



Comparaison des algorithmes de construction de groupes de cultures à partir du RPG

Thomas Poméon, Philippe Martin

► To cite this version:

Thomas Poméon, Philippe Martin. Comparaison des algorithmes de construction de groupes de cultures à partir du RPG : Rapport de synthèse projet PACS-AAC. [Rapport Technique] Office National de l'Eau et des Milieux Aquatiques. 2018, 50 p. hal-02437618

HAL Id: hal-02437618

<https://hal.science/hal-02437618>

Submitted on 13 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparaison des algorithmes de construction des séquences de groupes de cultures à partir du RPG

Approches comparée de RPG Explorer (SADAPT) et de la méthode T-AGIR (ODR)

Rapport de synthèse

**Thomas POMEON (INRA US ODR), Philippe MARTIN
(AgroParisTech UMR SADAPT)**

Mai 2018

Document élaboré dans le cadre de :

La convention ONEMA-INRA 2016-2018 – Action 6 « Pression agricole et
Coordinations socioéconomiques sur les AAC »

- **AUTEURS**

Thomas POMEON, Ingénieur de Recherches (INRA), thomas.pomeon@inra.fr

Philippe MARTIN, Professeur (AgroParisTech), philippe.martin@agroparistech.fr

- **CONTRIBUTEURS**

Pierre CANTELAUBE, Ingénieur d'Etudes (INRA), Pierre.Cantelaube@inra.fr

Nicolas PISKIEWICZ, Développeur (entreprise Piskiewicz), npiskiewicz@gmail.com

Pierre CASEL, Ingénieur d'Etudes (INRA), pierre.casel@inra.fr

Calypso PICAUD, Ingénieur d'Etudes (INRA), calypso.icaud@inra.fr

- **CORRESPONDANTS**

Partenaire : Philippe MARTIN, Professeur (AgroParisTech), philippe.martin@agroparistech.fr

Agence française pour la biodiversité : Claire BILLY, chargée de mission « pression/impact pollutions diffuses agricoles » (AFB), claire.billy@afbiodiversite.fr

Droits d'usage : accès libre
Niveau géographique : national
Couverture géographique : France
Niveau de lecture : professionnel et experts

- **COMPARAISON DES ALGORITHMES DE CONSTRUCTION DES SEQUENCES DE GROUPES DE CULTURES A PARTIR DU RPG, POMEON, T. ET MARTIN, P.**

- **RESUME**

La production de connaissances sur les successions de cultures au sein des parcelles agricoles constitue un enjeu de premier ordre. Elles permettent d'alimenter des travaux et des analyses à différentes échelles sur des enjeux agronomiques et environnementaux majeurs. C'est pour cela que des équipes de recherche ont développé des outils pour déterminer ces séquences, en utilisant les données du Registre Parcellaire Graphique. Le présent travail visait à explorer et activer les synergies entre deux approches de valorisation interannuelle des données issues du Registre Parcellaire Graphique (RPG) : le développement de RPG Explorer (RPG-E) au sein de l'UMR SADAPT et le traitement effectué à l'INRA de Toulouse par l'UMR AGIR avec l'appui de l'US ODR (T-AGIR). L'idée, concrétisée dans un complément à l'action ONEMA PACS-AAC (2017-2018) est de rapprocher les procédures et de développer des synergies pour favoriser l'accès à ces données pour les gestionnaires d'espaces agricoles (utilisateurs de RPG-E en particulier).

Une différence essentielle entre RPG Explorer et T-AGIR réside dans la relation aux utilisateurs des données produites. Pour RPG Explorer on a une approche paramétrable et multitâches par l'utilisateur qui nécessite de disposer de ses propres données RPG pour faire le traitement alors que pour T-AGIR le traitement est intégralement fait par les concepteurs et le résultat peut être transmis aux utilisateurs sans qu'ils ne disposent des données RPG source.

A partir de la description des procédures mise en œuvre par RPG-E et T-AGIR et une analyse comparée globale et détaillée des résultats obtenus testées sur un département (le Cher), nous avons identifié les principaux points communs et divergences entre les 2 approches. Elles s'articulent autour des 2 principales étapes : les méthodes de nettoyage des couches et de filiations des îlots ; et la reconstitution des séquences à partir des îlots affiliés et des groupes de cultures qui leur sont attribués.

Pour les deux méthodes, l'idée est de ne pas retenir les recouvrements trop faibles entre deux îlots de deux années (de type intersection entre îlot A de l'année n et îlot B de l'année n-1) lié uniquement à l'imprécision de la représentation d'un même objet d'une année à l'autre, mais la manière dont les recouvrements trop faibles sont éliminés de la base diffère fortement entre RPG Explorer et T-AGIR. L'approche de RPG Explorer peut conduire à éliminer des îlots réels de très petite taille ou très allongés. Cette perte reste très faible. Dans l'exemple du département 18 on perd ainsi environ 0.2% des surfaces du fait de la perte de 0.3% des îlots.

Pour la reconstitution, les deux méthodes s'appuient sur le même principe de base : pour chaque année on identifie l'ensemble des surfaces déclarées pour une filiation d'îlots donnée. On reconstitue ensuite les séquences en s'appuyant sur des correspondances de surfaces entre années. Mais plusieurs divergences liées à des façons différentes de mettre en œuvre ce principe commun et à des évolutions postérieures sont ressorties au cours de l'analyse : gestion des intersections multiples entre îlots de plusieurs années, prise en compte des séquences tronquées, règles de reconstitution appliquées, attribution des surfaces, etc. Les nouvelles versions de RPG Explorer issues du travail de comparaison permettent d'atteindre un plus grand taux de reconnaissance des séquences sur le cas du département du Cher (18) utilisé pour le test. Les différences persistantes conduisent à un recoupement non intégral dans les résultats. Les résultats sont toutefois globalement assez proches. Si on regarde au niveau des séquences élémentaires reconstituées, on constate que 79% des séquences reconstituées par T-AGIR le sont également par RPG-E sur les mêmes îlots, soit en surface une correspondance 78%. A l'inverse, si on part des séquences reconstituées par RPG-E, on retrouve ces séquences dans 56% des cas, et pour 66% en surfaces dans les résultats de T-AGIR. Si on ne regarde que les 4 années les plus récentes (2011-2014), on passe de 56% à 80% des séquences reconstituées par RPG-E qu'on retrouve pour les mêmes îlots avec T-AGIR. Cela confirme que l'écart dans les résultats s'explique en grande partie par des différences sur une ou quelques cultures en début de période, avec côté RPG-E une plus forte propension à reconstituer des séquences plus nombreuses (plus détaillées) et de plus faible surface élémentaire.

Les travaux menés dans cette action ont donc permis de :

- Préciser les points communs et différences entre les deux méthodes T-AGIR et RPG Explorer,

- Quantifier et qualifier les écarts quant aux résultats obtenus,
- Corriger un certain nombre d'erreurs dans RPG-Explorer qui en retour ont permis de réduire l'importance des écarts (par exemple la correspondance exacte entre îlots était de 50% en début d'étude contre 59% en fin d'étude, idem pour la correspondance des séquences qui est passée de 71% à 78% des surfaces) : ne pas prendre en compte des surfaces déjà utilisées (gestion des surfaces de culture incluse dans plusieurs filiations), reconstitution des séquences même pour les îlots avec une année manquante, mise en œuvre de la règle 10 des surfaces majoritaires.
- Rapprocher les travaux des deux équipes et mettre en commun des données et des procédures pour améliorer les outils et informations disponibles : base de correspondances des identifiants d'exploitation, trajectoire d'exploitations, etc.

Pour étayer plus solidement cette comparaison il serait très utile de disposer de vérités terrain, c'est-à-dire de données validées qui associent aux îlots les séquences à l'intérieur de ces îlots. Malheureusement ces vérités-terrain n'existent pas à l'heure actuelle, en tout cas pas sur une échelle de temps et d'espace correspondante aux données traitées via les RPG de 2006 à 2014 sur l'ensemble du territoire français. En revanche des perspectives très intéressantes s'ouvrent avec l'accès aux nouvelles données. En effet à partir de la campagne 2015 les données du RPG sont distribuées à la fois à l'échelle parcellaire et à l'échelle de l'îlot. Il sera alors possible de refaire le traitement par îlot, tel que réalisé sur les années antérieures, et confronter ces résultats avec les vérités terrain. Ce mode d'approche pourrait notamment permettre de mieux ajuster les paramétrages de RPG Explorer en fonction des départements. La poursuite des travaux, outre une plus forte connexion entre équipes que cette action a déjà permis de renforcer, implique une continuité à la fois dans les équipes de travail et dans les financements. Ce point constitue une difficulté et une limite non négligeable pour l'amélioration et la pérennisation des travaux sur les séquences RPG.

- **MOTS CLES (THEMATIQUE ET GEOGRAPHIQUE)**

Registre Parcellaire Graphique (RPG), Îlots, Séquences de culture, RPG Explorer, France métropolitaine

- **SOMMAIRE**

1	Introduction : rappel du contexte et des objectifs	7
2	Description, atouts et limites des approches mobilisées pour la construction des séquences de culture.....	9
2.1	Approche mise en œuvre par l'UMR AGIR (T-AGIR)	9
2.2	Approche mise en œuvre dans RPG-Explorer (RPG-E)	14
2.3	Premières différences identifiées entre les approches T-AGIR et RPG-E.....	17
3	Travaux de comparaison et analyse des résultats.....	20
3.1	Premiers résultats sur la comparaison entre les 2 approches	20
3.1.1	Présentation de la démarche	20
3.1.2	Résultats : approche globale	21
3.1.3	Résultats : exemples illustrant les procédures de reconstitution des séquences	25
3.1.4	Résultats : discussion sur les principales divergences entre les 2 approches	28
3.2	Point sur la validation de la filiation des exploitations à partir des données issues du traitement des données MAE-PAC	29
4	Synthèse des acquis et Perspectives.....	31
4.1	Synthèse des acquis.....	31
4.1.1	Filiation des îlots	31
4.1.2	Etablissement des séquences au sein des filiations d'îlots	32
4.2	Perspectives.....	33
	Annexe 1 : Déroulé des travaux en 2017	35
	Annexe 2 : méthode développée dans RPG Explorer	37
	Annexe 3 : méthode de rapprochement et sélection des identifiants d'exploitations	39
	Annexe 4 : comparaison du différentiel de traitement dans le cas d'îlots non stables dans le temps	40
5	Glossaire.....	47
6	Sigles & Abréviations	48
7	Table des illustrations	49

COMPARAISON DES ALGORITHMES DE CONSTRUCTION DES SEQUENCES DE GROUPES DE CULTURES A PARTIR DU RPG

1 Introduction : rappel du contexte et des objectifs

Le travail visait à explorer et activer les synergies entre deux approches de valorisation interannuelle des données issues du Registre Parcellaire Graphique (RPG) : le développement de RPG Explorer (RPG-E) au sein de l'UMR SADAPT et le traitement effectué à l'INRA de Toulouse par l'UMR AGIR avec l'appui de l'ODR (T-AGIR). L'idée, concrétisée dans un complément à l'action ONEMA PACS-AAC (2017-2018) est de rapprocher les procédures et de développer des synergies pour favoriser l'accès à ces données pour les gestionnaires d'espaces agricoles (utilisateurs de RPG-E en particulier).

Le RPG est un dispositif administré par l'Agence de Service et de Paiement (ASP) dans le cadre de la gestion des aides européennes surfaciques de la Politique Agricole Commune (PAC). C'est un système d'information géographique qui permet l'identification des parcelles agricoles et la caractérisation de l'occupation agricole du sol et des structures foncières. Il contient 6 millions d'îlots environ, soit plus de 27 millions d'hectares, déclarés annuellement par près de 400 000 agriculteurs. Jusqu'en 2016, l'ASP a diffusé une version anonymisée (et payante) du RPG pour des usages relevant d'une mission de service public, notamment la recherche ou encore la conduite de projets ou d'études sur l'aménagement du territoire ou la qualité des eaux. L'INRA a ainsi eu accès, dans le cadre d'une convention avec l'ASP, à la version du RPG dite de niveau 4 (pour les années 2006 à 2014), qui contient l'ensemble des informations sur les îlots recensés et associe chaque îlot avec un identifiant anonymisé de l'exploitation¹. Les données du RPG étaient livrées par département et par année. Des traitements sont donc nécessaires pour reconstituer un jeu de données intégrant plusieurs départements, et plus encore pour relier les informations entre années, par exemple pour reconstituer des séquences de culture par îlot.

Le module de RPG Explorer servant à l'établissement des séquences de groupes de cultures est proche de celui mis en œuvre par l'UMR AGIR, sans être rigoureusement identique. Chacune des deux procédures se caractérise par des objectifs et des approches sensiblement différents. Cette différence conduit nécessairement à des différences dans les résultats de séquences de groupes de cultures obtenues. Chacune des deux approches peut aussi comporter des erreurs conceptuelles ou de codage informatique. La confrontation présentée ici visait aussi à les mettre à jour et les corriger.

L'extraction et la diffusion des données issues du traitement AGIR est administré par l'ODR, qui facture au demandeur externe à l'INRA les coûts liés à la gestion et mise en œuvre de cette base de données (500 €/demande). RPG-Explorer est diffusé gratuitement par l'UMR SADAPT. Chaque utilisateur est en charge de récupérer les données RPG qui lui sont nécessaires et de réaliser les traitements via RPG-Explorer pour obtenir les indicateurs recherchés. L'enjeu est ici de valider les données produites et diffusées via chaque méthode sur la reconstitution des séquences et d'établir des pistes d'amélioration, qui pourront servir aux différents utilisateurs des données sur les séquences. Plus spécifiquement, les objectifs de chaque partenaire sont :

- Côté ODR : continuer à rendre disponible les séquences, avec la documentation et les métadonnées ; éventuellement dans les différentes approches (RPG-E et T-AGIR), chacune avec ses avantages et inconvénients, si cela vaut la peine (selon les utilisations : échelle, finesse des séquences, etc.). Si une approche ressort comme clairement supérieure (et pas juste différente), privilégier l'approfondissement et la diffusion de cette approche et de ses résultats comme référence.
- Côté RPG Explorer : continuer à faire de RPG Explorer un outil d'agrégation de méthodes d'analyse des données RPG. De ce fait RPG Explorer est en évolution continue (1 à 2 actualisations publiques par an) avec des enjeux forts en termes d'intégration des nouvelles données RPG (à compter des données 2015) ou de traitement au niveau national et plus seulement local. RPG Explorer ne disposant d'un budget de développement que jusqu'à 2019, une réflexion est en cours pour envisager la suite (développement collaboratif, ...).

Le travail a été organisé en deux volets :

- Volet 1 : Analyse comparée et convergence des algorithmes utilisés à AGIR-ODR et dans RPG Explorer pour l'établissement des séquences de groupes de cultures pour s'assurer que les

¹ Pour plus d'informations, voir par exemple :

https://www.asp-public.fr/sites/default/files/plaquette_mad_asp_2009_v5.pdf

deux services donnent des résultats proches et cohérents tout en identifiant leurs points forts et limites respectifs.

- Volet 2 : Définition du cahier des charges d'un nouveau système permettant à RPG Explorer d'interroger la base de AGIR-ODR pour télécharger les données RPG prétraitées (séquences de groupes culture).

Ce rapport dresse le bilan du volet 1. L'objectif de ce volet est de s'assurer que les deux services donnent des résultats proches et cohérents d'une part, et de mieux identifier leurs points forts et limites respectifs d'autre part. Les premiers échanges nous ont amenés à élargir le travail à l'ensemble des points de convergence entre SADAPT et l'ODR autour des données RPG. Cela a concerné en particulier le traitement des doublons des identifiants d'ilots et d'exploitations et le suivi interannuel des exploitations et de leur parcellaire (table permettant de recenser les identifiants anonymes successifs d'une même exploitation au cours du temps).

Les principales personnes impliquées dans les travaux ont été : P. Martin (UMR SADAPT), P. Cantelaube, T. Poméon et C. Picaud (US ODR), N. Piskiewicz (l'entreprise Piskiewicz assure le développement informatique de RPG Explorer), avec également la participation de P. Casel (UMR AGIR).

2 Description, atouts et limites des approches mobilisées pour la construction des séquences de culture

2.1 Approche mise en œuvre par l'UMR AGIR (T-AGIR)

Chaque année (2006-2014), les données du RPG sont intégrées département par département dans les bases de données de l'ODR (base SQL). Un nettoyage des données est réalisé. Il consiste, en plus d'une correction des erreurs de géométrie de base, à repérer et supprimer les **îlots en doublon** (îlots présents dans plusieurs fichiers départementaux, principalement sur les « bordures » de départements), et à générer des tables de données **France entière**. Ces îlots en bordure sont utilisés pour reconstituer les territoires d'exploitation, ce qui requiert de les associer à un seul numéro d'exploitation². Une procédure a été définie par l'ODR (P. Cantelaube) pour l'attribution de cet identifiant unique (voir annexe 3). L'identifiant d'exploitation conservé est celui intégrant le numéro de département auquel est rattaché le siège de l'exploitation (s'il n'existe pas alors il est créé, ce qui est le cas quand une exploitation a des îlots dans plusieurs départements mais aucun dans son département siège). Le département siège est une des variables présentes dans les données attributaires du RPG anonymisé niveau 4 par l'ASP. On souhaite ajouter à cette information la commune du siège d'exploitation, mais cette variable, n'est pas présente dans la version du RPG transmis par l'ASP. Elle est donc estimée en prenant pour commune siège la commune qui regroupe la plus grande surface des îlots de l'exploitation au sein du département siège. On aboutit ainsi à **une base nationale sans îlots en doublon** avec les données attributaires suivantes pour chaque îlot : n° d'îlot, n° de commune de l'îlot (vérifié via les données de l'IGN), n° d'exploitation unique, groupes de cultures et surfaces associées.

Les données attributaires des îlots qui se superposent pour tout ou partie, et qui ne correspondent pas à des répétitions en bordure de département, sont conservées comme telles. En effet elles correspondent généralement à des erreurs de géométrie mais pas de déclaration de surface (les surfaces déclarées existent bien, même si elles sont en l'occurrence mal géoréférencées). Nous verrons plus loin les implications possibles de ces îlots superposés pour tout et partie dans la procédure de reconstitution des séquences.

Concernant l'identification des séquences, l'UMR AGIR (P. Casel et O. Théron ; avec l'appui initial de R Chartier et V. Fuzeau, US ODR) a développé une **méthodologie de reconstitution des séquences de groupes de cultures**, à partir des données prétraitées par l'ODR. La première étape consiste à déterminer une **filiation entre les îlots des différentes années**, puisque chaque couche annuelle d'îlots est indépendante et qu'il n'y pas de continuité dans l'identification des îlots. On part de la géométrie des îlots (càd le fichier shape) de **l'année de référence n (càd la plus récente)**. Pour chaque îlot de n on recherche **le ou les îlots correspondants géométriquement dans chacune des années précédentes n-x** (avec $x = 1, 2, 3, \dots$), année par année (îlots n avec les îlots n-1, îlots n avec les îlots n-2, etc.), à partir de la condition d'une **intersection géométrique minimale de 10%**. Autrement dit, en partant des îlots de l'année de référence n, on regarde pour chaque année précédente par jointure géographique les îlots qui intersectent à au moins 10% de l'îlot le plus petit : $\text{surf_intersect} \geq 0.1 * \min(\text{surf_ilot_n} ; \text{surf_ilot_n-x})$. Le schéma suivant (Figure 1) illustre la logique de la procédure d'intersection des îlots dans T-AGIR.

