**Hugo SENGES**

**SOAD**  
Variables numériques en appui aux métiers de L’OFB

Vincennes, le 11 Octobre 2021

Sylvain BEORCHIA,

Ingénieur à Oslandia

Constitution d’un référentiel d’entraînement dérivé des données « Registre Parcellaire Graphique »

Note de pré-développement concernant une chaîne de traitement algorithmique visant la production d’une maille vectorielle dérivée des chroniques RPG.

1. Contexte et objectif

Dans le cadre de l’activité d’ajout d’une chaîne de traitement dans le package INRAE SISPPEO (dédié à la télédétection des CIPAN[[1]](#footnote-1)), il est nécessaire de disposer d’une maille d’entraînement. Celle-ci prendra la forme d’une donnée vectorielle dérivée des données annualisées « Registre Parcellaire Graphique » mis à disposition publiquement par l’IGN[[2]](#footnote-2) (GeoServices). Se basant sur l’AMI précédent (Intelligence Artificielle en appui à l’élaboration des plans de contrôle Eau & Nature), l’AMI « Variables Numériques » s’inscrit dans un environnement de développement établi. Ainsi les méthodes de collecte, tuilage et traitement des images Sentinel sont déjà implémentées et validées au sein du package python SISSPEO, hébergé sur les serveurs de l’INRAE. C’est dans ce cadre qu’il s’agit de développer une nouvelle branche de ce projet, qui permette l’élaboration d’un outil statistique d’aide à la décision sur la thématique des CIPAN. Cette activité se rapporte à l’objectif 1.1, inclue dans le document Descriptif de phase 1 du projet :

**Objectif 1.1. Mise à disposition d’une API de résultats statistiques aux mailles de référence INPN, sur la thématique du type de couverture de sol sur les parcelles agricoles mentionnées au RPG.**

**1.1.1. Accès aux images satellitaires et préparation des données issues**

**du registre parcellaire graphique** :

Correspond à la première étape, mais aussi à la préparation des données nécessaires aux étapes suivantes (notamment RPG). Ce sous-objectif poursuit la collecte des données initiée lors de la *phase 0*. *Cette phase est d’autant plus importante que les sources d’images satellitaires sont appelées à évoluer, en prenant en compte les améliorations attendues (nouveaux programmes satellitaires) mais aussi la durée de vie (7 ans et 3 mois) attendue des satellites Sentinel.*

**1.1.2. Consolidation des chaînes de traitement depuis les images**

**satellitaires vers des résultats de classification et intégration à**

**SISPPEO** :

correspond à l’ensemble des groupes d’étapes *Préparation des données* (hors première étape), *composition des bandes* et *Classification et scores.* Les classes de couverture de sol visées sont : Sol nu | Végétation sèche | Végétation mixte basse (sol sec / végétation peu couvrante) Végétation peu couvrante | Végétation mixte couvrante (végétatio peu couvrante / végétation couvrante) | Végétation couvrante | Végétation hétérogène | Couverture inconnue

Il s’agit donc, dans un premier temps, de constituer une maille vectorielle sur la base des chroniques [2015-2019] du RPG 2.0. C’est sur la base de cette donnée que seront ensuite développées les chaînes de traitement en télédétection. En outre, l’avantage des données vectorielles RPG consiste en la mise à disposition d’un profil de culture, pour l’année du millésime et pour chaque parcelle (entité). La table attributaire correspondante est la suivante :

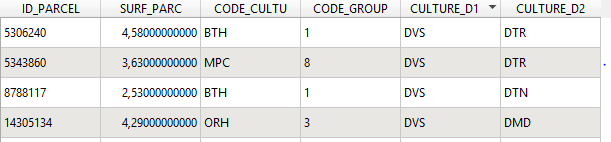


Figure 1: table attributaire d'un jeu de données RPG 2.0

Le code culture ainsi disponible permet de constituer une ressource de variables auxiliaires. Celles-ci ne peuvent se substituer à l’acquisition d’une « vérité-terrain » en bonne et due forme. Mais, dans une optique d’entraînement d’un programme de télédétection des CIPAN, elles fournissent des informations (sur une base déclarative) qui peuvent être confrontées dans un second temps avec les prévisions du modèle de classification des surfaces. De plus, la manipulation de données vectorielles ainsi renseignées permet une appréhension facilité des zones de cultures (possibilité de filtrer, trier, masquer les zones d’intérêt).

Il s’agira donc de produire un chaîne d’acquisition et de traitement de cette donnée vectorielle. Encapsuler le script SQL (extension spatiale PostGIS incluse) dans un script python (via les fonctionnalités de la librairie psycopg2) sera également nécessaire pour assurer la cohérence du programme et son inclusion dans le package python SISSPEO.

