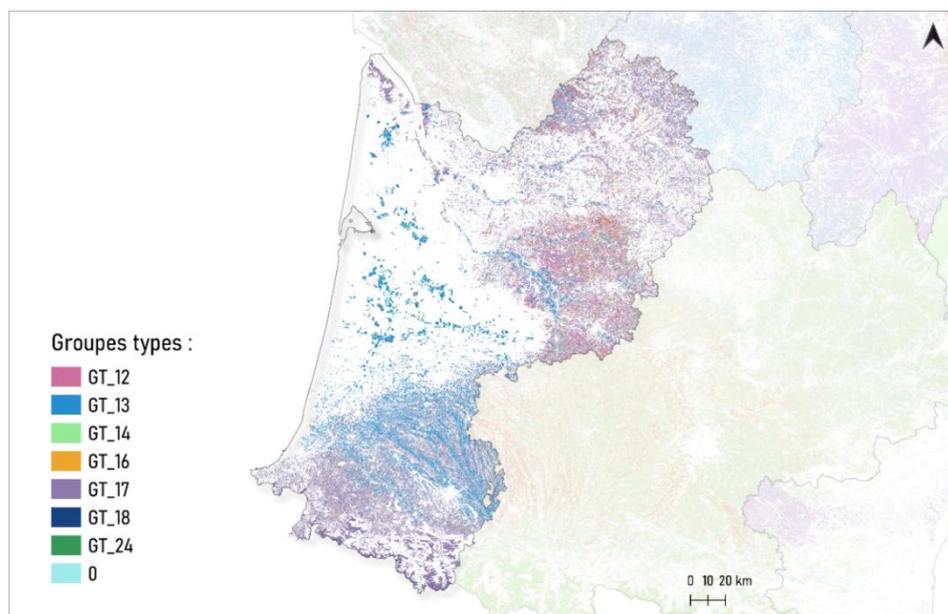


Figure 13 : Zoom sur les groupes types



Les groupes types de la région Aquitaine (72) en 2016

Nous observons aussi que souvent, des groupes types semblent majoritaires par département, et la présence de plusieurs groupes types à l'intérieur d'un département survient surtout lorsqu'il y a une différence d'altitude, comme nous pouvons le voir par exemple très bien sur le département 64 de la région Aquitaine.

Sources : RPG (2016), AROPA (2016), ISN-GEODLA  
Réalisation : Bermudes C, 2022

Figure 14 : Les groupes types d'Aquitaine

### 3.1.2. Des résultats plus justes avec le RPG

Comme nous l'avons exposé précédemment lors de l'explication du projet, cette spatialisation a déjà été opérée sur une carte des sols beaucoup moins précise. Nous pouvons donc, à présent, comparer visuellement les résultats. Nous avons ici un exemple avec l'Alsace, facile à appréhender car la région ne comporte que quatre groupes types.

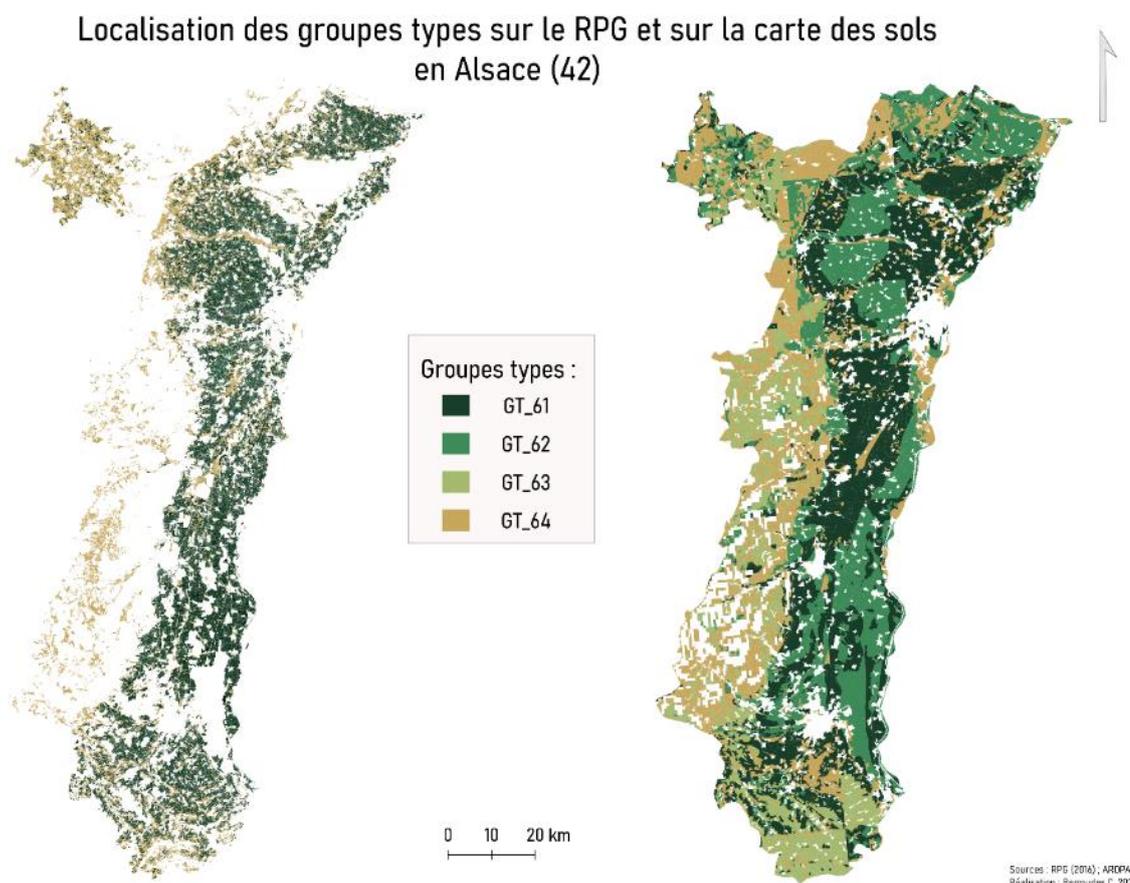


Figure 15 : Comparaison de la spatialisation des groupes types entre le RPG (à gauche) et la carte des sols (à droite)

Tout d'abord, nous voyons que le résultat de la spatialisation via la carte des sols (à droite), comparé au résultat obtenu via le RPG (à gauche), retourne beaucoup plus de surface. Comme nous l'avons déjà évoqué, le plus petit polygone de la base de données CLC représente 25 ha, la carte des sols, basée sur CLC, couvre donc plus de surface que le RPG, auquel, rappelons-le, nous avons retiré 35,06% des parcelles et qui déjà ne présente pas la totalité des surfaces agricoles du territoire français. Si nous regardons l'orientation technico-économiques des communes de Bourgogne Franche-Comté :

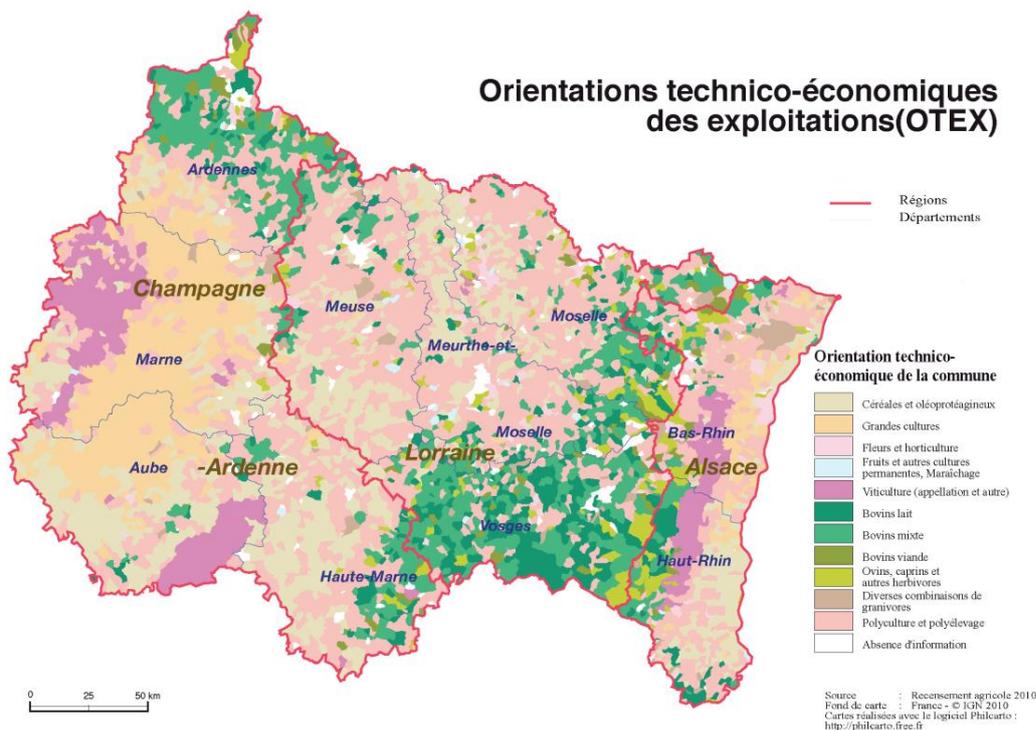


Figure 16 : Orientations technico-économiques des exploitations (OTEX) des régions Champagne-Ardenne, Lorraine et Alsace d'après le Recensement agricole de 2010.

Diagnostic de la région Alsace, Champagne-Ardenne, Lorraine (ACAL), 2015, Préfectures des régions Alsace, Champagne-Ardenne, Lorraine.

Même si les couleurs ne sont pas toujours très distinctes, nous remarquons, à l'Ouest de l'Alsace, une forte présence de la viticulture, vers le Nord Est, quelques communes dont l'OTEX sont les fleurs et l'horticulture, et au Nord, des communes avec « diverses combinaisons de granivores ». Pour les deux premiers, nous savons que ces parcelles ont été retirées du RPG car non traitées par le modèle. Le RPG semble donc être plus adapté à l'utilisation du modèle, puisque nous avons la possibilité de connaître les cultures présentes, et de les trier en fonction de celui-ci.

### 3.2. Les sorties du modèle à l'échelle de la parcelle

Les sorties du modèle sont multiples : pesticides, engrais, charges économiques des agents, entre autres. Précédemment, comme nous l'avons déjà indiqué, ces sorties étaient spatialisées sur une carte d'usage et d'occupation des sols construite à partir de multiples bases de données. Nous allons pouvoir observer ces sorties au niveau parcellaire, et quelles perspectives d'analyses cela nous ouvre.

#### 3.2.1. La diversité des sorties du modèle AROPAj

Les utilisations du modèle AROPAj sont diverses pour les champs d'application en économie de l'agriculture, de l'environnement ou des ressources. Souvent associé au modèle de cultures STICS (Simulateur mulTidisciplinaire pour les Cultures Standard) également développé à l'INRA en 1996 pour simuler le système sol-atmosphère-culture, il a été utilisé, par exemple, pour développer divers scénarios de taxation différentielle sur l'azote selon l'exploitation agricole et le plan d'eau concerné (Lungarska et Jayet, 2016).

Ce rapport ne fait évidemment pas l'objet d'études aussi poussées, puisque nous cherchons surtout à connaître le potentiel derrière l'association du modèle avec les données d'occupation du sol RPG.

J'ai pu recevoir un fichier de sorties avec plusieurs types de variables : les champs « *consengr* » qui indique la consommation totale du GT en engrais, et « *margbrut* » indiquant la marge brute totale du GT (en fonction des prix des produits, taxes, coûts des intrants, aides...). Pour connaître approximativement la consommation en azote, il est possible de multiplier par 0.22 les résultats de la consommation en engrais, les engrais ayant des compositions en azote variables. Nous avons également des données comme la *surface totale et la collecte totale de blé tendre*.

GT	SURF_TOT	CONSENGR	SURFBLET	COLLBLET	MARGBRUT
----	----------	----------	----------	----------	----------

Tableau 5 : Des sorties d'AROPAj

Il nous faut rappeler que ces sorties ne correspondent pas à la réalité, mais sont bien des scénarios, rentrant dans une démarche prospective selon les réformes de la PAC et autres influences économiques et environnementales. L'application de ces sorties sur les parcelles du RPG découle d'une extrapolation de données de sorties AROPAj disponibles à l'hectare. Nous faisons, sur R, une division entre les données de surfaces d'engrais et les données de surface totale en hectare des groupes types indiquées dans les sorties AROPAj :

```
fraengrais <- fraregion$consengr/fraregion$surf_tot
```

Pour les sorties en lien avec les rendements de blé, nous sélectionnons uniquement les parcelles dont le code groupe RPG est égal à 1, qui rassemble la grande majorité des parcelles de blé RPG. Nous confirmons encore une fois que le RPG est idéal pour les sorties du modèle, qui sont parfois spécifiques selon les cultures : la carte des sols

Avec ces résultats, nous faisons une multiplication de matrice : le tableau contenant les probabilités de présence des GT par groupement parcellaire et celui contenant les surfaces d'engrais par GT. Nous obtenons un nouveau tableau, nous indiquant la consommation d'engrais (tonnes à l'hectare) ou la marge brute par groupement parcellaire (réellement par groupe types, puisqu'à chaque groupement parcellaire est associé un groupe type).

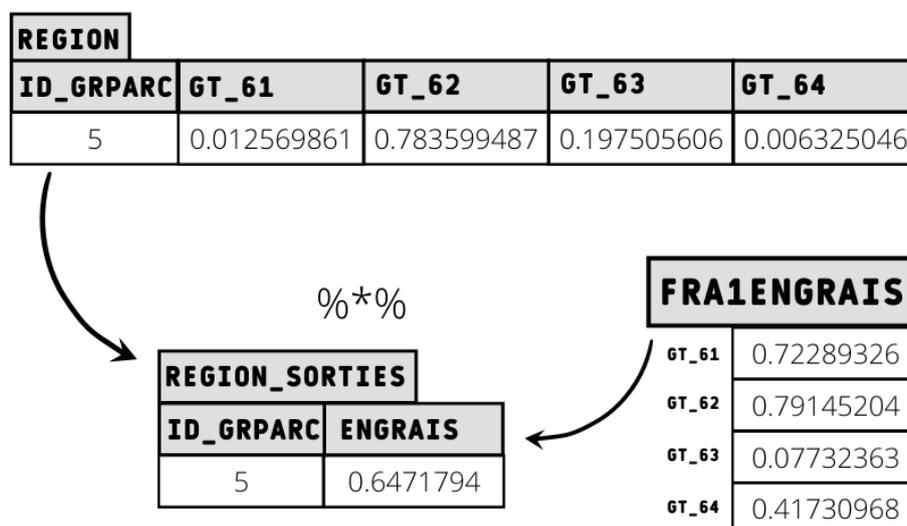


Figure 17 : Schéma du calcul matriciel effectué pour le groupement parcellaire 5 de la région 42 (Alsace)

Nous pouvons alors faire une jointure (sur le champ ID\_GRPARC) entre ce tableau et nos derniers fichiers RPG contenant la géométrie des parcelles.

J'ai fait le choix de visualiser ces sorties avec le logiciel QGIS, bien que j'aie également réalisé, à la suite de ces calculs matriciels et jointures, un script permettant de les cartographier avec R à l'aide la librairie map\_sf. Je trouvais la cartographie sur QGIS simplement plus rapide.