² A compter de 2010, les exploitations ayant des îlots sur plusieurs départements se voient attribuer un identifiant différent par département alors qu'elles avaient un identifiant unique (différent d'une année à l'autre) jusqu'en 2009

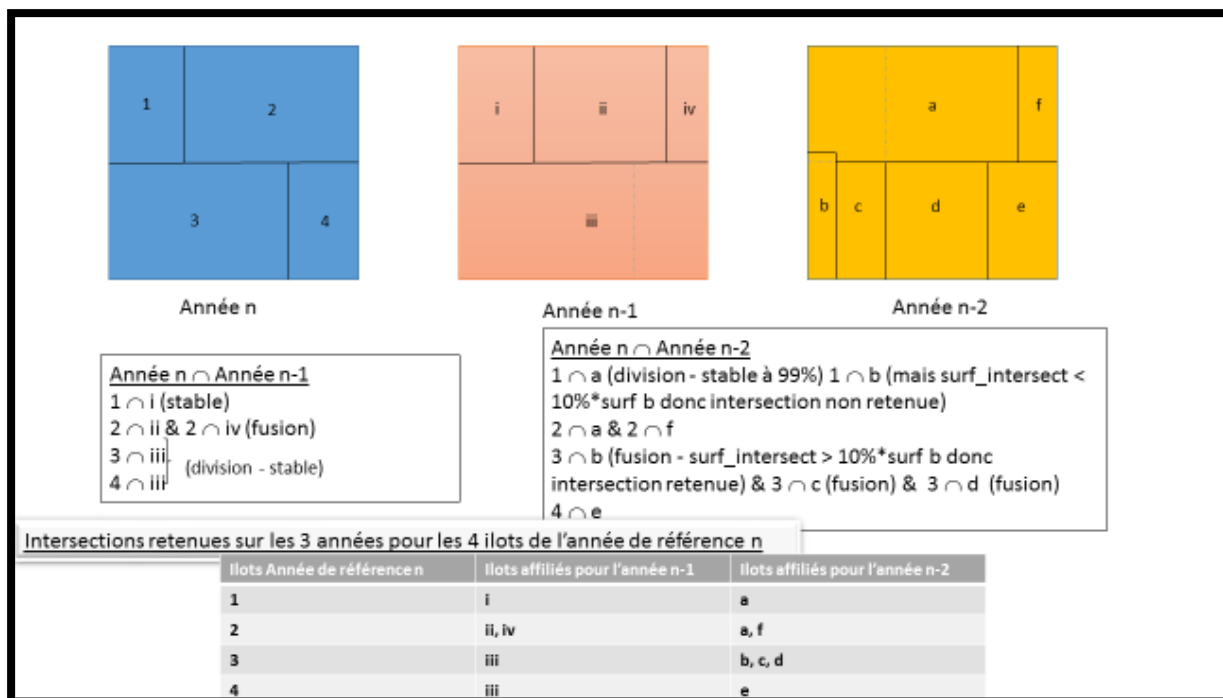


Figure 1 : logique de filiation interannuelle des îlots pour la méthode T-AGIR

Lors de cette étape de filiation, un **biais** est possible du fait de l'association entre îlots d'années précédentes qui ne s'intersectent pas géométriquement, mais dont les surfaces peuvent être intégrées dans une même séquence car relié au même îlot de référence. Par exemple, pour l'îlot de référence 2, les îlots ii (n-1) et f (n-2) d'une part et iv (n-1) et a (n-2) sont virtuellement associées à travers leur correspondance commune avec l'îlot 2, même s'ils ne s'intersectent pas l'un avec l'autre. Autrement dit, une séquence pourrait associer des cultures des îlots ii et f, alors que ces cultures sont localisées dans des parcelles différentes. Un autre biais possible avec cette méthode consiste à une affiliation d'une même surface à plusieurs îlots de l'année de référence. Par exemple on voit que les îlots de référence 3 et 4 sont tous les deux affiliés avec l'îlot iii, et que donc on pourrait potentiellement avoir des séquences pour les îlots 3 et 4 qui reprennent les mêmes surfaces de l'îlot a (voir ci-dessous pour un exemple de surfaces retenues plusieurs fois dans deux séquences différentes).

Un 3ème biais potentiel est lié aux erreurs de géométrie mentionnées plus haut. En effet, si plusieurs îlots qui se superposent une année donnée sont conservés (Figure 2), cela peut engendrer le croisement (fictif) de l'îlot superposé avec plusieurs îlots des autres années :

- Si l'erreur concerne les îlots de l'année de référence, alors chaque îlot superposé sera conservé, et croisé indépendamment avec les autres années. On aura donc plusieurs îlots, qui reprendront chacun la même information pour les années précédentes (risque de surfaces doublonnées, c'est-à-dire mobilisées dans plusieurs séquences).
- Si l'erreur concerne les années précédentes, alors seront associés à l'îlot de référence (unique lui) l'ensemble des surfaces des îlots « superposés » qui se superposent.

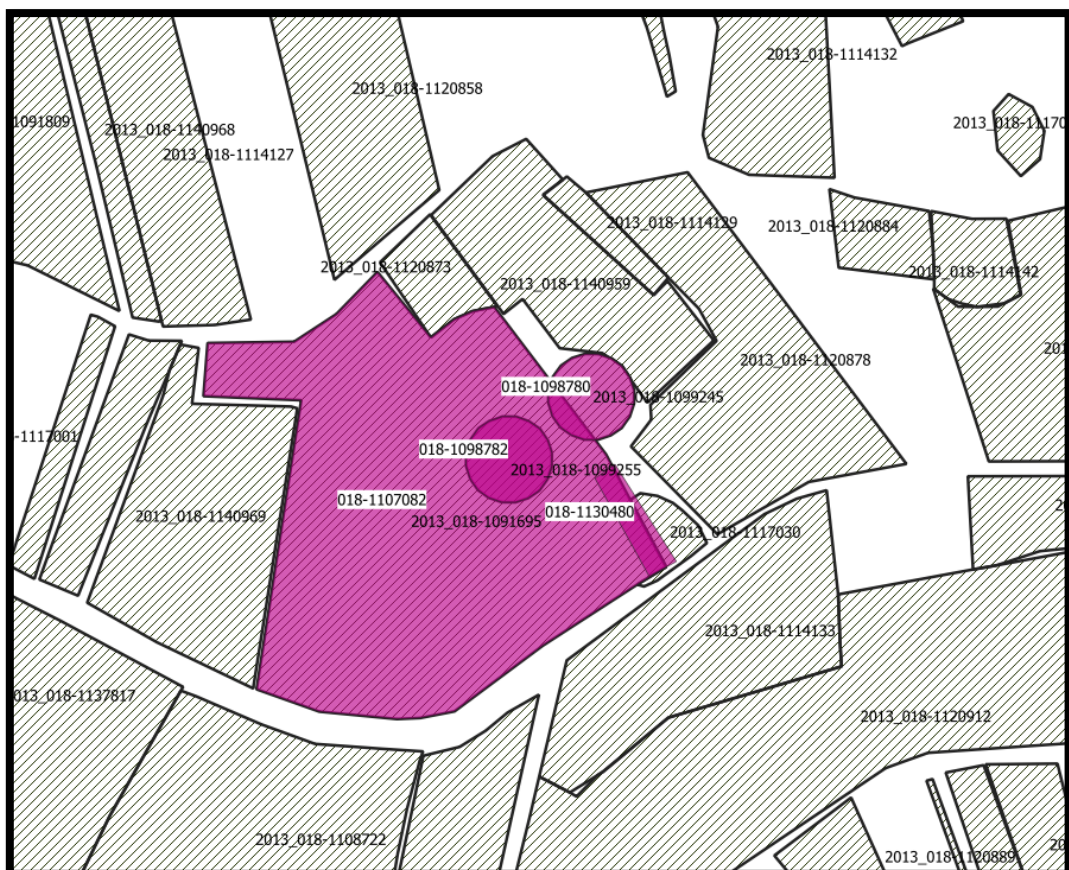


Figure 2 : exemple d'ilots qui se superposent pour une même couche annuelle (RPG 2014). On voit ici 4 ilots colorés : l'ilot 1098782 est complètement inclus dans l'ilot 1107082. Les deux autres ilots (1098780 et 1130480) se superposent partiellement avec ce même ilot 1107082. En hachuré est représentée la couche des ilots RPG en 2013 (identifiants d'ilots précédés du préfixe « 2013_ ») : on retrouve la même configuration sauf pour l'ilot 1130480 qui n'a pas d'équivalent. Dans la logique de la méthode T-AGIR, les 4 ilots 2014 seront affiliés à l'ilot 2013 correspondant à l'ilot le plus grand (1107082), avec donc potentiellement une répétition dans plusieurs séquences des surfaces de groupes de cultures considérées.

Une fois les filiations établies entre ilots par intersection, on obtient **pour chaque îlot de l'année de référence des surfaces de groupes de cultures par année** par îlot affilié (correspondant aux tables « culture_DPT », avec les champs *id_ilot* de l'année n, groupe de culture –gc, surface, année, et pourcentage de surface intersectée des ilots associés – *surf_intersect* telle que définie précédemment), sur lesquels un algorithme est appliqué pour déduire quels sont les groupes de cultures qui se succèdent. L'extrait ci-dessous (Tableau 1) illustre l'information disponible et mise en forme à ce stade de la procédure.

Tableau 1 : exemple des données pour un îlot de référence issues de l'intersection des couches des différentes années

id_ilot [identifiant de l'ilot pour l'année de référence – 2014 ici]	type_culture [libellé du groupe de culture RPG (code)]	Surface [surface en ha du groupe de culture pour l'année concernée]	ilot_pc [pourcentage de l'intersection avec l'ilot correspondant]	Annee [année de la couche RPG]
018-1090845	blet (1)	2.6	100	2014
018-1090845	blet (1)	1.3	100	2013
018-1090845	orge (3)	1.14	100	2013
018-1090845	autrescereales (4)	0.16	100	2013
018-1090845	prairiep (18)	2.6	100	2012
018-1090845	prairiep (18)	2.6	100	2011
018-1090845	prairiep (18)	2.6	100	2010
018-1090845	prairiep (18)	2.7	100	2008
018-1090845	prairiep (18)	2.7	100	2007
018-1090845	prairiep (18)	2.7	98	2006

Nota bene : L'ilot de référence est ici associé à 1 groupe de culture en 2014, 3 en 2013, 1 en 2012/11/10, aucun en 2009 puis 1 en 2008, 2007 et 2006. A noter que, pour l'exemple présenté ici, la procédure de filiation des ilots aboutit chaque année à un seul îlot affilié à l'ilot de référence 2014 (avec 3 groupes de culture différents en 2013), sauf pour 2009 où aucun îlot n'est identifié.

Des règles (filtres) successives (du plus sûr au moins certain) (Tableau 2) sont appliquées sur la table « culture_DPT » pour chaque îlot de l'année de référence. Les **séquences de groupes de cultures avec les surfaces associées³ sont ainsi reconstituées**. Toutes les combinaisons possibles **pour l'ensemble des années⁴** sont successivement testées avec les différents filtres, du plus sûr au moins certain. Quand avec un filtre l'algorithme parvient à reconstituer une séquence, avant de passer au filtre suivant, d'une part, la séquence est stockée, et d'autre part, les combinaisons id_îlot- culture-surface-année associées à cette séquence sont retirées de la liste des cultures associées à l'îlot (table « culture_DPT »). Ainsi, chaque filtre ne réutilise pas les surfaces de cultures affectées aux séquences générées par le filtre précédent. Toutes les règles sont donc appliquées successivement en prenant en compte d'abord l'ensemble des années, par exemple sur 9 ans pour 2006-2014. Si aucune combinaison possible n'est trouvée, ou s'il reste des cultures non attribuées, l'algorithme va ensuite appliquer successivement les mêmes filtres mais en retirant l'année la plus ancienne. Pour le cas des séquences 2006-2014, cela signifie que si la reconnaissance des séquences n'a pas pu être finalisée sur des séquences longues de 2006 à 2014, alors la première année (2006) est retirée et toutes les combinaisons possibles pour 2007-2014 sont alors établies. Dans ce cas, le filtre indiqué avec la séquence précise seulement la règle utilisée à cette étape, et ne garde pas en mémoire s'il y a eu auparavant des séquences déjà reconstituées. Une conséquence importante est que par exemple le filtre « une_culture_anne_1 » peut se trouver pour une séquence de n-1 années, alors qu'il y a plusieurs cultures dans l'îlot chaque année. La raison en est que dans ce cas des séquences de n années ont déjà été reconstituées et qu'il ne reste ensuite plus qu'une culture par année sur cette période⁵. Et ainsi de suite jusqu'aux 2 années les plus récentes (ici, 2013-2014) pour tenter de reconstituer des séquences de 2 ans avec les cultures restantes. Notons que **si plusieurs combinaisons équivalentes existent pour un même ensemble d'années et un même filtre au sein d'une même filiation d'îlots alors la méthode T-AGIR refuse d'en choisir une et la séquence n'est donc pas reconnue**. Au fur et à mesure, les surfaces de cultures reconstituées en séquences sont stockées dans des tables temporaires et retirées du pool de surfaces, pour éviter qu'elles soient mobilisées plusieurs fois.

Si on reprend l'**exemple** du cas de l'îlot 018-1090845 (voir Tableau 1), la méthode T-AGIR aboutit à la détermination de 3 séquences pour 2006-2014. Notons que vu qu'il n'y a aucune culture identifiée pour 2009, les trois séquences reconstituées ne couvrent que les années 2010 à 2014 : 0_0_0_0_18_18_18_4_1 // 0_0_0_0_18_18_18_3_1 // 0_0_0_0_18_18_18_1_1 (avec 0 = pas de cultures identifiées ; 18 = prairies permanentes ; 4 = autres céréales ; 3 = orge ; 1 = blé tendre). Une seule année (2013) présente plusieurs groupes de cultures, dont la somme correspond bien aux surfaces du groupe de culture identifié pour les autres années. Au final on a donc 3 séquences qui ne varient que par rapport à la culture identifiée en 2013.

Tableau 2 : règles mobilisées successivement pour la reconstitution des séquences de groupes de culture par T-AGIR

	Filtre de Reconstitution	Courte description
1	une_culture_annee_1	un seul groupe de cultures par îlot. Le filtre traite spécifiquement les filiations d'îlots ne comportant qu'un seul groupe de cultures chaque année.
2	surf_eq	surface égale. Le filtre traite les filiations d'îlots comportant plusieurs groupes de cultures chaque année. On met en correspondance les groupes de cultures de même surface d'une année à l'autre.
3	agdg_eq_1	agrégation-désagrégation à surfaces égales. Le filtre traite les filiations d'îlots comportant au moins certaines années plusieurs groupes de cultures. On met en correspondance les groupes de cultures d'une année n dont la somme est égale à la surface d'un seul groupe de cultures des années n+1 ou n-1.
4	surf_sim	surface similaire à 5% (différence maximale 1 ha). Le filtre est le même que surf_eq mais en acceptant une différence de surface de de 5% (max de 1 ha)

³ Si pour un îlot de référence on a obtenu 2 îlots intersectés avec un ou plusieurs mêmes groupes de culture (par exemple 2 ha de blé dans un îlot intersecté d'une part et 1 ha de blé dans l'autre îlot intersecté d'autre part), ces 2 surfaces sont chacune reportées séparément dans la table « culture_DPT » mais sont ensuite traitées comme un même ensemble (2+1 = 3 ha de blé) pour la reconstitution des cultures. Voir par exemple les îlots 018-1091008 ou 018-1095997, cas non détaillés ici.

⁴ Avec pour certains filtres (ceux avec agrégation-désagrégation) une approche progressive par paire d'année pour faciliter la reconstitution.

⁵ Voir les exemples du 3.1.3

5	agdg_eq_2	agrégation-désagrégation à 0,01 ha. Le filtre est le même que agdg_eq_1 en acceptant une différence de surface de 0.01 ha.
6	agdg_sim_0	agrégation-désagrégation à 5% (différence max = 1 ha). Le filtre est le même que agdg_eq_1 en acceptant une différence de surface de 5%.
7	agdg_sim	identique au filtre précédent mais sans compter les surfaces inférieures à 0,3 ha (permet de réduire la combinatoire à explorer)
8	une_culture_annee_2	une seule culture par îlot. Identique à « une_culture_annee_1 ». Ce filtre peut récupérer des surfaces de séquences suite à l'application des filtres précédents, qui peut avoir ramené les situations où les îlots comportaient plusieurs groupes de cultures à une situation avec un seul groupe de cultures par an.

La méthode retenue pour définir l'intersection entre îlots présente un biais potentiel pour la reconstitution des séquences. En effet le critère d'une intersection minimal de 10% peut conduire à **retenir plusieurs fois une même surface**, comme signalé plus haut. Si on a un seul îlot n-x pour plusieurs îlots n, il y a un risque qu'un groupe de culture de l'îlot n-x soit associé dans plusieurs îlots n, et donc soit ainsi fictivement démultiplié. Si par exemple en 2013 un îlot de 10 ha contient 4ha de maïs, 3ha de tournesol et 3ha de soja. Et en 2014 cet îlot est divisé en deux îlots de 5ha qui contiennent chacun 4 ha de maïs et 1 ha de tournesol. Alors l'algorithme T-AGIR détectera deux séquences de maïs de deux ans et de 4 ha chacune, mobilisant donc en 2013 deux fois plus de maïs que dans la réalité. Si on prend un autre exemple à partir de la configuration présentée dans la figure 1, on pourrait avoir un même 1 ha de blé de l'îlot iii inclus à la fois dans une séquence associée à l'îlot 3 et une autre associée à l'îlot 4 ; et donc un 1 ha de blé en plus pour l'assolement de l'année n-1 qui serait recalculé à partir des séquences.

Actuellement, il n'existe pas de moyen simple pour repérer et mesurer l'impact de cette erreur. Pour l'évaluer globalement, il faut pouvoir repérer les éventuels cas où la surface pour un groupe de cultures incluse dans des séquences reconstituées serait supérieure à la surface déclarée de cette culture pour une année donnée. La procédure de contrôle par le calcul des assolements à partir des séquences (évoqué plus haut) est néanmoins complexe à mettre en œuvre de manière systématique et reste incertaine. D'autres éléments peuvent intervenir ou manquer (pas de traçabilité des îlots des années précédentes associés à chaque culture, prise en compte des surfaces non reconstituées, etc.), et on ne peut donc pas (en tout cas dans un délai raisonnable) estimer précisément les surfaces répétées. On peut cependant faire l'hypothèse que ce biais a peu d'impacts sur les résultats, du fait de la faible probabilité des conditions de son occurrence (combinaison d'un cas d'une superposition d'îlots pour une même année ou d'une division d'îlot d'une année sur l'autre⁶ + présence dans les 2 îlots issus de la division de surfaces correspondantes chacune à la surface d'un même groupe de culture de l'année antérieure) : les règles de reconstitution des séquences et le nombre relativement limité d'erreurs de géométrie font que vraisemblablement ces biais n'ont qu'un impact limité sur les résultats. C'est ce que confirmera plus loin l'analyse des résultats issus des procédures de reconstitution des séquences, tant en termes de résultats globaux que pour les exemples qui seront examinés.

A l'issue de l'étape de reconstitution des séquences, un algorithme de T-AGIR réalise une **interpolation** à partir des surfaces de culture restantes et des séquences déjà reconstituées. Sur la base des séquences les plus fréquentes à l'échelle de l'exploitation et du segment Arvalis⁷ d'autre part, l'algorithme vérifie pour chaque îlot la possibilité de reconstituer des séquences identiques à partir des surfaces qui n'ont pas été utilisées pour l'étape de reconstitution. Par exemple, pour le département du Cher – 18, l'étape de reconstitution permet de caractériser les séquences sur la période 2006-2014 pour 87% des surfaces des îlots de l'année de référence 2014 (378 000 ha environ), et pour 98% (425 000 ha environ) après l'étape d'interpolation. A la fin de l'ensemble du processus, il reste un pool de cultures (par îlot) qui n'ont pu être ni reconstitué ni interpolé dans des séquences de culture. Il s'agit des surfaces de groupes de cultures qui ne correspondent pas à des séquences reconstituées existantes (dans l'exploitation ou le segment Arvalis), ou pour lesquelles il n'a pas été possible de trouver des surfaces

⁶ A titre d'exemple, si on regarde les couches RPG 2014 et 2013 pour un département (le Cher, analysé plus loin), on a 298 îlots 2013 qui sont associés à plus d'1 îlots 2014 (551 îlots 2014 au total, avec entre 2 et 4 îlots 2014 qui sont affiliés avec le même îlot 2013) (comme illustré dans la Figure 2), et donc potentiellement avec des séquences mobilisant les mêmes surfaces. Cela reste assez faible sur un total de plus de 50 500 îlots, d'autant plus qu'il n'y a pas forcément de séquences reconstituées pour ces îlots, ou alors des séquences de cultures différentes.

⁷ Zonage réalisé en 2011 par Arvalis qui regroupe des cantons homogènes en termes de conditions pédoclimatiques et d'itinéraires techniques. A cette échelle, l'hypothèse d'une similarité des systèmes de culture dans un même segment Arvalis permet d'interpoler les séquences de culture sur la base de celles déjà reconstituées.

correspondantes à combiner les autres années pour interpoler une séquence. A noter que cette étape d'interpolation n'est pas analysée ici car elle est spécifique à T-AGIR.

P. Casel de l'UMR AGIR a réalisé un traitement France entière pour l'année 2014 (séquences 2006-2014) qui a été rendu disponible en mars 2017 (la version précédente correspondait à l'année de référence 2012). Les données de séquences de groupes de cultures sont actuellement diffusées au cas par cas (sauf pour les demandes internes à l'INRA), comme pour les autres données issues du RPG hébergées par l'ODR. Seules sont transmises les données attributaires (séquences + surfaces) avec les numéros d'îlot pour l'année de référence.

2.2 Approche mise en œuvre dans RPG-Explorer (RPG-E)

L'UMR SADAPT a développé un outil de traitement en local des données du RPG. Cet outil nommé RPG Explorer et codé par N. Piskiewicz nécessite que l'utilisateur dispose des données RPG niveau 4 pour l'utiliser. Les données sont gérées par département et par année. **RPG Explorer⁸** est une **boîte à outils** qui permet une série de traitements sur le RPG, qui se font directement à partir des données brutes, selon les paramètres définis par l'utilisateur (années et territoires considérés, choix méthodologiques de traitement). L'approche développée dans RPG Explorer pour la filiation des îlots et la reconstitution des séquences est décrite en détail dans l'annexe 2. Nous ne donnons ci-dessous que les éléments principaux pour la comparaison avec la méthode T-AGIR.

