

Chapitre 2 - Intégration des données dans PostGIS

2.1 Présentation générale

En géomatique, deux grands types de formats de données sont utilisés : le format « vecteur » et le format « raster ».

Les vecteurs représentent la forme des objets à l'aide de points qui peuvent être reliés afin de définir des lignes ou des polygones. Ces points sont simplement définis par leurs coordonnées. Or celles-ci dépendent de l'opération mathématique faite pour passer de la Terre à la carte, c'est-à-dire du système de projection. Il est donc important de connaître celui-ci. Tout objet représenté ainsi est une « entité ». Les entités disposent par ailleurs d'attributs, qui sont les informations à caractère non spatial les caractérisant.

La « donnée vectorielle » représente le monde réel en utilisant ce lien entre la forme des objets et les informations qui les concernent.

Il est aussi possible de représenter une partie de la Terre en la découpant avec des pixels. On parle alors de « raster ». L'information est la valeur du pixel. Une image géoréférencée, c'est à dire dont les pixels sont positionnés sur une carte, est donc un raster. Lorsque cette image possède plusieurs bandes, il est équivalent de parler de raster multi-bandes ou de superposition de rasters à une bande. Dans le premier cas, le pixel contient un tableau d'autant de valeurs que de bandes, dans le second nous disposons d'autant de pixels se chevauchant.

La mise en ligne de données cartographiques peut avoir de multiples finalités. Mais les images dont dispose l'internaute sont toujours constituées d'une superposition de données vecteur ou raster. Dans cet ouvrage, nous utiliserons ces deux grands types de données afin d'en souligner les spécificités dans une architecture de webmapping.

Nous ferons usage d'un réseau routier au format vectoriel pour proposer un service en ligne de calcul d'itinéraire entre deux points. Par ailleurs, ces routes nous serviront comme fond cartographique dans la confection d'une page web de suivi des positions d'un client.

Nous disposerons aussi d'une couche d'entités polygonales représentant des forêts. Les contours de celles-ci étant d'assez mauvaise qualité, nous essaierons de les améliorer en utilisant des images satellites, ce qui nous permettra de mettre en ligne les corrections effectuées ainsi que des rasters d'activité photosynthétique. Nous

verrons que l'ensemble de ce processus de corrections géométriques à partir d'images satellites est automatisable : l'internaute peut, dès la publication de l'image satellite bénéficier de l'information qu'elle apporte.

2.2 Récupération des données utilisées

Toutes les données que nous utiliserons sont librement téléchargeables. Nous ferons usage de la donnée vectorielle d'OpenStreetMap (OSM) issue d'un projet collaboratif sur lequel vous trouverez des informations ici : <http://openstreetmap.fr/>.

Nous utiliserons aussi comme donnée raster des images prises par les satellites SENTINEL de l'ESA.

Nous les utiliserons, sans faire appel à des algorithmes complexes, afin d'effectuer les corrections des formes des forêts évoquées plus haut. Nous pourrions explorer ainsi les fonctionnalités des logiciels fraîchement installés.

2.2.1 Données OpenstreetMap (OSM)

Nous travaillerons sur la région Midi-Pyrénées. Les données disponibles sur cette région le sont sous forme de fichiers de formes (« shapefiles ») ou de fichiers « Protocolbuffers binary format » (pbf) plus compacts et rapides en écriture et lecture mais aussi plus difficiles d'utilisation sur les systèmes Windows. Pour nous concentrer sur le cœur de notre sujet nous utiliserons les historiques fichiers de formes disponibles ici :

<http://download.geofabrik.de/europe/france/midi-pyrenees-latest-free.shp.zip>

De multiples couches sont disponibles : nous utiliserons celle des routes (gis.osm_roads_free) et de l'occupation du sol (gis.osm_landuse_a_free) dans laquelle nous trouverons les forêts évoquées plus haut.