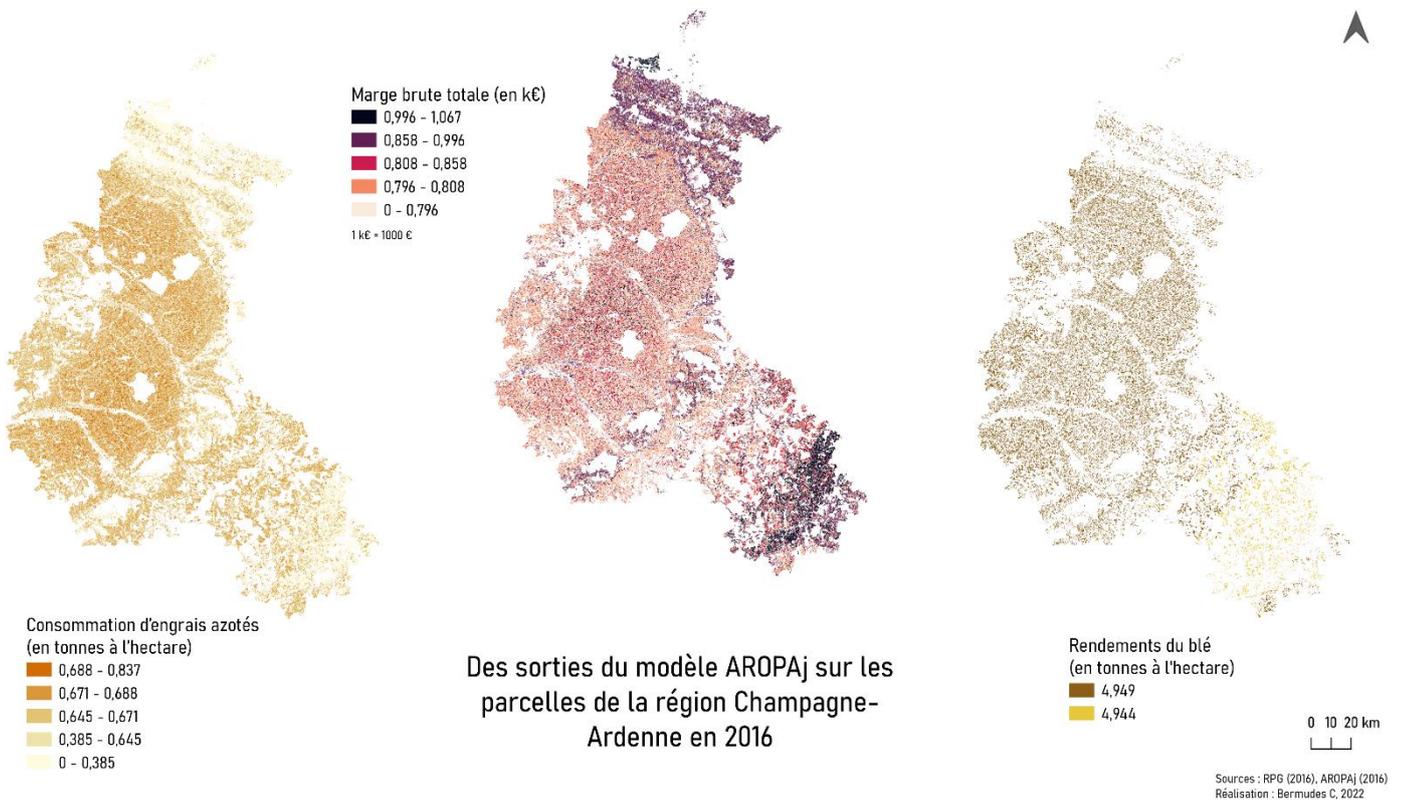


Figure 18 : Représentations de plusieurs sorties AROPAj sur la région Champagne-Ardenne

J'ai choisi de surtout me concentrer, pour les prochaines parties, sur la consommation d'engrais, étant la plus simple à analyser et à lier avec d'autres données. Cependant, je me pencherai sur les autres et éventuellement plus encore dans la suite de mon stage.

### 3.2.2. L'occupation du sol affectée à la parcelle : l'avantage du RPG

L'avantage principal de l'utilisation du RPG, en plus de permettre la visualisation des sorties du modèle à fine échelle, est bien de connaître la véritable occupation du sol présente sur la parcelle. Ayant décidé de me concentrer sur les sorties de la consommation d'engrais, je voulais donc savoir quels étaient les types de cultures les plus consommatrices, et si nous retrouvions les mêmes types de cultures consommatrices dans chaque région.

Néanmoins, afin de réaliser plus rapidement des cartes, j'ai préféré n'utiliser qu'une seule manière de discrétiser (réalisation d'un découpage en classes) pour chacune des régions :

les classes d'effectifs égaux (les quantiles), la plupart des distributions des données s'affichant dissymétriques et bi voire plurimodales sur un histogramme. Il est nécessaire de garder cela à l'esprit lorsque nous analysons les cartographies et que nous sélectionnons les types de cultures présents dans la classe de valeur la plus haute (j'aurais pu sélectionner la ou les classes de valeurs les plus hautes, certaines classes de valeurs étant très proches).

2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes et fleurs	4 : autres céréales	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
17 905	4 440	5 134	9 361	1 477	6 336	91

25 : légumes et fleurs	3 : orge	4 : autres céréales	5 : colza
23 520	38 107	18 952	12 004

15 : légumineuses à grains	2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes et fleurs	4 : autres céréales	5 : colza	6 : tournesol	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
1 105	37 524	985	6 524	14 112	9 977	39 765	12 468	6 454	17

Tableau 6 : Types de cultures (codes groupe du RPG) présents dans la classe de valeur la plus haute (classes d'effectifs égaux) pour les régions (de haut en bas) : 24 (Centre), 53 (Bretagne), 73 (Midi-Pyrénées)

Les groupes de culture dominants sont le maïs, l'orge, les légumes ou fleurs, le tournesol, le colza, ou encore les « autres cultures industrielles » : cela dépend de la région, puisque, par exemple, certaines abritent plus de parcelles de maïs que d'autres. D'autres tableaux pour d'autres régions se retrouvent en annexe (Annexe 8).

Plantes Cultivées	Quantité d'azote nécessaire pour produire une unité de rendement (q : quintal ; t : tonne)	Commentaire
Blé tendre	2,8 à 3,2 kg N/q de grain	En fonction des variétés
Blé dur	3,5 à 4,1 kg N/q de grain	En fonction des variétés
Colza	7 kg N/q de grain	
Lin oléagineux	4,5 kg N/q de grain	
Chanvre	15 kg N/t de paille et chènevis	
Maïs	2,1 à 2,3 kg N/q de grain	En fonction de l'objectif de rendement
Orge	2,5 kg N/q de grain	A minorer pour orge brassicole
Sorgho	2,1 à 2,9 kg N/q de grain	En fonction de l'objectif de rendement
Seigle	2,3 kg N/q de grain	

Tableau 7 : Définition des besoins azotés pour la production de quelques plantes cultivées. Sources : Instituts techniques Arvalis et Terres Inovia, 2016, utilisé ici : <https://www.encyclopedie-environnement.org/vivant/les-nitrates-dans-lenvironnement/>

Ce qui peut nous paraître étrange cependant est la présence de quelques parcelles de légumineuses à grains, oléagineux et protéagineux, des plantes n'ayant que peu ou pas besoin d'engrais azotés. Nous pouvons penser qu'il s'agit de parcelles sur lesquelles sont mises en place des rotations entre diverses céréales et ces cultures. Il s'agit peut-être, selon les sorties AROPAj, de groupes types concernés par de l'orge ou du blé.

Ces tableaux permettent de démontrer le potentiel d'utilisation du RPG. L'association du RPG avec les probabilités de localisation des groupes types a été fructueuse, puisque les résultats semblent pour le moins cohérents. Ces tableaux ne sont néanmoins qu'un petit exemple de ce que nous pouvons être en mesure de faire avec le RPG ; et il est aussi nécessaire de faire attention à ces résultats, la manière dont ils ont été obtenus et aux types de cultures présents en général sur la région.

### 3.3.3. Quels liens sont possibles avec d'autres types de données ?

En décembre 1991 entrait en vigueur la « directive nitrate », directive européenne avec deux objectifs principaux, réduire la pollution des eaux par le nitrate d'origine agricole, et prévenir l'extension de cette pollution. La présence de nitrate dans l'eau favorise l'eutrophisation de celle-ci : c'est-à-dire le développement de micro-organismes se nourrissant d'oxygène, en privant ainsi d'autres êtres vivants et entraînant un déséquilibre de l'écosystème. Néanmoins, le nitrate est l'une des formes minérales de l'azote, élément essentiel à la bonne croissance des plantes. Il est donc normal de retrouver de l'azote dans la majorité des matières fertilisantes (engrais) utilisées par les agriculteurs.

La déclinaison française de la directive nitrate passe par l'établissement de stations de mesure permettant de savoir où se trouvent les zones les plus affectées par ces pollutions (dites « zones vulnérables ») et de suivre leur évolution. \*

Afin d'évaluer le potentiel derrière des sorties d'AROPAj à échelle parcellaire confrontées à d'autres données, nous choisissons de télécharger les données surfaciques des zones vulnérables (révisions datant de 2022, ce qui n'est finalement pas gênant étant donné que les résultats du modèle représentent des scénarios allant jusqu'à 5 ans). Observons les résultats de la consommation d'engrais à l'hectare couplée à localisation des zones vulnérables pour quelques régions :

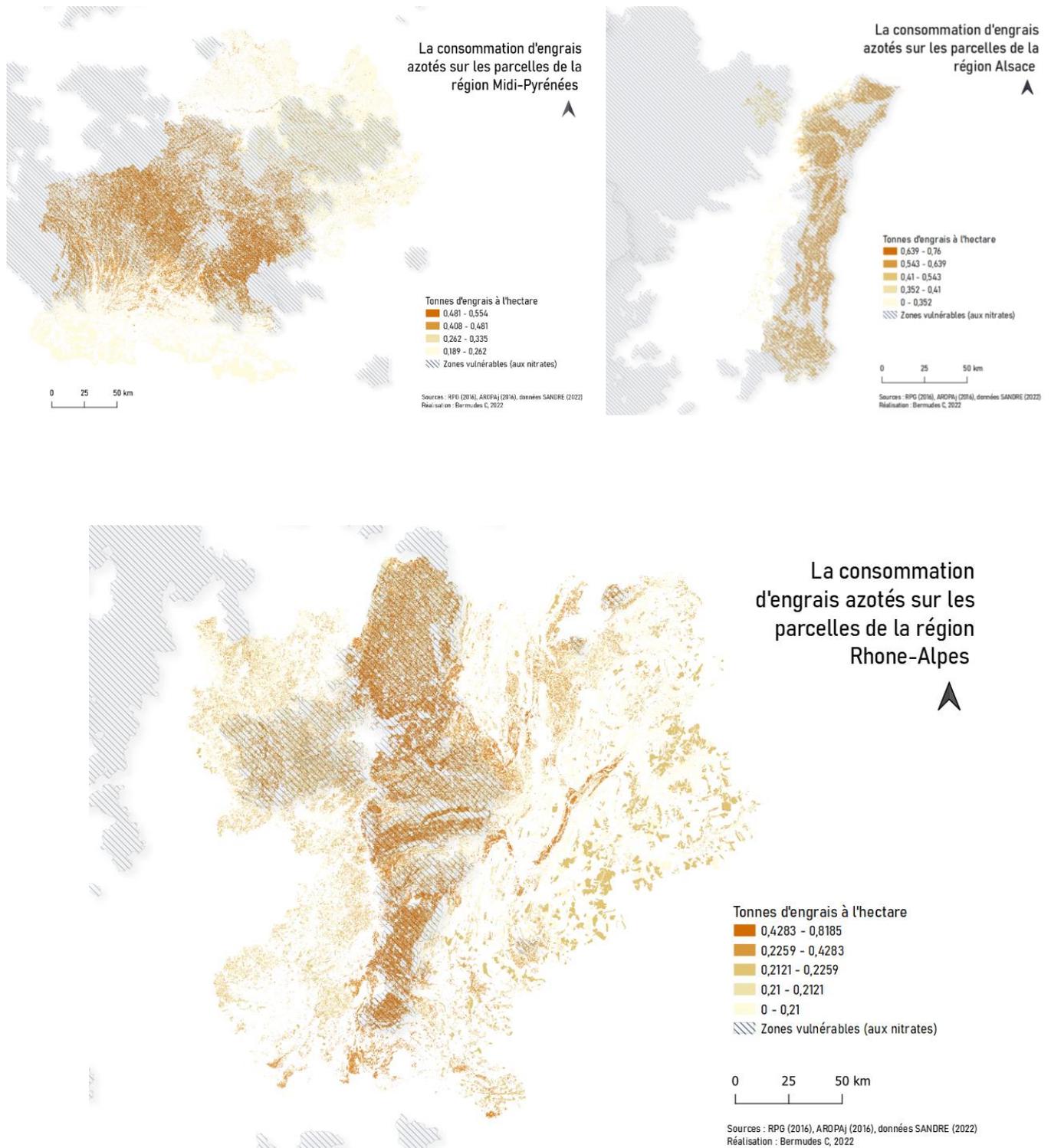


Figure 19 : La consommation d'engrais azotés à l'hectare confrontée aux zones vulnérables sur trois régions (Midi-Pyrénées, Alsace, Rhone-Alpes)

Il est remarquable de constater que les données semblent très bien coïncider : les parcelles avec le plus de tonnes d'engrais à l'hectare se superposent plus ou moins bien avec les

zones vulnérables. Nous remarquons alors que malgré l'instauration de zonages afin de limiter l'utilisation de ces produits, la pression sur les milieux persiste.

## \* LES ZONES VULNÉRABLES

Les zones vulnérables, d'après le **SANDRE** (Service d'Administration Nationale des Données et Référentiels sur l'Eau) est « *une partie du territoire où la pollution des eaux par le rejet direct ou indirect de nitrates d'origine agricole et d'autres composés azotés susceptibles de se transformer en nitrates, menace à court terme la qualité des milieux aquatiques et plus particulièrement l'alimentation en eau potable.* »

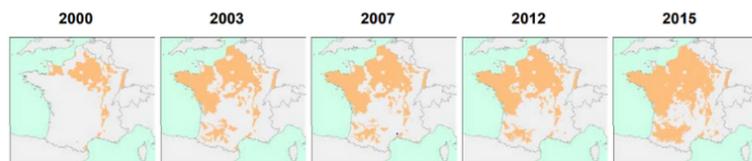
En 1992, **3 000 stations de mesure sont installées dans ou près de zones agricoles, et permettent de définir des zonages**, revus environ tous les quatre ans à l'échelle du bassin hydrographique après nouveaux résultats de qualité des eaux superficielles et souterraines. Chaque zone vulnérable correspond à un arrêté préfectoral (préfet coordonnateur de bassin) après avis du conseil départemental d'Hygiène, du conseil général et régional et du comité de bassin.