Lors de l'importation des données RPG dans la base locale, un premier traitement élimine les îlots ayant une forme géométrique non valide (auto-intersection, etc.)⁹. RPG Explorer va ensuite identifier les recouvrements importants entre îlots (>10% de la surface d'un îlot) d'une même couche annuelle et « raboter » de manière raisonnée les îlots pour supprimer ces recouvrements. On s'appuie pour cela sur les surfaces géométriques des îlots confrontées aux surfaces déclarées. Si on reprend l'exemple de la Figure 2, ce nettoyage aboutit à une élimination des îlots 01098782 et 1130480 et au maintien du plus grand (1107082) et de l'îlot 1098780 dont la surface d'intersection est faible avec ce dernier.

Pour la filiation des îlots il faut **définir l'espace** sur lequel on souhaite travailler. Cet espace peut correspondre à un département, un ensemble de départements ou une portion de un ou plusieurs départements (cas d'une Aire d'Alimentation de Captage sur plusieurs départements). Une fois cette sélection opérée, on effectue une **intersection géométrique** entre les îlots des différentes années visées pour définir les **filiations d'îlots** (voir la Figure 3 pour une description de la procédure). L'utilisateur choisit une **année de référence** et un sens temporel pour réaliser la filiation des îlots (**chronologique ou antéchronologique**). Rappelons que pour T-AGIR, l'année de référence n est la plus récente, et l'ordre de filiation est antéchronologique. Dans la suite nous prendrons le cas antéchronologique pour faciliter la comparaison avec T-AGIR.

⁸ RPG Explorer est actuellement diffusé à toute personne qui en fait la demande à P. Martin. L'outil est décrit sur le site suivant : https://tice.agroparistech.fr/coursenligne/courses/RPGEXPLORER/?id_session

⁹ Cela peut constituer une source de perte de donnée même si ne sont concernés qu'une dizaine d'îlots par département.

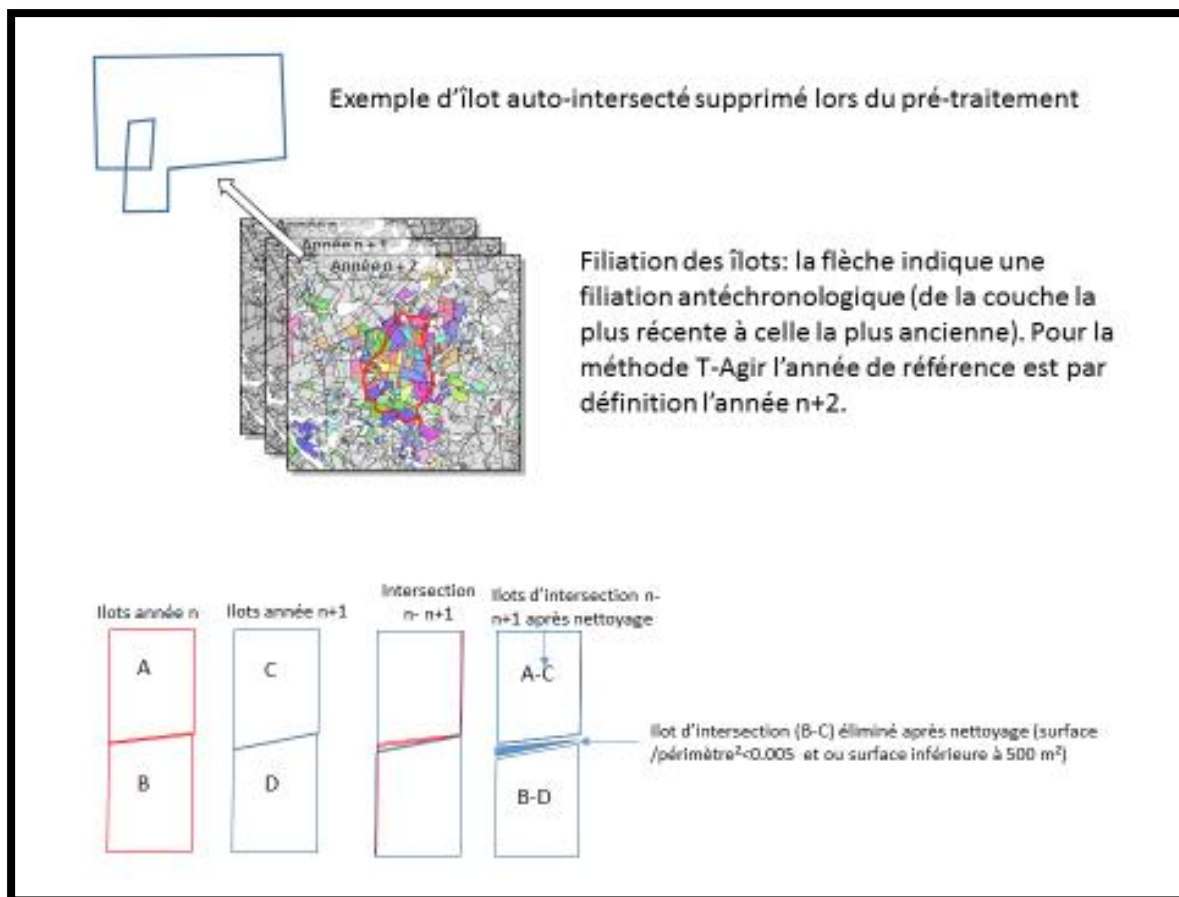


Figure 3 : logique d'intersection des îlots entre plusieurs couches annuelles dans l'approche RPG-E

Pour la filiation antéchronologique, RPG-E réalise **l'intersection géométrique des îlots de l'année n avec l'année précédente n-1**. D'une année à l'autre le tracé d'un îlot stable peut varier à la marge du fait d'une légère modification effectuée par l'agriculteur déclarant. Il en résulte que l'intersection va générer des « **micro-polygones** » qui ne correspondent à aucune réalité physique et que RPG Explorer doit **éliminer** avant de chercher à établir les séquences de groupes culture. L'élimination de ces micro-polygones s'effectue sur la base de deux critères : la **surface minimum du polygone** et son **indice de forme** (surface/périmètre²). Le polygone est éliminé si sa surface et/ou son indice de forme sont inférieurs aux valeurs de référence. Ces valeurs de références sont **paramétrables** par l'utilisateur sachant que des valeurs par défaut sont proposées - 500 m² pour la surface et 0,005 pour l'indice de forme (ce qui équivaut pour une surface de plus de 500 m² à éliminer toute forme rectangulaire de plus de 155 m de long et moins de 3 m de large). On élimine aussi de cette couche les éventuels doublons en bordure de département et on effectue plusieurs corrections géométriques (voir annexe 3 pour les détails). Cette couche nettoyée est ensuite intersectée avec les îlots de l'année n-2, la couche obtenue est nettoyée, puis croisée avec n-3, etc. On obtient ainsi à la fin une **couche d'îlots d'intersection** intégrant l'ensemble des années, avec pour chaque îlot d'intersection les id_îlot de chaque année, une surface, et une information de qualification de la filiation « **qualif_fil** » qui décrit **l'évolution annuelle de l'entité** îlot (stable, fusion, division, apparition, absence, disparition). La Figure 4 illustre la logique suivie par RPG-E pour établir une filiation entre les îlots de plusieurs années, en reprenant la même configuration-type que pour la présentation de la méthode T-AGIR (Figure 1).

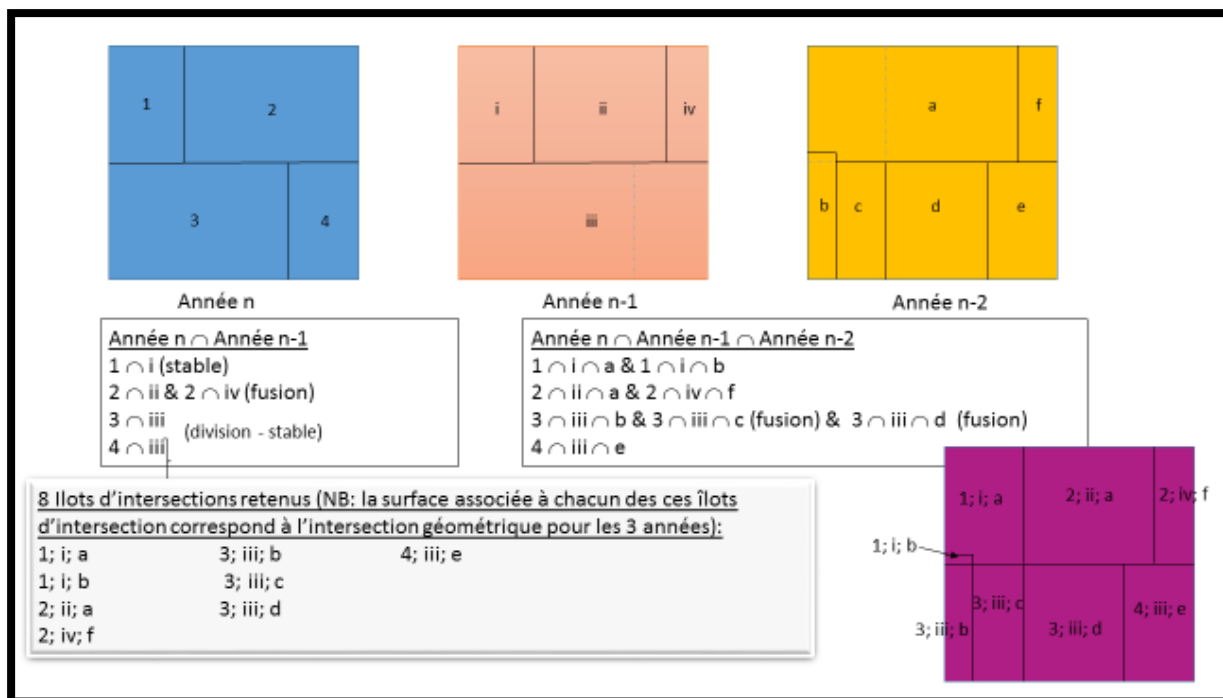


Figure 4 : logique de filiation interannuelle des îlots pour la méthode RPG-E

La base de travail pour la reconstitution des séquences de cultures par RPG-E est donc une couche de **polygones issus de l'intersection physique des couches îlots de chaque année**. Pour RPG-E, le passage par l'intersection a été rendu nécessaire du fait de l'intérêt porté également aux dynamiques de parcellaires d'exploitation et donc du besoin de disposer des liens entre îlots de 2 années successives (et pas uniquement entre l'année de référence et chacune des années de la succession). Cela requiert de connaître le nombre d'îlots et leur surface pour chaque année, ainsi que la façon dont ils se recoupent d'une année sur l'autre.

L'étape d'intersection permet donc d'établir des **îlots d'intersection**, et c'est sur cette base que **l'algorithme de reconstitution des séquences** est appliqué. Les séquences de groupes de cultures sont établies sur des **règles proches** de celles de l'approche T-AGIR, tout au moins pour les filtres 1 à 8. Toutefois deux filtres complémentaires ont été développés (9 et 10) et des ajustements ont été réalisés pour éviter de comptabiliser plusieurs fois les surfaces d'un même îlot. D'autre part, la règle « agdg_sim » de T-AGIR n'a pas d'équivalent dans RPG-E. Les résultats entre les deux approches devraient donc être sensiblement différents.

Tableau 3 : règles mobilisées successivement pour la reconstitution des séquences de groupes de culture par RPG-E

Filtre qualificatif de la séquence RPG-E	Courte description	Equivalent (exact ou approché) dans T-AGIR
1	une culture par ilot et par an	1 : une_culture_annee_1
2	surface égale entre les années des n gc	2 : surf_eq
3	agrégation/désagrégation à surface égale	3 : agdg_eq_1
4	surface similaire à X %	4 : surf_sim
5	agrégation/désagrégation à surface égale sur les restes après filtre 4	5 : agdg_eq_2
6	agrégation/désagrégation à surface égale à X%	6 : agdg_sim_0
7	une culture par ilot et par an (mais pas égale; après prise en compte des surfaces égales)	8 : une_culture_annee_2
9	identification des cultures pérennes (les cultures pérennes sont associés, même si les critères de similarité des surfaces d'une année à l'autre ne sont pas	/

	respectées. La surface attribuée est la surface annuelle minimale du groupe de culture correspondant)	
10	règles des surfaces majoritaires (si un îlot comporte une année une part majoritaire d'un groupe de cultures A (x ha) et une part minoritaire d'un autre B (y ha) et que l'année suivante on a un groupe de cultures C de z ha ($z > y$) alors on peut reconstituer a minima qu'on a une séquence A-C de surface = $z - y$ ha	/
8	non reconnue	
/	/	7 : agdg_sim

Pour **reconstituer les séquences sur un ensemble d'années**, RPG-E applique **une approche par couple d'année**, c'est-à-dire que l'outil établit **pour chaque paire d'année successives** (2014-13 ; 2013-12 ; 2012-11 ; etc.) **l'ensemble de combinatoires possibles** entre groupes de cultures sur la base de leur surface, en partant du filtre le plus probable/sûr au moins certain. Cela permet de réduire les combinatoires possibles et donc le temps et la puissance de calcul nécessaires pour l'analyse, par rapport à une approche qui prendrait en compte directement les combinatoires pour toutes les années. Ensuite, les données issues **des séquences sur 2 ans sont assemblées ensemble selon leur correspondance**, sur la base des mêmes filtres de reconstitution. Dans les cas où une règle détermine une séquence et que la surface n'est pas rigoureusement la même pour toutes les années, RPG-E retient comme surface de séquence identifiée la plus petite surface des différentes années, là où T-AGIR retenait la surface de l'année de référence. Cette règle des minimums s'applique sur chaque séquence mais aussi sur le total des séquences d'un îlot. C'est ainsi que l'on va calculer la surface minimum des déclarations faites chaque année pour l'îlot d'intersection considéré. Si une séquence unique permet d'atteindre cette surface alors RPG Explorer ne recherchera pas d'autres séquences même s'il reste des surfaces de groupes cultures certaines années. Notons que RPG Explorer partage avec T-AGIR le fait de refuser de choisir entre plusieurs séquences équiprobables au sein d'une même filiation d'îlots. La séquence n'est alors pas reconnue (mis dans le filtre 8).

Dans la version de RPG Explorer utilisée pour les premiers tests, quand un îlot manque alors l'outil considère qu'il ne peut pas générer de séquence (alors que cela serait possible sur la période entre l'année de référence et l'année avant l'absence de l'îlot). Dans la version de RPG Explorer utilisée au début du processus de comparaison des méthodes on avait aussi le même problème que pour T-AGIR, à savoir qu'on pouvait mobiliser plusieurs fois les mêmes surfaces dans la détermination d'une séquence (i.e. : cas d'un îlot unique en année n-1 segmenté en année n avec une surface d'un groupe de cultures identique dans l'îlot unique n-1 et dans chacun des îlots segmentés en année n. La même surface de l'îlot n-1 est alors mobilisée dans deux séquences différentes). Ce problème a été résolu en associant comme données attributaire aux îlots de segmentation l'information sur la surface des groupes de cultures qu'ils contiennent. RPG Explorer sait ainsi si plusieurs îlots n issus d'un même îlot n-1 affichent des groupes de cultures avec la même surface. En ce cas il s'interdira de reconstituer des séquences ne sachant pas quel groupe de cultures de l'année n associer au groupe de cultures de même surface de l'année n-1. La version de RPG Explorer utilisée dans la procédure numérique de comparaison des approches (voir plus loin) intégrait cette amélioration technique.

Si on reprend l'exemple de l'îlot 018-1090845 (Tableau 1), RPG-E reconstitue une seule filiation d'îlot, c'est-à-dire un seul îlot d'intersection couvrant les 2,6 ha de l'îlot de 2014, avec aucun îlot identifié pour 2009. En termes de séquences, on retrouve les 3 mêmes séquences (mêmes groupes de cultures et mêmes surfaces) que celles reconstituées par T-AGIR. A noter que ce résultat diffère de ce qui avait été trouvé avec la version de RPG-E utilisée lors des premiers tests, qui n'identifiait aucune séquence du fait de l'absence d'information pour l'année 2009.

2.3 Premières différences identifiées entre les approches T-AGIR et RPG-E

Comme mentionné précédemment, RPG-E et T-AGIR diffèrent dans leur approche de la filiation interannuelle des îlots qui sert de base à la reconstitution des séquences. La Figure 5 illustre un décalage possible dans le traitement de l'information du RPG.

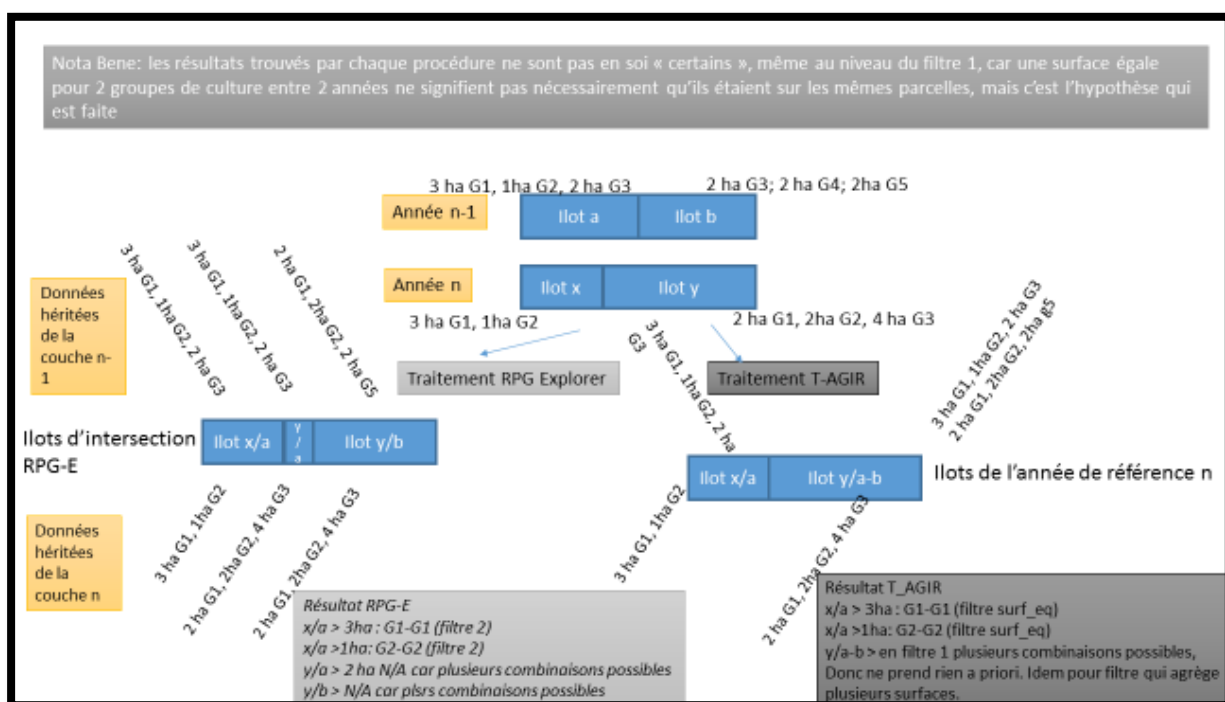


Figure 5 : illustration des différentes étapes de reconstitution selon les approches T-AGIR et RPG-E

Dans l'approche RPG-E, 3 îlots d'intersection sont identifiés pour chaque combinaison entre la couche n (année de référence) et n-1. Chacun se voit attribuer l'information complète des 2 îlots auxquels ils correspondent, mais leur surface totale limite est bien celle de l'intersection. Côté T-AGIR, l'information est agrégée pour chaque îlot de l'année de référence : on a donc 2 blocs, dont un qui correspond à la filiation entre un îlot (y) de l'année n et deux îlots de l'année n-1 (a et b). Dans ce cas les informations attributaires de tous les îlots sont intégralement affecté à cet îlot y. C'est l'année la plus récente (n) qui « commande » le découpage et l'attribution des groupes de culture. La dimension spatiale devient en quelque sorte implicite.

