

Télédétection Approfondissement

Projet 31 JANVIER 2025 / SANCHEZ Anna ~ POTTELET Thomas ~ VASSEL Noémie

Tables des matières

Introduction.....	3
Construction du masque raster.....	5
Pré-traitement.....	6
Analyse des échantillons.....	9
Production d'une d'essences forestières à l'échelle du pixel.....	14
Analyse des résultats.....	18
Conclusion.....	25

Introduction

Pour l'évaluation de l'UE de télédétection avancée, nous avons réalisé un projet comprenant différents traitements. L'objectif de ce projet est d'évaluer nos compétences de télédétection à l'issue de cette dernière année de master SIGMA. Pour ce faire, nous avons comme donnée primaire la BD Forêt® version 2.0, base de données contenant de nombreux éléments clés concernant l'espace forestier et les milieux naturels. L'objectif de ce projet est de savoir si cette BD peut être utilisée comme source d'échantillons de référence pour réaliser une classification supervisée de séries temporelles d'images Sentinel 2. Dans le cadre de ce projet, nous avons mis en place un dépôt Git avec une arborescence précise, conforme aux consignes imposées. Ce dépôt permet d'organiser les différentes étapes du traitement et d'assurer une traçabilité, une organisation et une répartition efficace du travail réalisé.

Le document présente l'ensemble des traitements effectués pour mener à bien cette étude. La première étape concerne la construction du masque raster, qui permet de limiter l'analyse aux zones forestières en s'appuyant sur la BD Forêt®. Cette étape a nécessité le filtrage des classes forestières pertinentes et l'application d'un masque spatial pour extraire uniquement les zones d'intérêt. Ensuite, un pré-traitement des images Sentinel-2 a été réalisé. Ce traitement a inclus la reprojection et le rééchantillonnage des bandes spectrales, le découpage des images en fonction de l'emprise d'étude et l'application du masque forêt. Un calcul des indices spectraux, notamment le NDVI, a également été effectué pour enrichir l'analyse des signatures spectrales.

Une fois les images préparées, nous avons procédé à la sélection des échantillons de référence. Cette phase a consisté à extraire et nettoyer les échantillons issus de la BD Forêt® afin d'assurer leur qualité et leur représentativité. Nous avons appliqué un filtrage des classes d'intérêt et

harmonisé les correspondances entre la BD Forêt® et les images Sentinel-2. L'analyse des échantillons a ensuite été menée afin d'évaluer leur cohérence et leur distribution spatiale. Nous avons généré des graphiques et des statistiques descriptives permettant de visualiser la variabilité spectrale des différentes classes forestières sur plusieurs dates d'observation Sentinel-2.

Sur la base des échantillons sélectionnés et validés, nous avons produit une première carte des essences forestières à l'échelle du pixel. La classification supervisée a été effectuée à l'aide d'un algorithme de Random Forest, entraîné avec les échantillons de référence, puis une matrice de confusion a été générée pour analyser les erreurs de classification.

Dans un second temps, nous avons généré une carte de classification à l'échelle des peuplements forestiers. Cette approche agrège les résultats obtenus au niveau du pixel pour produire une cartographie plus homogène et interprétable à l'échelle du territoire. L'utilisation des **statistiques zonales** et l'application de règles de décision ont permis de généraliser les résultats à partir des distributions des classes spectrales.

Enfin, nous avons conclu cette étude en analysant les résultats obtenus et en évaluant la pertinence de la BD Forêt® comme base d'échantillons de référence pour la classification des séries temporelles Sentinel-2. Nous avons également mis en avant les limites du projet, notamment les incertitudes liées à la qualité des échantillons et les variations spectrales intra-classe, ainsi que des pistes d'amélioration pour de futurs travaux.

Chaque partie du document détaille les étapes du code Python employé, accompagnées de cartes et de graphiques illustrant les résultats obtenus. Nous avons également intégré des diagrammes de flux permettant d'expliquer les méthodes utilisées et leur articulation tout au long du traitement.

Ci-dessous, nous présentons une arborescence récapitulant l'organisation des fichiers et des résultats attendus à la fin du projet.

Les scripts	<ul style="list-style-type: none"> ● my_function.py ● build_mask.py ● sample_curation.py ● sample_analysis_nb_sample.py ● sample_analysis_temp_signature.py ● sample_analysis_spectral_variability.py ● pre_traitement.py ● classification_pixel.py ● classification_stand.py
Les images	<ul style="list-style-type: none"> ● masque_foret.tif ● Serie_temp_S2_allbands.tif ● Serie_temp_S2_ndvi.tif ● Sample_BD_foret_T31TCJ.shp ● carte_essences_echelle_pixel.tif

Les graphiques

- diag_baton_nb_poly_by_class.png
- diag_baton_nb_pix_by_class.png
- violin_plot_nb_pix_by_poly_by_class.png
- temp_mean_ndvi.png
- diag_baton_dist_centroide_classe.png
- violin_plot_dist_centroide_by_poly_by_class.png

Construction du masque raster (build_mask)

Pour produire un masque raster des zones à classer et des zones à ne pas classer, nous avons suivi les étapes suivantes :

1. Définition des zones à inclure et à exclure :

- Les zones à classer incluent les polygones représentant des formations forestières de la BD Forêt.
- Les classes suivantes ont été exclues :
 - Lande
 - Formation Herbacée
 - Forêts ouvertes
 - Forêts fermées sans couvert arboré.

2. Caractéristiques du masque raster :

- Format : GeoTiff.
- Encodage : 8 bits.
- Emprise et résolution spatiale identiques à celles des images Sentinel-2 (appréciées après prétraitement).
- Chemin de sauvegarde : results/data/img_pretraitees/masque_foret.tif.
- Valeurs des pixels :
 - Zone de forêt : 1.
 - Zone hors forêt : 0.

3. Bibliothèques utilisées :

- GDAL et OGR de la bibliothèque OSGeo
- GeoPandas
- OS

4. Description du code :

- Nous avons chargé les fichiers shapefiles de la BD Forêt et de l'emprise d'étude à l'aide de Geopandas pour en extraire leurs données géographiques.
- Les limites géographiques (emprise) ont été calculées avec Geopandas, et les formations végétales correspondant aux codes TFV spécifiés ont été exclues grâce à des opérations de filtrage.
- Les dimensions du raster ont été déterminées en fonction de l'emprise et de la résolution des pixels, en utilisant les données extraites avec Geopandas.
- Un raster vide a été créé à l'aide de GDAL, avec une géotransformation et une projection adaptée à l'emprise d'étude.
- Une couche vectorielle temporaire a été initialisée avec OGR pour y insérer les polygones filtrés avec leurs attributs.
- La couche vectorielle temporaire a été rastérisée en utilisant GDAL pour assigner des valeurs aux zones incluses dans le masque.

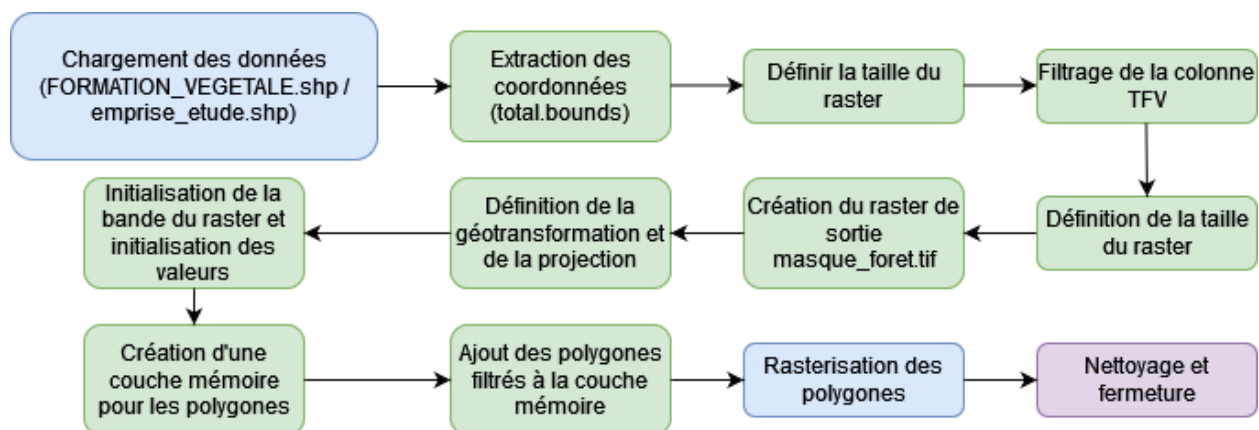


Figure 1 : Diagramme de flux du script "build_mask"

Pré-traitement

1. Prétraitement des bandes Sentinel-2

Les bandes Sentinel-2 sont reprojetées pour s'assurer qu'elles utilisent un système de coordonnées uniforme. Cela permet d'aligner les images et de garantir leur compatibilité pour les analyses ultérieures. Chaque bande est préparée individuellement avant les calculs.

2. Calcul des indices NDVI

À partir des bandes rouge et infrarouge proche, l'indice NDVI est calculé pour chaque date afin d'évaluer la végétation. Ce calcul permet de détecter les variations de densité et de santé des plantes. Les résultats sont sauvegardés sous forme d'images NDVI distinctes pour chaque période.

3. Découpage des images NDVI

Les images NDVI sont découpées selon l'emprise définie par un shapefile pour cibler une zone d'étude spécifique. Cette étape élimine les données hors de la région d'intérêt. Chaque fichier NDVI est réduit à la zone pertinente.

4. Fusion des images NDVI

Toutes les images NDVI découpées sont fusionnées en une seule image multi-bandes. Chaque bande représente le NDVI pour une date donnée. Cela permet d'étudier les variations temporelles de la végétation dans une seule structure de données.

5. Application d'un masque de forêt

Un masque forestier est appliqué pour ne conserver que les données relatives aux zones boisées. Les régions non-forestières sont exclues en les remplaçant par des valeurs NoData. Cela affine l'analyse en se concentrant sur l'écosystème cible.