Les algorithmes que nous mettrons en œuvre peuvent être coûteux en termes de ressources machine (CPU, RAM) ; aussi récupérerons-nous le contour des départements et communes français afin de travailler à l'échelle d'un département ou d'une commune plutôt que d'une région. Vous trouverez à l'adresse suivante un jeu de données des départements :

<https://www.data.gouv.fr/fr/datasets/contours-des-departements-francais-issus-d-openstreetmap/>

Et à cette adresse-ci les communes :

<https://www.data.gouv.fr/fr/datasets/decoupage-administratif-communal-francais-issu-d-openstreetmap/>

Toutes ces données ont leurs coordonnées en « WGS 84 » - comme pour les GPS -, d'identifiant n°4326 suivant l' *European Petrol Survey Group* (EPSG). La connaissance de cet identifiant est nécessaire et suffisante pour définir le système de projection d'une couche. Ainsi la gestion des projections est grandement simplifiée. Par ailleurs, l'encodage des attributs est en UTF-8, paramètre à ne pas oublier lors des imports dans PostGIS, sans quoi ils seront illisibles.

2.2.2 Données Sentinel 2

Les clichés pris par le satellite Sentinel 2A (ou B) sont composés de 13 bandes spectrales d'une résolution spatiale de 10 à 60 mètres sur une emprise de 290 kilomètres de large. La sensibilité spectrale va du visible à l'infrarouge court (SWIR) en passant par le proche infrarouge (PIR). Elles sont fournies géoréférencées en UTM 31, dont l'identifiant EPSG est 32631. Pour pouvoir croiser ces clichés avec la donnée OSM cette information est essentielle. La fréquence de balayage d'une même zone est d'un passage de satellite tous les cinq jours, ce qui permet de suivre des phénomènes, par exemple agronomiques, de manière très fine.

Les images sont téléchargeables gratuitement à la seule condition de créer un compte sur le portail Copernicus à l'URL suivante :

<https://scihub.copernicus.eu/dhus/#/home>

Ce site est de surcroît un moteur de recherche d'images qui permet d'effectuer des requêtes intégrant de nombreux critères : type de satellite, zone visée, date, couverture nuageuse, qualité du post-traitement, etc. Nous utiliserons ce moteur de recherche de manière programmatique au chapitre 8.

Pour illustrer les points abordés dans cet ouvrage, nous vous proposons de récupérer trois images de même emprise, centrées sur l'agglomération toulousaine et datant de 2017.

Attention les noms sont rébarbatifs, mais si vous êtes attentifs vous reconnaîtrez la date de la prise de vue.

- S2A_MSIL1C_20170705T105031_N0205_R051_T31TCJ_20170705T105605.SAFE
- S2A_MSIL1C_20170215T105121_N0204_R051_T31TCJ_20170215T105607.SAFE
- S2A_MSIL1C_20170824T105031_N0205_R051_T31TCJ_20170824T105240.SAFE

Pour obtenir ces images il vous suffit de saisir les noms ci-dessus dans le moteur de recherche Copernicus et de télécharger l'image.

Voici résumées les caractéristiques spectrales des bandes en fonction de leur résolution spatiale et spectrale.

- Bandes d'une résolution spatiale médiane (20m) mais balayant un large spectre (la largeur des rectangles indique le domaine de sensibilité des capteurs) :

