

Classification supervisée 'pixel'

La classification supervisée dite 'pixel' est l'une des techniques de classification les plus utilisées actuellement, notamment pour la cartographie de l'occupation des sols. Cette méthode est basée sur un apprentissage supervisé, c'est-à-dire qu'elle nécessite l'utilisation d'un ensemble de données d'entraînement préalablement étiqueté pour permettre à l'algorithme de classification de 'comprendre' les différents types de surfaces présents dans l'image.

Le processus de classification supervisée pixel en télédétection se déroule généralement en plusieurs étapes :

- **Acquisition des données** : cela implique la collecte des images satellites à partir desquelles la classification sera effectuée.
- **Sélection des données d'entraînement** : cette étape consiste à sélectionner un ensemble de pixels de l'image pour lesquels la classe à laquelle ils appartiennent est connue et étiquetée. Cette étape est soit manuelle (digitalisation) ou semi-automatique (sélection d'une données de référence d'occupation des sols).
- **Entraînement du modèle** : l'algorithme de classification est alors entraîné à partir de ces données d'entraînement. Différentes méthodes peuvent être utilisées pour cette étape, notamment les algorithmes de classification basés sur les statistiques, les réseaux de neurones ou encore les machines à vecteurs de support. Dans ce TD, nous comparerons une approche statistiques (Maximum de vraisemblance ou *Maximum Likelihood*) et une méthode d'apprentissage automatique basée sur des forêts aléatoires (*Random Forest*).
- **Classification de l'image** : une fois le modèle entraîné, il peut être utilisé pour classer l'ensemble de l'image. Chaque pixel est alors assigné à une classe en fonction de sa proximité avec les pixels d'entraînement.
- **Évaluation de la classification** : la qualité de la classification est évaluée en comparant les résultats obtenus avec les données d'entraînement et en effectuant des vérifications sur le terrain. Pour cela, le jeu de données est séparé en jeu d'entraînement (entraînement uniquement du modèle) et en jeu de test (indépendant du jeu d'entraînement qui sert à tester le modèle).

L'approche de classification supervisée dite pixel en télédétection est largement utilisée dans de nombreux domaines, notamment pour la cartographie des sols, la gestion des ressources naturelles, la surveillance de l'environnement, etc.

Application 1 : Classification 'pixel' par méthode supervisée d'une image Sentinel-2 pour la classification de l'occupation des sols. Comparaison de deux méthodes : *Maximum Likelihood* et *Random Forest*.

Étape 1 : Création d'une donnée de référence

Charger l'image *Sentinel2A_20170619_10m.tif* sur QGIS et la découper suivant l'emprise du vecteur *limite_EMS.tif*. La reprojeter en projection WGS84 (EPSG :4326) et la nommer *Sentinel2A_20170619_10m_EMS_WGS84.tif*.

Raster > Extraction > Clip Raster by Extent

Sinon en ligne de commande :

```
gdalwarp -cutline limite_EMS.shp -crop_to_cutline Sentinel2A_20170619_10m.tif  
Sentinel2A_20170619_10m_EMS_WGS84.tif
```

ATTENTION : ne pas écraser l'image fournie qui sert de *backup* en cas de problème avec les traitements.

Charger la nouvelle image. Combien contient-elle de bandes spectrales? Quelle est sa projection géographique? La visualiser en fausses couleurs et générer une transparence pour les pixels ayant une valeur nulle (0) pour les trois bandes affichées (Figure 1).