6. Nettoyage des fichiers intermédiaires

Tous les fichiers générés lors des étapes intermédiaires (reprojections, découpes, masques) sont supprimés pour optimiser l'espace de stockage. Seuls les fichiers finaux utiles sont conservés. Cela assure un environnement de travail propre et organisé.

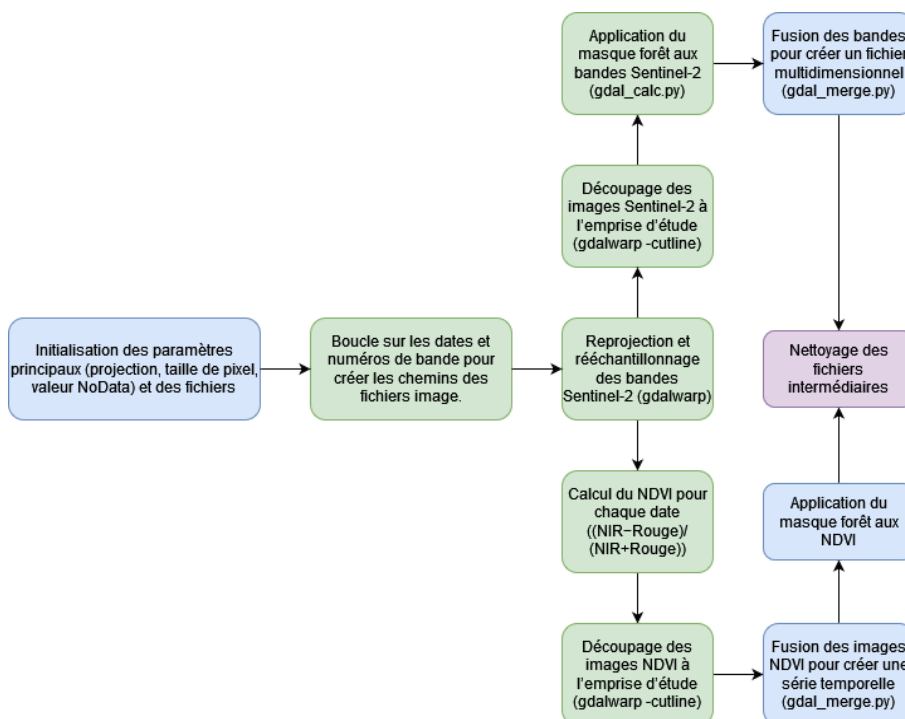


Figure 2 : Diagramme de flux du script "pre_traitements"

Sélection des échantillons (sample_curation)

1. Chargement des fichiers et définition des correspondances TFV

Le shapefile contenant les données de végétation (FORMATION_VEGETALE.shp) et celui représentant l'emprise d'étude (emprise_etude.shp) sont chargés dans des GeoDataFrames. Une correspondance entre les valeurs de la colonne TFV et leurs catégories spécifiques est définie à l'aide d'un dictionnaire.

2. Création et mappage des nouvelles colonnes

De nouvelles colonnes (Nom_pixel, Code_pixel, Nom_objet, Code_objet) sont ajoutées au shapefile en mappant les valeurs de TFV avec celles définies dans le dictionnaire. Les objets qui ne correspondent à aucune entrée dans le dictionnaire sont supprimés. Le fichier intermédiaire est ensuite sauvegardé.

3. Découpage des données selon l'emprise

Les données de végétation sont découpées pour ne conserver que celles qui intersectent avec l'emprise définie dans le shapefile d'emprise. Les géométries sont reprojetées si nécessaire pour assurer la cohérence entre les systèmes de coordonnées. Le résultat est sauvegardé comme shapefile final.

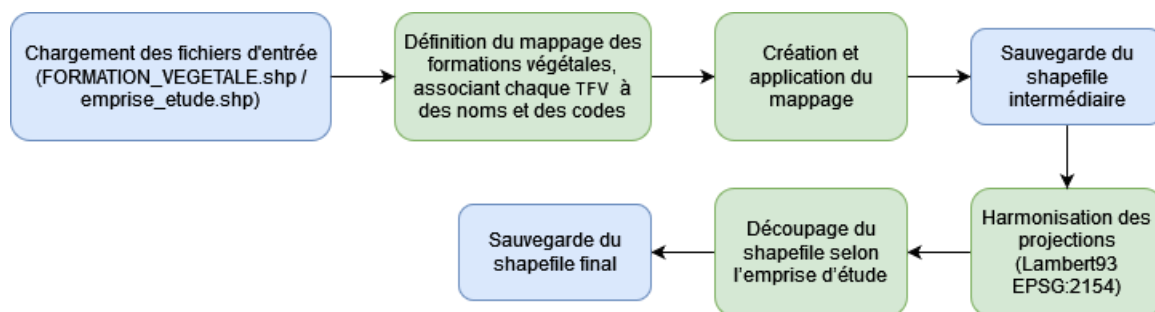


Figure 3 : Diagramme de flux du script "sample_curation"

Analyse des échantillons

1. Importation des bibliothèques

Nous avons importé les bibliothèques nécessaires à la manipulation des données géospatiales et raster, notamment geopandas, gdal et matplotlib. Cela nous a permis de gérer les fichiers shapefile, raster et de générer des visualisations.

2. Chargement des données

Nous avons chargé la couche vectorielle contenant les polygones et l'avons reprojetée en Lambert 93 afin de travailler avec des unités en mètres. Cette étape a assuré la cohérence des calculs spatiaux dans un système de coordonnées approprié.

3. Filtrage des données

Nous avons filtré les données en supprimant les polygones sans classe définie (NaN) ainsi que ceux ayant un code pixel nul. Cela nous a permis de conserver uniquement les entités pertinentes pour l'analyse.

4. Comptage des polygones par classe

Nous avons comptabilisé le nombre de polygones pour chaque classe (Code_pixel) en utilisant la fonction `value_counts()`. Ce comptage nous a donné une première indication de la répartition des classes dans la couche vectorielle.

5. Définition des dimensions du raster

Nous avons déterminé les dimensions du raster en nous basant sur l'étendue spatiale des polygones et une résolution de 10 mètres par pixel. Ces dimensions ont permis de créer une grille raster adaptée à nos données.

6. Création du raster temporaire

Nous avons créé un raster vide au format GeoTIFF avec une seule bande stockée en entiers 32 bits. Ce raster a servi de support pour la conversion des données vectorielles en données matricielles.

7. Lecture et comptage des pixels par classe

Nous avons ouvert le raster et parcouru ses données pour compter le nombre de pixels correspondant à chaque classe. Cette opération a permis d'établir une correspondance entre les polygones et les pixels dans le raster.

8. Calcul de la surface et estimation des pixels

Nous avons calculé la surface de chaque polygone et estimé le nombre de pixels qu'il devrait couvrir en divisant son aire par la surface d'un pixel. Cela a fourni une approximation du nombre de pixels attendus par polygone.

9. Visualisation des résultats

Nous avons généré un diagramme en boîte (boxplot) afin de visualiser la distribution du nombre de pixels estimés pour chaque classe. Cette représentation graphique a permis d'identifier les variations et éventuelles anomalies dans les données.

10. Nettoyage des fichiers temporaires

Nous avons supprimé le raster temporaire une fois le traitement terminé afin de libérer de l'espace disque et optimiser l'utilisation des ressources.

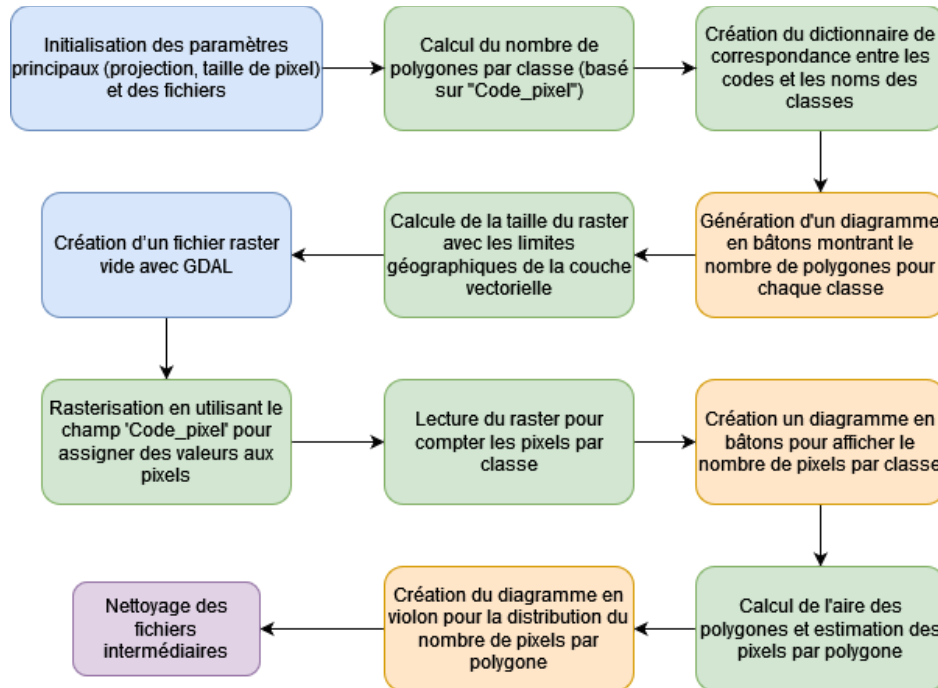


Figure 4 : Diagramme de flux du script "sample_analysis_nb_sample"

1. Importation des bibliothèques

Nous avons importé les bibliothèques nécessaires à la manipulation des données géospatiales et raster, notamment numpy, pandas, geopandas, gdal et matplotlib. Ces outils nous ont permis de gérer des fichiers raster, des données tabulaires et des géométries.

2. Définition des chemins des fichiers

Nous avons défini les chemins vers les fichiers de données, notamment un raster (Serie_temp_S2_ndvi.tif) et un shapefile (Sample_BD_foret_T31TCJ.shp). Cette étape a permis de centraliser l'accès aux données et d'assurer la cohérence des traitements.