Si au final on a **des différences sur les méthodes d'intersections entre couche annuelle et sur la manière de combiner les informations issues des différentes années considérées**, cela peut avoir un effet déterminant sur le résultat final (pas les mêmes séquences reconstituées pour un même îlot de référence), ou seulement sur les étapes intermédiaires comme dans la figure 2 ci-dessus (mêmes séquences reconstituées, même si l'intersection des îlots donne des associations différentes entre îlots et attributs en surface de groupe de culture des îlots). L'annexe 4 présente 3 cas différents où on a des biais possibles liés à la superposition d'îlots une même année et/ou à des phénomènes de fusion/division au cours du temps. Ces cas permettent de faire ressortir les différences en termes de logique d'intersection et de filiation des îlots d'une année à l'autre. RPG-E réalise un nettoyage des îlots superposés, et crée des îlots d'intersection correspondant au croisement de l'ensemble des années. Côté T-AGIR, l'unité de référence reste les îlots de l'année la plus récente. On remarque que cela peut ne pas avoir d'effets (cas 1 dans l'annexe 4) sur les séquences reconstituées, du fait des règles de reconstitution similaires (et qui limitent les biais potentiels identifiés plus haut). Dans certains cas, on note par contre des différences qui peuvent être liées à une surface de référence réduite par la construction d'îlot d'intersection par RPG-E, et qui limite la reconstitution de surfaces de groupes de cultures supérieures (cas 2 et 3). Néanmoins cela permet de limiter les incohérences (par exemple plusieurs îlots superposés – cf. cas 3 dans l'annexe 4).

Des différences résident aussi dans la façon dont est réalisée la reconstitution proprement dite. Dans les deux cas, les filtres sont appliqués successivement, dans l'ordre des Tableau 2 et Tableau 3. La comparaison entre les Tableau 2 et Tableau 3 montre que les filtres 1 à 6 sont conceptuellement les mêmes pour les deux approches (le paramétrage par défaut dans RPG Explorer pour le filtre 6 est $X=5\%$). **Toutefois T-AGIR applique un filtre supplémentaire (7 : agdg_sim) avant d'appliquer le 8** qui est équivalent au filtre 7 de RPG Explorer. Ce filtre 7 vise à éliminer des reconstitutions des petites surfaces de groupes de culture (inférieures à 0.3 ha), réduisant ainsi potentiellement le nombre de séquences à identifier et accélérant le processus de traitement. La mise en place de ce filtre supplémentaire peut avoir deux effets opposés. Un premier effet est l'identification de séquences supplémentaires par rapport à ce que fait RPG-E. Mais cet effet est contre balancé globalement par le fait qu'en supprimant des surfaces on réduit le nombre de combinaisons et donc de séquences

potentielles. Pour RPG-E on a aussi deux filtres absents de l'approche T-AGIR (9 et 10) qui injectent de l'expertise agronomique dans le raisonnement et peuvent conduire à des séquences supplémentaires. Par ailleurs, RPG-E identifie explicitement les situations où les séquences n'ont pas été reconnues (filtre 8). A noter que pour RPG Explorer, l'ordre de mise en œuvre des filtres est bien 9, 10 puis 8. Cette inversion apparente vient du fait que les filtres 9 et 10 permettant la reconnaissance de séquences initialement non repérées ont été créés après les 8 autres filtres. Tous les filtres sont donc mis en œuvre avant le 8 qui récupère toutes les surfaces d'îlots pour l'année de référence n'ayant pas pu être associées à des séquences.

En plus de la différence entre certaines règles, RPG-E part d'un ensemble de séquences par paire d'année avant de reconstituer une séquence complète. Au contraire, T-AGIR vise directement une reconstitution de l'ensemble des années (9 années par exemple pour 2006-2014). Pour mieux comprendre les implications de ces différences, nous avons réalisé une analyse comparée des résultats issus des deux approches.

Il est toutefois difficile, en l'état, d'analyser finement les différences qui se jouent entre RPG-E et T-AGIR, car comme illustré par les différents exemples, les configurations sont nombreuses et requièrent une analyse fine et complexe au cas par cas. Cette analyse serait intéressante à mettre en œuvre, mais très coûteuse en temps et pas forcément conclusif. Nous avons donc choisi de réaliser une analyse comparée plus globale des 2 procédures, qui doivent permettre à la fois d'en tester la robustesse et également d'en analyser les principaux points forts et points faibles.

3 Travaux de comparaison et analyse des résultats

Après une première analyse de chaque approche, il ressort en premier lieu des différences liées à la finalité de chaque procédure et au contexte dans lequel elles ont été développées, dont dépendent certains choix techniques et méthodologiques.

- L'approche AGIR vise à produire une base de données exhaustive au niveau France entière, selon des paramètres fixes, définis par les concepteurs en fonction des contraintes de calcul et de la vraisemblance des choix et des résultats produits. Cette base est destinée à être diffusée et mobilisée dans différents contextes, avec une focale sur les travaux de recherche sur des grands territoires (France entière, région, etc.) pour lesquels elle doit fournir une bonne représentation des principales séquences de culture. Certains traitements doivent être simplifiés pour limiter le temps de traitement sur l'étendue France. Pour la filiation des îlots, une intersection est considérée comme valide si elle représente au moins 10% de la surface du plus petit des 2 îlots intersectés. Les données attributaires sur les surfaces de chaque groupe de culture sont ensuite attribuées à chaque intersection retenue, sans autre suivi géométrique et en se calant sur les surfaces de l'année de référence (la plus récente). L'idée générale est d'optimiser la quantité de surface de séquences reconstituées calées sur les surfaces de l'année de référence. Cela peut conduire à mobiliser les mêmes surfaces de groupes de cultures dans plusieurs séquences de groupes de cultures, ou d'autres approximations qui rendent l'information moins précise à un niveau géographique plus fin (à l'exploitation ou à la commune par exemple).
- RPG Explorer privilégie une approche paramétrable par l'utilisateur, généralement plus locale, qui repose sur les données apportées par l'utilisateur et les paramètres qu'il définit en cours de traitement (par exemple pour le taux de recouvrement dans l'intersection des îlots, les années de référence pour établir les séquences, etc.). La reconstitution des séquences passe par des îlots d'intersection, qui correspondent à un croisement géométrique de l'ensemble des couches géométriques annuelles. La couche des îlots d'intersection fait l'objet d'un « nettoyage » avant détermination des séquences qui peut conduire à supprimer des situations d'intersections selon un protocole (surface minimum et indice de forme) différent de celui mis en place pour T-AGIR (moins de 10% de recouvrement entre îlots). Les surfaces de séquences retenues sont les surfaces minimales sur la durée de la séquence, là où T-AGIR considère les surfaces de l'année la plus récente. Les séquences ne sont pas construites si l'îlot manque complètement pour une année (dans la version de base). Enfin, RPG Explorer présente 2 filtres de reconnaissance de séquences supplémentaires par rapport à T-AGIR.

Au-delà de cette approche conceptuelle, nous avons souhaité évaluer la proximité des résultats obtenus par les deux méthodes. Les divergences existantes entre les 2 approches doivent être rendues transparentes, pour le cas échéant permettre d'améliorer une méthode en intégrant un protocole intéressant développé dans l'autre mais aussi pour pouvoir choisir quelle approche privilégier selon la situation, quand les évolutions méthodologiques ne sont pas possibles. Pour cela, un travail d'analyse comparée des résultats a été effectué à partir d'extractions de séquences reconstituées selon chaque procédure.

3.1 Premiers résultats sur la comparaison entre les 2 approches

3.1.1 Présentation de la démarche

Afin de comparer les 2 approches, une analyse des séquences reconstituées à partir des 2 approches a été réalisée. Le département du Cher a été sélectionné pour les tests, du fait de la présence significative de systèmes de grandes cultures (et donc l'intérêt et la pertinence d'une approche par séquences de culture) mais aussi d'autres systèmes (viticulture, élevages). Pour chaque département, les données utilisées sont celles des séquences établies par T-AGIR d'une part et par RPG-E d'autre part de 2006 à 2014 (9 campagnes agricoles). Les données primaires du RPG ont également été utilisées (version RPG niveau 4 anonyme). Le traitement des données a été fait dans un environnement PostgreSQL. Les requêtes utilisées peuvent être consultées sur demande aux auteurs.

L'analyse vise dans un premier temps à comparer les données sur les séquences par le département du Cher (18) sur plusieurs points :

- **Nombre d'îlots avec des séquences reconstituées et surface correspondantes.** Comparaison avec les données brute du RPG annuel de 2014. La suppression de tout ou partie des îlots, en particulier dans l'étape d'intersection des couches par RPG-E, peut réduire le nombre d'îlots et leur surface.
- **Nombre de séquences reconstituées différentes** totalement ou partiellement par îlot ; et la nature des séquences reconstituées par îlot et surface correspondantes. Similarité entre les 2 approches (nombre et surfaces des séquences complètes ou partielles correspondant parfaitement) et écart (séquences non reconstituées, séquences partiellement reconstituées, surface attribuée à la séquence, filtre mobilisée pour la reconstitution, etc.). Les différences en termes de méthode d'intersection et de reconstitution des séquences (filtres appliquées et leur mise en œuvre) est source de divergence potentielle. On peut notamment supposer que le caractère plus contraint des algorithmes de RPG-E conduit à une moindre surface de séquences reconstituées.

A noter que nous avons réalisé les premières analyses à partir de données issue d'une version de RPG Explorer telle que disponible début 2018 (version 1.8.148). Les résultats obtenus ont fait apparaître une spécificité dans l'algorithme, qui ne reconstituait aucune séquence si pour une année l'îlot était absent et donc non intersecté. Il est pourtant possible de reconstituer dans ce cas une séquence partielle. Ce point a été corrigé et le travail de comparaison a été repris avec une nouvelle extraction de séquences par RPG-E (version 1.9.5). Nous avons choisi de montrer dans certains cas les résultats des 2 comparaisons, pour indiquer le gain de performance ainsi obtenu. Les résultats issus de la version initiale de RPG-E seront identifiée par le suffixe « OLD », ou indiqué explicitement comme tel dans les tables et/ou le texte.

3.1.2 Résultats : approche globale

Nous présentons ici les principaux résultats analysés à partir des données du département 18 (Cher). Le Tableau 4 permet d'analyser le contenu global de chaque jeux de données (RPG brut, séquences T-AGIR et séquences RPG-E). A noter que pour T-AGIR, la procédure est initiée à partir des données brutes, et donc des 50 779 îlots du RPG de 2014. Parmi cet ensemble, T-AGIR reconstitue des données sur 49 256 îlots, pour un total de 378 223 ha de séquences reconstituées. Les autres îlots (sans séquence reconstituée) ne sont pas inclus dans l'extraction de données. Dans le cas de RPG-E, la reconstitution des séquences se fait sur la base d'une couche géographique retraitée en amont de la procédure et lors de la génération des îlots d'intersection, ce qui explique que l'ensemble des îlots « 018 » des données brutes ne sont pas inclus¹⁰. Certains sont hors emprise sélectionnées et/ou ont été supprimés suite aux différentes procédures de corrections des géométries (taille inférieure à 500 m² pour la plupart des cas, indice de forme, etc.)... Ainsi, 185 îlots (122 ha) ne sont pas pris en compte dans les données extraites à partir de RPG-E, et donc les résultats pour ces îlots ne sont pas comparables. C'est donc à partir de cette **sélection de 50 594 îlots, base commune** aux deux extractions issues des procédures RPG-E et T-AGIR, que nous avons développé nos analyses comparatives.

Tableau 4 : statistiques globales des résultats des procédures T-AGIR et RPG-E

Source mobilisée	nb lignes (= nb de séquences reconstituées)	nb d'îlots 2014 distincts concernés	Somme des surfaces concernées (ha)
RPG 2014 – données brutes – ts îlots dpt 018	/	50 779	438 400
RPG 2014 – données brutes hors îlots non inclus dans extrait RPG-E	/	50 594	438 278
<u>T-AGIR – séquences reconstituées</u>	<u>81 878</u>	<u>49 256</u>	<u>378 223</u>
RPG-E – total îlot2014 OLD	112 748	50 594	437 380

¹⁰ A noter également que par construction, RPG-E produit 8543 lignes qui correspondent à des îlots d'intersection sans îlot en 2014 (îlots d'intersection qui croisent des îlots des autres années mais aucun de 2014) et donc sans séquence reconstituée (cela représente au total 9 340 ha). Ces lignes ne sont pas prises en compte dans le tableau 4 ni dans les analyses suivantes, car c'est bien la couche des îlots 2014 qui sert de référence.

dt RPG-E séquences reconstituées OLD	96 526	45 464	353 387
dt RPG-E - surfaces îlots d'intersection sans séquence reconstituée OLD	16 222	12 835	83 994
RPG-E – total ilot2014	125 735	50 594	438 098
dt RPG-E séquences reconstituées	116 918	49 378	390 486
dt RPG-E - surfaces îlots d'intersection sans séquence reconstituée	8 817	6 368	47 612

On constate en premier lieu que les modifications apportées ont eu un effet significatif sur les performances de RPG-E (Tableau 4). Ainsi, avec la nouvelle version, RPG-E reconstitue un nombre de séquences plus important que T-AGIR, avec légèrement plus d'îlots différents concernés et une surface totale de séquences supérieures de 12 000 ha environ par rapport à T-AGIR. Comme une séquence est attribuée à un îlot d'intersection dans RPG-E vs. un îlot de référence 2014 pour T-AGIR, on peut avoir plusieurs fois la même séquence pour un îlot 2014 avec RPG-E (car plusieurs îlots d'intersection pour cet îlot avec la même séquence), ce qui n'est possible avec T-AGIR. Cette différence n'explique toutefois par l'écart, car si on la prend en compte, on aboutit à 115 900 séquences uniques par îlot 2014 avec RPG-E, ce qui reste significativement différent de T-AGIR.

Par construction, les îlots de la couche d'intersection utilisée pour RPG-E sont en surface (et en nombre) inférieurs à ce que l'on a avec la méthode T-AGIR qui travaille directement sur la couche de l'année de référence la plus récente. L'établissement de ces îlots d'intersection (qui sont 58 148 au total, dont 52 322 ont au moins une séquence reconstituée) peut en effet conduire à l'élimination fortuite de petits îlots s'ils rentrent dans le cadre des règles de « nettoyage » de la couche d'intersection si les paramètres du nettoyage ont été mal ajustés. Dans l'exemple pris ici du département 18, sur le même ensemble d'îlots considérés (donc hors biais lié à un différentiel de sélection des îlots), on perd ainsi **environ 0.2% des surfaces totales d'îlots du fait de l'intersection et 0.3% des îlots**, ce qui reste très faible

Sur les îlots pris en compte par les deux méthodes, RPG-E reconstitue un **nombre de séquences supérieur**, mais souvent avec une surface par séquence inférieure. La méthode T-AGIR aboutit à une diversité moindre de séquences reconstituées, mais tend à attribuer aux séquences reconstituées une surface plus importante ; cela explique qu'on ait 30% de nombre de séquences reconstituées en moins dans T-AGIR, mais seulement 3% de surface de séquences reconstituées en moins.

De fait le mode de construction des séquences conduit, avec la règle du minimum de surface pour RPG-E, à des séquences de surfaces plus réduites qu'avec la méthode AGIR, qui prend comme référence l'année 2014 pour attribuer les surfaces, même en cas de correspondances approximatives (filtre 4 – surf_sim par exemple). En ce sens, **pour les cas où on retrouve les mêmes séquences dans les mêmes îlots avec les 2 procédures**, la **comparaison des surfaces attribuées** à ces séquences aboutit aux résultats suivants :

- 75% des cas présentent des surfaces identiques
- Dans 23,5% des cas, la surface attribuée par T-AGIR aux mêmes séquences est supérieure d'au moins 0.01 ha (limite de précision). Cela représente **un cumul de différences de surfaces de 38 256 ha**. Ce décalage s'explique par le fait que T-AGIR ramène les surfaces à la surface totale de l'îlot de référence (ou d'un groupe de cultures de l'îlot de référence). D'autre part, RPG-E produisant plus de séquences différentes pour un même îlot, les surfaces reconstituées sont attribuées à plusieurs séquences légèrement différentes là où T-AGIR n'en identifie qu'une seule globale. Il faut donc être prudent avec ce chiffre, car il inclut une part importante de séquences où la correspondance est partielle (sur les années les plus récentes) et dispatchés sur plusieurs séquences côté RPG-E.
- Dans 1,5% des cas, RPG-E attribue une surface supérieure par rapport à T-AGIR. Cela représente un surplus cumulé de 314 ha. Les quelques cas analysés font apparaître des petites différences liées aux méthodes de filiation (îlots d'intersection vs. îlots de l'année de référence) mais aussi de reconstitution des séquences (reconstitution par couples bisannuels vs. reconstitution directes sur l'ensemble de 9 années la période¹¹). L'existence de micro bug informatique sur l'une ou l'autre des méthodes pour ce cas n'est pas non plus à exclure.

¹¹ Voir l'exemple de l'îlot 2014 018-1117633, non détaillé ici.

Au final, on constate des différences dans les résultats globaux, qui peuvent s'expliquer par les différences observées de plusieurs façons :

- au niveau des filtres/règles de reconstitution (filtre 9 et 10 de RPG-E en particulier) qui permettent de récupérer des séquences ; ou du filtre 7 de T-AGIR qui limite les combinaisons possibles en ne prenant pas en compte les surfaces inférieures à 0.3 ha ;
- par les règles d'attribution de la surface des séquences sur la base de l'année de référence dans T-AGIR ; ou encore la règle du minimum de surface appliqué par RPG-E qui à la fois réduit les surfaces associées à chaque séquences tout en augmentant potentiellement le nombre de séquences car il reste ainsi plus de surfaces résiduelles dans les îlots.
- les différences en termes d'intersection et de filiation des îlots qui sert de base à la reconstitution des séquences.

Les différences fines entre T-AGIR et RPG-E restent difficiles à cerner, analyser et juger, en l'absence de vérité-terrain. La multiplicité des cas différents, et l'ensemble des paramètres de variation possibles rendent complexes le diagnostic. Cette comparaison fait toutefois bien ressortir les grandes spécificités de chaque approche. Notons aussi que l'existence de bugs informatiques ponctuels n'est pas exclue mais leur recherche et leur correction n'a pas pu être menée de manière exhaustive faute de moyens.

Un autre indicateur pour vérifier les procédures consiste à **agréger pour chaque îlot de référence 2014 les surfaces par séquences reconstituées**. Côté T-AGIR, seuls 5 îlots présentent un dépassement (de 0,01 à 0,24 ha max) du total de surfaces des séquences par rapport à la surface déclarée de l'îlot. Par ailleurs, dans les résultats de RPG-E, on identifie 1200 îlots pour lesquels la surface de séquences reconstituées est supérieure de plus de 0,01 à la surface totale de l'îlot en 2014. Ce dépassement reste marginal : il varie de 0.02 à 2,19 ha max par îlot, avec un total de 100 ha environ pour tout le département. Mais si on prend en compte également les surfaces non reconstituées (que seul RPG-E répertorie), le dépassement est de plus forte ampleur et concerne plus de 1800 îlots et un différentiel de 285 ha. Cette situation qui peut paraître anormale correspond à des cas particuliers où un îlot de l'année n va se reconfigurer en année $n+1$ en perdant de la surface d'un côté et en gagnant de l'autre. Il en résulte des îlots d'intersection auxquels on va potentiellement associer des surfaces de groupes cultures plus importantes que la surface des îlots géométriques d'intersection. C'est ainsi que dans la Figure 6, RPG Explorer construit 4 îlots d'intersection (1-A, 1-B, 2-B et 2-C). Pour chacun de ces îlots d'intersection on peut établir facilement des séquences et on va ensuite associer une surface à chacune des séquences selon la règle du minimum. Cette règle ne pose pas de problème pour les îlots d'intersection 1-A et 1-C. En revanche pour 2-B et 2-C cela conduit à afficher des surfaces de séquence de 7 ha à des îlots d'intersection qui n'en font que respectivement 6 ha et 1 ha. Cela tient au fait que les limites des deux îlots d'intersection concernés ne correspondent pas aux limites d'un îlot une année donnée mais aux limites de deux années successives. On embarque donc des surfaces excédentaires. Pour corriger ce problème, l'algorithme de RPG Explorer compare la somme des surfaces des séquences obtenues à la surface géométrique de l'îlot d'intersection concerné. Tant que la différence ne dépasse pas 10% on maintient l'information de surface de séquences trouvée à l'étape précédente. Si la différence dépasse 10% (cas de l'exemple traité) alors de deux choses l'une. Soit on n'a qu'une seule séquence obtenue avec la règle 1 (une seule culture par îlot chaque année) et en ce cas on affecte à la séquence la surface géométrique de l'îlot d'intersection. Soit on a plusieurs séquences ou une séquence déterminée avec une autre règle que la 1 et on dit que les séquences n'ont pas pu être déterminées tout en disant qu'on a la surface géométrique de l'îlot associée à la règle 8 (séquence non déterminée sur l'îlot d'intersection). A noter que si on a plus de 10% de différence, pas de séquence déterminée par la règle 1 mais que la différence entre la surface géométrique et la surface des séquences ne dépasse pas 0.3 ha alors on maintient les séquences trouvées. Les séquences trouvées correspondent à une réalité mais c'est la surface qui est fautive. On a donc estimé que si la différence de surface ne dépassait pas 0.3 ha on pouvait conserver l'information, car elle représente bien une séquence existante même si de surface incertaine.