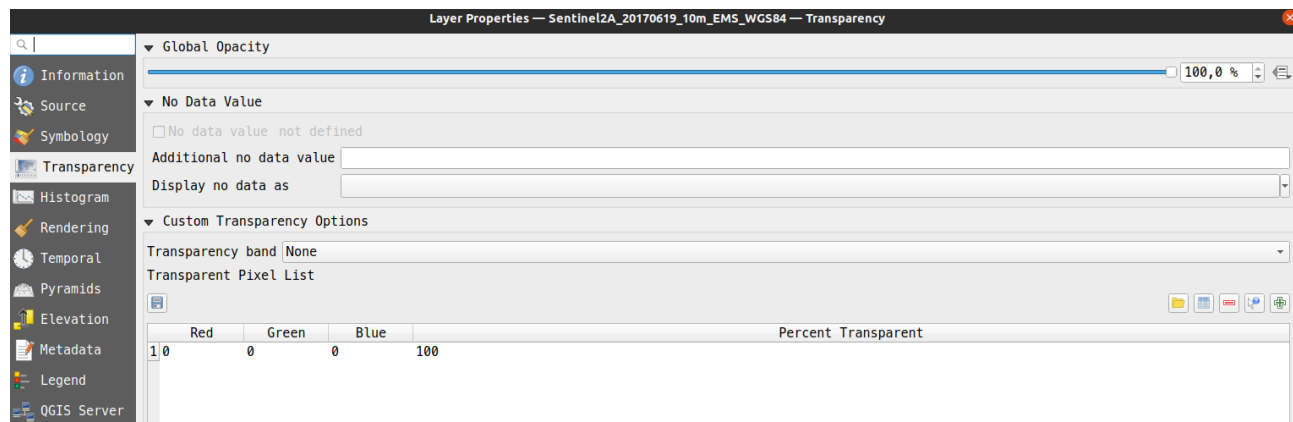


FIGURE 1 – Générer la transparence d'une image sur QGIS.

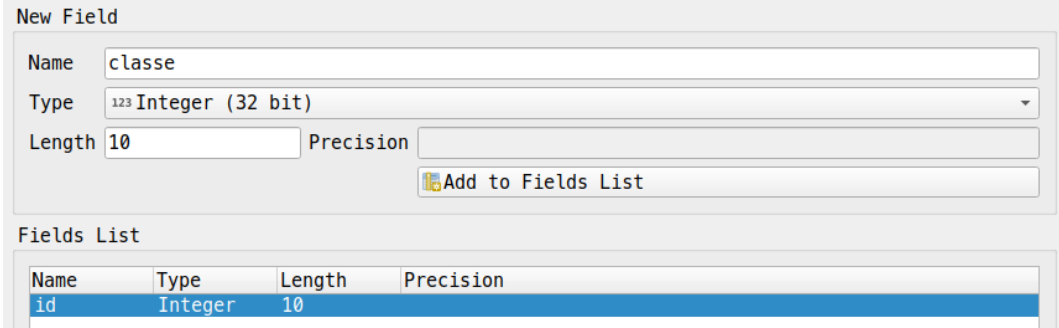
On peut aussi masquer automatiquement en rajoutant une bande *alpha* directement dans le raster. ajouter à la ligne de commande **-dstalpha** ou dans *Advanced Parameters* sur QGIS.

```
gdalwarp -cutline limite_EMS.shp -crop_to_cutline -dstalpha Sentinel2A_20170619_10m.tif  
Sentinel2A_20170619_10m_EMS_WGS84.tif
```

Maintenant, nous allons devoir créer un jeu d'entraînement pour nos modèles de classification. Créer un fichier vecteur (*shapefile*) ayant la même projection que l'image Sentinel-2 de

Strasbourg. Y ajouter une colonne (classe) pour obtenir l'information de la classe de chaque polygone (type entier) (Figure 2). Ne pas oublier de cliquer sur *Add to Fields List* et le type de géométrie (*Polygon*)! Ne pas se tromper dans la projection, SNAP souhaite du WGS84 (EPSG :4326).

Layer > Create Layer > New Shapefile layer



Name	Type	Length	Precision
id	Integer	10	

FIGURE 2 – Paramètres creation Shapefile.

L'objectif est maintenant de générer plusieurs échantillons d'entraînement pour le nombre de classes d'occupation des sols souhaitées. Pour ce TP, nous nous intéresserons à 4 classes d'occupation des sols : **urbain** (classe 1), **surfaces agricoles** (classe 2), **forêts** (classe 3) et **eau** (classe 4). Pour commencer la digitalisation, cliquer sur le crayon dans la barre d'outil et générer des zones d'entraînement pour chaque classe d'occupation des sols. Attention, respecter l'équilibre des classes, *c.a.d.* environ le même nombre de 'pixels' pour chaque classe (à l'oeil) (Figure 3).

Note : penser à activer *Shape Digitizing Toolbar* (clique droit sur la barre d'outil sur QGIS).