3. Chargement et filtrage des données vectorielles

Nous avons chargé le shapefile contenant les polygones représentant différentes classes forestières. Ensuite, nous avons filtré ces données pour ne conserver que certaines classes spécifiques, à savoir : "Chêne", "Robinier", "Peupleraie", "Douglas", "Pin laricio ou pin noir", et "Pin maritime".

4. Chargement du raster

Nous avons chargé le fichier raster contenant les données de télédétection associées à l'indice NDVI. Cette étape a permis d'extraire les valeurs des pixels aux emplacements des polygones sélectionnés.

5. Extraction des centroïdes des polygones

Nous avons calculé les centroïdes des polygones du shapefile afin d'obtenir un point unique pour chaque entité. Ces centroïdes ont ensuite été convertis en coordonnées raster, permettant d'extraire les valeurs du raster aux positions exactes des polygones.

6. Conversion des coordonnées en indices de pixels

Nous avons défini une fonction pour convertir les coordonnées du monde réel en indices de pixels du raster. Cela nous a permis d'associer chaque polygone à la valeur d'un pixel correspondant dans le raster NDVI.

7. Définition des dates d'analyse

Nous avons défini une série de dates correspondant aux acquisitions du satellite Sentinel-2 : "26-03-2022", "05-04-2022", "14-07-2022", "22-09-2022", "11-11-2022" et "19-02-2023". Ces dates ont été utilisées pour suivre l'évolution des valeurs NDVI dans le temps.

8. Calcul des statistiques par classe et par date

Nous avons calculé la moyenne et l'écart-type des valeurs NDVI pour chaque classe de forêt et pour chaque date. Ces statistiques ont permis d'analyser la variabilité des indices de végétation en fonction des espèces et des saisons.

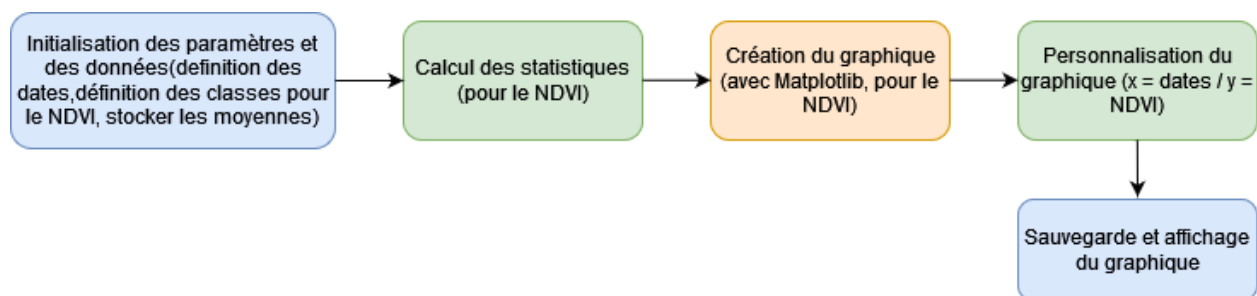


Figure 5 : Diagramme de flux du script "sample_analysis_temp_signature"

1. Importation des bibliothèques

Nous avons importé les bibliothèques nécessaires à la manipulation des données géospatiales et raster, notamment geopandas, gdal, numpy, pandas et matplotlib. Ces outils nous ont permis de charger, analyser et visualiser les données issues des fichiers shapefile et raster.

2. Chargement des données vectorielles et création des fichiers de sortie

Nous avons chargé le shapefile contenant les polygones forestiers et défini un répertoire de sortie pour enregistrer les figures et résultats.

3. Lecture et extraction des valeurs du raster NDVI

Nous avons chargé la première bande du raster NDVI et extrait ses valeurs sous forme d'un tableau (ndvi_array). Les dimensions et la géométrie du raster ont également été récupérées afin de faciliter les traitements spatiaux.

4. Définition des classes spectrales

Nous avons défini des classes spectrales spécifiques pour analyser la répartition des valeurs NDVI. Ces classes ont été utilisées pour caractériser différentes catégories de végétation et leurs signatures spectrales.

5. Calcul de la distance moyenne au centroïde par classe

Nous avons calculé la distance moyenne entre les polygones de chaque classe et leur centroïde. Cette mesure a permis d'analyser la distribution spatiale des entités appartenant à une même catégorie forestière.

6. Analyse des variations spectrales à l'échelle des polygones

Nous avons étudié les variations spectrales à l'échelle de chaque polygone en comparant les valeurs NDVI autour de leur centroïde. Cette analyse a permis d'observer les disparités internes à chaque entité géographique.

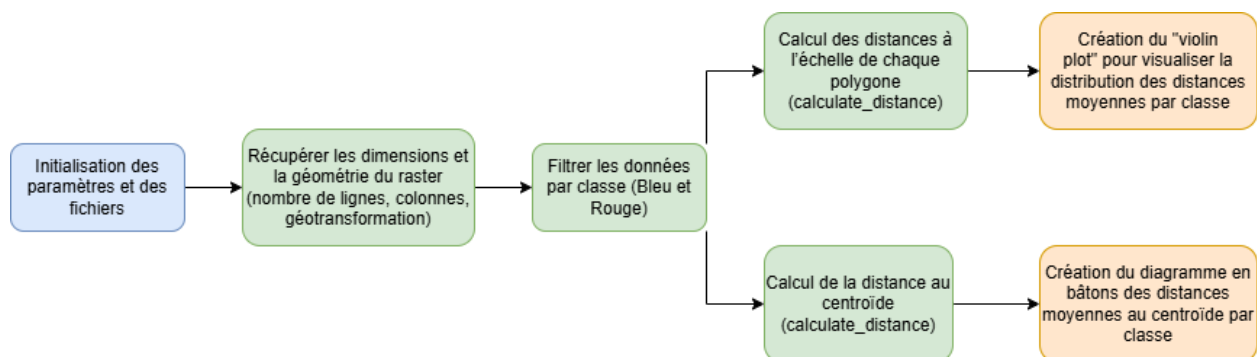


Figure 6 : Diagramme de flux du script "sample_analysis_spectral_variability"

Production d'une d'essences forestières à l'échelle du pixel

Importation des bibliothèques

Le script commence par l'importation des bibliothèques nécessaires, notamment NumPy, Pandas, GeoPandas, GDAL, Matplotlib, ainsi que Scikit-learn. Ces bibliothèques permettent de manipuler les données géospatiales et d'entraîner un modèle de classification.

Définition des paramètres et chargement des données

Le chemin vers le shapefile contenant les données d'échantillons (Sample_BD_foret_T31TCJ.shp) et celui du raster (Serie_temp_S2_allbands.tif) sont spécifiés. Un raster temporaire est également

créé pour stocker les valeurs des classes (raster_code_pixel.tif). Un champ ID est généré dans le shapefile pour identifier chaque échantillon de manière unique.

Extraction des échantillons d'apprentissage

Les échantillons sont extraits en associant les pixels du raster aux classes définies dans la BD Forêt®. Les caractéristiques (X) et les labels de classification (Y) sont ensuite préparés. Ces données sont divisées en ensemble d'apprentissage et de test à l'aide de la fonction `train_test_split()`.

Entraînement du modèle de classification

Un Random Forest Classifier est utilisé pour la classification des pixels. Le modèle est entraîné avec une profondeur maximale de 50 et un échantillonnage à 75 % par arbre. L'option `oob_score=True` est activée pour estimer la performance hors échantillon.

Validation croisée et évaluation des performances

Une validation croisée est mise en place avec une Stratified K-Fold (5 divisions répétées 30 fois). La précision globale (accuracy) et le F1-score sont enregistrés pour évaluer la robustesse du modèle.

Calcul et analyse des performances

Une matrice de confusion est générée pour identifier les erreurs de classification. Un graphique d'estimation de la qualité par classe est également affiché, avec des barres d'erreur représentant la variabilité des performances sur les différentes itérations de validation croisée.

Prédiction et production de la carte classifiée

Le modèle est appliqué à l'ensemble de l'image raster afin de prédire les classes d'essences forestières à l'échelle du pixel. Les prédictions sont ensuite enregistrées sous forme de raster GeoTIFF (carte_essences_echelle_pixel.tif).

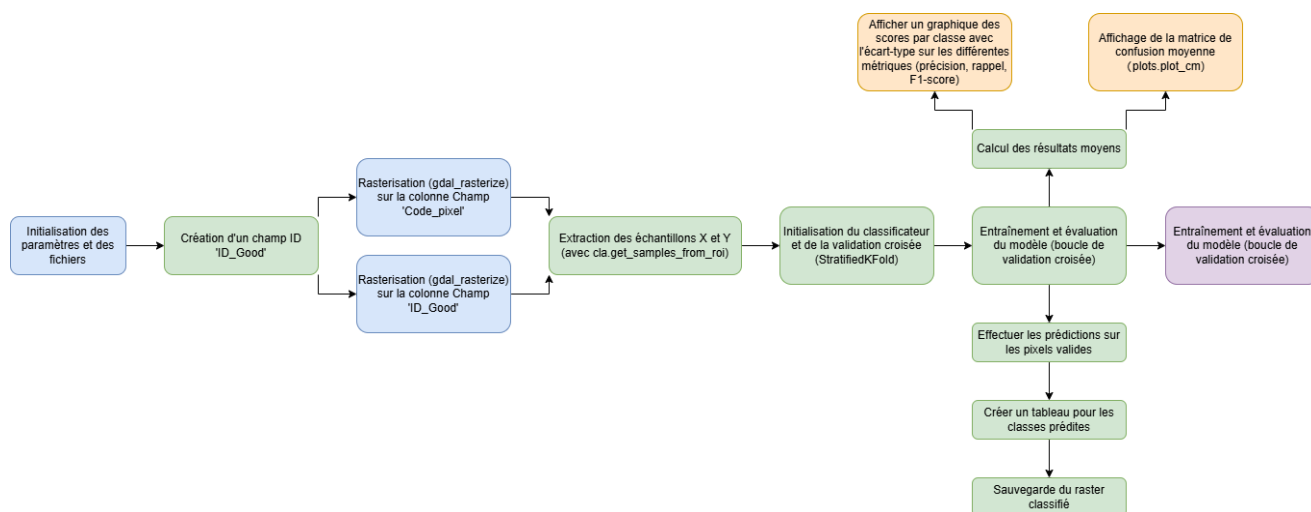


Figure 7 : Diagramme de flux du script "classification_pixel"

Production d'une carte à l'échelle des peuplements

Importation des bibliothèques

Nous avons importé les bibliothèques nécessaires à l'analyse des données géospatiales et à la classification des polygones. Ces bibliothèques incluent geopandas pour la manipulation des données géospatiales, numpy pour les opérations mathématiques, rasterstats pour effectuer des statistiques zonales, gdal pour le traitement des rasters, et sklearn.metrics pour évaluer la qualité de la classification à travers des métriques telles que la matrice de confusion.