**POUR EN SAVOIR PLUS, IL EXISTE LES DOCUMENTS SUR LES "CONCENTRATIONS EN NITRATES D'ORIGINE AGRICOLE DANS LES COURS D'EAU ET LES EAUX SOUTERRAINES EN FRANCE" PUBLIÉS APRÈS CHAQUE NOUVELLE CAMPAGNE DE RÉCOLTE DE DONNÉES DE QUALITÉ DES EAUX.**

[https://www.oieau.fr/eaudoc/publications/syntheses-techniques?title=&field\\_dc\\_langue\\_value=All&field\\_date\\_publication\\_value=&field\\_tematiques\\_synth\\_oiea\\_value=agriculture](https://www.oieau.fr/eaudoc/publications/syntheses-techniques?title=&field_dc_langue_value=All&field_date_publication_value=&field_tematiques_synth_oiea_value=agriculture)

+  
[https://odr.inra.fr/intranet/cartocartowiki/index.php/Zones\\_vuln%C3%A9rables](https://odr.inra.fr/intranet/cartocartowiki/index.php/Zones_vuln%C3%A9rables)

Evolution des zones vulnérables :



Source des données : Sandre

### Conclusion de la partie 3 :

En comparant les localisations des groupes types entre la carte des sols et le RPG, nous remarquons déjà que le RPG fait ressortir des résultats plus justes par rapport au modèle AROPAj, puisque nous avons pu choisir les parcelles du RPG correspondant aux cultures traitées par le modèle.

Maintenant que les probabilités de groupes types sont connues, l'application des sorties est possible, et celles-ci peuvent être nombreuses. Pour ce rapport, nous nous sommes concentrés sur la consommation d'engrais azotés à l'hectare, en mettant en avant l'avantage principal du RPG, celui de connaître précisément des occupations du sol, et les possibilités de couplage des sorties à d'autres données.

## Conclusion

A travers ce rapport, nous avons pu voir les étapes de la spatialisation des sorties du modèle AROPAj à l'échelle de la parcelle du Registre Parcellaire Graphique. Bien que le RPG soit une base de données déjà très riche en informations, ces dernières ne suffisent pas pour établir les probabilités de localisation des groupes types AROPAj. Il a donc fallu lui adjoindre des données d'altitude et d'irrigation, pouvant varier en précision, et dont les attributions à l'échelle parcellaire peuvent être variables voire questionnables (notamment pour ce qui est de l'irrigation). En outre, afin d'obtenir une unité spatiale plus proche du modèle et de simplifier les calculs de probabilités, les parcelles sont regroupées en "groupements parcellaires" selon les caractéristiques d'altitude, d'irrigation et de groupe de culture AROPAj.

Néanmoins, nous pouvons tout d'abord faire remarquer que le RPG représente les parcelles liées à la PAC, puisque dessinées afin de recevoir les aides. Les scénarios développés par le modèle AROPAj sont principalement alimentés par les réformes de la PAC : les parcelles du RPG sont donc un réceptacle idéal pour ces scénarios.

Avec le RPG, nous pouvons également choisir de conserver la précision géographique de départ de la base de données, ou alors d'utiliser les groupements parcellaires pour réduire le nombre de parcelles (et ainsi faciliter la visualisation, la manipulation des polygones). Le modèle de simplification parcellaire permettrait également de regrouper les parcelles selon d'autres paramètres comme la commune ou encore l'OTEX si l'information est disponible : en clair, partir d'une telle précision permet d'observer des résultats sur plusieurs types d'échelles si nous le souhaitons (agrégation à partir d'une échelle très fine).

L'atout principal du RPG est de connaître pour chaque parcelle l'exacte occupation du sol (précision thématique). Nous sommes donc en mesure de trier celles-ci selon leur pertinence face au modèle : par exemple les surfaces concernées par la viticulture ou l'arboriculture, qui ne sont pas traitées dans le modèle, peuvent être aisément retirées. Mais cet atout est aussi utile lors de la visualisation des résultats, des sorties du modèle : un grand nombre d'analyses sont réalisables avec ces informations, que l'on se contente uniquement des sorties du modèle, qui sont déjà très variées, ou qu'on les confronte avec de nombreux autres types de données.

## Perspectives

Les objectifs du stage étaient doubles :

1. Spatialiser le modèle AROPAj à l'échelle de la France en se basant sur les données du RPG afin d'affiner la cartographie des résultats pour des études environnementales portant sur la pollution diffuse des eaux par les nitrates ou des molécules phytopharmaceutiques ;
2. Etablir, au niveau du RPG, une correspondance entre les données issues de la Banque Nationale des Ventes des produits phytopharmaceutiques (BNVd), spatialisées à l'échelle de la parcelle RPG et les sorties du modèle AROPAj.

A travers ce rapport, nous avons pu, en un sens, finaliser cette première partie, bien que d'autres résultats pourraient être possibles, d'autres données pourraient être mises en lien, d'autres représentations pourraient être faites. Qu'il s'agisse de R ou de Python, les scripts utilisés pourraient être davantage commentés mais aussi être améliorés, tout comme les explications faites dans ce rapport de leur fonctionnement. J'aimerais également pouvoir faire plus de recherches sur les utilisations du modèle AROPAj.

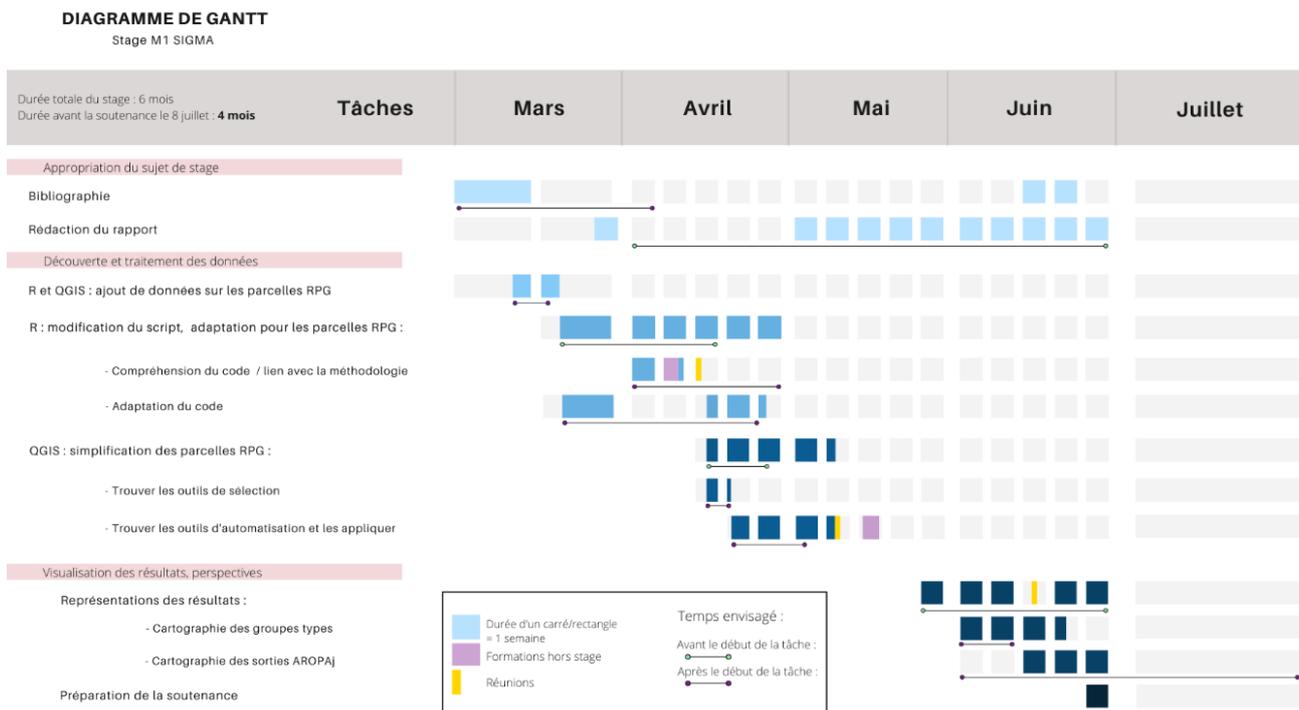


Figure 20 : Diagramme de Gantt : durées estimées et réelles des tâches réalisées

Pour la troisième partie (celle sur les résultats), j'aurais pu trouver d'autres manières de mettre en avant l'atout principal du RPG, celui de connaître les occupations du sol : j'aurais pu me concentrer sur une région et observer plus en détails les liens entre sorties AROPAj et types de culture. Ce qui aurait pu venir compléter cette troisième partie va être réalisé à la suite du rendu de ce rapport ; comme par exemple faire des représentations avec d'autres types de données comme des données ponctuelles sur la concentration en nitrates d'origine

agricole dans les eaux souterraines, des données de la BD Topage (référentiel hydrographique français remplaçant la BD Carthage, accessible depuis le catalogue SANDRE) et/ou intégrant la pente (calculée à partir d'un MNT).

La suite de ce stage sera consacrée à des analyses plus environnementales, à l'aide d'un lien que nous souhaitons faire entre les parcelles du RPG et la BNVD. Instaurée en 2008 par la Loi sur l'Eau et les Milieux Aquatiques, la Banque Nationale des Ventes de produits phytopharmaceutiques, dits « phytosanitaires », par les Distributeurs agréés (BNVD) regroupe des informations sur les ventes de produits localisées, depuis 2008, au niveau de la commune des distributeurs, et depuis 2013, du code postal du siège social des acquéreurs.

Ces données BNVD vont être liées au RPG par le type de culture, ou alors par la commune. Avec d'éventuelles sorties complémentaires d'AROPAJ, ainsi que ces données BNVD spatialisées, l'objectif serait éventuellement de mettre en place un indice de toxicité par type de culture ou bien par groupement parcellaire, pour rester dans la même unité spatiale.

## Bibliographie

Cantelaube P. et Carles M. (2014), « Le registre parcellaire graphique : des données géographiques pour décrire la couverture du sol agricole ». Le cahier des techniques de l'INRA : GPS & SIG Pour la conduite de dispositifs expérimentaux, Vers l'émergence de la communauté métier « GéoExpé » : INRA, 2014.

Cantelaube P., Ducharne A., Habets F., Jayet P.A, Mignolet C, Philippe E., Schott C., Viennot P., Zakharov P. (2008). « Modélisation économique des relations entre agriculture et environnement à l'échelle du bassin de la Seine ». Colloque 2008 du PIREN-Seine Paris France, février 2008.

Cantelaube P., Jayet P.A., Carre F., Bamps C., Zakharov P., (2012). « Geographical downscaling of outputs provided by an economic farm model calibrated at the regional level ». Land Use Policy, Elsevier, (2012), 29 (1), pp.35-44.

EUROPEAN COMMISSION (DIRECTORATE-GENERAL FOR AGRICULTURE AND RURAL DEVELOPMENT), Eu Farm Economics Overview FADN 2018. Brussels, 2021.

Chakir R., (2009). « Spatial Downscaling of Agricultural Land-Use Data : An Econometric Approach Using Cross Entropy ». Land Economics, 85.2 (2009), 238-251.

IGN (2021), RPG Version 2.0 – Descriptif de contenu et de livraison, Décembre 2021.

Jayet P.A., Cantelaube P., Zakharov P., Génin M., Bourgeois C., Polard A., Schott C., Mignolet C., Habets F., Viennot P., Philippe E., Souhar O., (2009). « Modélisation économique des relations entre l'agriculture et l'environnement à l'échelle du bassin de la Seine ». Programme PIREN-Seine, INRA, UMR Economie Publique, Thiverval-Grignon, février 2009.

Jayet P.A., Petsakos A., Chakir R., Lungarska A., De Cara S., Petel E., Humblot P., Godard C., Leclère D., Cantelaube P., Bourgeois C., Bamière L., Ben Fradj N., Aghajanzadeh-Darzi P., Dumollard G., Ancuta I., Adrian J., (2015). « The European agro-economic AROPAj model ». INRA, UMR Economie Publique, Thiverval-Grignon. [https://www6.versailles-grignon.inra.fr/economie\\_publique\\_eng/Research-work](https://www6.versailles-grignon.inra.fr/economie_publique_eng/Research-work)

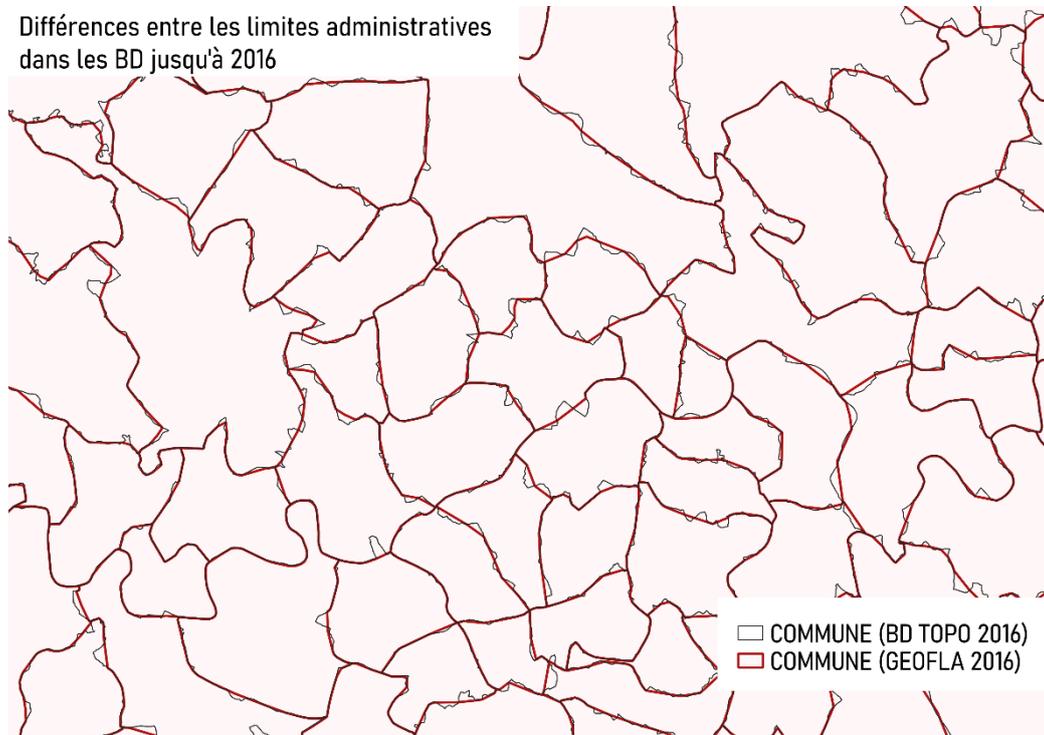
Lardot B., Cantelaube P., Carles M., Séard C. et Truche C., (2021). « Construction d'une base de données géographiques exhaustive à échelle fine sur l'occupation agricole du sol : Le « RPG complété ». Partie 1 : Production de la couche géographique des parcelles susceptibles d'accueillir les surfaces agricoles hors RPG ». Note Méthodologique US ODR, INRAE, février 2021, version 2 revue octobre 2021.