Les géométries attendues en sortie seront donc sous forme d’une collection de parties (post-traitement) des entités « Parcelles\_graphiques » des jeux de données RPG. C’est qu’en croisant les divers milésiemes de ce jeu de données, une analyse en première instance révèle un taux de correspondance des entités d’une année sur l’autre inférieur à 25%. Il y a donc nécessité de produire une variante de ces couches qui permettent le découpage des entités au niveau le plus précis possible, en donnant toujours la priorité à la donnée la plus récente. Ainsi la résolution des cas de superposition sera traitée de la façon suivante :

* L’intersection entre deux entités d’années différentes est isolée (ST\_Intersection) et on réalise un test logique sur la base de ses attributs géométriques :
  + Si l’aire de l’entité dépasse un seuil absolu, et un seuil relatif à l’aire de son entité parente (dans le jeu de données le plus récent), elle est établie comme nouvelle entité dans le jeu de données traité.
  + Dans le cas contraire, sa fusion (ST\_Union) avec son entité parente (dans le jeu de données le plus récent) remplace cette dernière entité dans le jeu de données traité.
* Les différences symétriques entre deux entités d’années différentes sont isolées (ST\_SymDifferences) et on réalise un test logique sur la base de leurs attributs géométriques :
  + Si l’aire de l’entité dépasse un seuil absolu, et un seuil relatif à l’aire de son entité parente, elle est établie comme nouvelle entité dans le jeu de données traité.
  + Dans le cas contraire, sa fusion (ST\_Union) avec son entité parente remplace cette dernière entité dans le jeu de données traité.

L’opération permet de produire une nouvelle couche issue du croisement des deux années les plus récentes du RPG. Les années précédentes (jusqu’en 2015) servent alors, de façon itérative, à recouper la couche ainsi obtenue. Lorsque les opérations morphologiques sont toutes accomplies, l’enrichissement sémantique de la couche finale est réalisé en SQL :

* Formatage d’une table attributaire vierge de (8 + **n**\*4) colonnes (où **n** est le nombre de jeu de données passés en entrée, soit pour 2015-2019, un total de 24 colonnes)
* Création de la clé primaire dans la première de ces colonnes
* Repeuplement de la table attributaire par Statistiques Zonales depuis les couches du RPG
* Calcul des attributs dérivées qui sont :
  + Géométriques : aire, périmètre et indice de compacité[[3]](#footnote-3). Ce dernier indice pourrait, lors d’une seconde phase du développement, permettre d’appréhender la pertinence d’une procédure de télédétection, au regard de segments qui pourraient avoir une forme particulièrement allongée
  + Sémantiques : vote majoritaire faisant remonter le profil cultural RPG majoritaire sur le segment durant la période d’observation.

1. Logigramme

Récupération des couches RPG 2.0 [2015-2019]  
(téléchargement par API Geoservices)

Compilation dans une base de données PostGis

(psycopg2)

Mise en forme de la table à 24 attributs et clé primaire

Statistiques zonales sur les profils de culture des années 2015-2019 depuis les couches RPG

PREPARATION DU JEU DE DONNEES

ECLATEMENT PARCELLAIRE

ENRICHISSEMENT ATTRIBUTAIRE DE LA COUCHE OBTENUE

Calcul des attributs dérivés (géométriques) et vote majoritaire sur les profils de culture antérieurs

MISE A DISPOSITION

Affichage en sortie des paramètres d’accès à la ressource

Push de la couche dans une table postgis dédiée

Relance du bloc avec la couche RPG obtenue et la couche immédiatement plus ancienne

Fusion des entités trop petites avec leur entité voisine reliée

Tests logiques (surface relative et absolue) sur les nouvelles entités

Sous-segmentation des intersections (ST\_Intersection) et des différences symétriques (ST\_SymDifference)

Sélection d’une couche de travail (la plus récente), et d’une couche de croisement   
(la plus récente après celles déjà utilisées)

1. Points de précision
2. Préparation du jeu de données  
   1. Récupération des données RPG :

* L’automatisation de l’acquisition des données RPG via une requête http auprès de l’API GéoServices[[4]](#footnote-4) permettrait une intégration de cette étape dans la chaîne de traitement de façon pérenne. Toutefois pour les besoins du développement, cette intégration pourrait se faire *a posteriori*, une fois les fonctionnalités de traitement implémentées et validées.
* La question est liée à celle du format de réception depuis l’API, qui pourrait ou non supposer une conversion vers un format plus pratique pour les fonctionnalités de traitement. Selon la documentation à disposition, les formats de sorties de la requête http sont les suivants : GML2 ; KML ; géoJSON ; gml13 ; gml32 ; SHAPE-ZIP (par défaut en EPSG : 4326). Pour faciliter les traitements, et permettre de mieux brancher en entrée l’étape d’acquisition depuis l’API, un format directement manipulable (comme KML ou géoJSON) serait à préférer devant le format SHAPE, nécessitant un dé-ZIPage.   
  1. Construction de la base de données :
* Deux questions se posent alors :   
  - quel format est le plus efficient en termes de temps de calcul au regard des fonctions de recouvrement ?   
  - les clés d’identifications ne se recoupant pas entre les différentes chroniques, leur extraction au sein d’un dictionnaire dédié devrait-il se faire dès l’acquisition, avant tout traitement ?