Chargement des données

Nous avons chargé un fichier shapefile (Sample_BD_foret_T31TCJ.shp) contenant les polygones forestiers en utilisant geopandas. Ce fichier contient des informations géométriques et attributaires des zones d'étude, telles que les types de couverture forestière. Le raster représentant les classes d'essences a été également chargé à l'aide de gdal.

Définition de la fonction de classification

Une fonction de classification des polygones a été définie pour assigner un code de peuplement à chaque polygone en fonction des statistiques zonales et de sa surface. Cette fonction analyse la proportion des différentes classes de couverture du sol à l'intérieur des polygones et applique une série de règles basées sur la surface et la composition des classes spectrales. Les classes de feuillus et de conifères sont notamment prises en compte pour attribuer un code à chaque polygone.

Extraction des statistiques zonales

Grâce à la méthode zonal_stats(), nous avons extrait les statistiques zonales pour chaque polygone en fonction du raster des classes d'essences. Ces statistiques indiquent la distribution des classes spectrales dans chaque polygone et servent de base à la classification des peuplement forestiers.

Calcul de la surface des polygones

La surface de chaque polygone a été calculée en hectares en divisant l'aire des polygones (exprimée en mètres carrés) par 10 000. Cette surface a été utilisée comme critère dans les règles de décision de la fonction de classification, permettant de différencier les petits et grands peuplements.

Application de la classification aux polygones

La fonction de classification a été appliquée à chaque polygone du shapefile en s'appuyant sur les résultats des statistiques zonales et la surface des polygones. Un code de peuplement a été attribué à chaque polygone en fonction des règles définies dans la fonction classif_polygone.

Mise à jour du shapefile et sauvegarde

Une fois la classification appliquée, le champ `code_predi` a été ajouté au shapefile avec les codes de peuplement prédit. Le shapefile mis à jour a été sauvegardé, permettant de visualiser les résultats de la classification au sein de l'environnement SIG.

Évaluation de la classification

Pour évaluer la qualité de la classification, nous avons calculé la matrice de confusion en comparant les prédictions de classification (`code_predi`) avec les codes réels issus du champ `Code_objet`. Cette évaluation a permis de mesurer la précision de la classification et de détecter d'éventuelles erreurs ou incohérences.

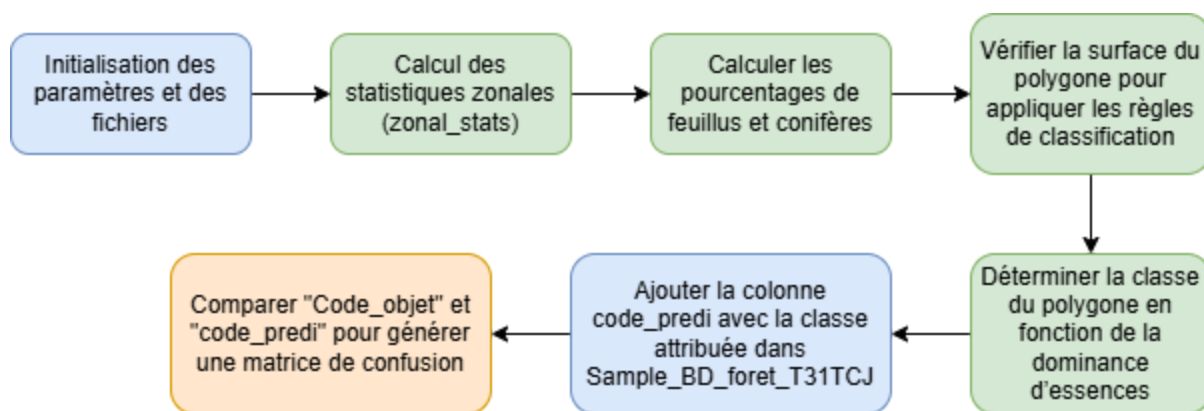


Figure 8 : Diagramme de flux du script "classification_stand"

Analyse des résultats

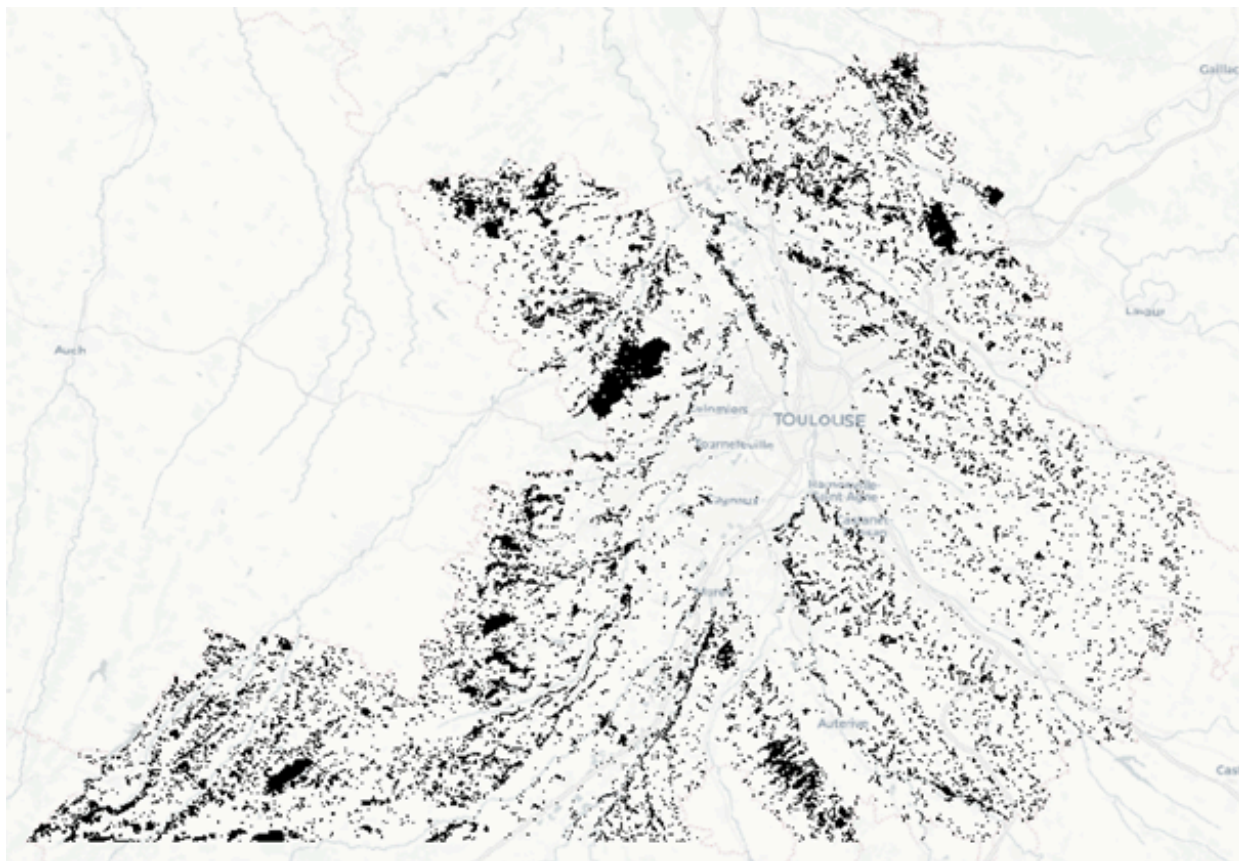


Figure 9 : Masque de la forêt (en noir)

Le résultat correspond à un masque (*figure 9*) excluant les zones qui ne seront pas traitées en premier lieu. Les zones sombres correspondent aux zones préservées pour la suite de notre étude, et les zones claires sont évincées de cette dernière.

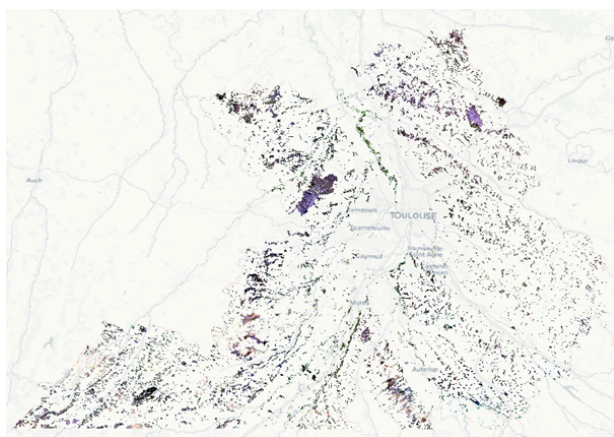


Figure 10 : Carte du raster regroupant les 60 bandes

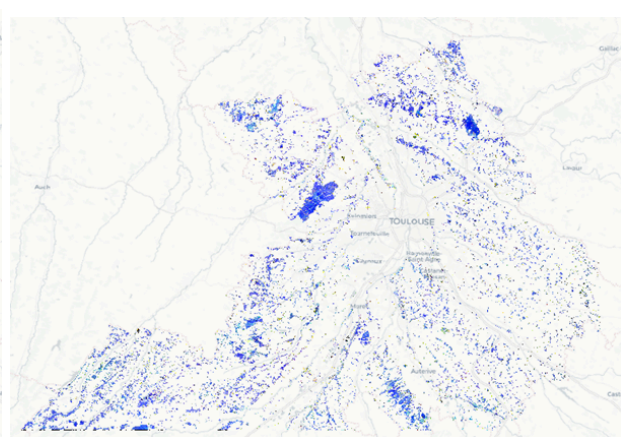


Figure 11 : Carte du raster NDVI pour les 6 dates

Le résultat des pré-traitements effectués est simple : deux cartes (*figure 10 et 11*) , l'une présentant l'image avec chacune des bandes qu'elle possède, et la seconde montrant la bande NDVI, qui indique la présence plus ou moins dense de végétation dans les zones sélectionnées.