Lungarska A., Jayet P.A, (2016). « Impact of Spatial Differentiation of Nitrogen Taxes on French Farms' Compliance Costs ». Environmental and Resource Economics : The Official Journal of the European Association of Environmental and Resource Economists, ISSN 0924-6460 Environ Resource Econ, DOI 10.1007/s10640-016-0064-9.

Neuenfeldt, S. et Gocht, A., (2014). « A Handbook on the use of FADN Database in Programming Models ». Technical report, Thünen Working Paper.

Ramalanjaona L., Poméon T., Ballot R., Barbu C., Bougon N., Fernandez F., Martin P., Mohamed H., (2020). « Mise à jour du calcul des coefficients de répartition spatiale des données de la BNVD ». Note méthodologique ODR, 2020.

## Annexes



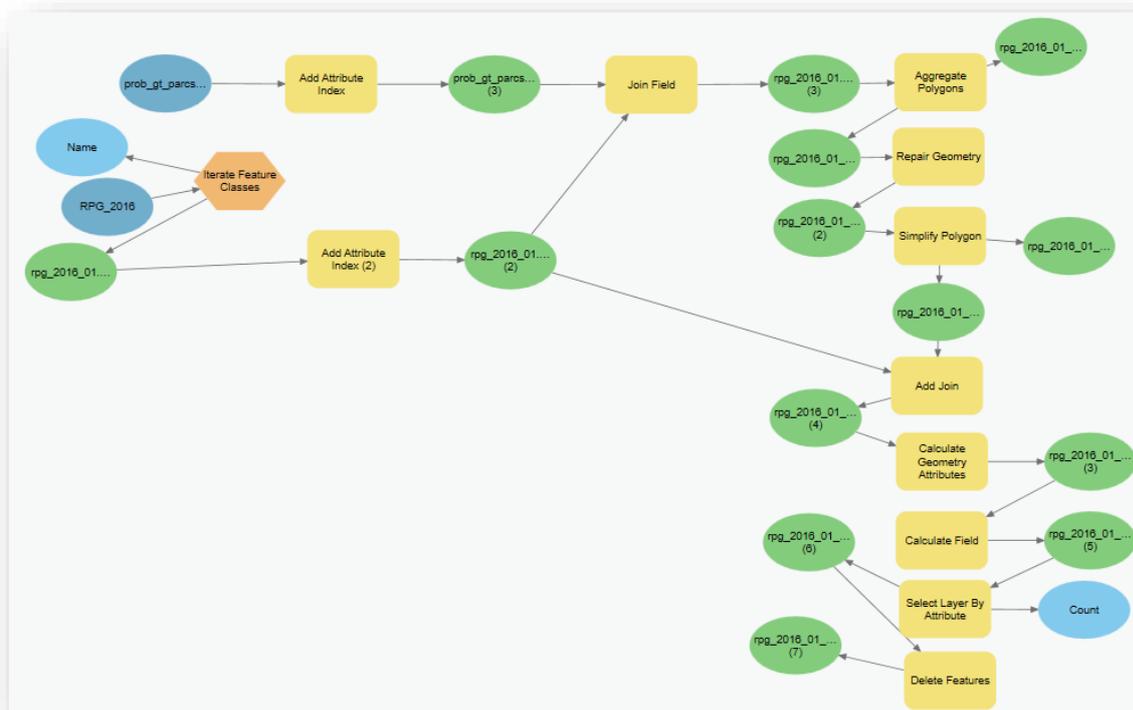
*Annexe 1 : Jusqu'à 2016, deux bases de données procuraient des limites administratives : GEOFLA et la BD TOPO*

$$T_{ik} = \sum_j P_{ij} \left( \alpha \frac{S_{jk}}{\sum_{k'} S_{jk'}} + \beta \frac{S_{jk}^{irr}}{\sum_{k'} S_{jk'}^{irr}} \mathbb{1}_{\mathbb{I}}(i) \right) + \sum_a \gamma_a U_{ak} \mathbb{1}_{\mathbb{A}(a)}(i)$$

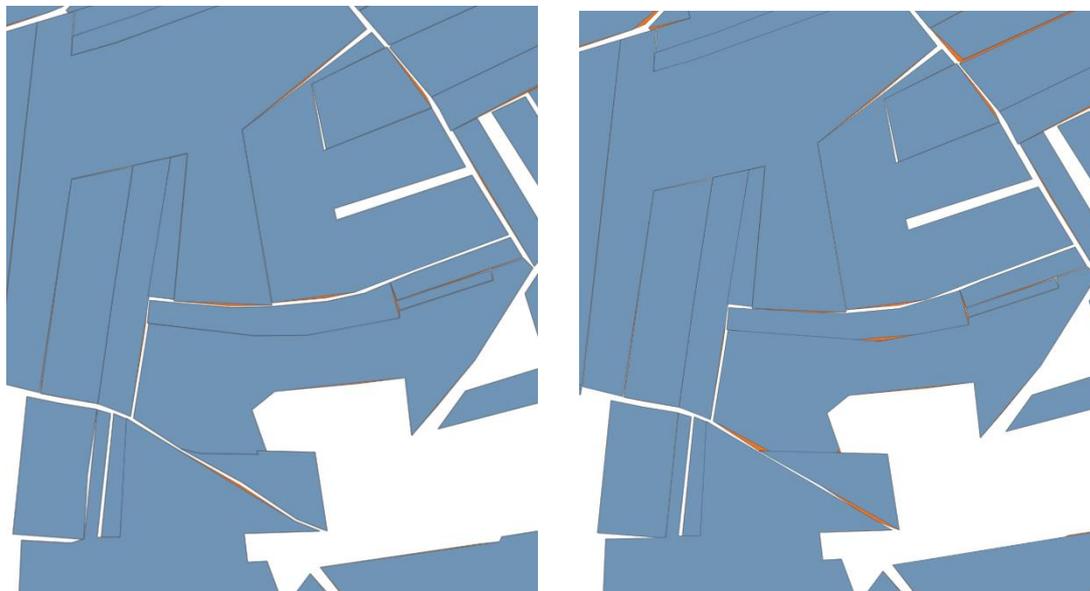
$$T_{ik} = \alpha \sum_j P_{ij} \frac{S_{jk}}{\sum_{k'} S_{jk'}} + \gamma \sum_a \frac{U_{ak}}{\sum_{k'} U_{ak'}} \mathbb{1}_{\mathbb{A}(a)}(i)$$

*Annexe 2 : Les équations déterminant les probabilités*

A gauche, le cas de figure de l'homogénéité altitudinale avec alpha et beta comme variables de distorsion. A droite, le cas de figure des altitudes multiples, avec la variable gamma changeant selon la classe s'il n'y pas d'irrigation.



Annexe 3 : ModelBuilder construit sur ArcGIS pour la simplification parcellaire



Annexe 4 : Tests pour la simplification géométrique des parcelles

A gauche : 5 mètres, à droite : 2 mètres



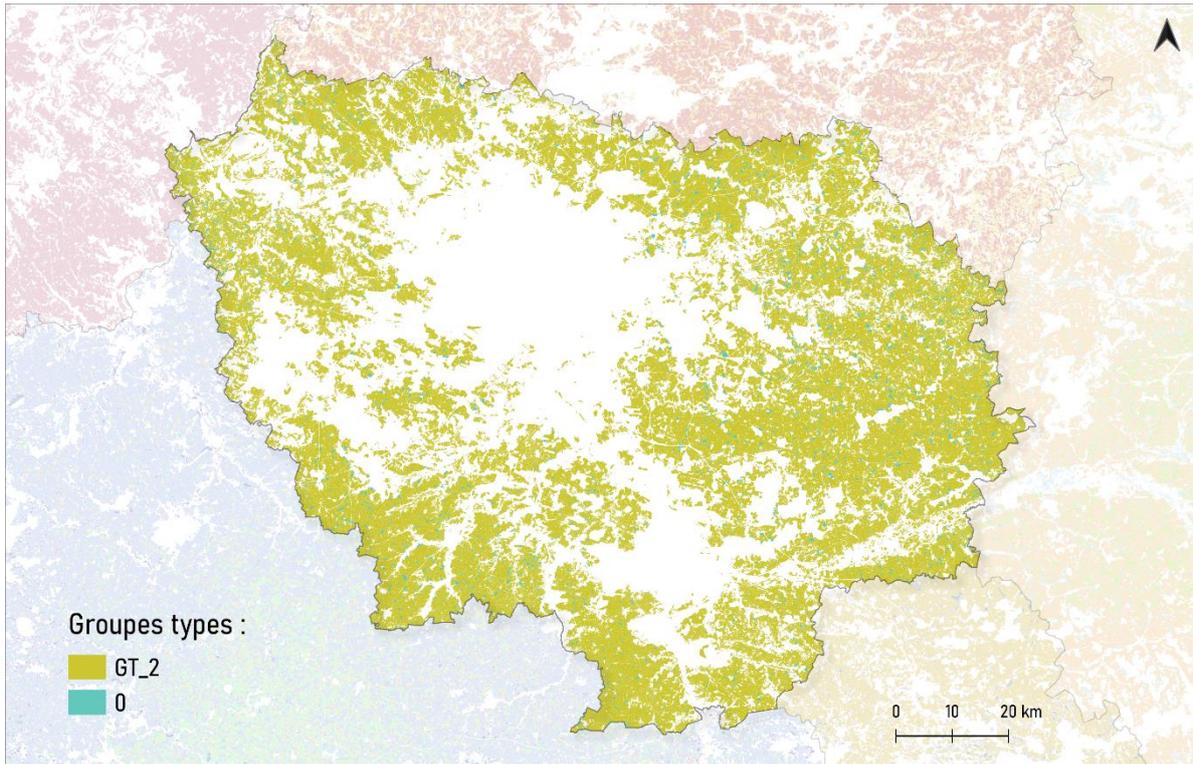
### Annexe 5 : L'historique des traitements

```

1 from qgis.core import *
2 import processing
3 import glob
4 import pandas as pd
5
6 reg = pd.read_csv(///tls-dyn-nas/odr/cbermudes/Documents/arorpg/data/dep_regs.csv")
7 regions = set(reg['CODE_REG'])
8 code_regs = [str(x) for x in regions]
9
10 pathw = '///tls-dyn-nas/odr/cbermudes/Documents/OUTPUT_AGR/agr_'
11
12 files = glob.glob(///tls-dyn-nas/odr/cbermudes/Documents/arorpg/output/*.dbf)
13 shapes = glob.glob(///tls-dyn-nas/odr/cbermudes/Documents/arorpg/data/RPG_2016/*.shp)
14
15 for code in code_regs :
16     for file in files :
17         if code in file :
18             code_i = int(code)
19             depts = list(reg.loc[reg['CODE_REG'] == code_i, 'CODE_DEPT'])
20
21             for shape in shapes:
22                 for dept in depts:
23                     if dept in shape:
24                         processing.run("model:simplification_parcelles", {
25                             'fichiercsv': file,
26                             'parcelles': shape,
27                             'simplificationgeom': 2,
28                             'slection': '(\"ind_miller\">\"0.12\" and \"surf_parce\">\"0.1\") or (\"ind_miller\"<\"0.12\" and \"surf_parce\" >\"0.2\") ',
29                             'native:savesselectedfeatures_1:parcelles_simp': pathw+element+".shp"})

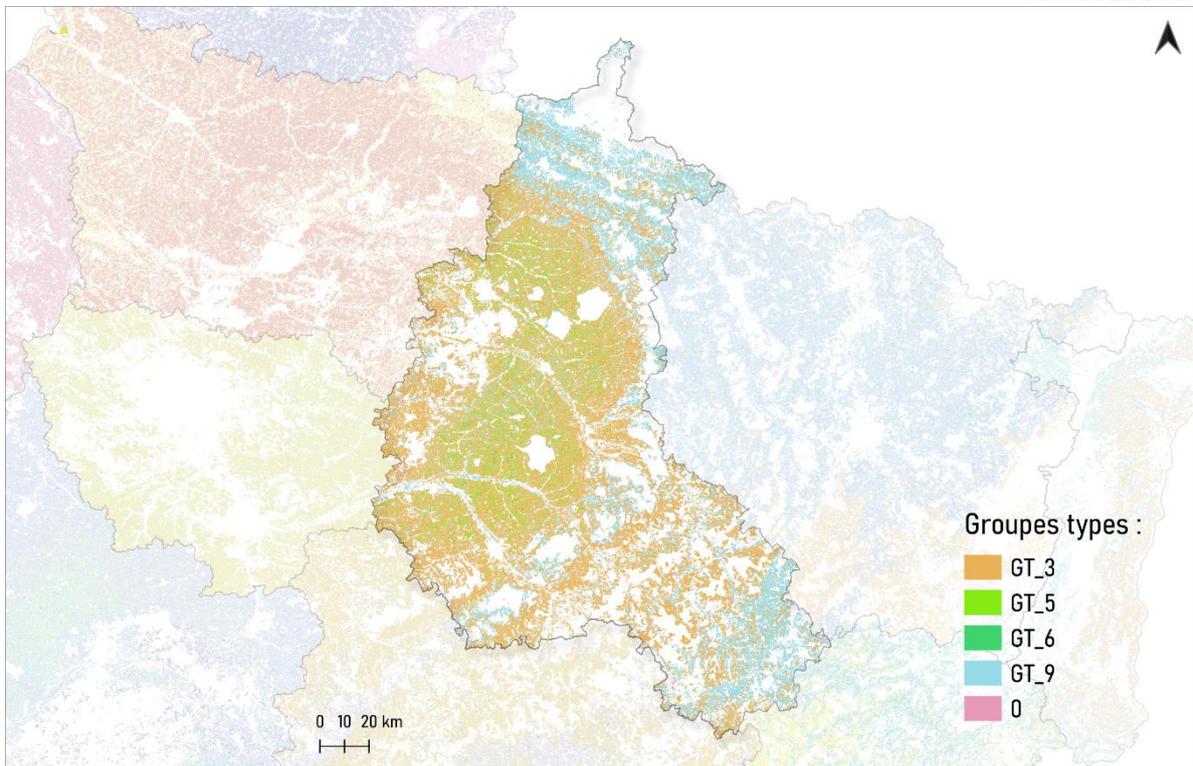
```