1. Eclatement parcellaire  
   1. Sous-segmentation :

* Une alternative est posée entre la convocation de fonctions natives de Postgres/PostGIS pour ces traitements (*a priori* ST\_Intersection et ST\_SymDifferences) et la possibilité de procéder par extraction-processing-reload dans la base données. Dans ce cas ce seraient plutôt les fonctions de traitement issues du package python GDAL qui seraient utilisées, dans un environnement de dialogue avec la base de données dirigé par le package psycopg2. Ces fonctions seraient notamment issus du package shapely (.intersection, .differences).   
  1. Tests logiques :
* Ici c’est la question de l’étape durant laquelle réaliser les tests portant sur la surface minimale des entités sous-segmentées qui est posée. Doit-on lancer une boucle de test sur toutes les entités issues de la sous-segmentation, et effectuer les modifications morphologiques associées dès qu’une entité ne satisfait pas aux critères de surface minimale ? Ou bien-doit-on effectuer le test à chaque fois qu’une entité est sous-segmentée à l’étape précédente ?

Dans le deuxième cas, il faudrait alors conférer deux attributs aux entités détectées comme « à agréger » : un attribut booléen « To\_merge » en modalité « TRUE », et un attribut numérique « Id\_entity\_to\_fuse » renvoyant à la clé d’identification de l’entité parente avec laquelle l’agréger.

* 1. Fusion/agrégation des petites entités :
* Outre le choix de l’environnement à préférer pour ce traitement (même questionnement entre PostGIS et Shapely qu’en 2-a.), c’est la gestion des clés d’identification des entités depuis les données RPG qui demande précision. Comme elles ne se recoupent pas d’une année sur l’autre, il faut les enregistrer comme attributs dans la couche sous-segmentée. La question étant encore l’étape à laquelle réaliser cet enregistrement :   
  si on procède « au fur et mesure » de la sous-segmentation on peut s’appuyer sur ces clés enregistrées pour faciliter l’étape d’agrégation (puisque l’on connaît directement l’id de l’entité parente). Dans le cas contraire, les données sont plus légères, mais pour retrouver l’entité parente il faut relancer une étape de test de recouvrement puis, une fois la donnée vectorielle sous-segmentée-agrégée obtenue, récupérer les clés d’identifications des couches parentes durant l’étape d’enrichissement attributaire (par Statistique Zonale).

1. Enrichissement attributaire :

* Dans la continuité de la question précédente, au fur et à mesure de l’itération des sous-segmentation-agrégation, les clés d’identification s’accumulent. En plus d’une clé par année de la chronique RPG (soit 5 entre 2015 et 2019 inclus), une nouvelle clé s’ajoute à chaque étape (le recoupement 2019-2018 produit une nouvelle clé ; son propre recoupement avec le RPG 2017 en produit une nouvelle … etc jusqu’au recoupement avec la dernière donnée RPG 2015).   
  En l’état, on se retrouverait donc avec 5+4 = 9 clés d’identifications ne se recoupant que partiellement. La question étant alors la sélection des clés à conserver dans le produit final.
* L’hypothèse de travail actuelle est de conserver les clés d’identification des entités parentes dans toutes les années de la chronique RPG (avec éventuellement des NULL pour les entités n’ayant pas d’intersection avec les entités de certaines années) et la clé issue du dernier recoupement, qui devient définitivement clé primaire de cette couche finalisée.

1. Cultures Intermédiaires Piège à Nitrate [↑](#footnote-ref-1)
2. Voir l’adresse : https://geoservices.ign.fr/services-web-experts-agriculture [↑](#footnote-ref-2)
3. indice de compacité : 𝑰𝒄= 𝑷/√𝑺 [↑](#footnote-ref-3)
4. L’URL type est donnée dans la documentation du site Géoservices.   
   De forme générique : https://wxs.ign.fr/agriculture/geoportail/wmts?SERVICE=WMTS&VERSION=1.0.0&REQUEST=GetCapabilities [↑](#footnote-ref-4)