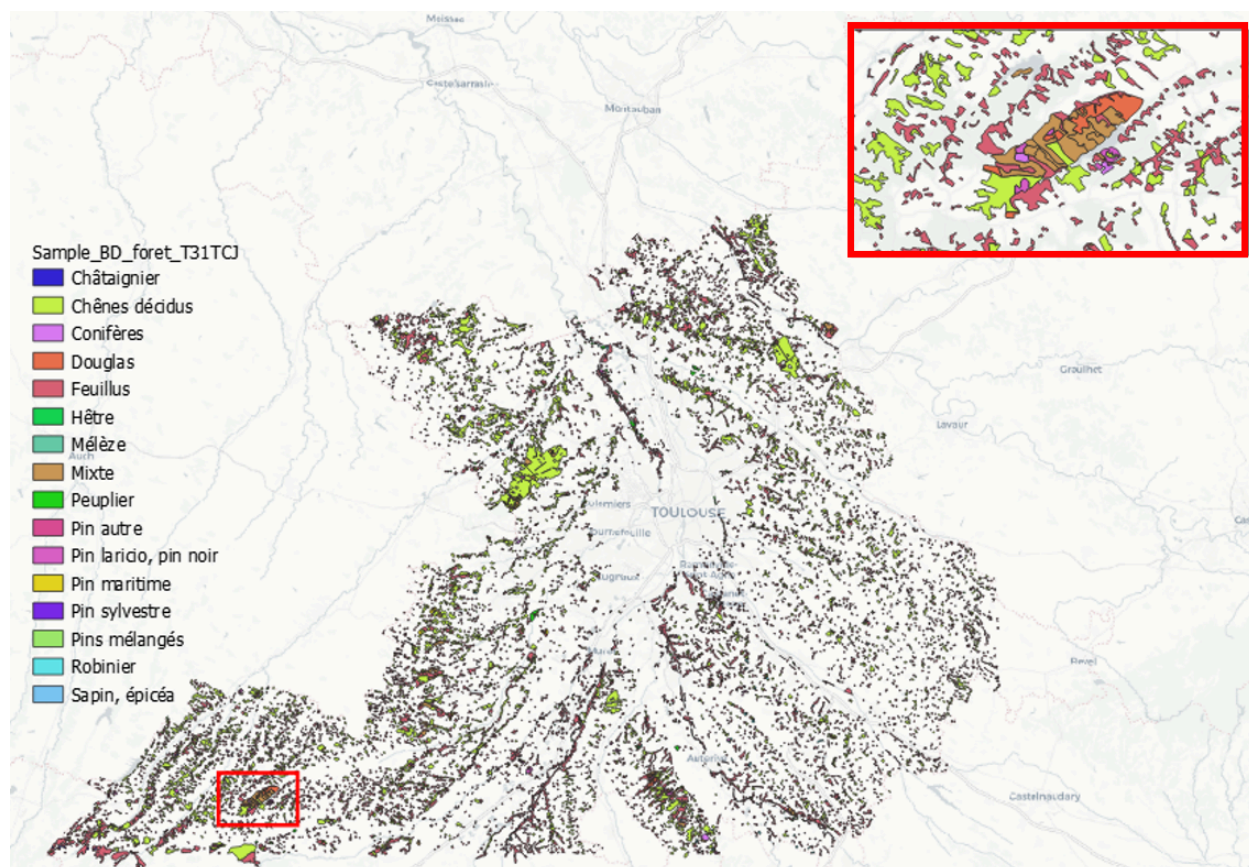


Figure 12 : Carte des échantillons

Le résultat de cette analyse se manifeste par une carte (*figure 12*), représentant chacune des essences présente sur notre territoire précédemment défini, avec une symbologie différente pour chacune d'elles. En haut à droite, un cadre présente le zoom effectué sur une partie précise de la carte; cette zone à été sélectionnée afin d'effectuer différentes analyses. Elle à été choisie pour la diversité des classes représentées en son sein, afin d'obtenir des résultats plus précis et donc plus fiables.

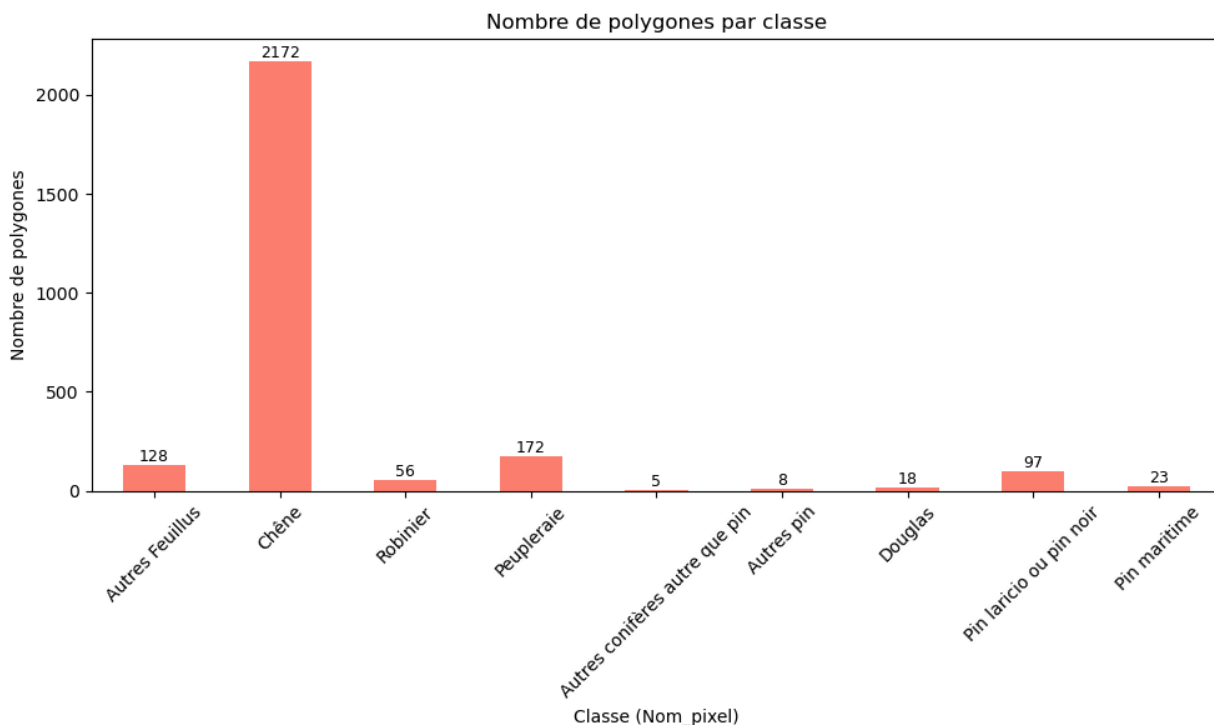


Figure 13 : Diagramme du nombre de polygones par classe

Ce graphique représente le nombre de polygones par classe de végétation. La classe "Chêne" possède 2 172 polygones, ce qui correspond à la plus grande valeur parmi toutes les autres classes. Cela signifie que cette espèce est très fragmentée spatialement, avec de nombreux petits polygones et une couverture très morcelée.

Les classes "Peupleraie" (172), "Autres feuillus" (128) et "Pin laricio ou pin noir" (97) sont les plus présentes après le chêne. Certaines classes sont plus rares, comme "Autres conifères autre que pin" (5 polygones), "Autres pin" (8 polygones) et "Douglas" (18 polygones).

Ce graphique révèle que le chêne est non seulement dominant en surface (comme vu sur les précédents graphiques), mais aussi en nombre de polygones. Les autres classes ont des occurrences bien plus faibles, suggérant soit une couverture limitée, soit une cartographie plus homogène (moins de polygones pour une même surface).

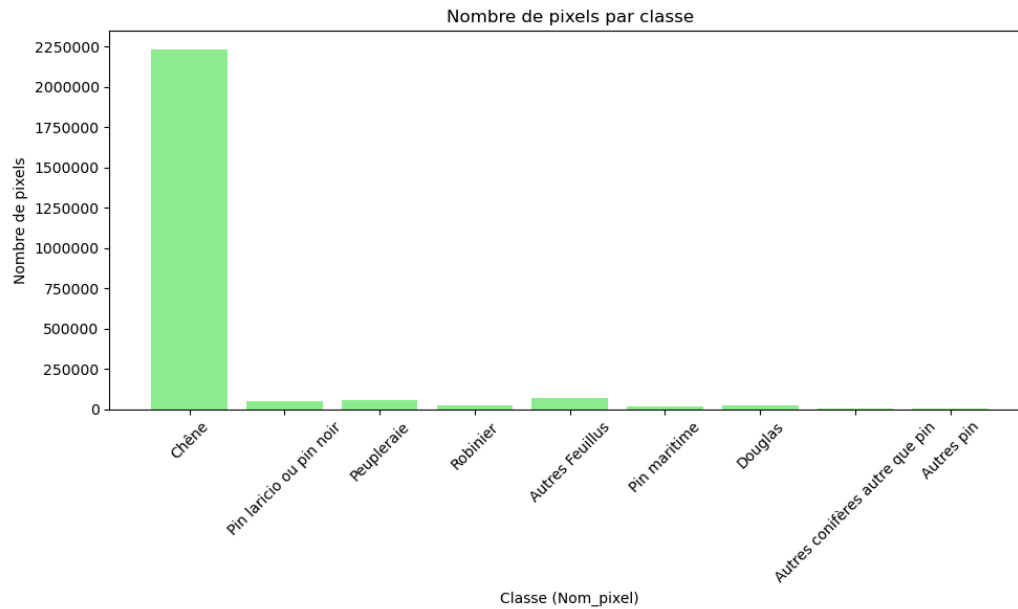


Figure 14 : Diagramme du nombre de pixels par classe

Le graphique ci-dessus nous montre que la végétation étudiée est principalement composée de chêne (plus de 2 250 000 pixels), avec une très faible diversité et répartition des autres espèces.