### Annexe 6 : Boucle sur les régions



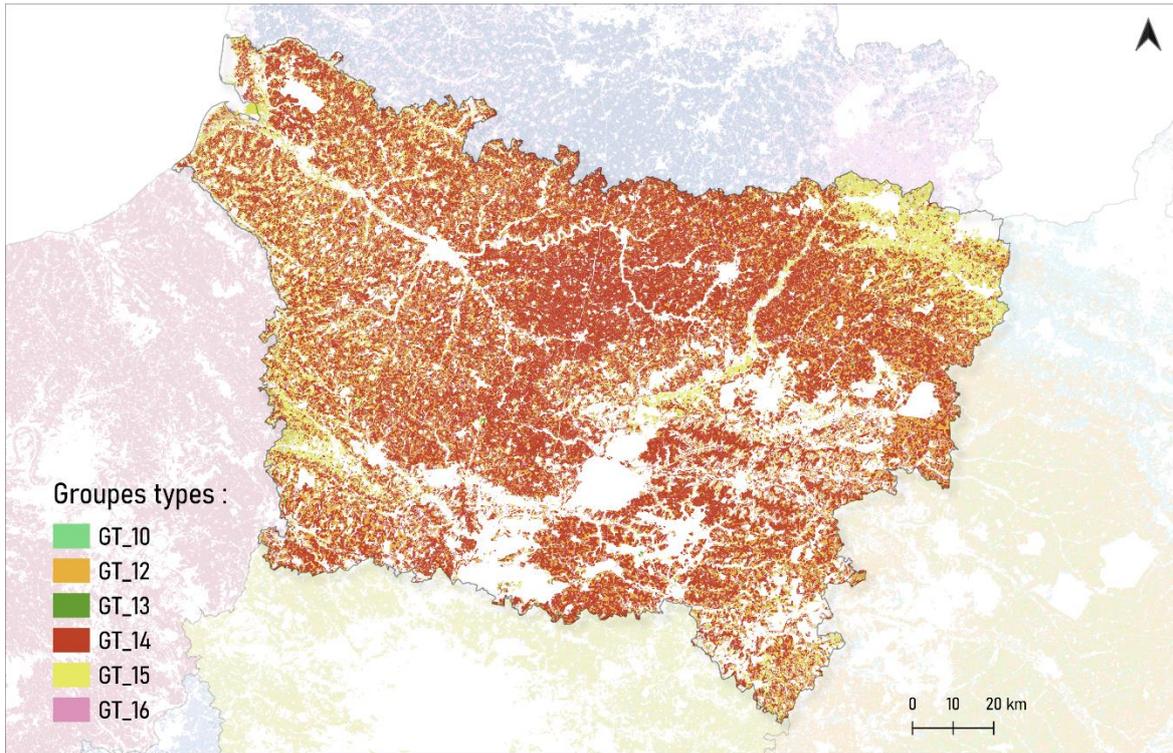
Les groupes types de la région Ile-de-France (11) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



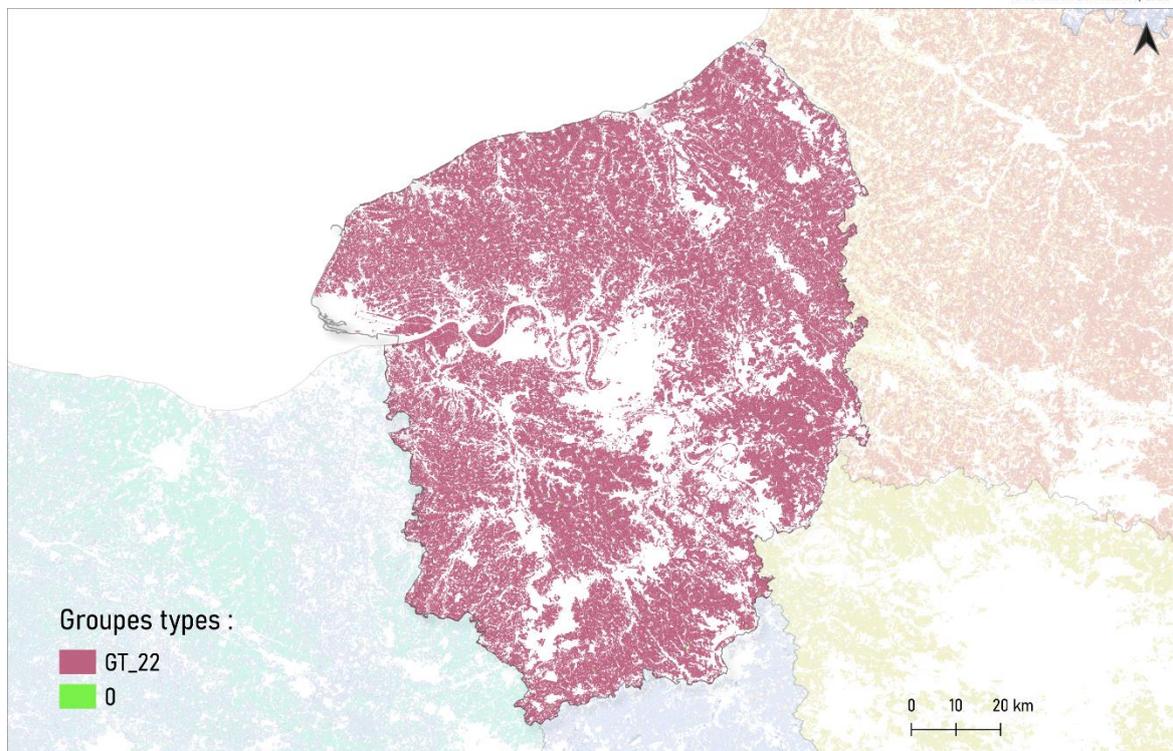
Les groupes types de la région Champagne-Ardenne (21) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



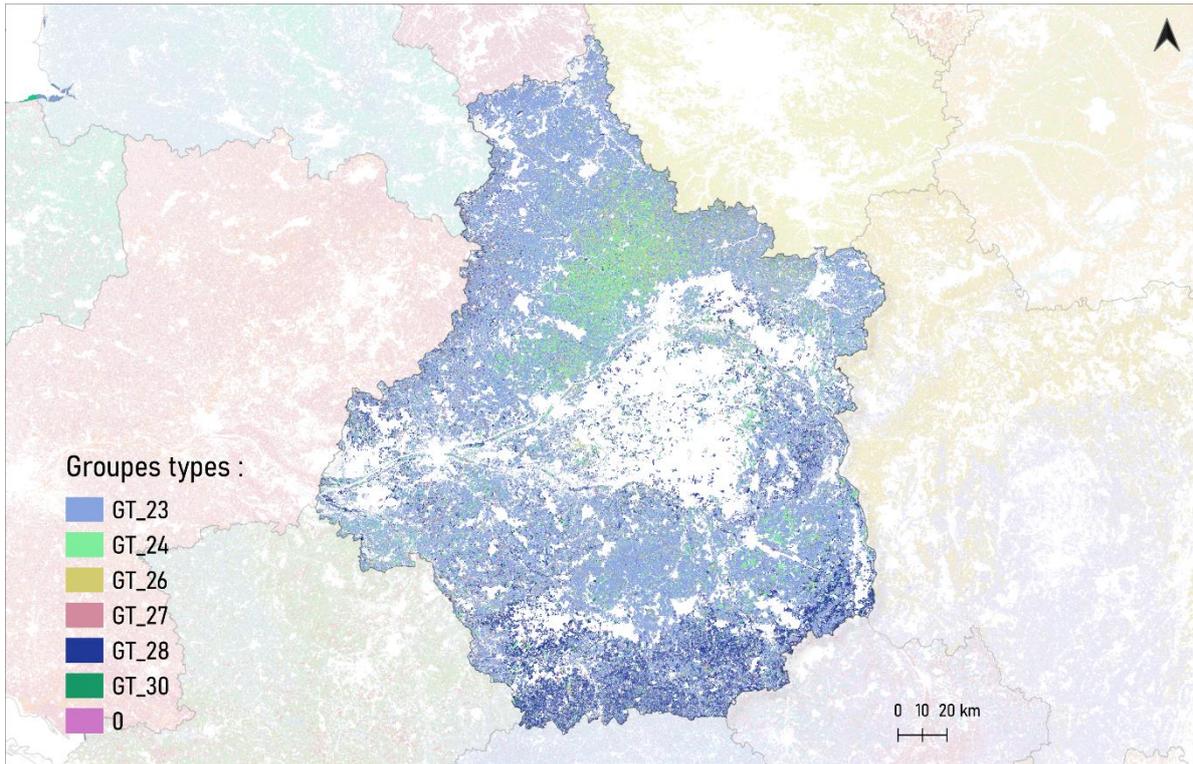
Les groupes types de la région Picardie (22)  
en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



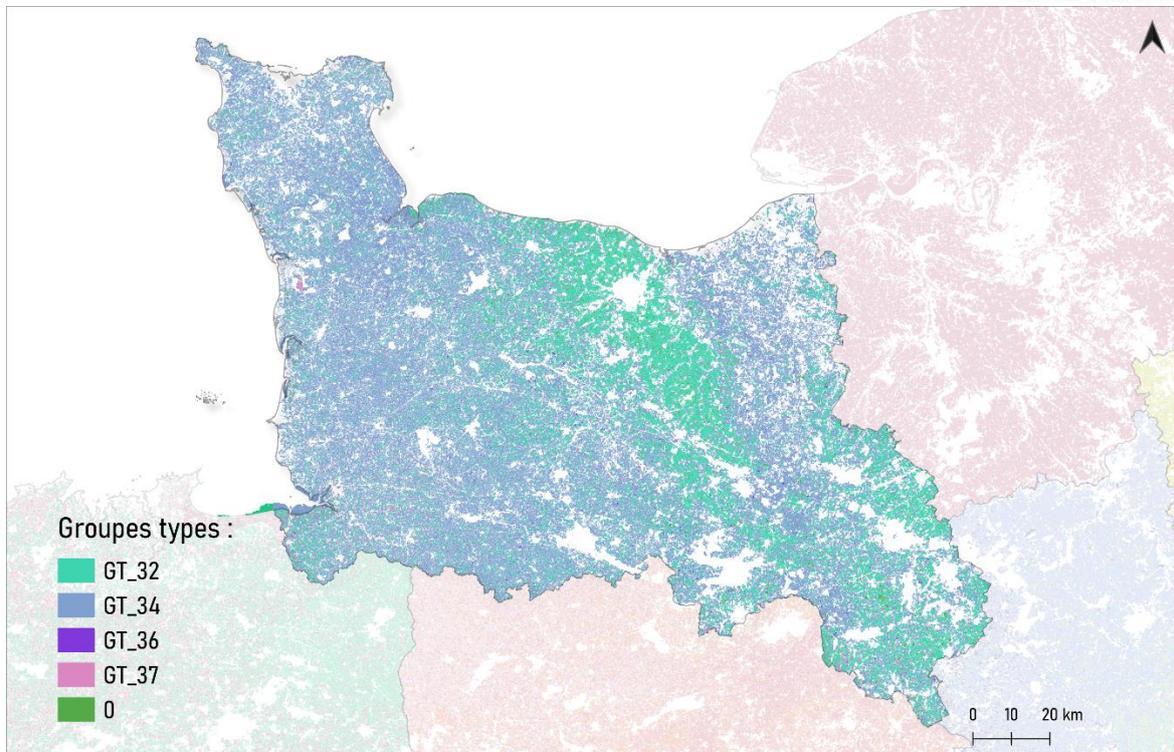
Les groupes types de la région Haute-  
Normandie (23) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



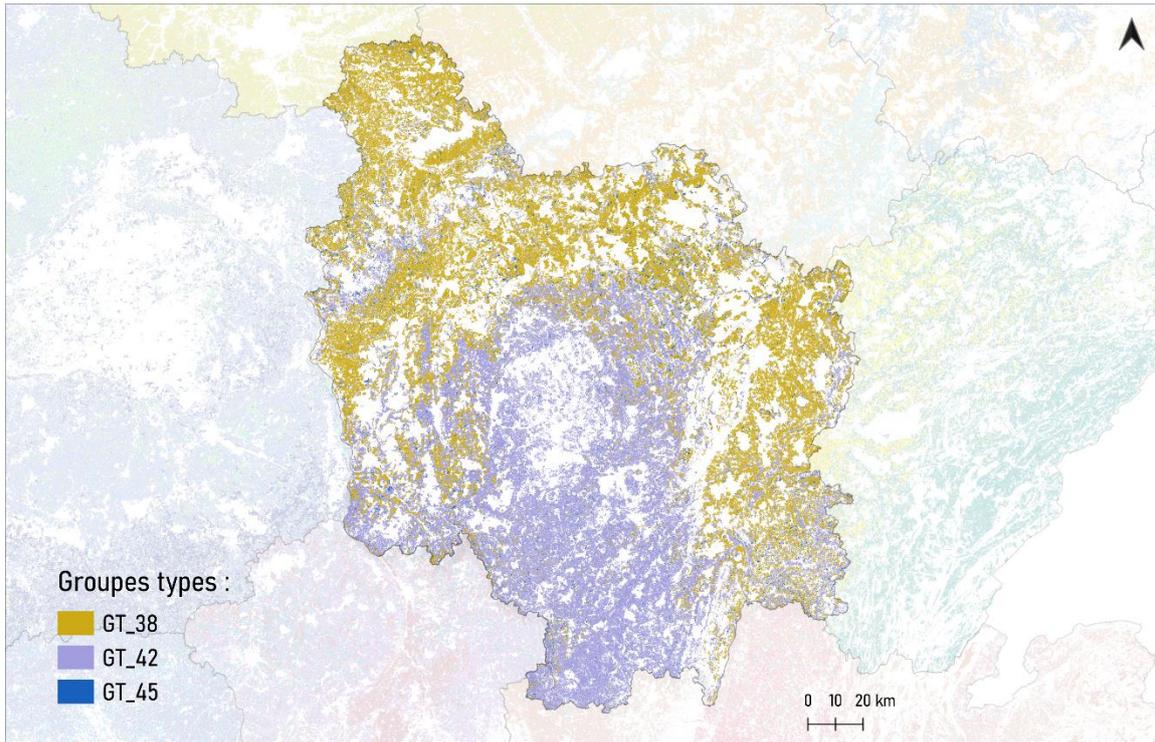
Les groupes types de la région Centre (24)  
en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



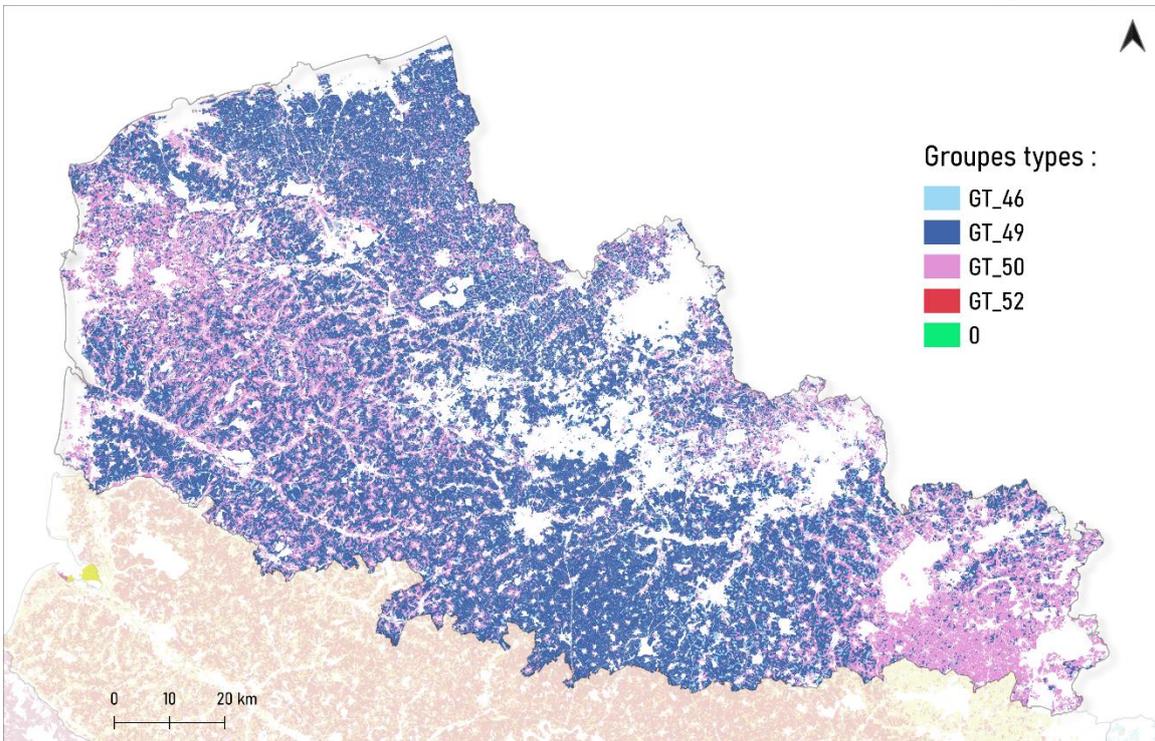
Les groupes types de la région Basse-  
Normandie (25) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