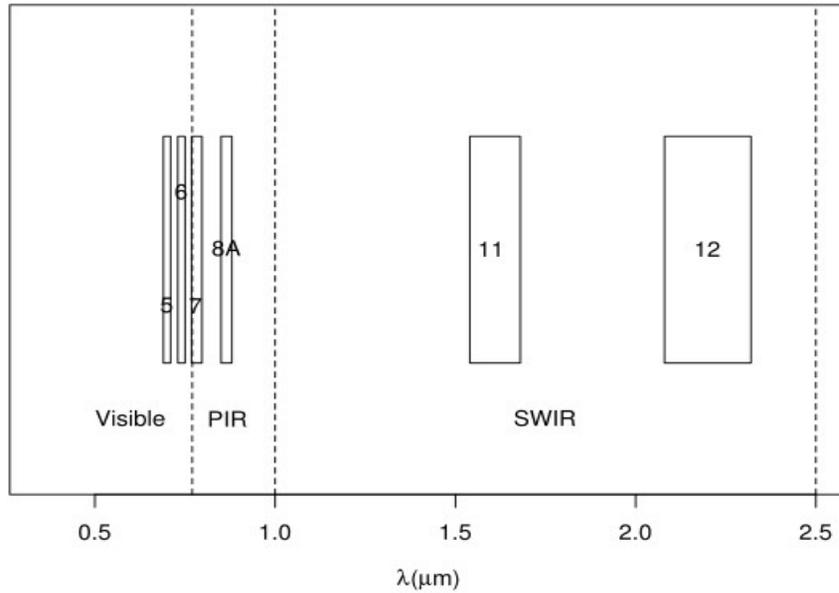


figure 2.1 : Schéma des résolutions spectrales des capteurs (numérotés) de Sentinel 2 dans le visible, le proche infrarouge et l'infrarouge court.

- Bandes à basse résolution spatiale :

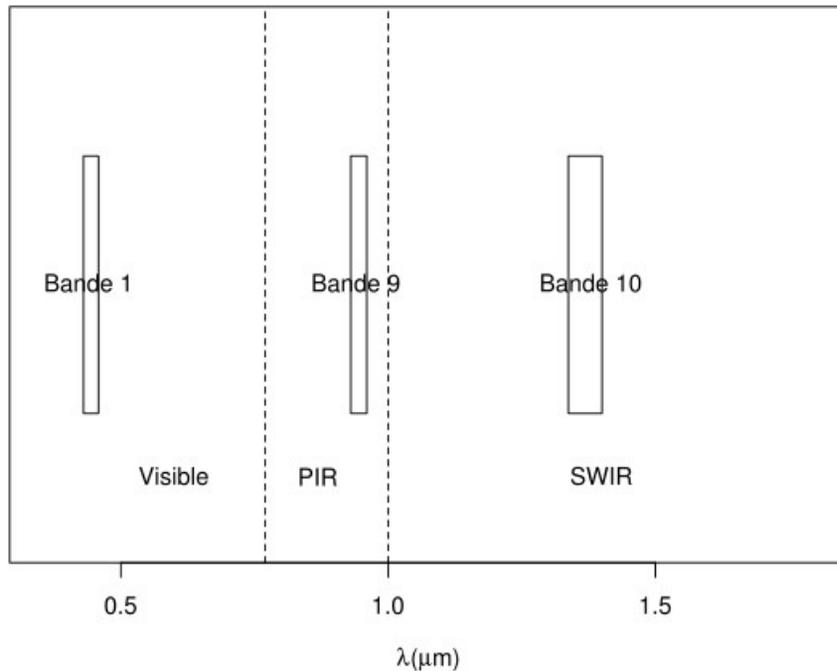


figure 2.2 : Domaine de sensibilité des trois bandes de 60m de résolution

Nous allons en partie exploiter par la suite les bandes à haute résolution spatiale (10m, figure 2.5) afin de nous essayer à la détection de forêts.

Si l'on compare les forêts numérisées dans OSM à la réalité, bien des défauts apparaissent : forêts manquantes, contours inexacts. Nous essaierons donc de voir si, avec des méthodes simples et automatisables, il serait possible d'améliorer la qualité de ces données vectorielles.

Pour tenter cette détection de forêts, il est important d'avoir en tête une caractéristique spectrale particulière des végétaux chlorophylliens.

La figure 2.3 présente le spectre de réflectance de différents végétaux. En première lecture nous pouvons interpréter un spectre comme « une composition de couleurs ».

Le saut qui apparaît entre le domaine du visible (de 0,4 à 0,7 μm) et celui du proche infrarouge (autour de 1 μm) sera notre point de départ.