Il résulte de la procédure décrite ci-dessus deux causes possibles pour les excès de surface constatés. Cela peut correspondre à des îlots où la somme des surfaces de séquences dépasse la surface déclarée en 2014 sans que ce dépassement ne dépasse 10% de la surface géométrique. On peut aussi avoir des cas où le dépassement est supérieur à 10%. On retient alors la surface géométrique de l'îlot d'intersection mais cette surface peut être elle-même supérieure aux surfaces déclarées à l'intérieur de l'îlot.

Pour l'exemple de la Figure 6 il est aussi intéressant de comparer les résultats obtenus entre les deux méthodes. Comme nous l'avons vu, RPG Explorer conduit à 4 séquences (une par îlot d'intersection). De son côté T-AGIR n'en donne que 3. Les deux méthodes sont en accord pour les îlots A et C. En revanche elles divergent pour l'îlot B où T-AGIR conduit à une séquence sans dépassement de surface

mais fausse puisqu'elle relie directement et intégralement les îlots B et 2 du fait des identités de surface déclarées.

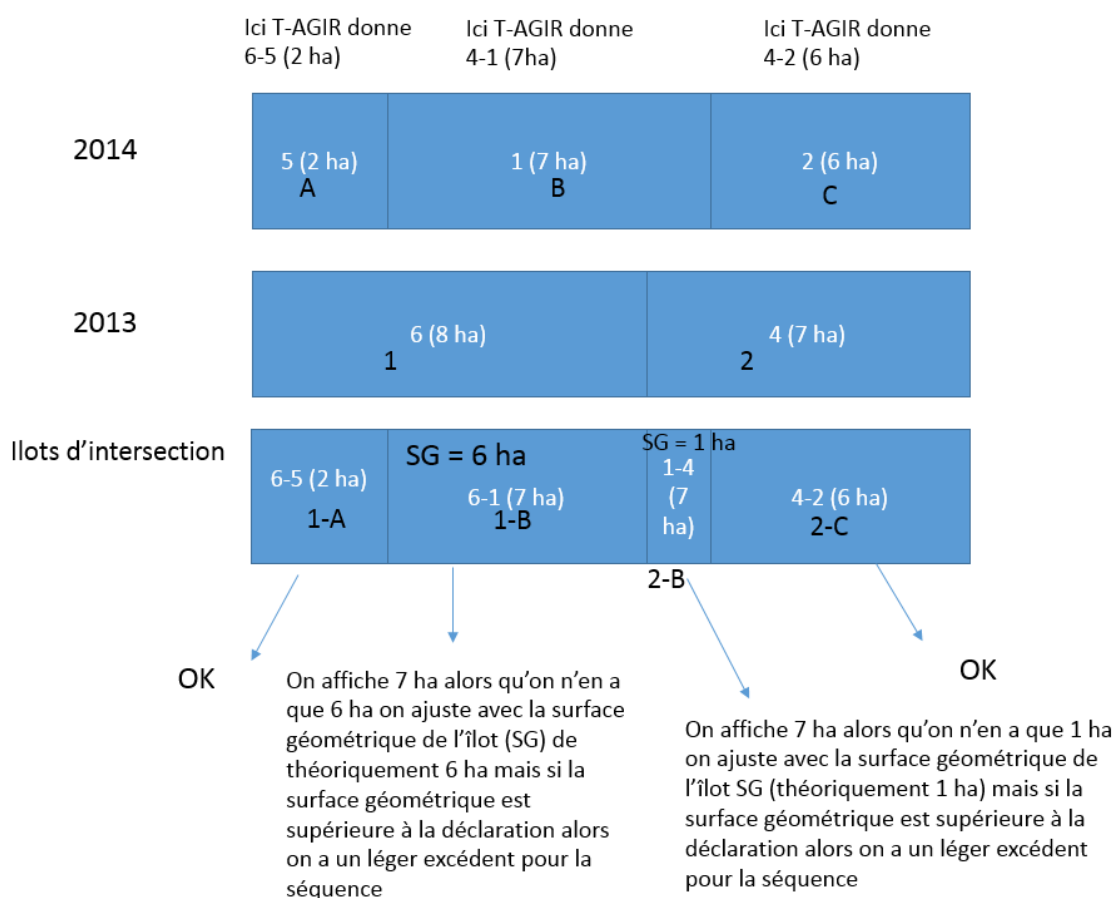


Figure 6: Explication des surfaces excédentaires générées par RPG Explorer

Au niveau des filtres qualifiant le mode de reconstitution des séquences, on ne retrouve pas de correspondance stricte dans les résultats, même pour des filtres a priori similaires dans les 2 approches (par exemple filtre 1 de RPG-E et « une_culture_annee_1 » de T-AGIR). Ceci rend compte notamment des différences déjà mentionnée entre RPG-E et T-AGIR : pour RPG-E, c'est le filtre le plus « complexe » mobilisé pour la séquence qui doit être indiqué dans les données (car le filtre est par couple d'année), alors que pour T-AGIR, le filtre qualifie la règle qui a permis de reconstituer directement l'ensemble de la combinaison (sur les x années de séquences reconstituées). On a vu par ailleurs que le filtre « une_culture_annee_1 » pouvait être appliqué dans des cas où on avait plusieurs cultures dans le même îlot et des durées de séquences identifiées différentes (cf. 2.1). Nous avons aussi pu identifier que dans certains cas RPG-E pouvait afficher des filtre 1 sur des îlots occupés par plusieurs cultures. Il s'agissait là d'un bug en cours de correction. Concernant les filtres 9 (cultures pérennes) et 10 (surface majoritaire) de RPG-E, qui n'ont pas d'équivalents dans T-AGIR, ils permettent de reconstituer respectivement 5 900 ha et 18 300 ha de séquences (soit environ 6 100 séquences différentes au total). Si certaines de ces séquences sont reconstituées également par T-AGIR via d'autres règles (10% des séquences), cela explique en partie les meilleures performances de RPG-E.

Si on regarde les résultats globalement, sur 50 594 îlots distincts dans les données traitées par RPG-E et T-AGIR, on trouve la répartition suivante (entre parenthèse les résultats avec la version ancienne de RPG-E de début 2018) :

- Pour 59,5% (50%) des îlots, on trouve exactement les mêmes séquences
- Pour 36% (38%) des îlots, on trouve des résultats différents
- Pour 1,5% (9%) des îlots, on n'a pas de séquence déterminée par RPG-E et une ou des séquences T-AGIR
- Pour 1,8% (1,6%) des îlots, on n'a pas de séquence T-AGIR et une ou des séquences RPG-E
- Pour 1,2% (1,4%) des îlots, on n'a aucune séquence reconstituée, quelle que soit l'approche

Le chiffre de 36% d'îlots avec des résultats différents est toutefois à relativiser. En effet, en regardant plus finement les données on constate que les résultats diffèrent tout en étant souvent assez proche : différence sur une ou plusieurs cultures de la séquence en début de période (ou des cultures « inconnues » en début de période pour une des approches), divergences du nombre de séquences trouvées (par exemple une séquence majoritaire en termes de surface est commune, mais une des 2 approches reconstitue d'autres séquences plus petites), etc.

Si on compare non pas au niveau des îlots (comme ci-dessus où on regarde si l'ensemble des séquences reconstituées sont exactement les mêmes pour chaque îlot) mais séquence par séquence (c'est-à-dire qu'on regarde si on retrouve les mêmes séquences dans les mêmes îlots, indépendamment les unes des autres), les similarités sont plus fortes que les 59.5% de l'approche par îlot. En effet, si on prend comme référence les séquences reconstituées par T-AGIR, 79% sont également reconstituées par RPG-E sur les mêmes îlots, soit en surface une correspondance de 78%. A l'inverse, si on part des séquences reconstituées par RPG-E, on retrouve ces séquences dans 56% des cas, et pour 66% en surfaces. Cela confirme l'hypothèse que RPG-E retrouve plus de séquences différentes, ces séquences étant avec des surfaces moins importantes. Si on ne regarde que les 4 années les plus récentes (2011-2014), on passe de 56% à 80% des séquences reconstituées par RPG-E qu'on retrouve pour les mêmes îlots avec T-AGIR. L'écart dans les résultats s'explique donc en grande partie par des différences sur une ou quelques cultures en début de période, avec côté RPG-E une plus forte propension à reconstituer des séquences plus nombreuses et plus petites.

3.1.3 Résultats : exemples illustrant les procédures de reconstitution des séquences

Pour comprendre les différences entre les résultats obtenus au niveau de la procédure de reconstitution des séquences, regardons un exemple, celui de l'îlot 2014 018-1090938. Cet îlot est stable, c'est-à-dire associé à un et un seul îlot entre 2006 et 2014, et la filiation donne bien les mêmes résultats pour T-AGIR et RPG-E. On a donc le même potentiel en termes de groupes de cultures successifs à assembler. Le Tableau 5 présente les données disponibles pour les 2 algorithmes.

Tableau 5 : groupes de culture associés à l'îlot 2014 018-1090938

type_culture	surface	annee
gelautre	2.08	2014
autrescereales	3.16	2014
gelautre	2.08	2013
orge	3.16	2013
blet	3.16	2012
gelautre	2.08	2012
colza	3.16	2011
gelautre	2.08	2011
gelautre	2.08	2010
orge	3.16	2010
blet	3.17	2009
gelsansprod	2.08	2009
colza	3.05	2008
gelsansprod	2.2	2008
orge	3.05	2007
gelsansprod	2.2	2007
mais	2.03	2006
gelsansprod	0.17	2006
blet	3.05	2006

Les 2 approches obtiennent des **résultats proches mais différents**. Comme on le voit dans le Tableau 6, malgré la similarité a priori des traitements réalisés (filtre de reconstitution, etc.), la mise en œuvre est différente et aboutit donc à des résultats légèrement différents. Ce cas illustre notamment une approche plus globale par T-AGIR, qui reconstitue moins de séquences différentes, tout en attribuant

l'ensemble de la surface aux séquences identifiées. RPG-E reconstitue plus de séquences différentes, avec des surfaces plus petites par séquences (par exemple ici le maïs de 2006 est reconstitué, alors qu'il n'est pas dans les séquences T-AGIR du fait de l'ambiguïté en 2006). On remarque également une non-correspondance des filtres, déjà vu dans le 2.1 dans la mesure où T-AGIR peut générer des séquences de type « une_culture_annee_1 » alors qu'on a bien de fait plusieurs cultures par îlot par année dans les données initiales utilisées pour la reconstitution.

Tableau 6 : résultat de la reconstitution des séquences pour l'îlot 018-1090938 par les approches T-AGIR et RPG-E

	Séquences	Surfaces	Filtre/règle
T-AGIR	<u>blet orge colza blet orge colza blet orge autrespereales</u>	3.16	surf_sim
T-AGIR	<u>inconnu_gelsansprod_gelsansprod_gelsansprod_gelautre_gelautre_gelautre_gelautre_gelautre</u>	2.08	une_culture_annee_1
RPG-E	<u>blet orge colza blet orge colza blet orge autrespereales</u>	3.04	3 (agrégation/désagrégation à surface égale)
RPG-E	<u>inconnu_gelsansprod_gelsansprod_blet_orge_colza_blet_orge_autrespereales</u>	0.12	3
RPG-E	<u>inconnu_gelsansprod_gelsansprod_gelsansprod_gelautre_gelautre_gelautre_gelautre_gelautre</u>	0.05	3
RPG-E	<u>maïs_gelsansprod_gelsansprod_gelsansprod_gelautre_gelautre_gelautre_gelautre_gelautre</u>	2.03	4 (surface similaire à X %)

Dans les cas où une règle détermine une séquence et que la surface n'est pas rigoureusement la même pour toutes les années, RPG-E retient la surface minimum des 2 surfaces associées. Il s'agit là d'une différence importante avec la méthode T-AGIR qui, dans la même situation va retenir la surface de l'année de référence. Plus on a une séquence de longue durée, plus on a de chances d'avoir des valeurs de surfaces qui diffèrent de celles de l'année de référence tant en plus qu'en moins. Pour une même séquence identifiée, l'approche RPG E conduit donc logiquement à des surfaces de séquences plus réduites que celles obtenues par l'approche AGIR. C'est ce qu'on constate par exemple pour la séquence principale dans l'exemple ci-dessus : 3.16 ha pour T-AGIR contre 3.04 pour RPG Explorer. Comme évoqué plus haut il en résulte qu'on aura aussi plus facilement des surfaces résiduelles pouvant ensuite être intégrées dans de nouvelles séquences absentes des résultats de T-AGIR.

Un exemple illustre un **cas où T-AGIR reconstitue une séquence qui n'est pas dans RPG-E**. L'exemple de l'îlot 2014 018- 1092518 est décrit dans le Tableau 7. Dans ce cas également, on a une situation **stable avec un seul îlot par année** précédente qui intersecte avec l'îlot de 2014, et qui plus est une configuration assez simple avec une culture par an. Néanmoins les deux approches vont aboutir à un **résultat légèrement différent** comme l'indique le T-AGIR reconstitue 2 séquences, en considérant l'écart de surface entre 2006 et les autres années comme une nouvelle séquence de 0.12 ha, qui ne diffère que pour la première culture (inconnu vs. prairie permanente). En revanche côté RPG-E, une seule séquence est retenue, avec la surface minimale pour laquelle l'information est valide (0,88 ha en l'occurrence). Ce résultat non intuitif s'explique à nouveau par la règle du minimum (cf 2.2). L'îlot est stable et correspond donc à un unique îlot d'intersection pour RPG Explorer. On va alors calculer la surface minimale sur l'ensemble des années. Dans l'exemple traité cette surface minimale est de 0.88 ha. On va alors rechercher les séquences qui entrent dans le cadre de cette surface de 0.88 ha. On trouve une première séquence d'une durée de 9 ans qui correspond à 0.88 ha ce qui fait que le processus de reconnaissance s'arrête. Le fait que RPG Explorer ne reconnaisse pas la deuxième séquence minoritaire tient à une incohérence de déclaration qui fait qu'un même îlot donne lieu à des déclarations de surfaces différentes d'une année à l'autre. A noter que si la surface de 0.88 ha avait été déclarée en 2014 et pas en 2006 la méthode T-AGIR aurait conduit aussi à une seule séquence de 0.88 ha sur l'ensemble de la période.

Tableau 8.

Tableau 7 : groupes de culture associés à l'îlot 2014 018-1092518

type_culture	surface	annee
prairiep	1	2014
prairiet	1	2013

prairiet	1		2012
prairiep	1		2011
prairiet	1		2010
prairiet	1		2009
prairiep	1		2008
prairiep	1		2007
prairiep	0.88		2006

T-AGIR reconstitue 2 séquences, en considérant l'écart de surface entre 2006 et les autres années comme une nouvelle séquence de 0.12 ha, qui ne diffère que pour la première culture (inconnu vs. prairie permanente). En revanche côté RPG-E, une seule séquence est retenue, avec la surface minimale pour laquelle l'information est valide (0,88 ha en l'occurrence). Ce résultat non intuitif s'explique à nouveau par la règle du minimum (cf 2.2). L'îlot est stable et correspond donc à un unique îlot d'intersection pour RPG Explorer. On va alors calculer la surface minimale sur l'ensemble des années. Dans l'exemple traité cette surface minimale est de 0.88 ha. On va alors rechercher les séquences qui entrent dans le cadre de cette surface de 0.88 ha. On trouve une première séquence d'une durée de 9 ans qui correspond à 0.88 ha ce qui fait que le processus de reconnaissance s'arrête. Le fait que RPG Explorer ne reconnaisse pas la deuxième séquence minoritaire tient à une incohérence de déclaration qui fait qu'un même îlot donne lieu à des déclarations de surfaces différentes d'une année à l'autre. A noter que si la surface de 0.88 ha avait été déclarée en 2014 et pas en 2006 la méthode T-AGIR aurait conduit aussi à une seule séquence de 0.88 ha sur l'ensemble de la période.

Tableau 8 : résultat de la reconstitution des séquences pour l'îlot 018-1092518 par les approches T-AGIR et RPG-E

	Séquences	Surfaces	Filtre/règle
T-AGIR	<u>prairiep prairiep prairiep prairiet prairiet prairiep prairiet prairiet prairiep prairiep</u>	0.88	une_culture_annee_1
T-AGIR	<u>inconnu prairiep prairiep prairiet prairiet prairiep prairiet prairiet prairiet prairiep</u>	0.12	une_culture_annee_1
RPG-E	<u>prairiep prairiep prairiep prairiet prairiet prairiep prairiet prairiet prairiep prairiep</u>	0.88	1 (une culture par îlot et par an)

Dans certains cas, RPG-E et T-AGIR reconstituent une ou plusieurs séquences pour un îlot, avec les mêmes proportions et les mêmes règles de reconstitution. Mais ils divergent en même sur d'autres séquences pour le même îlot. Par exemple, côté T-AGIR, s'il ne reste qu'un groupe de culture résiduelle par année, une séquence sera automatiquement traitée, avec comme la surface de l'année de référence comme surface maximale (règle une_culture_annee_2). Ce n'est pas systématique avec RPG-E, ce qui pose la question de la mise en œuvre de la règle 7 de RPG-E, qui est censé être identique à une_culture_annee_2 (cf. le cas de l'îlot 018-1095932).

Le **cas des îlots non stables** (c'est-à-dire avec des divisions fusion, apparition, etc. dans le temps) conduit à plus de différences entre approches, du fait des différences dans les procédures d'intersection et filiation interannuelles des couche d'îlots. Dans l'exemple de l'îlot 2014 018-1113788, RPG-E permet d'identifier 4 îlots d'intersection. Il reconstitue une séquence pour chaque îlot d'intersection, dont 2 sont identiques (Tableau 9). On aboutit ainsi à 3 séquences différentes associées à l'îlot 2014 mentionné.

Tableau 9 : résultat du traitement RPG-E pour l'îlot 018-1113788

seq_surf	qualif_fil	idilot2_014	idilot2_013	idilot2_012	idilot2_011	idilot2_010	idilot2_009	idilot2_008	idilot2_007	idilot2_006	seq_06_14
7.62	stable -> division -> stable -> fusion -> stable..	018-1113788	018-1130123	018-1110374	018-27631	018-36088	947711	724269	778200	5360940	18_18_18_18_18_18_18_18
7.44	stable -> division -> disparition	018-1113788	018-1130123	018-1110374	018-27631	018-36088	abs	abs	778200	5360940	0_0_0_0_18_18_18_18_18

	absenc e -> ...										
0.44	absenc e -> absenc e -> absenc e -> apparit ion -> S...	018- 11137 88	018- 11301 23	018- 11103 74	018- 27631	018- 36088	abs	abs	abs	abs	0_0_0_0_18_18_18_1 8_18
0.24	absenc e -> apparit ion -> stable - > fusion -> sta...	018- 11137 88	018- 11301 23	018- 11103 74	018- 27631	018- 36088	94771 1	72426 9	abs	abs	0_0_18_18_18_18_18 _18_18

L'approche de T-AGIR ne prend en compte que l'information sur les groupes de culture associés aux îlots qui s'intersectent (tout ou partie), quelle que soit l'évolution des îlots dans le temps (division, fusion, etc.). L'information de base est donc celle du Tableau 10, à partir de laquelle T-AGIR reconstitue dans ce cas 2 séquences, avec le filtre 1 (une seule culture par an) : 7,62 ha de prairies permanentes successives de 2006 à 2014 (comme RPG-E, cf. première ligne du Tableau 9), et 7,95 ha de « inconnu » de 2006 à 2009 puis prairies permanentes de 2010 à 2014 (contre 7,88 ha de même séquence avec RPG-E, et 0,24 ha avec de la prairie également en 2008 et 2009). Au total, RPG-E reconstitue donc 15,75 ha de séquences associée à l'îlot 2014 018-1113788, ce qui dépasse les 15,57 ha de l'îlot. Pour une première séquence (18_18_18_18_18_18_18_18), la surface est la même entre RPG-E et T-AGIR, alors que pour une deuxième séquence (0_0_0_0_18_18_18_18) la surface diffère légèrement (7,88 vs. 7,95 pour T-AGIR). Ce cas s'explique par la prise en compte par les îlots d'intersection par RPG-E, qui limite dans ce cas la reconstitution de la séquence à la surface de l'îlot d'intersection (7,44 + 0,44), alors que T-AGIR maximise l'utilisation des surfaces de groupes de cultures, en ne prenant en compte que la surface de l'année de référence (2014) comme maximum (15,57 – 7,62 = 7,95 ha). Enfin RPG-E reconstitue seule une 3^{ème} séquence (0_0_18_18_18_18_18_18), sans qu'on puisse vérifier l'origine des groupes de cultures pour 2009 et 2008, qui semblent surnuméraire par rapport à la surface de l'îlot 2014 et des surfaces de groupes de culture identifiées dans l'approche T-AGIR.