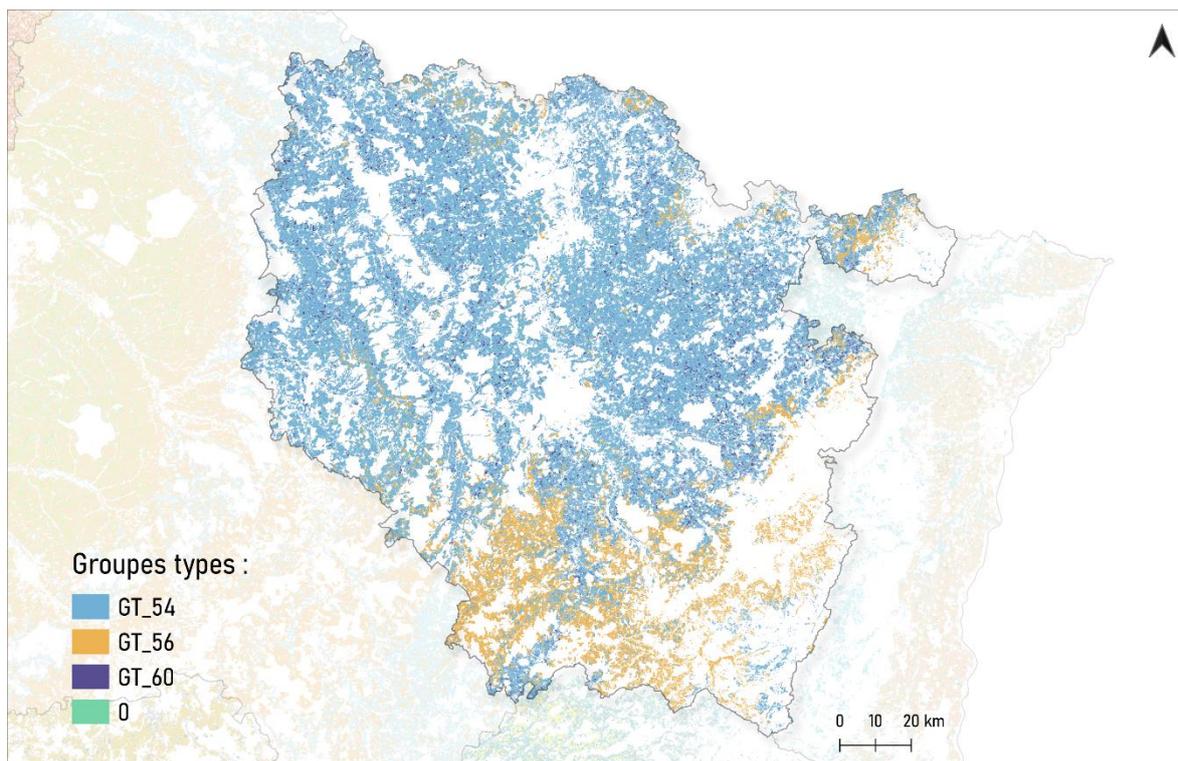
Les groupes types de la région Bourgogne  
(26) en 2016

Sources : RPG (2016), AROPA<sub>j</sub> (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



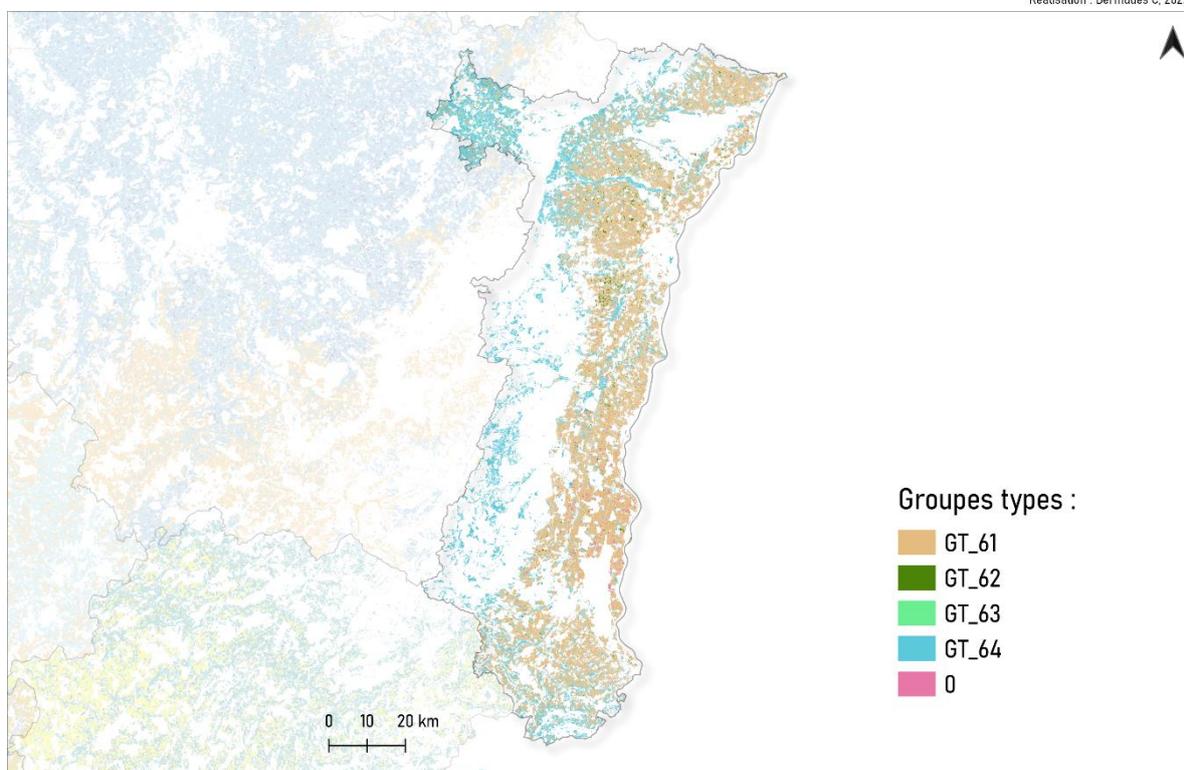
Les groupes types de la région Nord-Pas-  
de-Calais (31) en 2016

Sources : RPG (2016), AROPA<sub>j</sub> (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



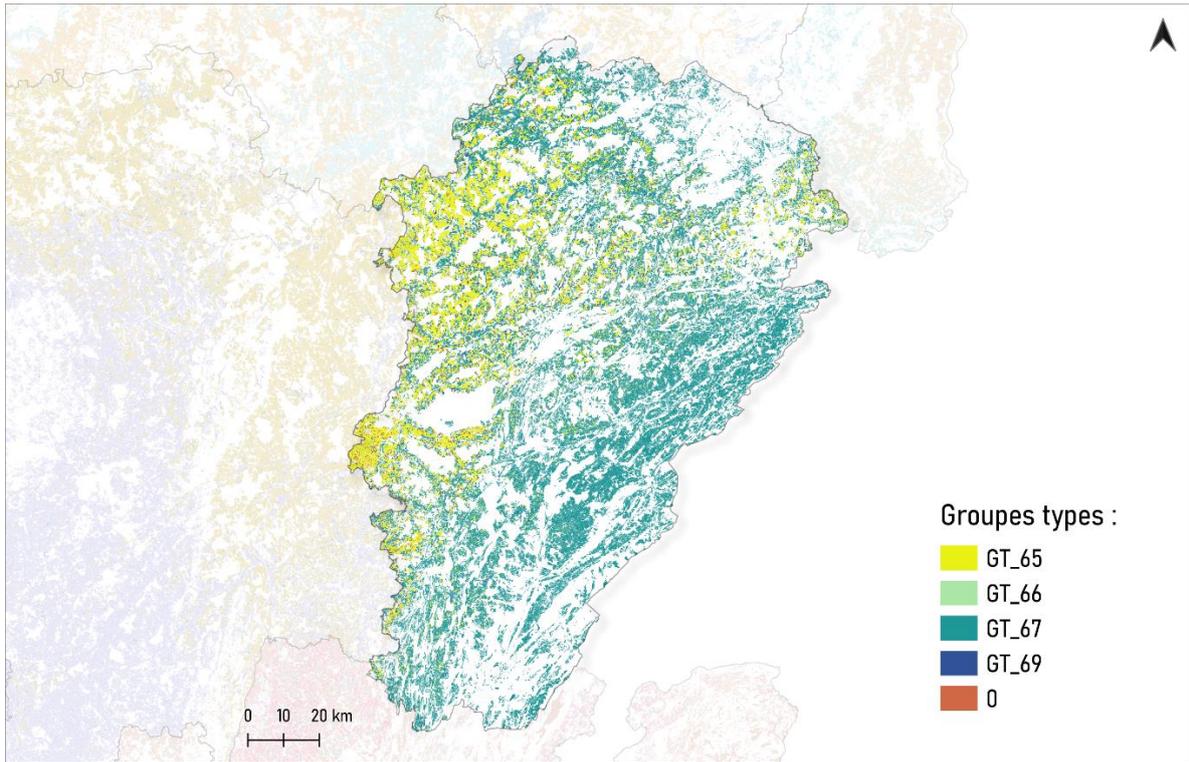
Les groupes types de la région Lorraine (41)  
en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



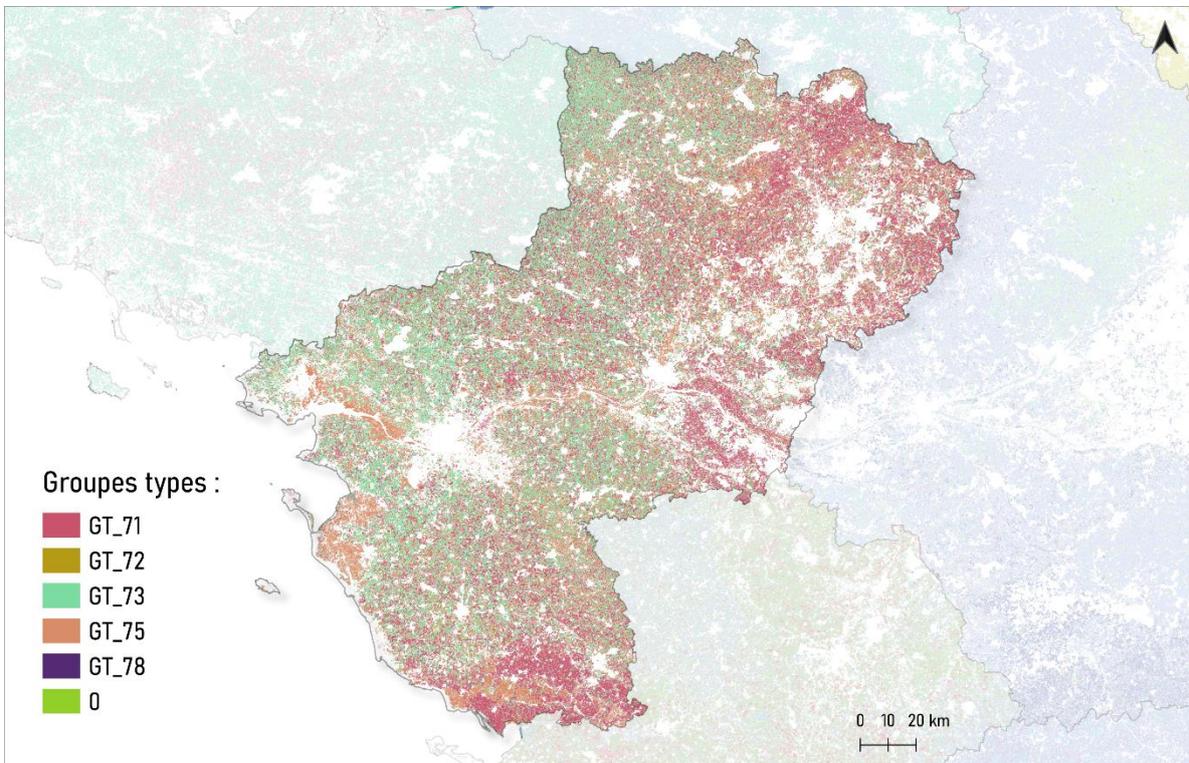
Les groupes types de la région Alsace (42)  
en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



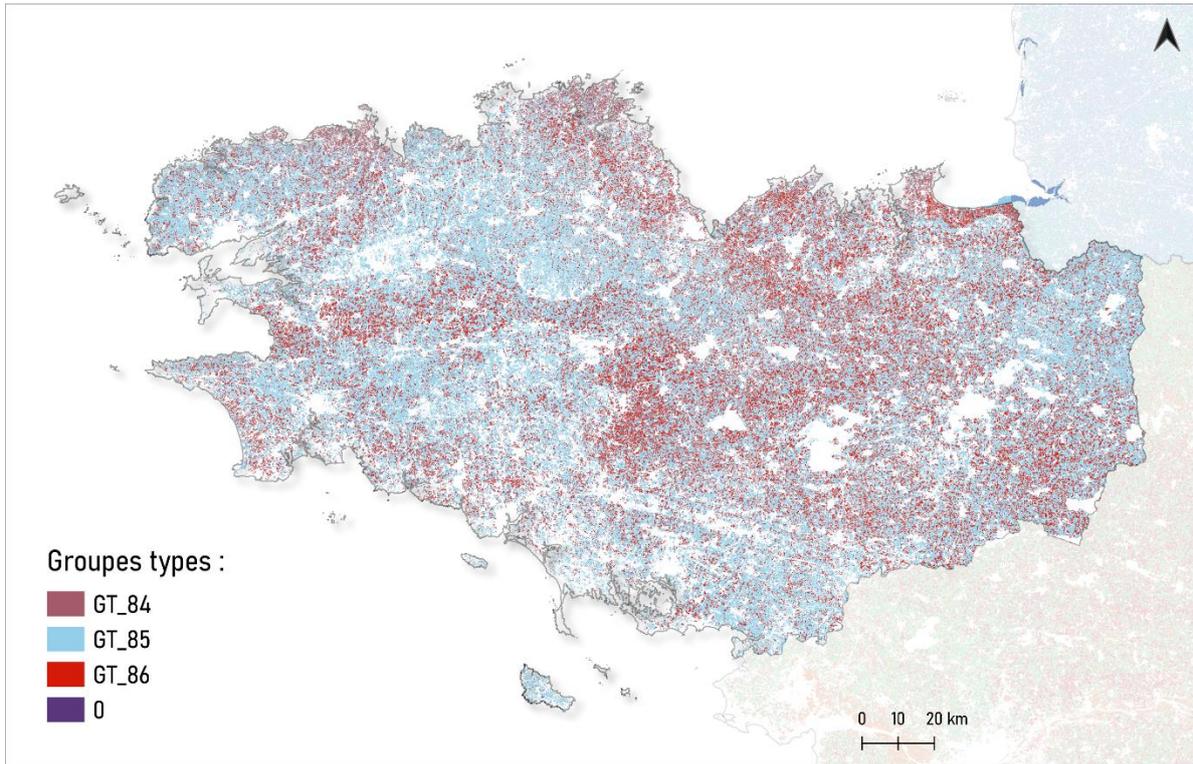
Les groupes types de la région Franche-Comté (43) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