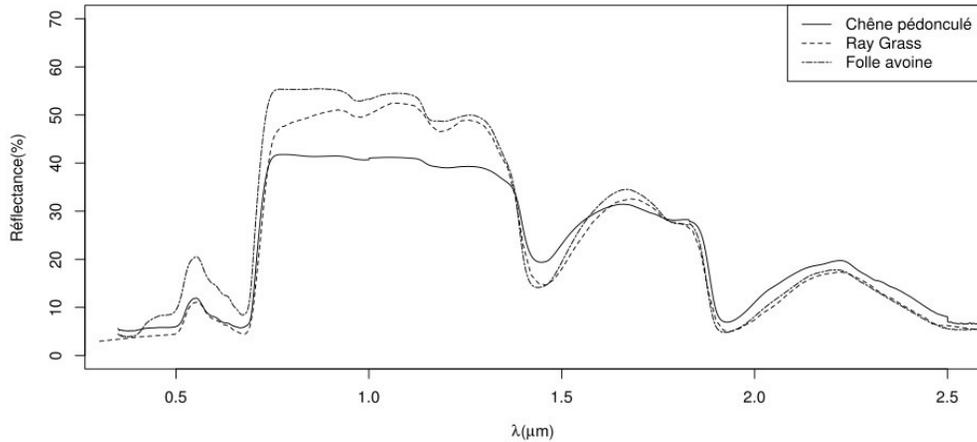


figure 2.3 : Signatures spectrales de végétaux

Ce saut signe l'activité de la photosynthèse. Il existe un indice dont la finalité est la mise en évidence de ce saut caractéristique, le NDVI (Normalized Difference Vegetation Index), défini par :

$$NDVI = \frac{(PIR - R)}{(PIR + R)}$$

où R désigne la radiométrie d'une bande centrée sur une longueur d'onde du visible correspondant en général au rouge (entre 0,4 et 0,7 nm) et PIR une bande centrée sur une valeur comprise entre 0,7 et 1,1 nm (Proche infrarouge) comme l'indique la figure 2.4.

Nous calculerons le NDVI de chacune des images téléchargées en associant, comme la figure 2.5 nous y invite, le PIR à la bande 8 et le R à la bande 4, afin de travailler à la meilleure résolution spatiale offerte par Sentinel 2 (10 m). En effet, le calcul est aussi envisageable avec les bandes 9 et 1 mais à une moindre résolution spatiale. Notons d'ores et déjà que la correction des parcelles boisées sera limitée par cette résolution.