Tableau 10 : groupes de culture associés à l'îlot 2014 018-1113788

type_culture	surface	ilots_pc	annee
prairiep	15.57	100	2014
prairiep	15.57	100	2013
prairiep	15.57	100	2012
prairiep	15.57	100	2011
prairiep	15.57	100	2010
prairiep	7.62	100	2009
prairiep	7.62	100	2008
prairiep	15.04	98	2007
prairiep	15.04	98	2006

3.1.4 Résultats : discussion sur les principales divergences entre les 2 approches

Les analyses montrent donc à la fois une forte similarité des résultats entre les procédures T-AGIR et RPG-E, et des différences assez nombreuses mais souvent mineures (décalage de surfaces, différence sur quelques groupes de cultures, etc.). Leur performance en termes de nombre d'îlots, de séquences et de surfaces reconstituées sont proches, RPG-E ayant a priori une meilleure capacité à identifier plus de séquences différentes et donc des meilleurs résultats quantitatifs et qualitatifs. Chaque approche a toutefois ces spécificités, liées notamment au choix de traiter plus ou moins finement les plus petites surfaces, ou au contraire de préférer une approche plus globale mais de fait plus rapide et plus « efficace » en termes de surfaces reconstituées. Les différences dans les procédures de filiation des

ilots interannuelle conduisent à ce que certaines filiations soient retenues d'un côté et pas de l'autre mais par définition ce ne seront pas de grandes surfaces. La deuxième différence majeure tient au fait que T-AGIR travaille pour la détermination des surfaces de séquences sur les surfaces de l'année la plus récente (2014), qui sert de référence, ce qui peut amener à garder des surfaces pour des séquences en plus ; là où RPG-E prend les surfaces minimum sur la séquence (selon la surface de l'îlot d'intersection, ou la surface minimale du groupe de culture inclus dans la séquence, quand les surfaces ne sont pas exactement égales). **Il est probable que la méthode T-AGIR soit plus pertinente pour les premières années (amélioration progressive de la qualité des déclarations) mais que sur des données stabilisées elle masque une partie de la complexité des séquences.** La mobilisation de règles supplémentaires par RPG-E (9-cultures pérennes et 10-surface majoritaire) explique également certaines différences et les meilleures performances de RPG-E. Enfin l'analyse préliminaire des procédures mises en œuvre et les différents exemples analysés mettent en évidence des différences dans la façon de mettre en œuvre la reconstitution des séquences même pour des règles a priori identiques. Cela renvoie à des éléments précis dans la rédaction du code qui permet de réaliser cet étape (langage différent, code élaboré sur la même logique de départ mais indépendamment, approche par paire d'années avant de recombinaison sur l'ensemble des années côté RPG-E vs. des combinaisons directement testées sur l'ensemble des années pour T-AGIR, etc.).

La comparaison pourrait être poussée plus loin, mais c'est une procédure longue et complexe, et difficilement automatisable. En l'absence de vérité-terrain, on ne peut pas vérifier et comparer les performances « réelles » des 2 approches par rapport à ce qui a été effectivement fait dans les parcelles agricoles. Par ailleurs, le temps d'analyse et la complexité qui apparaît au fur et à mesure des analyses nous a conduit à limiter le travail approfondi sur un département. Les spécificités des données du RPG dans d'autres départements (taille des îlots, formes, diversité des cultures rencontrées et de leur distribution, etc.) pourraient être à l'origine de d'autres écarts dans les résultats des approches de T-AGIR et RPG-E.

Ces travaux ne permettent donc pas de conclure sur la pertinence comparée des 2 approches. Toutefois, en révélant et explicitant les logiques à l'œuvre, elle permet d'améliorer les outils existants. C'est déjà ce qui a été fait pour RPG-E, où certaines améliorations ont pu être intégrées : prise en compte des doublons d'îlots, correction d'erreur dans l'algorithme de constitution, etc. Il s'est également avéré intéressant de travailler sur la validation et l'amélioration d'autres sections de RPG-E, notamment sur la procédure de filiation des identifiants des exploitations agricoles dans le temps. Côté T-AGIR, le développement n'est plus pris en charge par l'équipe initiale, ce qui rend difficile l'intégration de nouvelles améliorations. La production et l'utilisation de procédure complexe et évolutive (les données et leur structures évoluant également, comme les besoins) requiert en effet à la fois des étapes de validation et un suivi dans le temps, qu'il est malheureusement souvent difficile d'inclure dans le temps de projets de recherche ou d'expertise de court ou moyen terme. Cela renvoie au défi de construire et donc de financer dans le temps des outils collectifs nécessaires pour exploiter toutes les possibilités offertes par la mise à disposition de grands jeux de données brutes.

3.2 Point sur la validation de la filiation des exploitations à partir des données issues du traitement des données MAE-PAC

Dans la partie précédente nous avons comparé les approches T-AGIR et RPG E quant à l'établissement des séquences de groupes de cultures. En complément RPG Explorer propose d'autres traitements des données dont la reconstitution des évolutions de parcellaires d'exploitation. La collaboration entre SADAPT et ODR a permis de proposer un protocole de test de ce module en s'appuyant sur la base de données sur les données surfaciques liées aux Mesures Agroenvironnementales (MAE) de la PAC gérée par l'ODR¹². Ces données parcellaires, non anonymes, contiennent les parcelles agricoles qui accueillent des engagements MAE ; c'est un extrait du RPG (les parcelles MAE s'inscrivent dans les îlots RPG). Les 2 sources de données, RPG et MAE parcellaires, sont parfaitement superposables et intersectables (issues de la même source : déclaration des agriculteurs sur orthophotos IGN). Nous pouvons donc établir une correspondance entre les identifiants anonymes successifs des exploitations dans le RPG des campagnes successives et le numéro Pacage (stables dans le temps si pas de

¹² Parcellaire MAE : source ASP et MAAF, fournie suivant convention entre l'ASP, le ministère et l'INRA US. ODR pour l'évaluation des programmes communautaires de développement rural 2007-2013 (PDRH).

changement¹³), pour les exploitations qui reçoivent des aides aux titres d'un dispositif MAE surfacique. Les données sont disponibles pour les campagnes 2007 à 2014, mais pas exhaustives (elles couvrent seulement les exploitations ayant des parcelles recevant effectivement une MAE surfacique au cours de la campagne). On a donc au maximum des informations sur 5 ans pour une même parcelle.

L'idée était de constituer des successions de numéros RPG anonymes d'exploitations établis à partir de la réalité du terrain et de confronter cela aux résultats produits par RPG Explorer. L'ODR (P. Cantelaube) a donc construit une table de suivi des exploitations sur la période 2007-2014. A partir des numéros Pacage présents en 2014 (c-a-d les exploitations qui ont des MAE en 2014), l'information est reconstituée pour les campagnes précédentes tant que des MAE sont présentes sur des parcelles de ces exploitations initiales (identifiées via le n. Pacage). Une ligne (=une exploitation) qui est renseignée jusqu'en 2012 uniquement signifie que les engagements MAE de cette exploitation ont débutés en 2012. Il peut y avoir des "trous" : il y a des engagements MAE entre 2007 et 2011 (5 ans), puis de nouveau en 2013. La colonne pour 2012 est alors vide mais 2007 à 2011 et 2013 à 2014 sont remplies. La table est construite en se basant sur le numéro Pacage : s'il y a eu un changement de ce numéro nous perdons le suivi (exactement comme si l'exploitation n'avait pas/plus de MAE). Pour chaque identifiant anonyme d'exploitation RPG, la surface déclarée a été ajoutée ; c'est un contrôle pour repérer des erreurs "flagrantes" (la surface d'une exploitation peut en effet évoluer au fil des campagnes). Enfin pour la livraison finale des données à l'équipe de RPG-E pour la réalisation des tests, le numéro Pacage a été effacé, afin de respecter la confidentialité et les conditions d'utilisation de la base de données sur les MAE. Ne reste donc dans la base que les numéros RPG attribués par l'ASP de manière anonyme.

Le travail mené sur les extractions a permis d'affiner les algorithmes de RPG Explorer pour prendre en compte certaines spécificités du RPG. Une fois ces modifications effectuées, on aboutit à un taux de reconnaissance des trajectoires de parcellaires de plus de 97%. Certaines trajectoires très rares mettent en défaut les règles de reconnaissances de RPG Explorer. D'autres situations restent inexpliquées, mais ce travail a permis d'améliorer significativement et de valider les procédures de traitement des données par RPG-E. Cet élément n'avait pas été prévu au départ dans la collaboration, mais il constitue une avancée significative pour rendre RPG-E plus fonctionnelle et sûre et garantir une information de qualité pour les utilisateurs (collectivité, bureaux d'études, etc.).

Au-delà de l'amélioration des algorithmes de reconstruction des trajectoires de parcellaires d'exploitation, la collaboration SADAPT-ODR a permis d'intégrer à RPG Explorer une base d'information essentielle constituée par l'ODR. Pour en comprendre l'intérêt rappelons que l'utilisation de RPG Explorer se fait sur un lot restreint de départements. L'utilisateur n'est donc pas certain d'avoir l'intégralité des données pour les exploitations qui l'intéressent, celles-ci pouvant avoir des îlots dans d'autres départements. De 2007 à 2009, une exploitation qui a des îlots dans plusieurs départements restait identifiée par un numéro unique anonyme d'exploitation. A compter de 2010 il y a autant de numéros différents que de départements concernés pour la même exploitation. Conscient de cette situation, l'ODR a compilé pour chaque exploitation française les départements dans lesquels elle exploite des îlots. A partir de 2010 l'information est complétée par l'attribution d'un numéro d'exploitation unique en lieu et place des différents numéros issus des différents départements concernés par une même exploitation. Cette base d'information nationale a donc été intégrée dans RPG Explorer ce qui lui permet d'indiquer à l'utilisateur désireux de travailler sur des parcellaires d'exploitation complet, la liste du ou des départements nécessaires pour compléter l'information de telle ou telle exploitation. Cela permet donc en retour à l'utilisateur d'injecter des données dans l'outil en fonction des besoins.

¹³ Par exemple si l'exploitation change de statut juridique, elle va changer de numéro Pacage ; le lien entre parcellaire de l'année n et celui de l'année n + 1 ne sera donc pas possible en ce cas. A noter qu'à partir de cette source, il n'est pas possible a priori de suivre l'évolution d'un îlot et l'enchaînement des séquences, car on ne dispose pas d'information sur les cultures présentes sur les parcelles dans les données MAE.

4 Synthèse des acquis et Perspectives

Dans cette partie, dans un premier temps, nous reprenons de manière synthétique les principaux résultats de la comparaison des algorithmes ainsi que actions correctives mises en œuvre. Dans un deuxième temps nous donnons quelques pistes pour d'éventuels approfondissements.

4.1 Synthèse des acquis

Une différence essentielle entre RPG Explorer et T-AGIR réside dans la **relation aux utilisateurs des données produites**. Pour RPG Explorer on a une approche paramétrable par l'utilisateur qui nécessite de disposer de ses propres données RPG pour faire le traitement alors que pour T-AGIR le traitement est intégralement fait par les concepteurs et le résultat peut être transmis (sous réserve de respect des conditions de diffusion, et avec éventuellement une participation aux coûts engendrés par ce travail) aux utilisateurs sans qu'ils ne disposent des données RPG source. Par ailleurs T-AGIR ne permet de transmettre aux utilisateurs finaux que l'établissement des séquences alors que RPG explorer intègre des modules complémentaires de traitement des données RPG (calcul d'assolement, de dispersion d'îlots, de dynamique de parcellaire d'exploitation...).

A partir de la description des procédures mise en œuvre par RPG-E et T-AGIR et une analyse comparée globale et détaillée des résultats obtenus, nous avons identifié les principaux points communs et divergences entre les 2 approches. Elles s'articulent autour des 2 principales étapes : les méthodes de nettoyage des couches et de filiations des îlots ; et la reconstitution des séquences à partir des îlots affiliés et des groupes de cultures qui leur sont attribués.

4.1.1 Filiation des îlots

La question qui se pose dans les deux approches est la même, à savoir quels îlots et quels liens entre îlots retenir d'une année à l'autre pour établir les « filiations d'îlots ». T-AGIR retient tous les îlots de l'année de référence (la plus récente) et va ensuite rechercher les **correspondances spatiales avec les îlots des années antérieures**, année par année (îlots 2014 avec îlots 2013, puis îlots 2014 avec îlots 201, etc.). Un **critère de 10% de surface minimale intersectée** détermine les cas où la filiation est prise en compte ou pas. Sont au final considérés comme **affiliés tous les îlots qui intersectent à au moins 10% de la surface l'îlot de l'année de référence**.

RPG-E fonctionne différemment dans la mesure **où il n'y a pas d'îlots de référence, chaque année de données ayant le même statut. La filiation des îlots est réalisée par une intersection spatiale** des couches de données des différentes années considérées : l'unité de référence est donc **l'îlot d'intersection**, formé par tout ou partie de l'îlot de référence (le plus récent par exemple) **redécoupé en fonction des intersections successives avec les îlots des autres années**. L'imprécision du RPG fait que les îlots ne se correspondent pas toujours exactement d'une année à l'autre. La représentation d'un îlot A de l'année n peut partiellement recouvrir son voisin B l'année n-1.

Pour les deux méthodes, l'idée est de ne pas retenir les recouvrements trop faibles entre deux îlots de deux années (de type intersection entre îlot A de l'année n et îlot B de l'année n-1) lié uniquement à l'imprécision de la représentation d'un même objet d'une année à l'autre, mais la **manière dont les recouvrements trop faibles sont éliminés de la base diffère fortement entre RPG Explorer et T-AGIR**. Pour T-AGIR les recouvrements non pris en compte sont ceux qui sont inférieurs à 10% de la surface de l'îlot le plus petit des deux années. Pour RPG Explorer, on élimine les îlots d'intersection qui sont inférieurs à une certaine surface et/ou à une valeur du rapport (surface / périmètre²) (valeur par défaut respectivement 500m² de surface min et un rapport de 0.005 qui équivaut pour une surface de plus de 500 m² à éliminer toute forme rectangulaire de plus de 155 m de long et moins de 3 m de large). L'approche de RPG Explorer peut conduire à éliminer des îlots réels de très petite taille ou très allongés. Cette perte reste très faible. Dans l'exemple du département 18 on perd ainsi environ 0.2% des surfaces du fait de la perte de 0.3% des îlots.

Au-delà des biais potentiels propres à chaque procédure, nous n'avons pas testé **l'effet de cette différence procédurale sur les filiations d'îlots obtenues**. Ce test consisterait à comparer les numéros uniques de filiation d'îlots obtenus par chacune des méthodes afin de vérifier quelle filiations d'îlots ont pu être obtenues de manière identique de part et d'autre et quelle différences peuvent être constatées. Cela impliquerait de reconstituer cette filiations dans T-AGIR, qui n'explique pas cette filiation dans la procédure et dans les résultats intermédiaires et définitifs produits. On pourrait ainsi voir si l'une ou l'autre des méthodes est plus ou moins sensible à la sélection de certaines situations.

4.1.2 Etablissement des séquences au sein des filiations d'îlots

Les deux méthodes s'appuient sur le **même principe de base : pour chaque année on identifie l'ensemble des surfaces déclarées pour une filiation d'îlots donnée**. On reconstitue ensuite les séquences en s'appuyant sur des **correspondances de surfaces entre années**. Mais plusieurs **divergences** liées à des façons différentes de mettre en œuvre ce principe commun et à des évolutions postérieures sont ressorties au cours de l'analyse:

- 1- Les deux méthodes avaient en début du travail de comparaison le même défaut à savoir **que si un îlot se fragmentait en deux îlots** et qu'on avait une même surface de groupe de cultures de x ha que l'on retrouvait dans l'îlot d'origine et dans chacun des deux îlots résultants, alors **l'algorithme pouvait mobiliser deux fois les mêmes x ha de l'îlot d'origine** (une fois dans chaque îlot résultant). Cette erreur de type « double-compte » a pu être corrigée dans RPG Explorer en informant chacun des îlots résultant des surfaces présentes sur l'autre ; sachant qu'une surface identique est présente dans deux îlots d'intersection distincts, l'algorithme décide de ne pas générer de séquences sur les x ha car il n'est pas en mesure de l'attribuer à l'un ou l'autre des deux îlots résultants.
- 2- Les deux méthodes se **différencient sur le sens d'établissement des séquences**. Pour **RPG Explorer** ces séquences peuvent être établies indifféremment de manière **chronologique** (de l'année la plus ancienne vers l'année la plus récente) ou de manière **antéchronologique** (de l'année la plus récente vers la plus ancienne). **T-AGIR ne fonctionne que de manière antéchronologique**. Le positionnement de RPG Explorer répond à une volonté de plus grande souplesse pour mieux répondre aux attentes variées des utilisateurs. Celui de T-AGIR correspond à une volonté de standardisation pour diffusion à l'échelon national.
- 3- Les deux méthodes différaient initialement sur les **modalités de rupture des séquences**. La méthode T-AGIR permettait de reconnaître des **séquences tronquées** du fait qu'un îlot de filiation était absent une année donnée. Pour RPG Explorer aucune séquence n'était reconnue tant que tous les îlots de la filiation n'étaient pas présents chaque année. La comparaison a permis d'intégrer à RPG Explorer la même logique que celle de T-AGIR. Les **séquences tronquées par l'absence d'un îlot dans la filiation sont donc maintenant reconnues**, avec des inconnus à partir du moment où l'information est manquante.
- 4- Les deux méthodes mobilisent successivement une **série de filtres permettant d'extraire successivement des séquences de groupes de cultures** en commençant par les plus certaines et en finissant par les plus hypothétiques. Ces filtres ont commencé à être développés par T-AGIR. Le principe en avait été transmis aux développeurs de RPG Explorer mais sans le détail exact des algorithmes utilisés, et donc avec possiblement des différences dans le détail leur mise en œuvre. Par la suite, les développeurs de RPG Explorer ont mis en place **deux filtres absents de la méthode T-AGIR** (filtres dits « cultures pérennes » et « surfaces majoritaires »). Ces différentes combinaisons de filtres aboutissent donc à des **résultats proches mais en partie différents** pour un même pool de surfaces de groupe de culture à combiner. Inversement RPG-E n'intègre pas le filtre 7 qui élimine les surfaces inférieures à 0.3 ha au sein des îlots avant de faire la recherche de séquence, limitant ainsi le nombre de combinaisons possibles pour reconstituer des séquences.
- 5- Les méthodes diffèrent aussi quant à **la manière d'associer une surface à une séquence reconnue dans une filiation d'îlots**. Dans la **méthode T-AGIR, la surface associée est toujours celle de l'année de référence qui est aussi l'année la plus récente** (mode de reconnaissance de séquences antéchronologique cf point 2). Pour **RPG-Explorer, la surface associée à la séquence est la surface minimum déclarée sur la durée de la séquence**. Il en résulte que les **séquences reconnues par RPG Explorer sont associées à des surfaces inférieures ou égales à ce qu'on peut avoir avec T-AGIR**. En effet si la surface minimum déclarée l'est pour l'année la plus récente, les séquences auront la même surface pour T-AGIR et RPG-Explorer. Si cette surface l'est sur l'une ou l'autre des années antérieures, les séquences obtenues avec RPG-Explorer seront inférieures à celles obtenues avec T-AGIR.

Malgré la perte de certains îlots (cf filiation des îlots) et une minimisation des surfaces associées aux séquences (point 5), l'intégration des séquences tronquées par absence d'îlot (point 3) ainsi que la présence de filtres supplémentaires non présents dans la méthode T-AGIR (point 4) permet aux nouvelles versions de RPG Explorer issues du travail de comparaison d'atteindre un plus grand taux

de reconnaissance des séquences sur le cas du département du Cher (18) utilisé pour le test. Les différences persistantes sur les modalités d'élimination du « bruit » lors de la filiation des îlots ainsi que des différences persistantes dans le détail des filtres utilisés dans les deux méthodes pour l'établissement des séquences conduit à un recoupement non intégral dans les résultats. C'est ainsi que pour le département du Cher, sur 50 780 îlots distincts dans les données initiales (2014), on trouve la répartition suivante :

- Pour 59,5% des îlots, on trouve exactement les mêmes séquences
- Pour 36% des îlots, on trouve des résultats différents
- Pour 1,5% des îlots, on n'a pas de séquences déterminées par RPG-E et une ou des séquences T-AGIR
- Pour 1,8% des îlots, on n'a pas de séquences T-AGIR et une ou des séquences RPG-E
- Pour 1,2% des îlots, on n'a aucune séquence reconstituée, quelle que soit l'approche

Les résultats sont toutefois plus proches que ce que suggèrent les chiffres précédents. Si on regarde au niveau des séquences élémentaires reconstituées et non plus au niveau de l'îlot complet, on constate que 79% des séquences reconstituées par T-AGIR le sont également par RPG-E sur les mêmes îlots, soit en surface une correspondance 78%. A l'inverse, si on part des séquences reconstituées par RPG-E, on retrouve ces séquences dans 56% des cas, et pour 66% en surfaces dans les résultats de T-AGIR. Si on ne regarde que les 4 années les plus récentes (2011-2014), on passe de 56% à 80% des séquences reconstituées par RPG-E qu'on retrouve pour les mêmes îlots avec T-AGIR. Cela confirme que l'écart dans les résultats s'explique en grande partie par des différences sur une ou quelques cultures en début de période, avec côté RPG-E une plus forte propension à reconstituer des séquences plus nombreuses (plus détaillées) et de plus faible surface élémentaire.