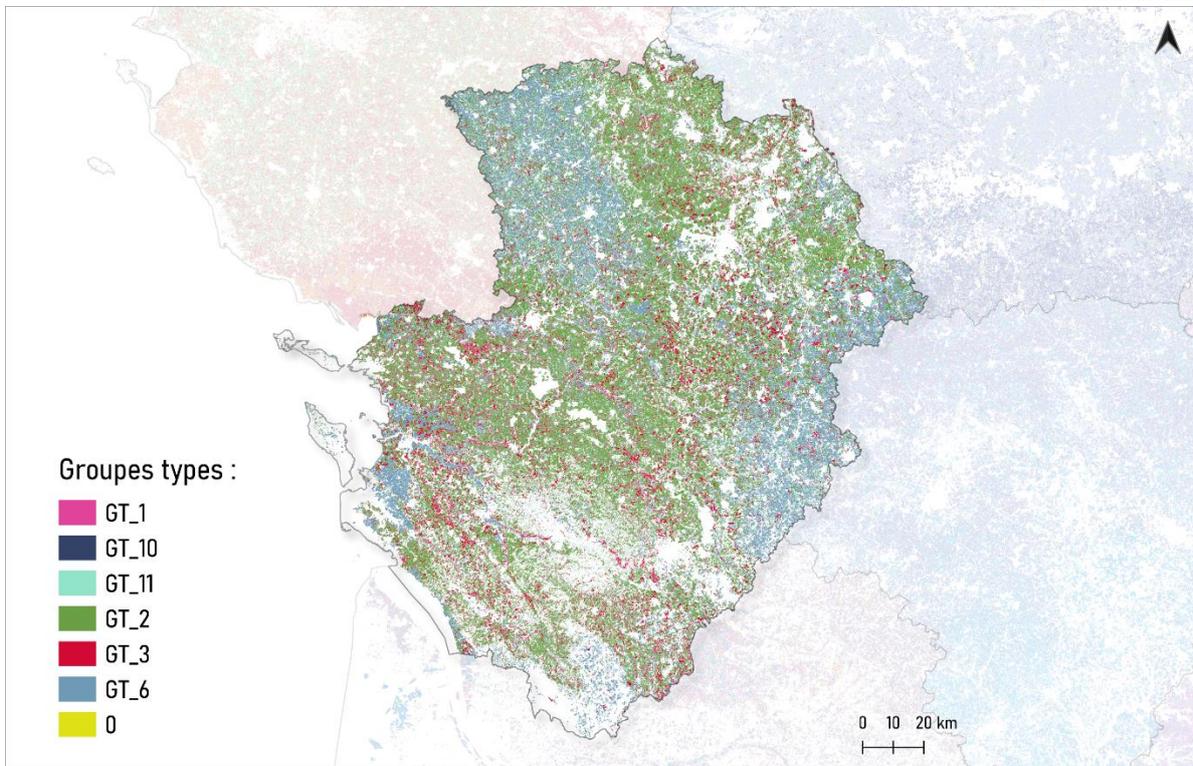
Les groupes types de la région Pays de la Loire (52) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



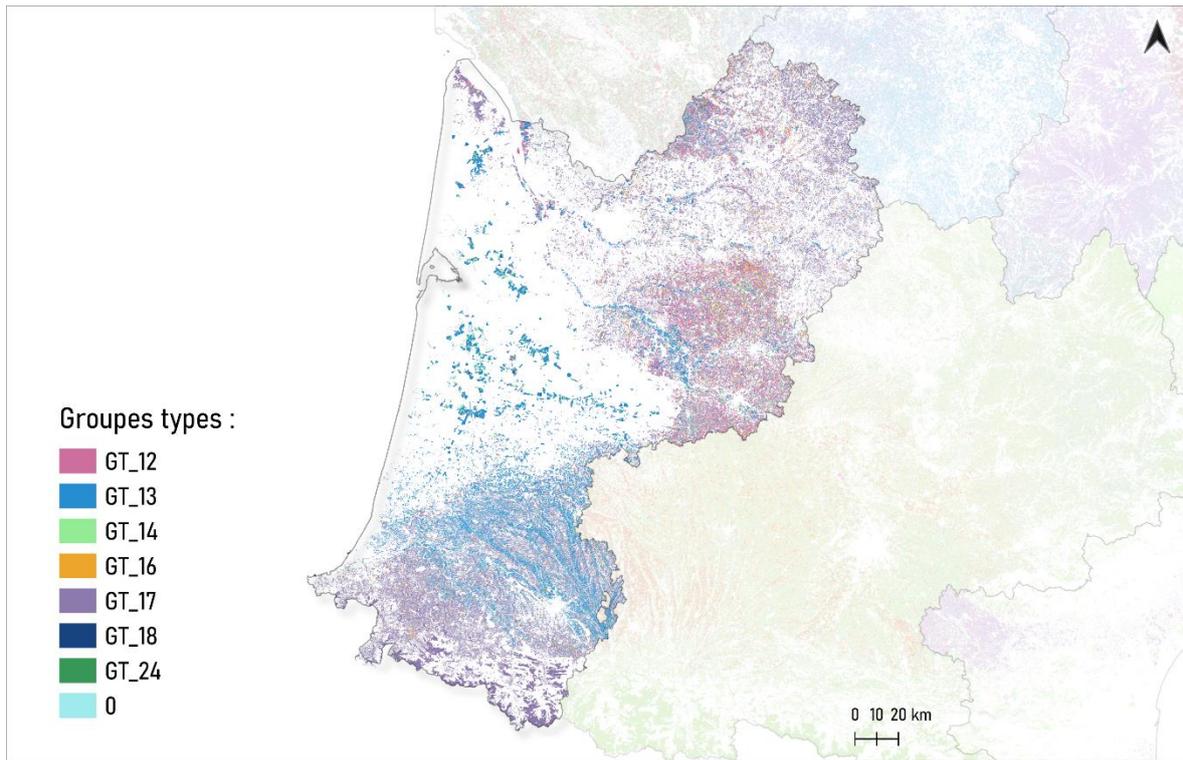
Les groupes types de la région Bretagne  
(53) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



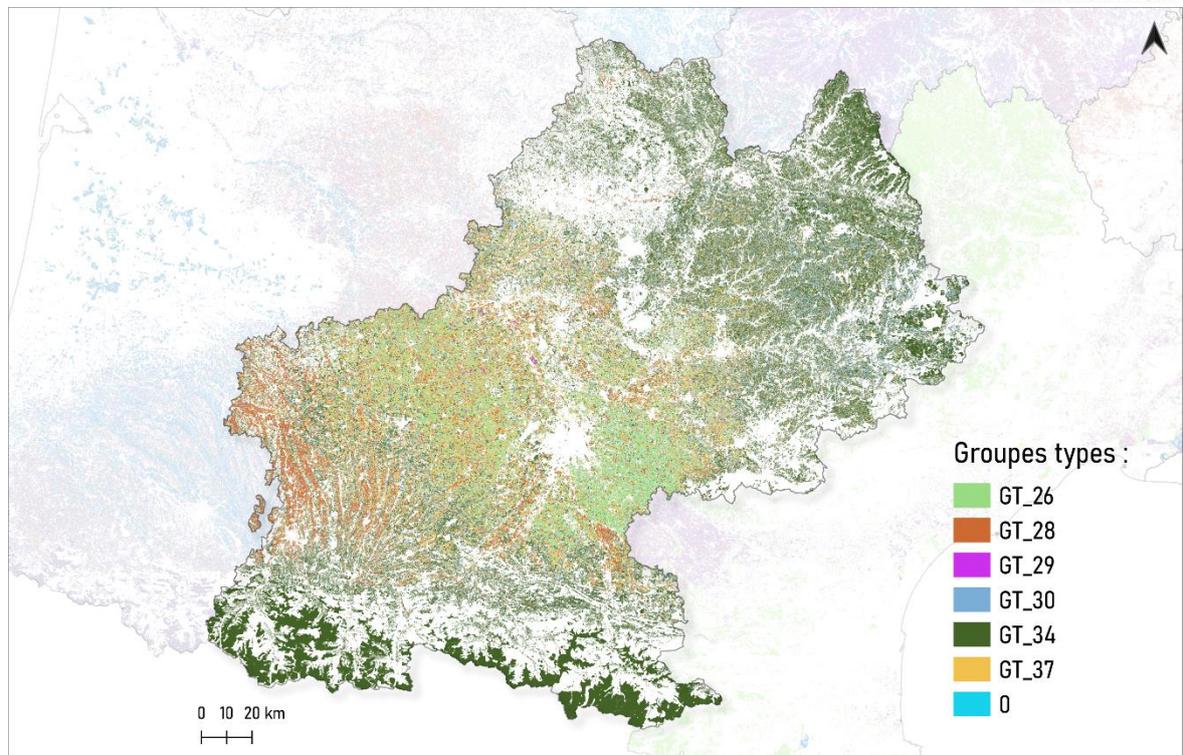
Les groupes types de la région Poitou-  
Charentes (54) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



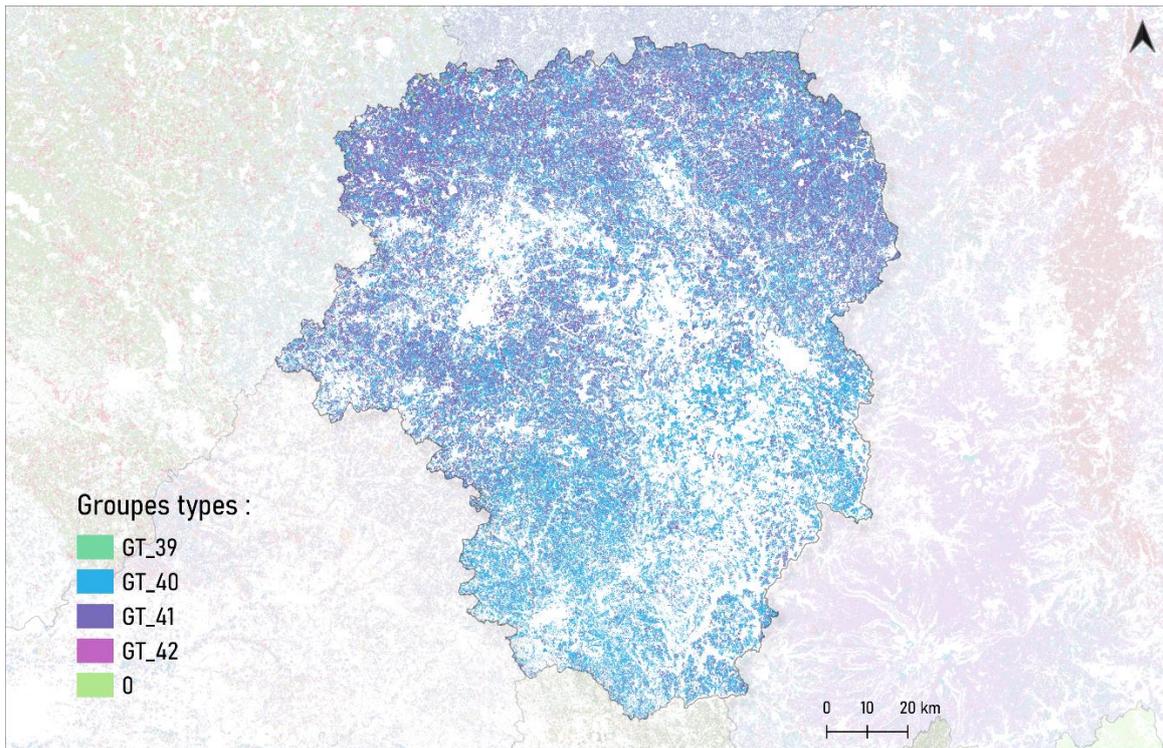
Les groupes types de la région Aquitaine  
(72) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



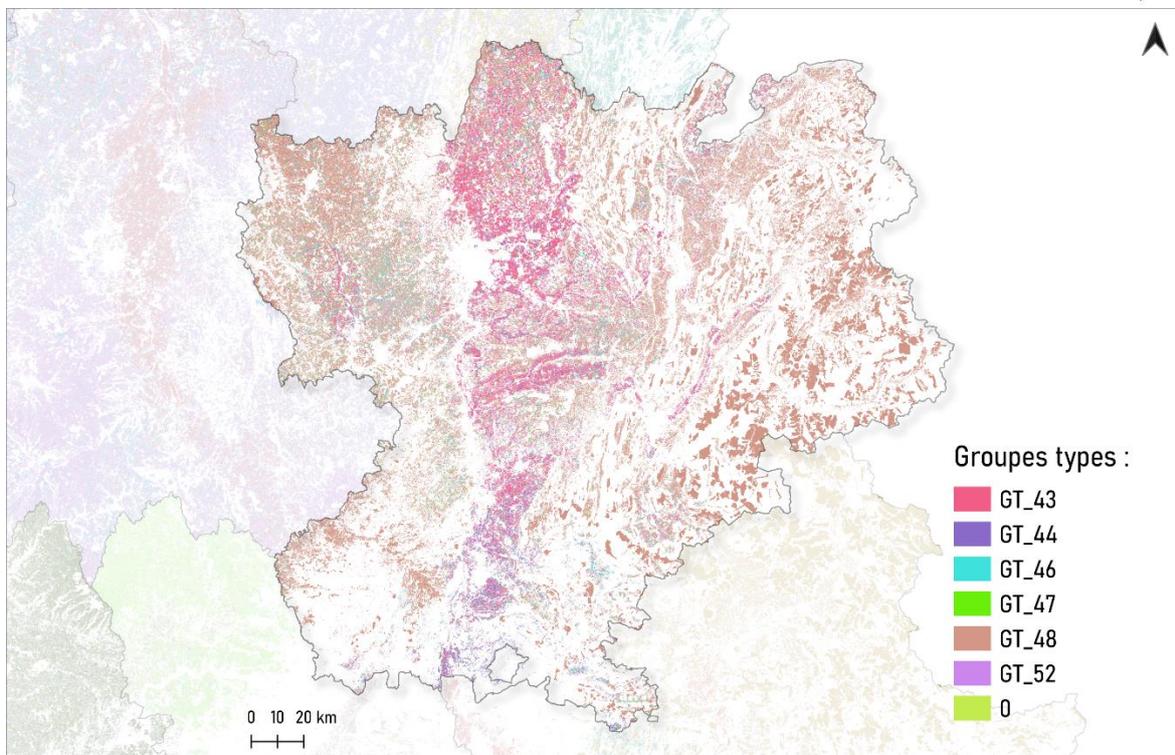
Les groupes types de la région Midi-  
Pyrénées (73) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