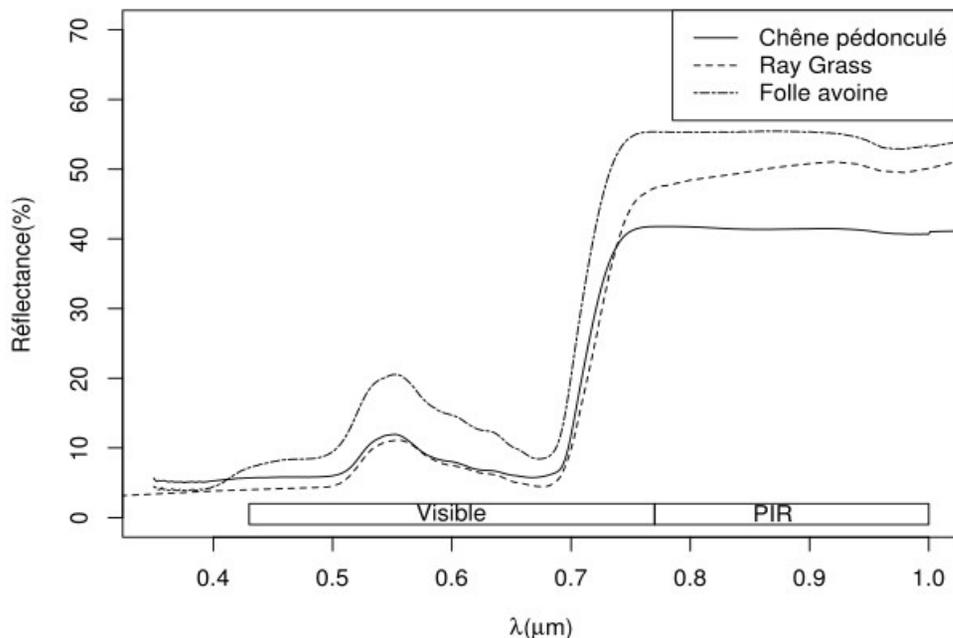


figure 2.4 : Saut spectral de la végétation entre le visible et le PIR

Nous effectuerons cette opération à trois dates différentes pour discriminer les forêts des autres végétaux dont le cycle n'est pas identique : ainsi en hiver les champs de cultures automnales ou les prairies ont déjà une forte activité photosynthétique quand les feuillus n'en ont plus ; à l'inverse en juillet ces derniers sont en pleine activité alors que les cultures d'automne sont moissonnées. Il nous faut aussi essayer de discriminer les cultures de printemps (maïs, tournesol) de nos forêts de feuillus.

Ce type d'analyse est qualifiée de « multi-temporelle ». Le lecteur imaginera facilement à l'issue de nos manipulations la qualité des résultats qu'il est possible d'obtenir en mettant à contribution l'ensemble des images disponibles sachant qu'une même zone terrestre est scannée tous les cinq jours, en ne considérant que les images Sentinel.

Même si dans le cadre de cet ouvrage nous nous restreindrons aux forêts, l'occupation du sol en général peut faire l'objet de telles détections.

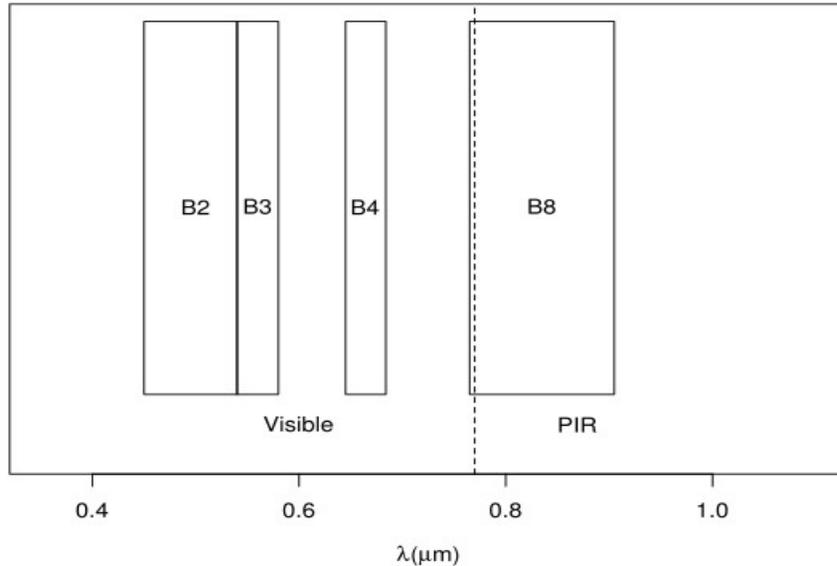


figure 2.5 : Les bandes du visibles et du PIR d'une résolution de 10m

Nous n'utiliserons donc que deux des treize bandes proposées dans chaque image, laissant aux chercheurs le soin d'exploiter les données spectrales dans leur ensemble.

2.3 Création d'une base de données spatiale

Bien qu'il soit possible de créer cette base en langage SQL, nous allons ici utiliser l'interface pgAdmin 4, installée en même temps que PostGIS.

Un simple clic droit sur l'onglet « Database » suivi du choix « Create » nous amène à renseigner les onglets montrés en figure 2.6.

Nommez la base « webmapping », en évitant, pour les dénominations, l'usage de majuscules ou autres accents qui créent de mauvaises surprises à l'occasion. Il est absolument nécessaire dans la page « Definition » de choisir le « Template », soit *modèle* en anglais, « PostGIS_2x_sample » qui va installer les fonctionnalités spatiales dans votre nouvelle base données en les copiant depuis cette dernière. Celle-ci doit être déconnectée pour l'opération et il est conseillé de ne jamais la modifier. Son existence permet justement la création d'une base spatiale en un clic.