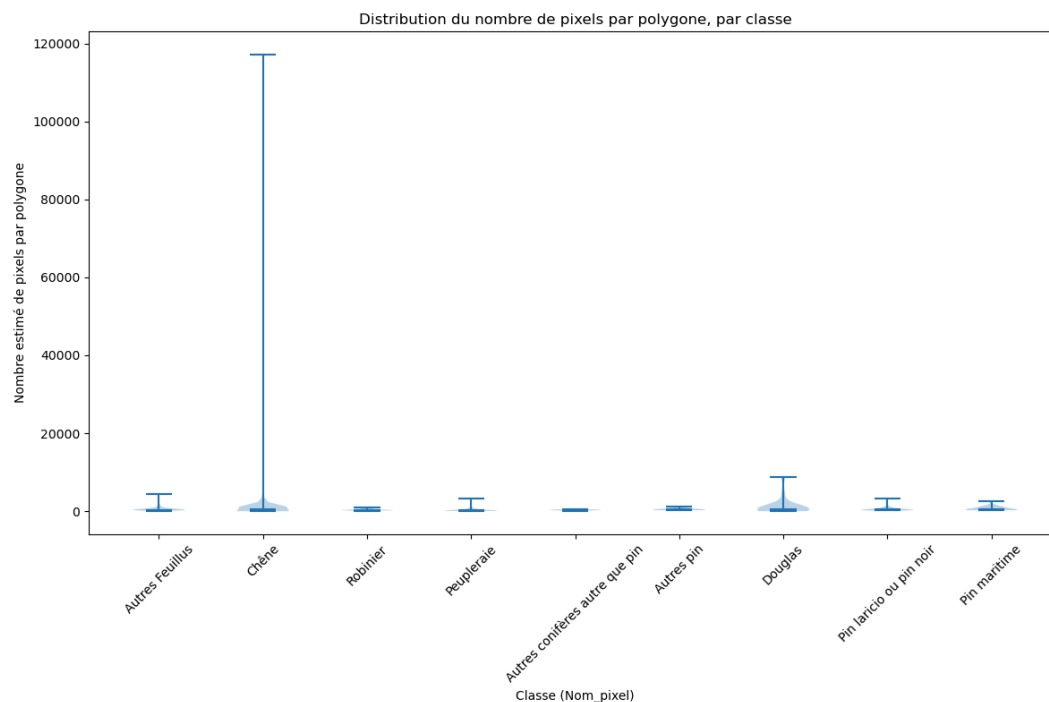


Figure 15 : Violin plot du nombre de pixels par polygones

Ce graphique est un violin plot représentant la distribution du nombre de pixels par polygone, pour chaque classe de végétation. Nous pouvons constater que la classe "Chêne" a une distribution très étendue avec une valeur maximale qui atteint environ 120 000 pixels par polygone; cela signifie que les polygones sont très étendus pour cette classe, ce qui peut indiquer

une dominance spatiale du chêne sur de vastes zones. Comparées au chêne, les autres classes ont des distributions plus compactes, avec un nombre de pixels nettement inférieur par polygone. Par exemple, les classes comme "Douglas", "Peupleraie" et "Autres feuillus" présentent une légère dispersion, mais restent bien en dessous du chêne.

Ce résultat confirme la dominance du chêne, mais nous informe surtout qu'il existe une très grande variabilité dans la taille des polygones de cette classe. Les autres classes semblent plus homogènes en termes de répartition spatiale.

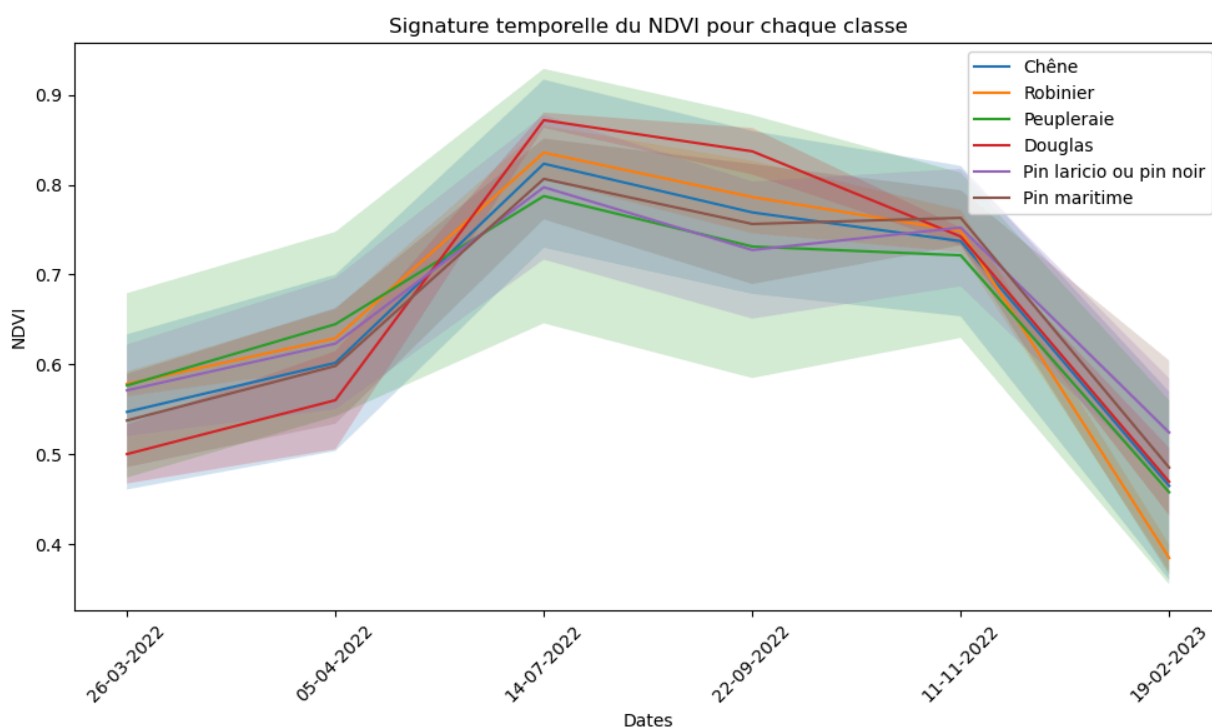


Figure 16 : Diagramme de la signature temporelle du NDVI pour chaque classe

Ce graphique nous montre l'évolution du NDVI (indice de végétation normalisé) au fil du temps pour différentes classes d'arbres (ex. Chêne, Robinier, Peuplier, etc.). Nous pouvons observer une tendance générale où le NDVI augmente au printemps et en été (avril à juillet), atteint un pic en septembre, puis diminue en automne et hiver (novembre et février). Les tendances générales sont plutôt similaires pour chacune des essences présentes, avec une légère démarcation de la classe 'Douglas' entre le mois d'avril et le mois de novembre, où son NDVI est plus élevé que la moyenne des autres classes présentes.

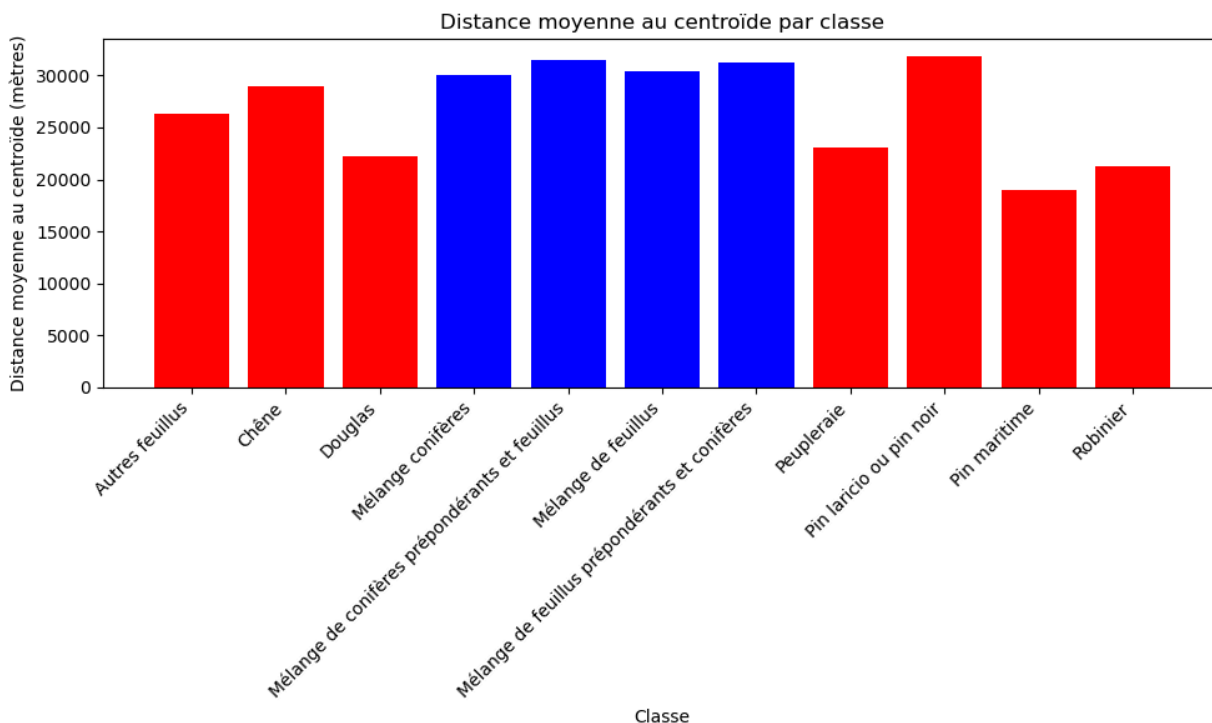


Figure 17 : Diagramme de la distance moyenne au centroïde, par classe

Le graphique nous montre la distance moyenne des éléments de chaque classe par rapport à leur centroïde (point moyen de répartition dans l'espace). Les classes de couleur rouge ont des distances plus faibles que celles en bleu, ce qui suggère qu'elles sont plus homogènes, et donc spatialisées de manière compacte. Les classes "Mélange de conifères" et "Mélange de feuillus" présentent des distances plus élevées, ce qui peut indiquer que ces dernières sont plus espacées. Cela signifie que certaines classes sont plus regroupées géographiquement, tandis que d'autres sont plus dispersées dans l'espace.