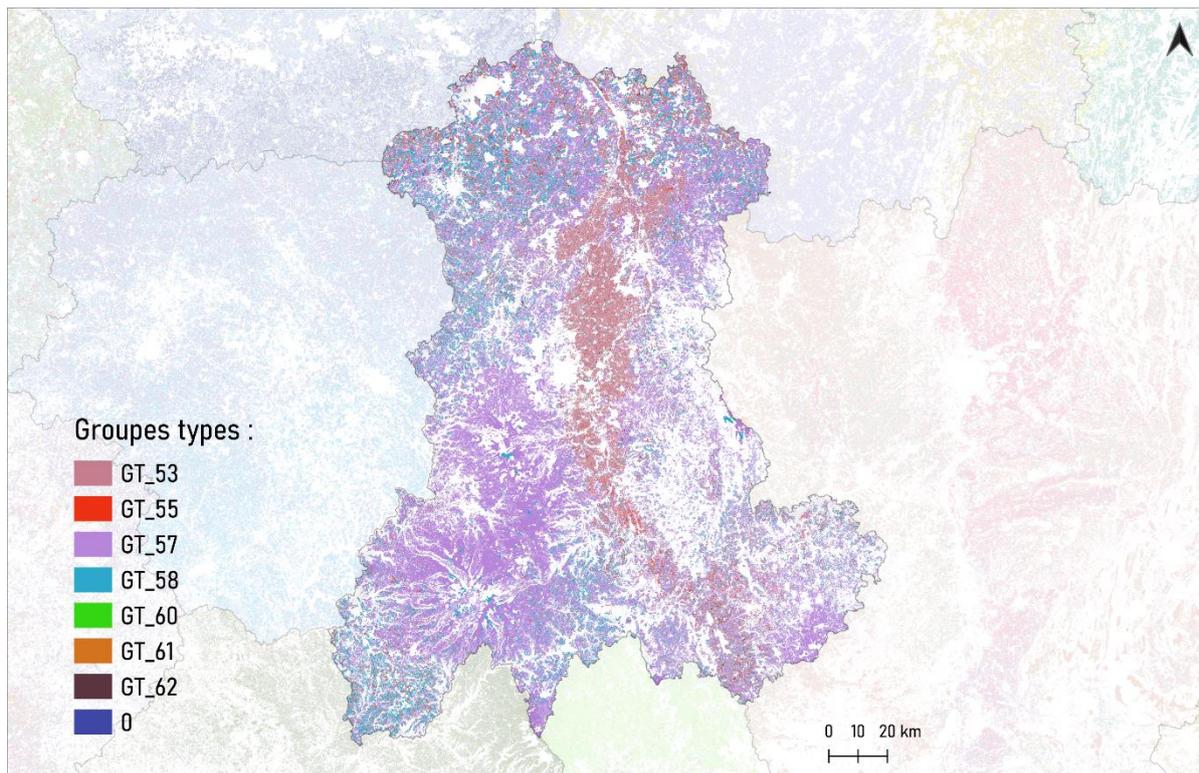
Les groupes types de la région Limousin  
(74) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



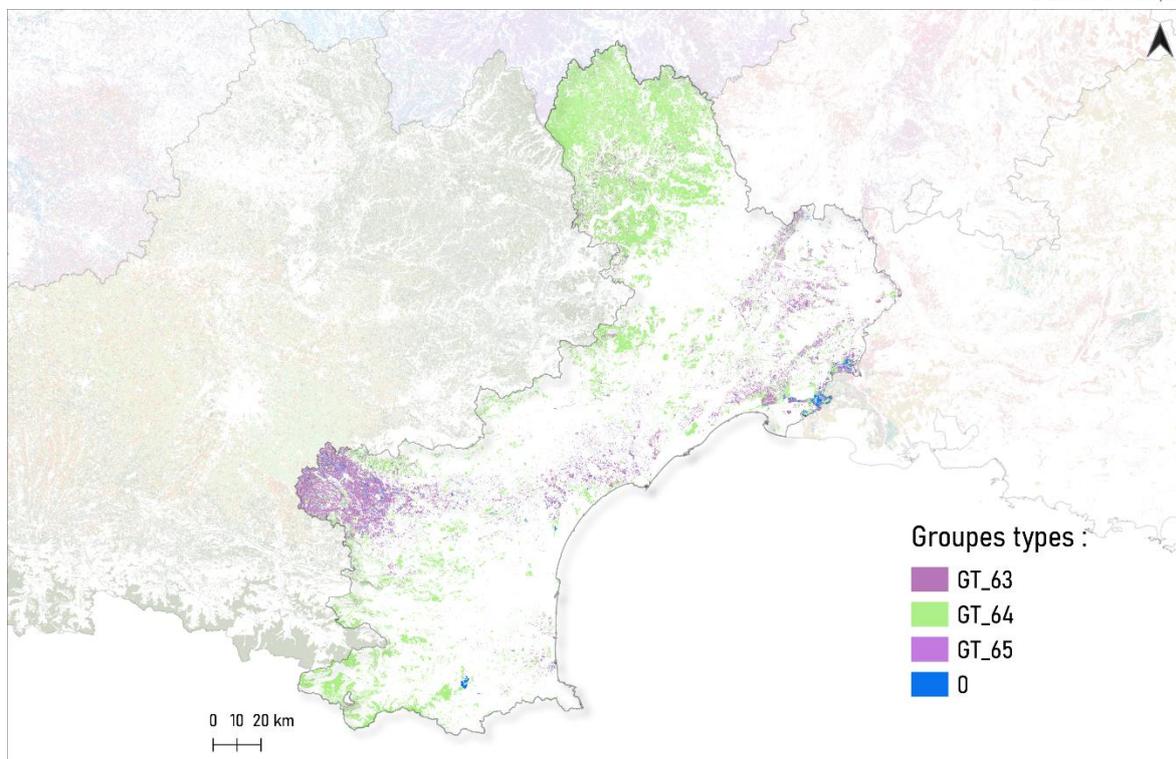
Les groupes types de la région Rhone-Alpes  
(82) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



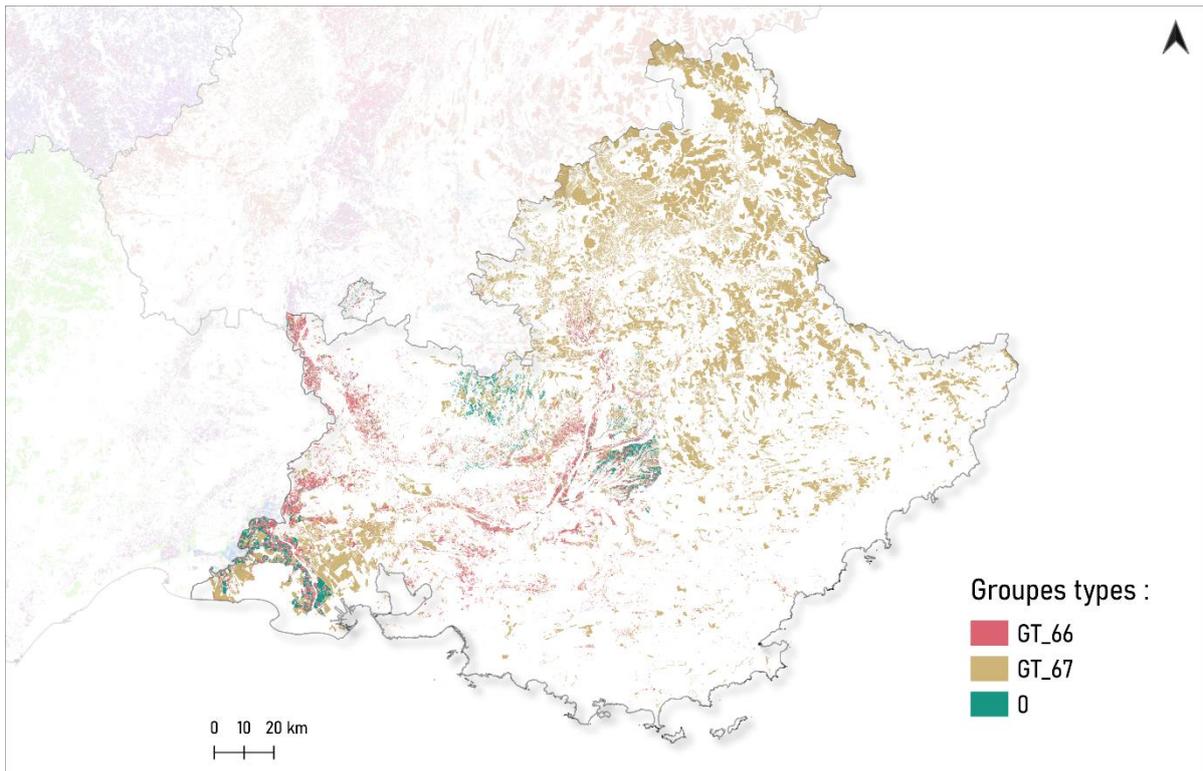
**Les groupes types de la région Auvergne  
(83) en 2016**

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



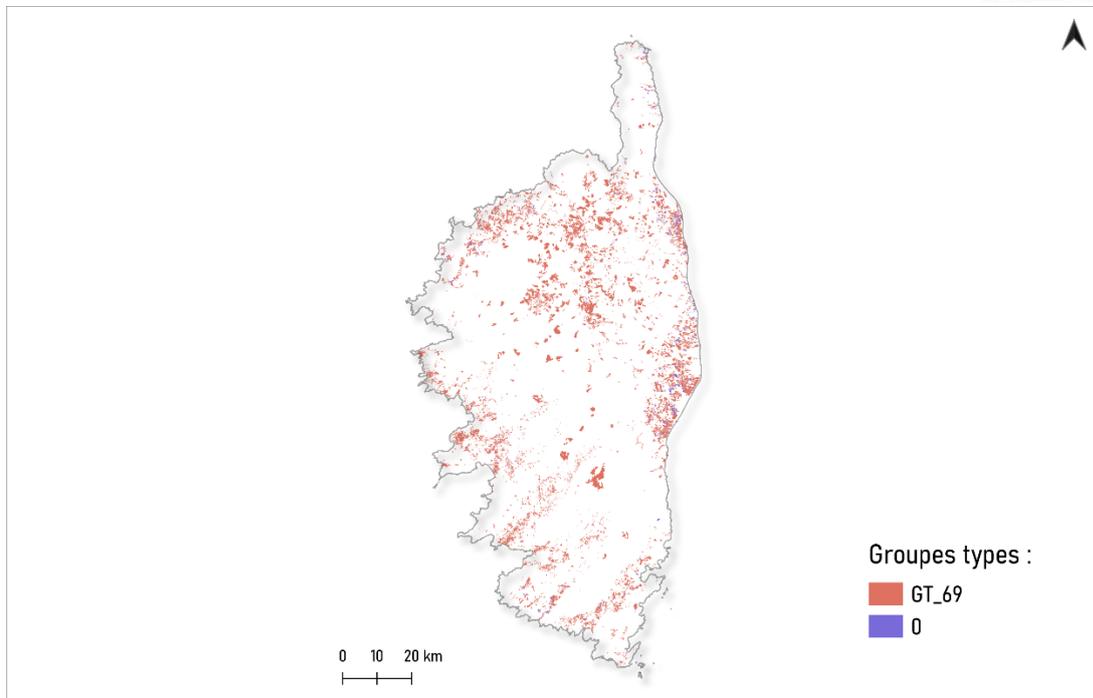
**Les groupes types de la région Languedoc-  
Roussillon (91) en 2016**

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



### Les groupes types de la région Provence-Alpes-Côte-d'Azur (93) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022



### Les groupes types de la région Corse (94) en 2016

Sources : RPG (2016), AROPAJ (2016), IGN-GEOFLA  
Réalisation : Bermudes C, 2022

*Annexe 7 : Atlas des groupes types sur chaque ancienne région française*

## R22

2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes ou fleurs	4 : autres céréales	6 : tournesol	9 : plantes à fibres
8 812	15 496	8 656	1 891	52	2 084

## R25

15 : légumineuses à grains	24 : autres cultures industrielles	25 : légumes ou fleurs	3 : orge	4 : autres céréales	5 : colza	6 : tournesol	9 : plantes à fibres
5	1 273	489	12 632	1 517	7 675	60	1 170

## R26

15 : légumineuses à grains	24 : autres cultures industrielles	25 : légumes ou fleurs	5 : colza	6 : tournesol	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
410	1 216	704	23 150	1 832	3 311	3 679	152

## R31

24 : autres cultures industrielles	25 : légumes et fleurs	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
11 243	18 778	63	1 526	3 024

## R41

1 : blé tendre	15 : légumineuses à grains	2 : maïs grain et ensilage	25 : légumes et fleurs	3 : orge	5 : colza	6 : tournesol
9 574	58	3 475	398	5 927	15 299	807

## R42

6 : tournesol	24 : autres cultures industrielles	25 : légumes ou fleurs
210	2 763	3 203

## R43

1 : blé tendre	15 : légumineuses à grains	2 : maïs grain et ensilage	5 : colza	6 : tournesol
13 021	7	5 976	4 764	231

## R52

1 : blé tendre	15 : légumineuses à grains	2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes ou fleurs	3 : orge	4 : autres céréales	5 : colza	6 : tournesol	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres

53	225	27 623	1 387	6 174	19 017	6 958	12 842	5 229	968	4 071	745
----	-----	-----------	-------	-------	-----------	-------	-----------	-------	-----	-------	-----

#### R54

15 : légumineuses à grains	2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes et fleurs	4 : autres céréales	5 : colza	6 : tournesol
670	19	135	2 540	11 661	21 868	31 869

#### R72

24 : autres cultures industrielles	25 : légumes et fleurs	4 : autres céréales	6 : tournesol	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
1 042	4 988	572	16 865	7 137	2 001	33

#### R74

1 : blé tendre	16 :	2 : maïs grain et ensilage	25 : légumes et fleurs	3 : orge	4 : autres céréales	5 : colza
8 960	4 831	11 885	1 535	4 164	15 356	851

#### R82

1 : blé tendre	15 : légumineuses à grains	2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes et fleurs	4 : autres céréales	5 : colza	6 : tournesol	7 : autres oléagineux	8 : protéagineux	9 : plantes à fibres
22 256	173	36 914	7 016	5 535	3 907	6 476	6 227	4 422	1 183	26

#### R83

1 : blé tendre	15 : légumineuses à grains	16 :	2 : maïs grain et ensilage	24 : autres cultures industrielles	25 : légumes et fleurs	3 : orge	4 : autres céréales	5 : colza	6 : tournesol
33 407	2 710	5 886	12 232	1 102	1 708	11 026	1 569	3 302	2 352

#### R91

15 : légumineuses à grains	2 : maïs grain et ensilage	25 : légumes et fleurs	4 : autres céréales	5 : colza	6 : tournesol
1 566	975	4 114	18 182	857	5 786

*Annexe 8 : D'autres tableaux avec les types de cultures de la classe la plus élevée par région*