Les travaux menés dans cette action ont donc permis de :

- Préciser les points communs et différences entre les deux méthodes T-AGIR et RPG Explorer,
- Quantifier et qualifier les écarts quant aux résultats obtenus,
- Corriger un certain nombre d'erreurs dans RPG-Explorer qui en retour ont permis de réduire l'importance des écarts (par exemple la correspondance exacte entre îlots était de 50% en début d'étude contre 59% en fin d'étude, idem pour la correspondance des séquences qui est passée de 71% à 78% des surfaces) : ne pas prendre en compte des surfaces déjà utilisées (gestion des surfaces de culture incluse dans plusieurs filiations), reconstitution des séquences même pour les îlots avec une année manquante, mise en œuvre de la règle 10 des surfaces majoritaires.
- Rapprocher les travaux des deux équipes et mettre en commun des données et des procédures pour améliorer les outils et informations disponibles : base de correspondances des identifiants d'exploitation, trajectoire d'exploitations, etc.

4.2 Perspectives

Le travail relaté dans ce rapport a permis de comparer les approches entre elles d'un point de vue conceptuel et pratique. Toutefois pour étayer plus solidement cette comparaison il serait très utile de disposer de vérités terrain, c'est-à-dire de données validées qui associent aux îlots les séquences à l'intérieur de ces îlots. Malheureusement ces vérités-terrain n'existent pas à l'heure actuelle, en tout cas pas sur une échelle de temps et d'espace correspondante aux données traitées via les RPG de 2006 à 2014 sur l'ensemble du territoire français. Une piste a été explorée par des collègues de l'UMR SADAPT pour comparer les résultats des séquences avec les données issus de Teruti-Lucas, directement issus d'observation de terrain. Toutefois ce travail n'a pas pu être complètement finalisé du fait de certaines incohérences flagrantes entre les deux jeux de données RPG/TERUTI et est donc difficilement mobilisable pour répondre pleinement à nos besoins de validation. En revanche des perspectives très intéressantes s'ouvrent avec l'accès aux nouvelles données. En effet à partir de la campagne 2015 nous disposons des données à la fois à l'échelle parcellaire et à l'échelle de l'îlot. Dès que nous aurons 3 ou 4 années de disponibles (actuellement seules 2015 et 2016 sont disponibles) nous pourrions faire tourner les algorithmes sur les données par parcelle pour constituer une « vérité terrain ». Il sera alors possible de refaire le traitement par îlot, tel que nous le faisons sur les années antérieures, et confronter ces résultats avec les vérités terrain. Ce mode d'approche pourrait notamment permettre de mieux ajuster les paramétrages de RPG Explorer en fonction des départements.

Annexe 1 : Déroulé des travaux en 2017

Pour rapprocher les algorithmes plusieurs réunions d'échanges entre les porteurs des deux démarches ont eu lieu. Ces réunions avaient pour objectif de :

1. Mettre à plat le détail des procédures mises en place dans chaque approche pour prendre en compte les nombreuses erreurs présentes dans le RPG¹⁴.
2. Lancer et discuter le développement d'algorithmes de comparaison et d'analyse des différences des résultats obtenus par les deux approches.
3. Analyser la sélection et l'implémentation des meilleurs choix méthodologiques (volet 2).

Comité de pilotage de l'action ONEMA PACS-AAC – volet 1 – 08/11/2016

Présentation par P. Martin et T. Poméon des travaux prévus de comparaison et de rapprochement entre RPG-E et T-AGIR.

Réunion de lancement – 04/01/2017

Participants :

ODR : P. Cantelaube, T. Poméon, C. Picaud (CDD ODR)

SADAPT : P. Martin

Entreprise Piskiewicz : N. Piskiewicz

Objet : échange sur les traitements mis au point dans les deux équipes afin de mieux cerner les points à traiter en priorité.

Décisions prises :

- Transmission à l'équipe de RPG-E du code et du pseudo-code développé par P. Casel. Partant de là N. Piskiewicz fait des propositions de modifications de RPG Explorer (prise en compte des surfaces majoritaires, etc.) et éventuellement des propositions de modification des procédures AGIR.
- Transmission par l'ODR d'une table sur les identifiants d'exploitation retenus en cas de doublon, pour utiliser le même référentiel dans RPG-E.
- Nécessité d'anticiper les protocoles d'échange de données entre les bases de données hébergées à l'ODR et RPG Explorer : modalités pour un accès aux bases de données hébergées par les serveurs de l'ODR pour RPG-E (pour des données stockées en cache, selon les droits des utilisateurs) ; et application des algorithmes développée dans RPG-E sur les serveurs de l'ODR.
- P. Casel finalise l'établissement des séquences de groupes culture pour la France entière de 2006 à 2014. NB : à la date du 04/01/2017 les séquences n'étaient disponibles que sur 2006-2013

Réunion de suivi #1– 13/03/2017

Participants :

ODR : P. Cantelaube, T. Poméon

SADAPT : P. Martin

Entreprise Piskiewicz : N. Piskiewicz

AGIR : Pierre Casel

Objet : la précédente réunion avait eu lieu en visioconférence sur une ½ journée, et malgré l'échange de documents effectué entre temps comme prévu, un certain nombre de points méritait encore clarification (notamment du fait de l'absence de représentants d'AGIR à la visioconférence). Le groupe a donc décidé de se réunir physiquement le 13/03/2017 à Toulouse. On visait à améliorer l'interconnaissance des logiques mises en œuvre dans RPG-Explorer (RPG-E) et les travaux de l'ODR (rapprochement des exploitations à cheval sur plusieurs dpt, etc.) ainsi que ceux d'AGIR (algo de reconstitution des séquences – T-AGIR) sur le RPG. Définir les actions et moyens à mobiliser pour la suite des travaux.

Décisions prises :

- Transmission de la dernière version du document explicatif de l'approche de construction des séquences mise en œuvre par AGIR.

¹⁴ Rappelons que la mauvaise qualité du RPG a amené un rappel à l'ordre de la France par la commission européenne et la reprise en main du dossier par l'IGN.

- Intégration dans RPG Explorer du travail réalisé par l'ODR sur les identifiants d'exploitation à l'échelon national. Ce travail permet notamment de rapprocher les identifiants anonymes d'exploitations qui bien que différents correspondent de fait à la même exploitation. L'intégration de cette information dans RPG explorer permet à l'utilisateur qui ne travaille que sur certains départements de se focaliser sur les exploitations dont le parcellaire est complet (pas d'ilots dans d'autres départements) ou d'avoir l'information sur les départements dont il faudrait avoir les données pour compléter les parcellaires de la zone d'étude.
- Intérêt côté ODR pour développer un suivi des exploitations d'une année à l'autre. Possibilité de constituer une BdD sur ce point, qui pourrait en retour alimenter RPG-E (mais du coup sans possibilité de paramétrage des intersections, donc avec un paramétrage fixe – ou 2-3). L'ODR explore l'utilisation des données sur les MAE pour créer un référentiel de suivi temporel des identifiants d'une même exploitation entre plusieurs années.
- Extraction des séquences à partir de RPG-E pour 4 départements (18, 52, 55, 89) et envoi à l'ODR qui met en place protocole de comparaison des résultats entre sorties AGIR et RPG Explorer.

Réunion de suivi #2– 13/11/2017

Participants :

ODR : P. Cantelaube, T. Poméon

SADAPT : P. Martin

Entreprise Piskiewicz : N. Piskiewicz

Objet : point sur les avancées dans la procédure d'analyse des données et des différences entre les 2 méthodes (RPG-E et T-AGIR). Discussion autour de quelques cas particuliers illustratifs.

Décisions prises :

- Sur les surfaces, un bug a été identifié dans RPG-E, qui rajoute des surfaces artificiellement dans certains cas (ilots avec fusion/intersection). SADAPT produit une première version corrigée, qui sera bientôt stabilisée. Une nouvelle extraction des séquences sera alors réalisée dans RPG-E. Par ailleurs, l'ODR peut continuer à travailler sur les données actuellement disponibles, en repérant les cas où la somme des surfaces de séquences dépasse celle de l'ilot en 2014 ce qui permettra de quantifier l'effet du bug. Réciproquement peut-être est-il possible d'identifier les situations où la méthode AGIR générerait aussi des surfaces comptées 2 fois mais sans dépasser la surface de l'ilot 2014 (à voir si c'est possible de le repérer de manière systématique).
- L'ODR construit un fichier répertoriant l'ensemble des id ilot 2014 pour le dpt 18, et leur présence/absence dans les données RPG-E et AGIR. SADAPT vérifiera l'origine des décalages (divergence de l'emprise/échantillon considérée, bug ou autre)
- Les ilots où aucune séquence n'est reconstitué doivent aussi être pris en compte, car il témoigne de la similitude (ou de la divergence) des résultats obtenus des 2 algo. L'ODR inclut la prise en compte de ces éléments dans le processus de comparaison, en s'appuyant notamment sur la table répertoriant les ilots présents/absents (point ci-dessus).
- L'ODR regardera ce qui se passe pour les ilots stables avec le filtre 1, pour lequel on est censé retrouver le même résultat dans AGIR et RPG-E.
- Il est nécessaire de poursuivre les investigations en recherchant d'autres comparaisons possibles, en regardant le nb d'ilots intersectés selon les années, le nb de séquences par ilot, la durée des séquences reconstituées, ilots où $\text{sum}(\text{surf_seq}) > \text{surf_ilot}$, etc.
- L'ODR regarde ce qu'il est possible de faire pour avoir une table de suivi des numéros d'exploitations, d'abord pour les 4 dpts étudiés (18, 52, 55, 89). Peut être on peut en tirer une info sur les nb de parcelles par ilot ? A voir si faisable et utile.

1 : Import des données Asp

Successions d'étapes permettant l'import des données shape transmises par l'Asp :

- Reconnaissance du système de projection du shape (lambert 2 étendu, Lambert 93), par lecture du fichier .prj.
- Lecture dans la table PostgreSQL « spatial_ref_sys », de la correspondance au système de projection détecté, par comparaison des paramètres du système de projection
- Affichages à l'utilisateur des systèmes de projection et de leurs correspondances détectées par l'outil.
- Création d'un fichier .sql, via une ligne de commande faisant appel à l'utilitaire shp2pgsql, contenant l'ensemble des requêtes permettant l'import des données.
- Lecture de ce fichier .sql par Postgresql : import des données dans une table
- Suppression des ilots ayant une forme non valide : Delete From table where not st_isvalid(geom) (une dizaine de cas/département)
- Modification des noms de colonnes, et du type associé, afin de répondre aux exigences de l'outil.
- Modification de polygones dans les données, pouvant entraîner des problèmes dans la suite de l'utilisation de l'outil :
 - Sélection des ilots au sein d'une même couche (d'une même année), dont l'intersection fait plus de 100m². Ils se recouvrent donc, et seront problématiques pour les opérations géométriques liés à la filiation.
 - On vérifie pour chacun des ilots, la proximité entre sa surface géométrique, et les données déclarées dans le fichier Asp des groupes de cultures (somme des surfaces des groupes de culture).
 - En fonction des résultats on rabote la surface commune, à l'un, à l'autre, au bien aux deux
 - NB : cette opération engendre parfois une erreur géométrique, qui n'est actuellement pas gérée.

2 : Filiation des données

Options de sélections proposées à l'utilisateur :

- Choix parmi 3 types de territoires :
 - Ensemble des données Asp d'un département
 - Sélection d'une ou plusieurs exploitations agricoles
 - Sélection d'un Espace Ecologique spécifique (par import d'un Shape), et choix parmi 2 types d'intersection :
 - Sélection des ilots dans l'Espace Ecologique
 - Sélection des ilots des exploitations qui possèdent au moins un ilot dans l'Espace Ecologique
 - choix d'une ou plusieurs années de référence
 - Possibilité d'une sélection dynamique
- Seuils de suppression des ilots d'intersection superflus (petits ilots, ilots allongés)

Successions d'étapes permettant la Filiation des données de l'Espace Ecologique :

2.1 : Sélection de l'empreinte de l'Espace Ecologique souhaité :

- Selon le type de sélection, création d'une table contenant les ilots concernés
- Suppression des ilots dupliqués entre départements
- Vérification des doublons :
 - Vérification des ilots appartenant à plusieurs Espaces Ecologiques, modification des colonnes de références aux territoires
 - Suppression des vrais doublons
- Suppression des ilots ne répondant pas aux critères
 - Dump de la géométrie pour les multi-polygones, afin de désagréger les multi-polygones (st_dump)
 - Buffer trick de la géométrie, afin de redessiner le contour de géométries non valides (st_buffer)
 - Suppression des polygones non valides, ou de types (géométrie Collection, Linestring)
 - Suppression des ilots trop petits
 - Suppression des ilots allongés
 - Utilisation de la fonction PostgreSQL pour rendre les géométries valides (st_makeValid)

- Relance du processus dans le cas d'une filiation dynamique, avec pour Espace Ecologique, la sélection effectué par les étapes précédentes
- L'ensemble des ilots en sortie du processus de la sélection des données de l'Espace Ecologique forment **l'empreinte de l'Espace Ecologique**

2.2 : Intersection de l'empreinte de l'Espace Ecologique, avec les données Asp :

- Sélection des ilots de chacune des années des données Asp qui se superposent à l'empreinte de l'Espace Ecologique.
- Suppression des ilots ne répondant pas aux critères (sauf seuils)
- Suppression des ilots dupliqués entre départements
- L'ensemble des ilots en sortie de processus de la sélection annuelle des données de l'Espace Ecologique forment **l'empreinte annuelle de l'Espace Ecologique**

2.3 : Intersection successives des données Annuelles de l'Espace Ecologique :

- Intersection Géométrique des données Annuelles de l'Espace Ecologique
- Suppression des ilots ne répondant pas aux critères
- Ajout des ilots qui apparaissent ou disparaissent en partie, par différence symétrique des différentes couches
- Ajout des ilots qui apparaissent ou disparaissent intégralement, afin de retrouver les potentielles erreurs de suppression précédentes
- Ajouts des colonnes utilisées par Rpg Explorer :
 - Informations sur les Eas : id_ea, id_ea des années de références
 - Informations sur les Espaces Ecologiques : id_territory, id_territory_maj, id_terr_ea
 - Identifiants des ilots : Id_unique, il_in_ea
 - Informations sur la stabilité des ilots : qualif_fil

3 : Reconnaissance des séquences :

Construction de la correspondance de la filiation de L'Espace Ecologique par un objet métier spécifique à Rpg Explorer. Il permet un traitement facilité de l'objet, afin de ne pas passer par requête SQL pour Reconnaître les Séquences.

Reconnaissance des Séquences :

- Reconnaissance des séquences par pas de 2 ans
 - Etapes similaires à la Méthode de Toulouse, mais il faut regarder dans les détails.
 - Ajout d'une étape sur l'identification des cultures pluri-annuelles : elle est intégrée pour le moment entre l'étape 3 et 4.
- Construction progressive des séquences par identification des id Asp entre les différentes Reconnaissances effectuées par pas de 2 ans
 - On retire à chaque séquence de 2 ans, la part que l'on affecte à la séquence totale.
 - Affectation à la séquence de l'étape de Reconnaissance la moins-disante

Annexe 3 : méthode de rapprochement et sélection des identifiants d'exploitations

Voir la note de P. Cantelaube de août 2016 : *Rapprochement des exploitations du RPG anonymisées avec des identifiants dépendants du département de ses îlots (Données du RPG Anonyme à partir de la campagne 2010)*

Par ailleurs, un wiki est dédié au RPG

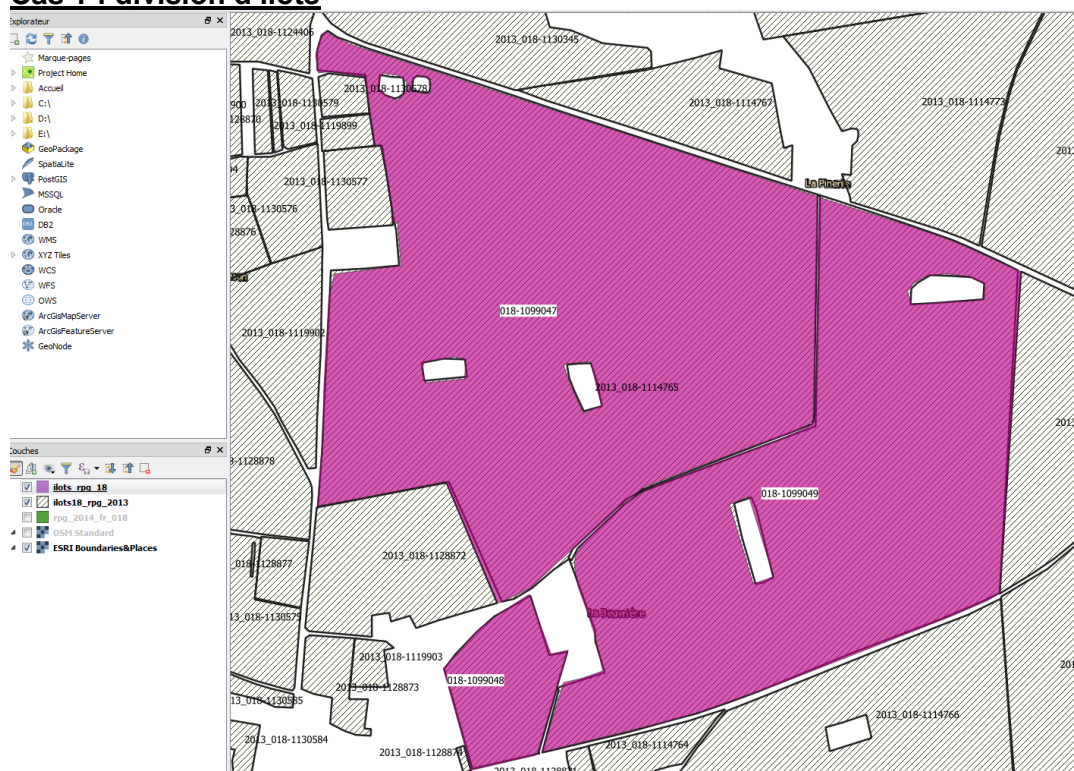
(<https://esrcarto.supagro.inra.fr/intranet/carto/cartowiki/index.php/RPG>), avec notamment des articles sur les problèmes rencontrés et leur résolution

([https://esrcarto.supagro.inra.fr/intranet/carto/cartowiki/index.php/RPG : _probl%C3%A8mes connus](https://esrcarto.supagro.inra.fr/intranet/carto/cartowiki/index.php/RPG:_probl%C3%A8mes_connus)).

Attention toutefois, ce wiki est un document de travail, susceptible d'être modifié selon les mises à jour, et pas quelque chose de systématiquement finalisé et destiné à être diffusé en externe.

Annexe 4 : comparaison du différentiel de traitement dans le cas d'îlots non stables dans le temps

Cas 1 : division d'îlots



Dans ce cas, on a un îlot, stable entre 2006 et 2013, qui est divisé en 3 îlots en 2014 (1099047, 1909048, 1099049). De fait, toutes les surfaces des années précédentes 2014 sont associées de la même façon aux 3 îlots en 2014. C'est ce qu'on retrouve donc dans les extraits de la table « culture_018 » de T-AGIR pour chacun des 3 îlots, dont des extraits sont inclus ci-dessous (limité aux 3 années les plus récentes pour faciliter la lecture).

ilot2014	gc	surf	ilot_pc	annee
018-1099047	gelautre	3.86	100	2014
018-1099047	tour	76.62	100	2014
018-1099047	blet	59.65	100	2013
018-1099047	autrescereales	31.06	100	2013
018-1099047	tour	29.36	100	2013
018-1099047	colza	19.35	100	2013
018-1099047	gelautre	7.31	100	2013
018-1099047	blet	71.71	100	2012
018-1099047	autrescereales	38.06	100	2012
018-1099047	colza	23.34	100	2012
018-1099047	gelautre	7.31	100	2012
018-1099047	tour	6.31	100	2012

ilot2014	gc	surf	ilot_pc	annee
018-1099048	tour	6.2	100	2014
018-1099048	blet	59.65	98	2013
018-1099048	autrescereales	31.06	98	2013
018-1099048	tour	29.36	98	2013
018-1099048	colza	19.35	98	2013
018-1099048	gelautre	7.31	98	2013
018-1099048	blet	71.71	98	2012
018-1099048	autrescereales	38.06	98	2012
018-1099048	colza	23.34	98	2012
018-1099048	gelautre	7.31	98	2012
018-1099048	tour	6.31	98	2012

ilot2014	gc	surf	ilot_pc	annee
018-1099049	blet	48.41	100	2014
018-1099049	Tour (6)	7	100	2014
018-1099049	gelautre	4.63	100	2014
018-1099049	blet	59.65	99	2013
018-1099049	autrescereales	31.06	99	2013
018-1099049	tour	29.36	99	2013
018-1099049	colza	19.35	99	2013
018-1099049	Gelautre (13)	7.31	99	2013
018-1099048	blet	71.71	98	2012
018-1099049	autrescereales	38.06	99	2012
018-1099049	colza	23.34	99	2012
018-1099049	Gelautre (13)	7.31	99	2012
018-1099048	tour	6.31	98	2012

Il est donc possible que les mêmes surfaces pour les années 2006 à 2013 soient reprises pour les 3 îlots de 2014 lors de la procédure de reconstitution. Toutefois, du fait de la non correspondance entre surfaces sur la base des règles de reconstitution, **T-AGIR** ne reconstitue qu'une seule séquence de 7 ha (0_0_0_0_0_13_13_13_6) pour l'îlot de référence 1099049.