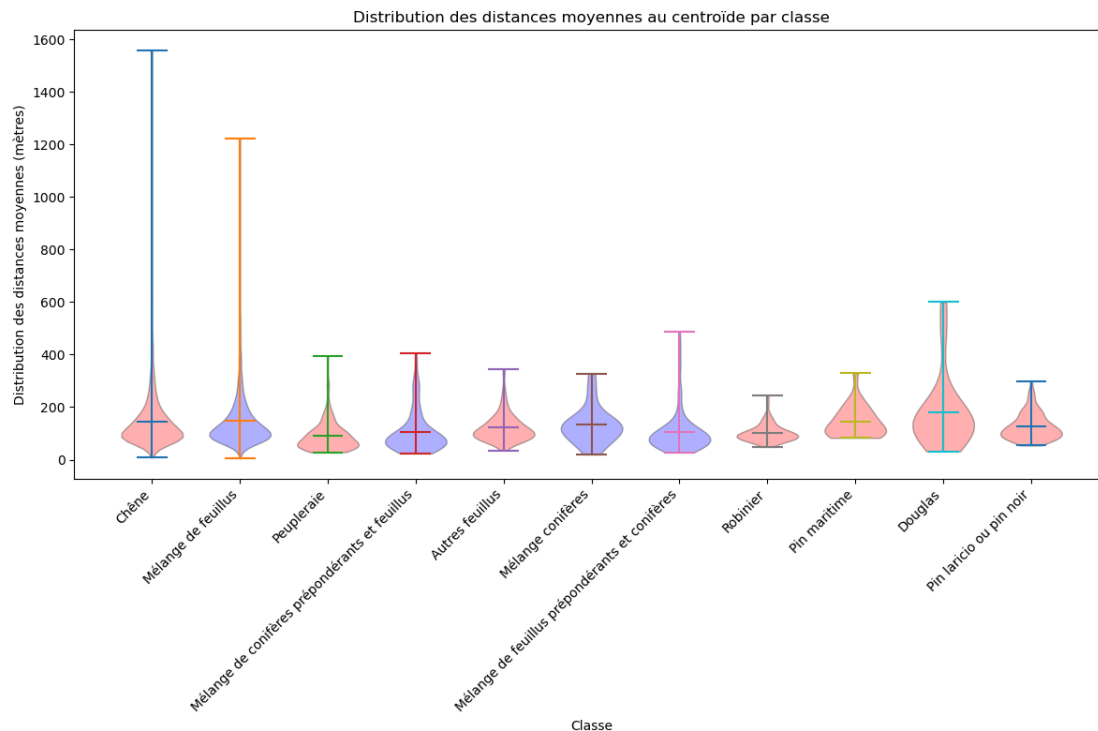


Figure 18 : Diagramme de la distribution des distances moyennes au centroïde, par classe

Ce graphique nous montre la distribution des distances moyenne au centroïde pour chaque classe sous forme de diagramme en "violin" (violin plot). Un violon large indique une grande variabilité des distances au centroïde, tandis qu'un violon étroit montre une répartition plus homogène. Nous pouvons remarquer que certaines classes comme Chêne et Mélange de feuillus ont une large distribution avec des valeurs extrêmes, indiquant une grande diversité spatiale. D'autres classes ont une distribution plus resserrée, suggérant qu'elles sont plus uniformément réparties autour de leur centroïde.

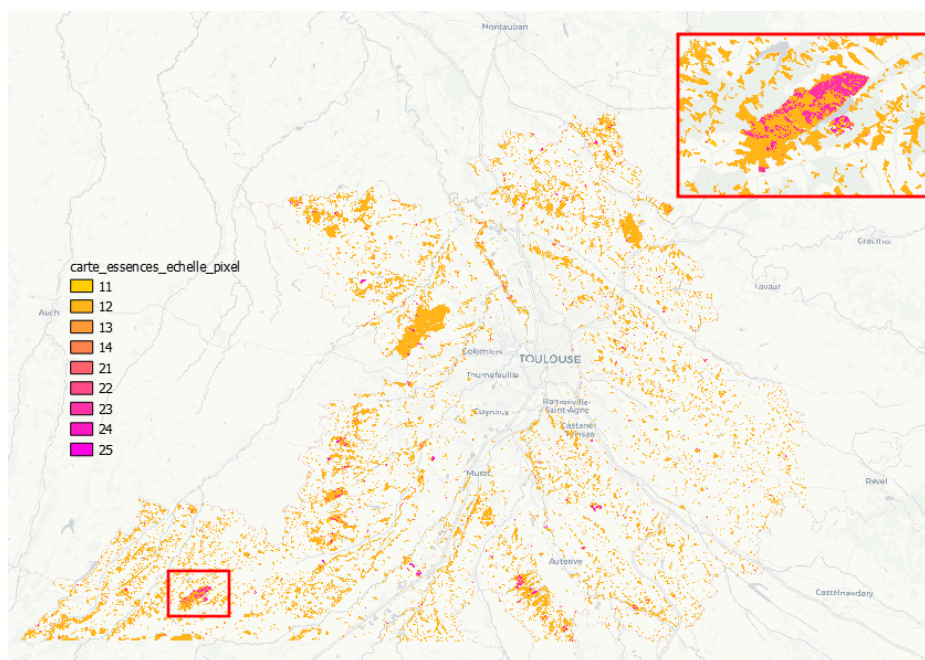


Figure 19 : Carte des essences d'arbres, à l'échelle d'un pixel

Le résultat de cette analyse (figure 19) vient corroborer les différents résultats précédemment obtenus sur nos graphiques : la classe 12, correspondant au chêne, est bel et bien dominante sur l'ensemble des zones étudiées, malgré une certaine dispersion de ses nombreux polygones. Nous pouvons également noter la représentation significative de la classe 23 (douglas), avec moins de fragmentation et de plus grands polygones compacts.

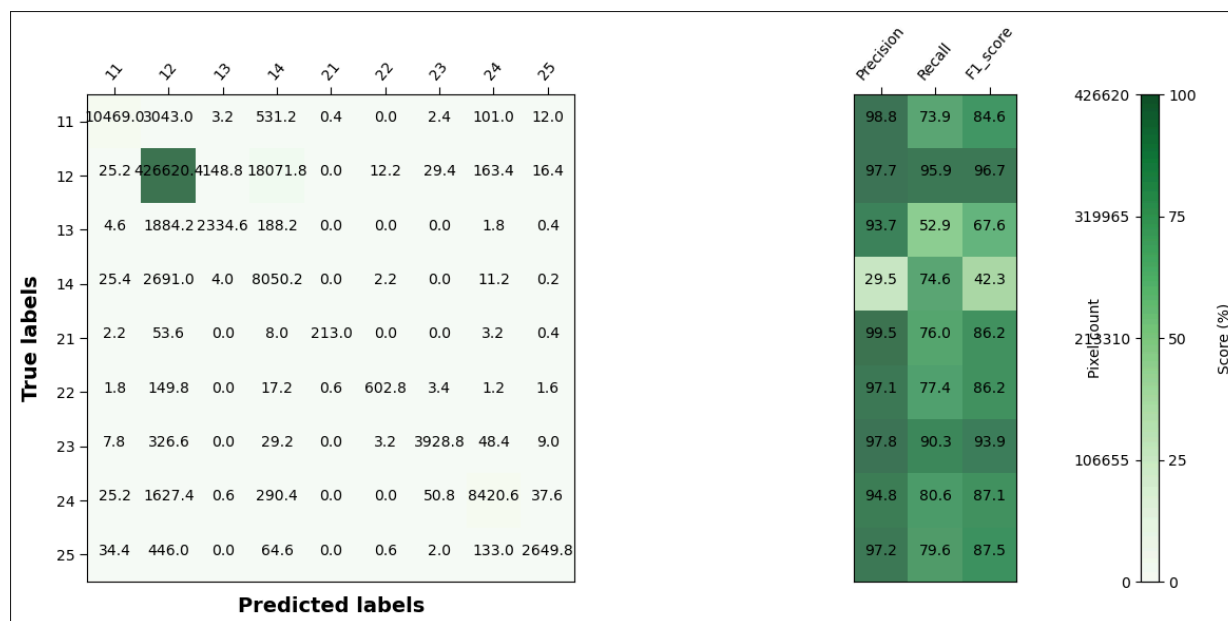


Figure 20 : Matrice de confusion des essences forestières

Cette matrice de confusion représente la classification des essences forestières, avec les labels réels en ligne et les prédictions en colonne. La diagonale principale indique les classifications correctes, tandis que les valeurs hors diagonale montrent les erreurs. Une forte concentration sur une case (426620) suggère qu'une essence est bien identifiée. À droite, une carte de chaleur affiche la précision, le rappel et le F1-score par classe, où certaines essences sont bien distinguées (~98% de précision), tandis que d'autres montrent une confusion accrue (F1-score bas de 42.3). La barre latérale indique la répartition des pixels classifiés, illustrant les classes dominantes. Cette analyse permet d'évaluer la performance du modèle et d'identifier les essences nécessitant des ajustements pour améliorer leur classification.