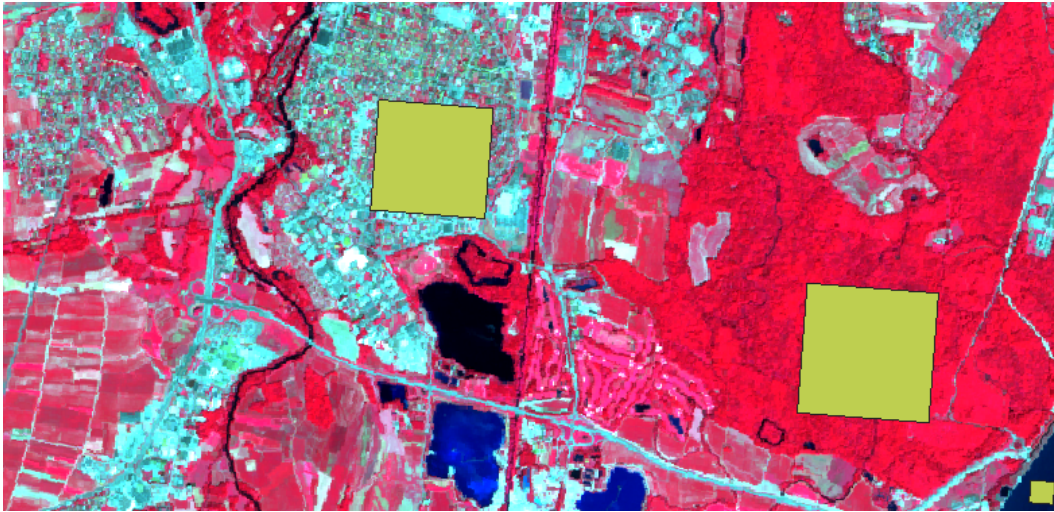


FIGURE 3 – Digitalisation de 'zones' d'entraînement sur Strasbourg.

Sur la digitalisation finale, appliquer un *Dissolve* sur la colonne *classe* pour regrouper les polygones des même classes en un polygone unique (avoir donc 4 multi-polygones pour 4 classes).

Autre solution : digitaliser directement sur SNAP. Sélectionner l'image à classer, *Vector > New Vector Data Container > le nom de votre classe à digitaliser (urbain, surfaces agricoles, forêts ou eau)*. Pas besoin d'importer de fichier vecteur dans SNAP pour la suite du TD.

Étape 2 : Entraînement des modèles

Lancer le logiciel **SNAP ESA**. Importer l'image *Sentinel2A_20170619_10m_EMS_WGS84.tif* et le *shapefile* avec les données de référence.

File > Import > Generic Formats > GeoTIFF / BigTIFF

File > Import > Vector Data > ESRI Shapefile

Accepter l'importation séparée des polygones (Figure 4).