Pour sa part, RPG-E repère 5 îlots d'intersection pour ce même ensemble d'îlots de référence. Mais **RPG-E** aboutit **au même résultat**, c'est-à-dire une seule séquence reconstituée de 7 ha (0_0_0_0_0_13_13_13_6) pour l'îlot 1099049.

id_unique	seq_surf	qualif_fil	seq_06_14
018-1099049_abs_abs_abs_abs_abs_abs_abs_abs	0.07	absence -> absence -> absence -> absence -> absence -> absence -> absence -> apparition	0_0_0_0_0_0_0_0_0
018-1099049_018-1114765_018-1104853_018-11342_018-1078_961619_760201_777174_5400091	7	stable -> stable -> stable -> stable -> stable -> stable -> stable -> division	0_0_0_0_0_13_13_13_6
018-1099049_018-1114765_018-1104853_018-11342_018-1078_961619_760201_777174_5400091	53.04	stable -> stable -> stable -> stable -> stable -> stable -> stable -> division	0_0_0_0_0_0_0_0_0
018-1099048_018-1114765_018-1104853_018-11342_018-1078_961619_760201_777174_5400091	6.2	stable -> stable -> stable -> stable -> stable -> stable -> stable -> division	0_0_0_0_0_0_0_0_0
018-1099047_018-1114765_018-1104853_018-11342_018-1078_961619_760201_777174_5400091	80.48	stable -> stable -> stable -> stable -> stable -> stable -> stable -> division	0_0_0_0_0_0_0_0_0

Cas 2 : fusion/division d'un ilot



Dans cet exemple, 2 îlots de 2014 (1120145 et 1133846) sont issus d'un même et unique îlot 2013, lui-même issu d'une succession de fusion-division depuis 2006. De fait, RPG-E identifie 7 îlots d'intersection, 5 associés à l'îlot 1120145 et 2 à l'îlot 1133846. Aucune séquence n'est ensuite reconstituée.

id_unique	seq_surf	qualif_fil	seq_06_14
018-1120145_018-1102916_018-1112893_018-15517_018-20850_959155_756304_777288_5364683	14.14	fusion -> division -> fusion -> division -> stable -> division -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1120145_018-1102916_018-1112893_018-15517_018-20850_959155_759924_777288_5364683	1.11	fusion -> division -> fusion -> division -> stable -> division -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1120145_018-1102916_018-1112893_018-8803_018-20850_959155_759924_777288_5364683	5	fusion -> division -> fusion -> division -> division -> fusion -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1120145_018-1102916_018-1112893_018-15517_018-10387_959155_756304_777288_5364683	0.94	fusion -> division -> fusion -> division -> fusion -> division -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1120145_018-1102916_018-1112893_018-15517_018-20850_959155_756304_777288_5367080	0.09	fusion -> division -> fusion -> division -> stable -> division -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1133846_018-1102916_018-1128586_018-15517_018-20850_959155_759924_777288_5364683	6.59	fusion -> division -> fusion -> division -> stable -> division -> fusion -> division	0_0_0_0_0_0_0_0_0
018-1133846_018-1102916_018-1128586_018-15517_018-20850_959155_756304_777288_5364683	0.9	fusion -> division -> fusion -> division -> stable -> division -> fusion -> division	0_0_0_0_0_0_0_0_0

Dans la procédure T-AGIR, les 2 îlots 2014 sont bien affiliés au même îlot 2013. Pour les années précédentes les surfaces affiliées sont parfois communes aux deux îlots et parfois attribuées à un seul, du fait de la grande instabilité des îlots analysés ici. On obtient les tables ci-dessous comme base avant reconstitution des séquences. Contrairement à RPG-E, T-AGIR reconstitue 2 séquences, uniquement pour l'îlot 1120145 : 15,14 ha de 0_0_0_0_0_0_1_2_1 sur la base de surface équivalente, puis 6,32 ha

de 0_0_0_0_0_0_2_1_2 du fait qu'il ne reste ensuite plus qu'une culture en 2014/2013/2012, et donc la possibilité de générer une autre séquence.

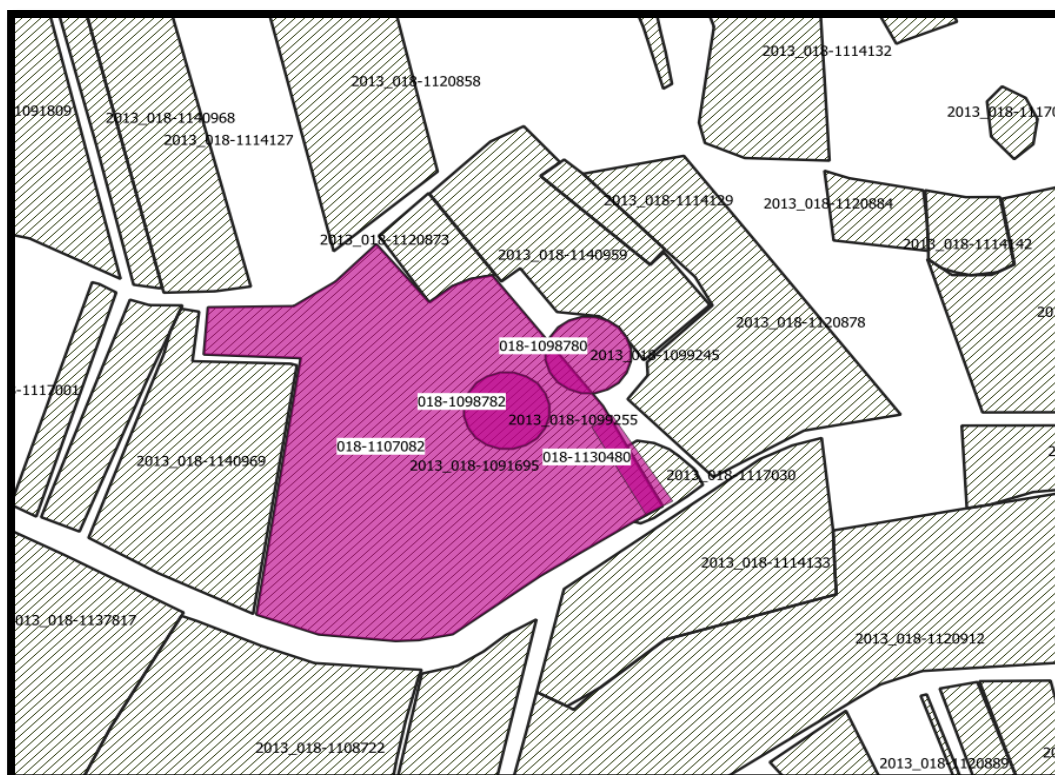
ilot2014	gc	surf	ilot_pc	annee
018-1120145	blet	15.14	100	2014
018-1120145	mais	6.32	100	2014
018-1120145	blet	14.68	100	2013
018-1120145	mais	15.13	100	2013
018-1120145	blet	15.14	100	2012
018-1120145	mais	6.68	100	2012
018-1120145	mais	15.65	77	2011
018-1120145	blet	9.16	77	2011
018-1120145	blet	5	99	2011
018-1120145	autrescereales	15.38	95	2010
018-1120145	mais	12.64	95	2010
018-1120145	mais	1.4	67	2010
018-1120145	gelautre	0.39	95	2010
018-1120145	autrescereales	14.08	100	2009
018-1120145	semence	15.4	100	2009
018-1120145	gelsansprod	0.38	100	2009
018-1120145	autrescereales	13.2	47	2008
018-1120145	gelsansprod	0.29	92	2008
018-1120145	autrescereales	0.88	92	2008
018-1120145	blet	15.49	92	2008
018-1120145	semence	14.08	100	2007
018-1120145	mais	15.49	100	2007
018-1120145	gelsansprod	0.29	100	2007
018-1120145	mais	15.52	100	2006
018-1120145	autrescereales	14.08	100	2006
018-1120145	gelsansprod	0.29	100	2006

ilot2014	gc	surf	ilot_pc	annee
018-1133846	legfleur	8.31	100	2014
018-1133846	mais	15.13	100	2013
018-1133846	blet	14.68	100	2013
018-1133846	mais	8	99	2012
018-1133846	mais	15.65	98	2011
018-1133846	blet	9.16	98	2011
018-1133846	autrescereales	15.38	95	2010
018-1133846	mais	12.64	95	2010
018-1133846	mais	1.4	27	2010
018-1133846	gelautre	0.39	95	2010
018-1133846	semence	15.4	100	2009
018-1133846	gelsansprod	0.38	100	2009
018-1133846	autrescereales	14.08	100	2009
018-1133846	autrescereales	13.2	84	2008
018-1133846	gelsansprod	0.29	16	2008
018-1133846	blet	15.49	16	2008
018-1133846	autrescereales	0.88	16	2008

018-1133846	mais	15.49	100	2007
018-1133846	semence	14.08	100	2007
018-1133846	gelsansprod	0.29	100	2007
018-1133846	mais	15.52	100	2006
018-1133846	autrescereales	14.08	100	2006
018-1133846	gelsansprod	0.29	100	2006

Cas 3 : superposition d'îlots intra et interannuelle

Ce cas reprend l'exemple de la figure 2, avec les 4 îlots qui se superposent en partie en 2014. Cette superposition implique de fait des filiations multiples des îlots des années précédentes, avec donc des filiations interannuelles entre les îlots qu'on retrouve tels quels dans les années précédentes, ainsi que des filiations entre îlots qui se superposent partiellement.



Côté T-AGIR cela implique la reprise des mêmes surfaces plusieurs fois pour différents îlots 2014.

ilots2014	gc	surf	ilots_pc	annee
018-1098780	divers	0.02	100	2014
018-1098780	vignes	2.62	18	2013
018-1098780	divers	0.02	100	2013
018-1098780	vignes	2.62	18	2012
018-1098780	divers	0.02	100	2012
018-1098780	divers	0.02	100	2011
018-1098780	divers	0.02	100	2010
018-1098780	divers	0.02	100	2009
018-1098780	divers	0.02	100	2008
018-1098780	divers	0.02	100	2007
018-1098780	divers	0.02	100	2006
018-1098782	divers	0.02	100	2014
018-1098782	vignes	2.62	100	2013
018-1098782	divers	0.02	100	2013

018-1098782	divers	0.02	100	2012
018-1098782	vignes	2.62	100	2012
018-1098782	vignes	2.15	27	2011
018-1098782	divers	0.02	100	2011
018-1098782	vignes	2.15	27	2010
018-1098782	divers	0.02	100	2010
018-1098782	divers	0.02	100	2009
018-1098782	vignes	2.15	27	2009
018-1098782	divers	0.02	100	2008
018-1098782	vignes	2.15	27	2008
018-1098782	vignes	2.15	27	2007
018-1098782	divers	0.02	100	2007
018-1098782	divers	0.02	100	2006
018-1098782	vignes	2.15	27	2006
018-1107082	vignes	2.62	100	2014
018-1107082	divers	0.02	18	2013
018-1107082	divers	0.02	100	2013
018-1107082	vignes	2.62	100	2013
018-1107082	divers	0.02	100	2012
018-1107082	divers	0.02	18	2012
018-1107082	vignes	2.62	100	2012
018-1107082	divers	0.02	100	2011
018-1107082	vignes	2.15	100	2011
018-1107082	divers	0.02	18	2011
018-1107082	divers	0.02	100	2010
018-1107082	divers	0.02	18	2010
018-1107082	vignes	2.15	100	2010
018-1107082	divers	0.02	18	2009
018-1107082	divers	0.02	100	2009
018-1107082	vignes	2.15	100	2009
018-1107082	divers	0.02	18	2008
018-1107082	divers	0.02	100	2008
018-1107082	vignes	2.15	100	2008
018-1107082	divers	0.02	18	2007
018-1107082	vignes	2.15	100	2007
018-1107082	divers	0.02	100	2007
018-1107082	vignes	2.15	100	2006
018-1107082	divers	0.02	100	2006
018-1107082	divers	0.02	18	2006
018-1107082	divers	0.07	66	2006
018-1130480	vignes	0.04	100	2014
018-1130480	divers	0.03	100	2014
018-1130480	vignes	2.62	72	2013
018-1130480	divers	0.07	22	2013
018-1130480	vignes	2.62	72	2012
018-1130480	divers	0.07	22	2012
018-1130480	vignes	2.15	72	2011
018-1130480	divers	0.07	22	2011

018-1130480	divers	0.07	22	2010
018-1130480	vignes	2.15	72	2010
018-1130480	vignes	2.15	72	2009
018-1130480	divers	0.07	22	2009
018-1130480	vignes	2.15	72	2008
018-1130480	divers	0.07	22	2008
018-1130480	divers	0.07	22	2007
018-1130480	vignes	2.15	72	2007
018-1130480	vignes	2.15	72	2006
018-1130480	divers	0.07	71	2006

Alors que potentiellement cette situation pourrait générer des erreurs et double-compte, le jeu des filtres de reconstitution limite cette possibilité, et T-AGIR produit au final 3 séquences pour ces 4 ilots 2014 (3 ilots avec une séquence, un ilot sans séquence reconstitué).

id_ilot	surface	annee	filtre	name_code
018-1098780	0.02	2014	surf_eq	28_28_28_28_28_28_28_28
018-1098782 (le petit au milieu du 082)	0.02	2014	surf_eq	28_28_28_28_28_28_28_28
018-1107082 (le grand qui inclut le 782)	2.62	2014	surf_eq	0_0_0_0_0_0_21_21_21

RPG-E fonctionne sur une autre logique avec les ilots d'intersection, et l'élimination des ilots aberrants géométriquement, comme ceux qui se superposent complètement. Sur l'ilot 018-108780, le résultat est le même, en termes de séquences et de surface. Par contre, les ilots 018-108782 et 018-1130480 ne sont plus présent dans la base RPG-E car ils ont été supprimés, au profit de l'ilot 018-1107082 (suppression des plus petits ilots en cas d'intersection forte entre 2 ilots une même année). Les surfaces liés aux groupes de culture des ilots affiliés à 018-1098782 se retrouve donc affilié au plus grand ilots, qui présente donc 2 séquences dans ce cas, avec des surfaces limitées à l'ilot d'intersection qui permettent de reconstituer une séquences de vignes complètes d'une part, et une séquence vignes/divers d'autre part avec les surfaces résiduelles de vignes.

id_unique	qualif_seq	seq_surf	qualif_fil	seq_06_14
018-1098780_018-1099245_018-1117448_018-40028_018-10654_932285_729386_825820_5398566	1	0.02	stable -> stable -> stable -> stable -> stable -> stable	28_28_28_28_28_28_28_28
018-1107082_018-1091695_018-1097348_018-2849_018-11566_944677_735628_777247_5368216	1	2.15	stable -> stable -> stable -> stable -> stable -> fusion -> stable -> stable	21_21_21_21_21_21_21_21_21
018-1107082_018-1091695_018-1097348_018-40026_018-10652_932287_729384_825822_5398654	1	0.02	stable -> stable -> stable -> stable -> stable -> fusion -> stable -> stable	28_28_28_28_28_28_21_21_21

5 Glossaire

Agence de Services et de Paiement (ASP) : l'ASP est chargé de verser les aides de la politique agricole commune (PAC) aux agriculteurs. A ce titre, elle gère un système d'information géographique informatisé (avec depuis 2015 l'appui de l'IGN) pour l'identification des îlots et la description de leur occupation culturale

Groupes de cultures (gc) : regroupement généré par l'ASP d'après les cultures déclarées sur l'îlot et les compositions des groupes de la nomenclature utilisée. La version disponible pour les travaux de recherche comprend 28 groupes de cultures différents.

Ilot: ensemble de parcelles culturales contiguës, entières ou partielles, portant une ou plusieurs cultures, exploitées par vous-même, limité par des éléments facilement repérables et permanents, comme un chemin, une route, un ruisseau...ou par d'autres exploitations. L'ensemble de ces parcelles sont stables d'une année sur l'autre. Ils constituent l'unité géographique de référence du RPG.

Recensement parcellaire graphique (RPG) : Le RPG est constitué de l'ensemble des parcelles culturales géolocalisées par les déclarants eux-mêmes (sur fond ortho-photographique fourni) lors de la campagne annuelle de "déclarations PAC" lancée au printemps sur l'ensemble du territoire français. Leur tracé sert de base légale à l'Agence de Services et de Paiement (ASP) pour déterminer le montant de diverses aides directes aux agriculteurs

RPG-E/ RPG-Explorer : procédure développée par l'UMR SADAPT pour le traitement des données du RPG et permettant notamment de reconstituer les séquences de culture dans les ilots RPG. RPG-Explorer inclut également d'autres fonctionnalités pour traiter et analyser les données du RPG (trajectoire d'exploitation, assolement de rotations, etc.)

Séquences de cultures : succession de cultures sur plusieurs années dans un ilot donné.

T-AGIR : procédure développée par l'UMR AGIR pour le traitement des données du RPG et permettant de reconstituer les séquences de culture dans les ilots RPG

6 Sigles & Abréviations

ASP : Agence de Services et de Paiement

gc : groupe de cultures

IGN : Institut Géographique National

Inra : Institut National de la Recherche Agronomique

RPG : recensement parcellaire graphique

UMR AGIR : Unité mixte de recherche Agronomie, Innovations et Territoires

UMR SADAPT : Unité mixte de recherche Sciences Action Développement – Activités Produits Territoires

US ODR : Unité de service Observatoire du Développement Rural

7 Table des illustrations

Figure 1 : logique de filiation interannuelle des îlots pour la méthode T-AGIR.....	10
Figure 2 : exemple d'îlots qui se superposent pour une même couche annuelle (RPG 2014). On voit ici 4 îlots colorés : l'îlot 1098782 est complètement inclus dans l'îlot 1107082. Les deux autres îlots (1098780 et 1130480) se superposent partiellement avec ce même îlot 1107082. En hachuré est représentée la couche des îlots RPG en 2013 (identifiants d'îlots précédés du préfixe « 2013_ ») : on retrouve la même configuration sauf pour l'îlot 1130480 qui n'a pas d'équivalent. Dans la logique de la méthode T-AGIR, les 4 îlots 2014 seront affiliés à l'îlot 2013 correspondant à l'îlot le plus grand (1107082), avec donc potentiellement une répétition dans plusieurs séquences des surfaces de groupes de cultures considérées.	11
Figure 3 : logique d'intersection des îlots entre plusieurs couches annuelles dans l'approche RPG-E	15
Figure 4 : logique de filiation interannuelle des îlots pour la méthode RPG-E.....	16
Figure 5 : illustration des différentes étapes de reconstitution selon les approches T-AGIR et RPG-E	18
Tableau 1 : exemple des données pour un îlot de référence issues de l'intersection des couches des différentes années	11
Tableau 2 : règles mobilisées successivement pour la reconstitution des séquences de groupes de culture par T-AGIR	12
Tableau 3 : règles mobilisées successivement pour la reconstitution des séquences de groupes de culture par RPG-E	16
Tableau 4 : statistiques globales des résultats des procédures T-AGIR et RPG-E	21
Tableau 5 : groupes de culture associés à l'îlot 2014 018-1090938.....	25
Tableau 6 : résultat de la reconstitution des séquences pour l'îlot 018-1090938 par les approches T-AGIR et RPG-E	26
Tableau 7 : groupes de culture associés à l'îlot 2014 018-1092518.....	26
Tableau 8 : résultat de la reconstitution des séquences pour l'îlot 018-1092518 par les approches T-AGIR et RPG-E	27
Tableau 9 : résultat du traitement RPG-E pour l'îlot 018-1113788	27
Tableau 10 : groupes de culture associés à l'îlot 2014 018-1113788.....	28

Avec le soutien financier de

**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**
ÉTABLISSEMENT PUBLIC DE L'ÉTAT

www.agence-francaise-biodiversite.fr