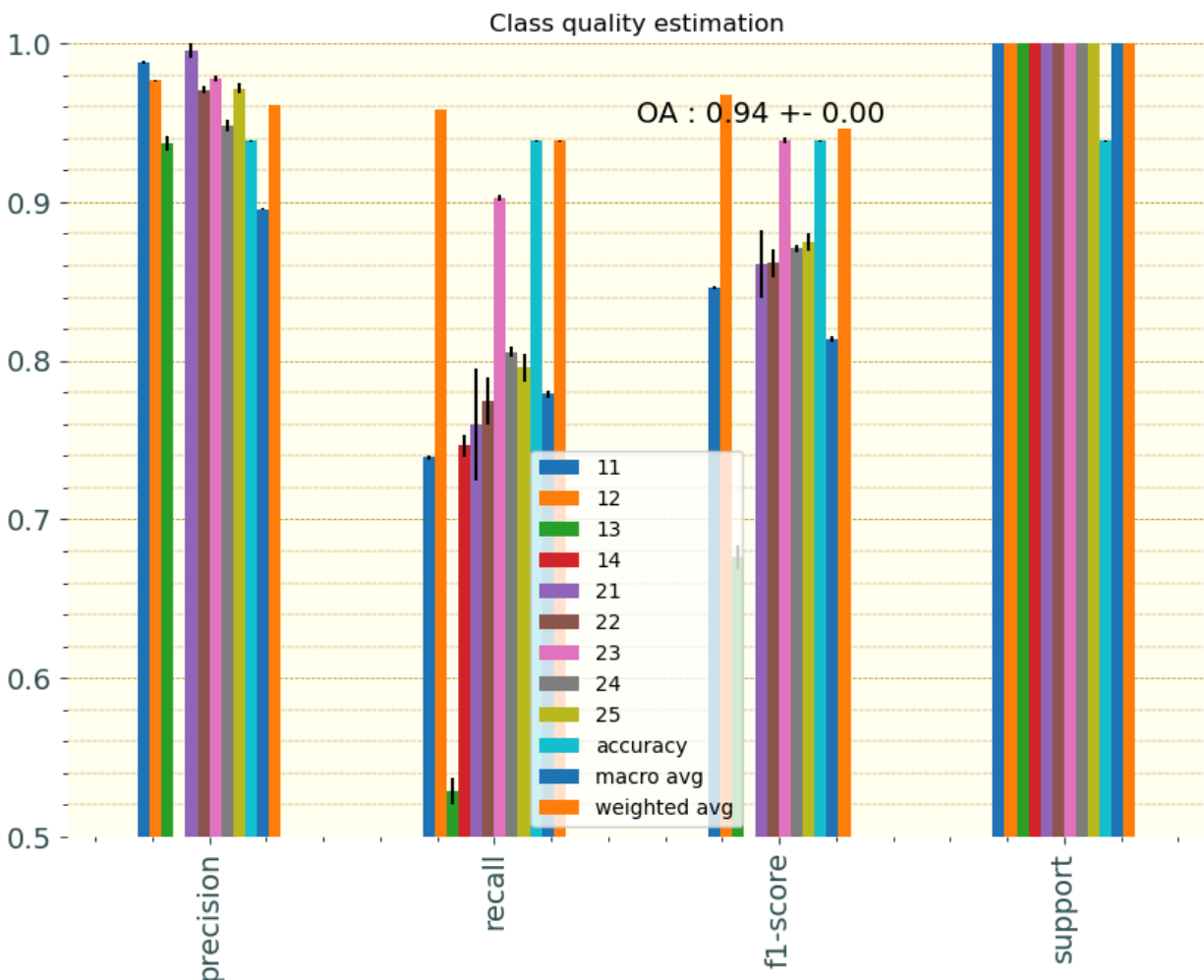


Figure 21 : Diagramme de la qualité des essences forestières

Le diagramme ci-dessus représente une estimation de la qualité de classification des essences forestières selon plusieurs métriques. L'OA (Overall Accuracy) : 0.94 ± 0.00 , nous indique une précision globale élevée du modèle. Les différentes métriques évaluées (sur l'axe des abscisses) sont les suivantes :

- Precision (Précision) : Proportion des prédictions positives correctes.
- Recall (Rappel) : Proportion des éléments pertinents correctement identifiés.
- F1-score : Moyenne harmonique entre la précision et le rappel.
- Support : Nombre d'exemples par classe dans les données.

Nous pouvons donc noter que les performances varient légèrement selon les classes : en effet, plusieurs classes sont représentées (comme les classes 11, 12, 13, ..., 25). Des statistiques globales sont aussi affichées : accuracy, macro avg, et weighted avg. La plupart des classes ont une précision proche de 1.0, ce qui signifie que peu de faux positifs sont présents à l'issue de l'étude. Nous pouvons également noter que le support est élevé et uniforme pour toutes les classes. Néanmoins, nous pouvons remarquer une certaine variabilité du rappel et du F1-score : par

exemple, certaines classes montrent une légère baisse en rappel et F1-score. Il y a également de petites barres d'erreur indiquant des variations mineures, mais globalement, la performance reste stable. Toutefois, la classe 12 se démarque; en effet, celle-ci semble avoir des scores parfaits en précision, rappel et F1-score. Pour conclure ce diagramme montre que le modèle de classification des essences forestières est très performant (OA = 94%), avec une excellente précision et un rappel élevé, bien que quelques classes puissent avoir une légère variabilité en rappel et F1-score.

Discussion

Matrice de confusion de la carte des essences forestières

Le modèle affiche une précision élevée, dépassant 95 % pour plusieurs classes. Cela signifie que les prédictions du modèle sont en grande majorité correctes lorsque celui-ci identifie une classe donnée. Certaines classes ont un rappel plus faible (~40-70%), ce qui signifie que le modèle ne détecte pas toutes les occurrences réelles de ces classes.

Le F1-score est globalement satisfaisant (~80-90%) pour plusieurs classes, ce qui indique un bon équilibre entre précision et rappel. Cependant, une classe en particulier (la 14) affiche un F1-score faible (39,7%), ce qui indique un problème de classification pour cette catégorie.

Des erreurs hors diagonale indiquent que le modèle confond certaines classes proches. Si ces confusions concernent des catégories critiques (ex: classification médicale ou industrielle), cela peut poser problème.

La classe 14 est problématique :

- Précision : 27,9 % → signifie que lorsque le modèle prédit cette classe, il se trompe souvent.
- Rappel : 69 % → montre que le modèle identifie une bonne partie des occurrences réelles de cette classe, mais avec de nombreuses erreurs de classification.
- Cela peut être dû à un déséquilibre des données ou à un manque de différenciation avec d'autres classes.

Les données d'entrée, soit les images Sentinel-2 et la BD Forêt® version 2.0, présentent plusieurs limites. Bien que Sentinel-2 offre une couverture régulière, sa résolution de 10 mètres ne permet pas une analyse précise des peuplements forestiers, notamment ceux contenant plusieurs essences. De plus, la BD Forêt® n'étant pas mise à jour régulièrement, elle peut ne plus refléter la réalité actuelle des terrains, introduisant ainsi des erreurs dans la classification.

L'agrégation des pixels à l'échelle des peuplements peut également lisser des différences importantes, réduisant la précision de la cartographie. Pour améliorer ces résultats, il serait pertinent d'intégrer des données plus précises comme le LiDAR, d'utiliser des images à plus haute résolution, et d'élargir la zone d'étude pour accroître la diversité des échantillons de référence.

Malgré ces limites, les résultats obtenus sont satisfaisants. La BD Forêt® peut être utilisée comme source d'échantillons pour une classification supervisée avec Sentinel-2. Les performances montrent qu'il est possible de cartographier les peuplements à grande échelle, bien qu'il soit nécessaire d'affiner les modèles pour mieux capturer la variabilité intra-peuplement.

Ces cartes sont utilisables pour la gestion forestière, la cartographie de la biodiversité et la prévention des risques, bien qu'elles nécessitent des données complémentaires pour des analyses plus fines.

Conclusion

Ce projet de télédétection avancée nous a permis d'explorer le potentiel des séries temporelles Sentinel-2 pour la classification des essences forestières en s'appuyant sur la BD Forêt® comme source d'échantillons de référence. À travers un processus comprenant le prétraitement des images, la sélection et l'analyse des échantillons, ainsi que la classification à l'échelle du pixel et du peuplement, nous avons pu évaluer la faisabilité et la précision d'une telle approche.

Les résultats obtenus montrent que la BD Forêt® peut être utilisée pour entraîner un modèle de classification supervisée, bien que certaines limites aient été identifiées. En effet, les erreurs de classification observées, notamment pour certaines essences, soulignent la nécessité d'améliorer la qualité et l'homogénéité des données d'apprentissage. De plus, la résolution spatiale de Sentinel-2, bien qu'adaptée aux analyses à large échelle, peut s'avérer insuffisante pour distinguer avec précision certaines classes d'essences à fine échelle.

Pour améliorer ces résultats, plusieurs pistes peuvent être envisagées : l'intégration de données complémentaires, telles que l'imagerie LiDAR ou des images à plus haute résolution, l'affinement des algorithmes de classification, et l'élargissement des échantillons pour couvrir une plus grande variabilité des formations forestières.

En dépit de ces défis, cette étude démontre l'intérêt des méthodes de télédétection pour la gestion et la cartographie des forêts. Nos résultats ouvrent la voie à des applications concrètes dans le suivi des écosystèmes forestiers, la gestion des ressources naturelles et la modélisation des dynamiques de végétation. Une approche combinant plusieurs sources de données et des méthodologies avancées permettrait d'affiner encore davantage ces analyses pour une meilleure compréhension des dynamiques forestières à différentes échelles.