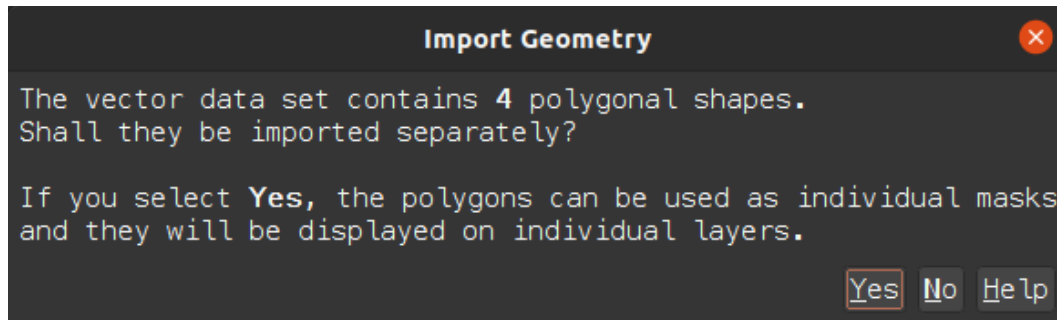
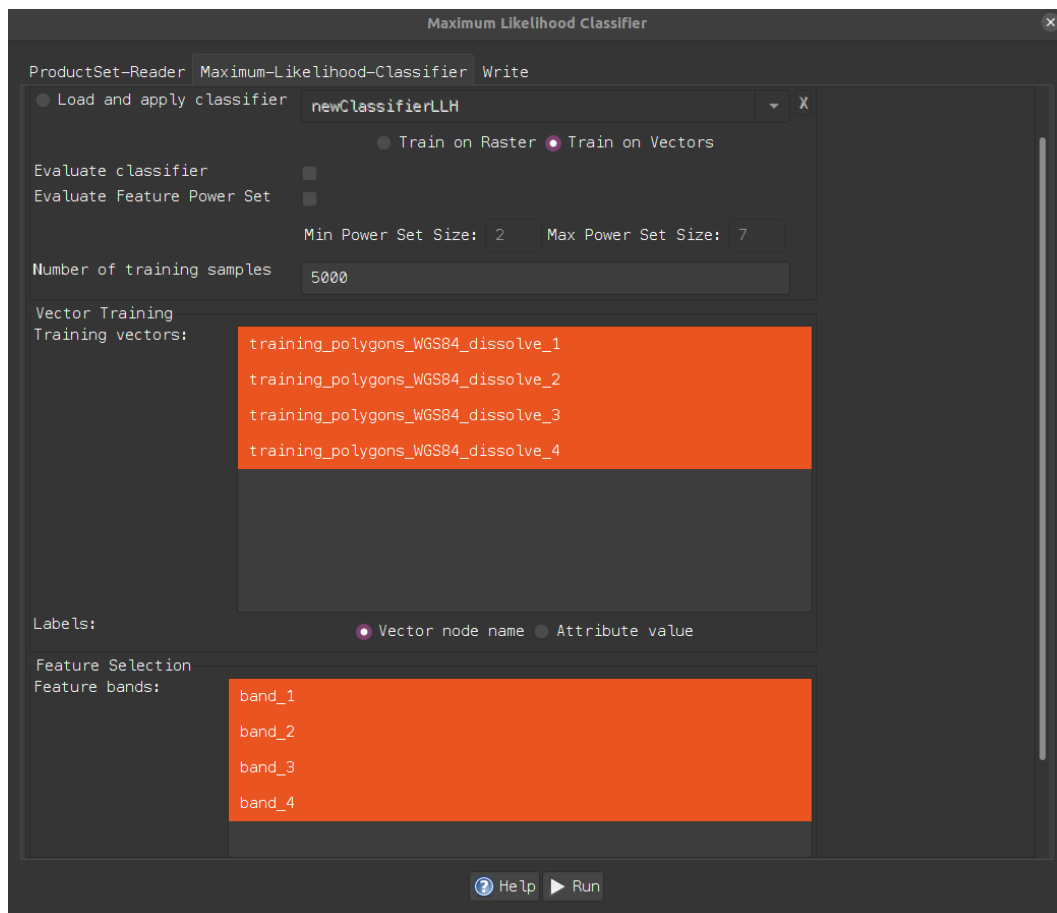


FIGURE 4 – Import des polygones.

Raster > Classification > Supervised Classification

Lancer l'apprentissage du modèle *Maximul Likelihood classification*. Ajouter l'image Sentinel-2 de Strasbourg. Sélectionner les paramètres, les 4 classes (*Vector training*) et les 4 bandes spectrales (Figure 5).

FIGURE 5 – Paramètres *Maximum Likelihood classification*

Note : Conserver le fichier d'évaluation du modèle. Le stocker sur le disque dur. Faire la même démarche pour *Random Forest*.

Étape 3 : Évaluer et comparer les classifications

Comparer les résultats. Quelle est le meilleur modèle ? Qualitativement ? Quantitativement ? Calculer un NDVI sur SNAP et relancer les classifications pour les deux modèles. Avons-nous de meilleurs ou de moins bons résultats ? Quelles sont les *features* (ou caractéristiques) les plus utilisées par le classifieur ? Exporter les résultats et faire des cartes d'occupation des sols sur QGIS en comparant les deux méthodes.

Essayer d'ajouter une ou deux classes supplémentaires.

BONUS 1 : Relancer l'algorithme *Random Forest* mais sur toute l'image Sentinel-2. Ajouter le NDVI, NDWI et BI (*Brightness Index*). Pour calculer le NDWI, il est nécessaire de rééchantillonner la bande de Moyen Infrarouge 1 à 10m de résolution spatiale.

BONUS 2 : Comparer avec une approche non-supervisée (K-Means) avec n clusters. Quelle est la meilleure approche ? Qu'est-ce qu'un cluster ?