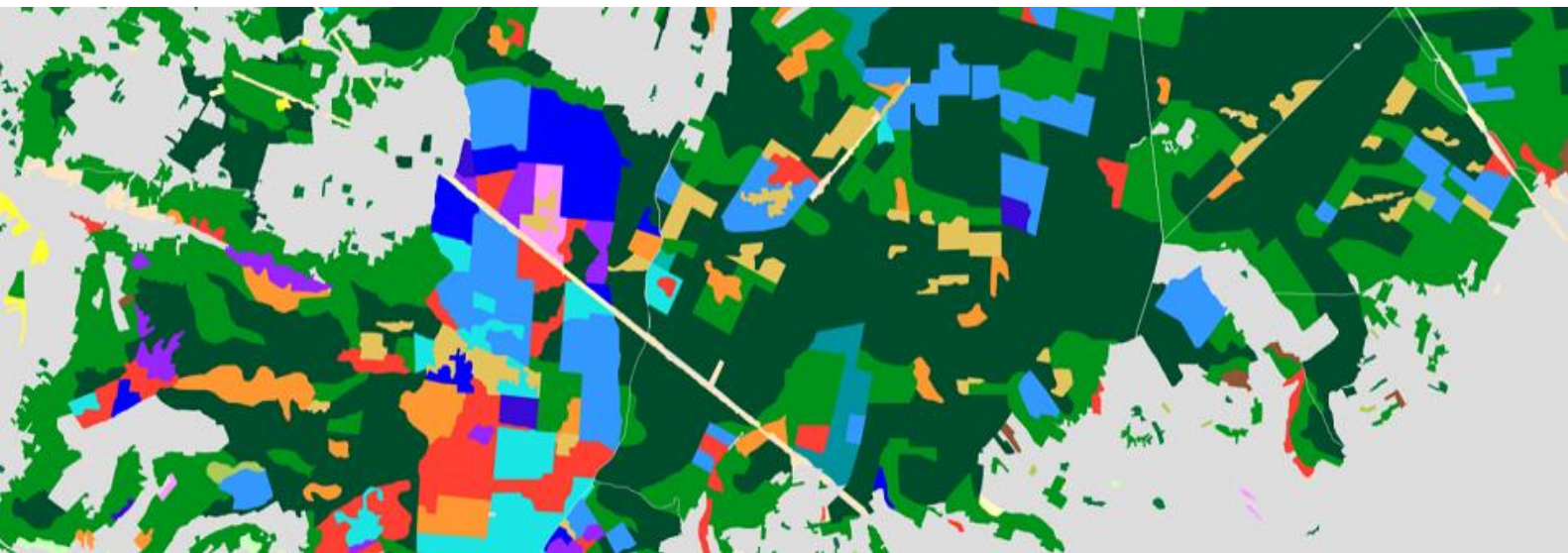


ANALYSE AVANCÉE D'IMAGES MULTISPECTRALES : MÉTHODES ET APPLICATIONS



901_21 - Traitement d'image satellite : approfondissement

La BD Forêt version 2.0 comme source d'échantillons
de référence pour réaliser une classification supervisée
de séries temporelles d'images Sentinel-2

Réalisé par Hugo CADARIO, Mélanie ECHEVARRIA et Lucas MAISSIN

Sous la supervision de Marc LANG

Promotion SIGMA 2023/2025

Table des matières

I.	Rappel d'objectifs.....	3
II.	Description de la méthode	3
2.1.	Construction d'un masque à partir de la BD Forêt (5.2)	3
2.2.	Sélection des échantillons (5.3).....	4
2.3.	Analyse du nombre d'échantillons (5.4.1)	4
2.4.	Pré-traitement des images (5.5).....	5
2.5.	Etude de la phénologie des classes de la BD Forêt (5.4.2).....	6
2.6.	Analyse de la variabilité spectrale de la BD Forêt (5.4.3).....	7
2.7.	Production des cartes finales (5.6.3).....	9
III.	Analyse des résultats	9
3.1.	Analyse des échantillons (5.4.1).....	10
3.2.	Analyse de la phénologie des peuplements purs	11
3.3.	Analyse de la variabilité spectrale de la BD Forêt	12
IV.	Discussion des résultats.....	13
	Bibliographie	16

Figure 1 -	Diagramme de flux de la construction d'un masque à partir de la BD Forêt	3
Figure 2 -	Diagramme de flux de la sélection des échantillons à partir de la couche formation végétale.....	4
Figure 3 -	Diagramme de flux de l'analyse du nombre d'échantillons	5
Figure 4 -	Diagramme de flux du prétraitement des images	6
Figure 5 -	Diagramme de flux de l'analyse de la phénologie des peuplements purs	7
Figure 6 -	Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle de l'image entière.....	8
Figure 7 -	Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle du polygone.....	8
Figure 8 -	Diagramme de flux de la classification par pixel	9
Figure 9 -	Histogramme du nombre de polygone par classe	10
Figure 10 -	Histogramme du nombre de pixels par classe	10
Figure 11 -	Distribution logarithmique du nombre de pixels par polygone et par classe	10
Figure 12 -	Signature temporelle de la moyenne et l'écart-type du NDVI par classe	11
Figure 13 -	Distance moyenne au centroïde des NDVI par classe à l'échelle de l'image entière	12
Figure 14 -	Distances moyennes au centroïde des NDVI par classe à l'échelle des polygones.....	13

I. Rappel d'objectifs

Ce projet s'inscrit dans l'UE 901_21, "*Traitement d'image satellite : approfondissement*" du Master 2 SIGMA. L'objectif est de déterminer si "*la BD Forêt® version 2.0 peut être utilisée comme source d'échantillons de référence pour réaliser une classification supervisée de séries temporelles d'images Sentinel 2*". Pour cela, nous allons détailler la méthode avec des diagrammes de flux, puis analyser les résultats et proposer une discussion de ces derniers.

Ce travail a été réalisé sur VSCode en ligne dans Onyxia. Il est possible de récupérer les informations du git en entrant la commande suivante dans le terminal :

```
git clone "https://github.com/Bernardodufutur65/901_21_CEM.git"
```

Pour récupérer les images, entrer cette seconde commande dans le terminal :

```
mc cp -r s3/mecva/diffusion/images /home/onyxia/work/data
```

II. Description de la méthode

Ce projet nécessite de nombreuses étapes correspondant à des scripts python dans l'objectif de répondre à la problématique initiale. Par conséquent, la démarche proposée sera représentée par chaque étape réalisée, et sera accompagnée de diagrammes de flux reprenant les traitements effectués.

2.1. Construction d'un masque à partir de la BD Forêt (5.2)

Premièrement, il faut construire un masque permettant de définir l'emprise des zones à classer, à savoir tous les polygones de la BD Forêt sauf les polygones de type lande et formation herbacée qui ne sont pas de la forêt.

Les premières étapes de ce code sont la lecture du fichier vecteur. Nous sélectionnons les zones de forêt grâce au code TFV et à une liste des codes à exclure. On crée un champ dans le gdf appelé "forest_zon" qui contient les valeurs 1 ou 0. La sélection prend une valeur de 1 et le reste 0. Pour ne pas modifier le fichier source je me permet de l'enregistrer en shape. Faire un enregistrement de la couche me permettra peut-être plus tard d'être réutilisé. En utilisant GDAL il est possible de rasteriser notre vecteur avec les mêmes paramètres que l'image souhaitée. Le champ qu'il faut utiliser est donc "forest_zon" pour avoir des valeurs de pixel à 1 pour les forêt et 0 pour le reste.

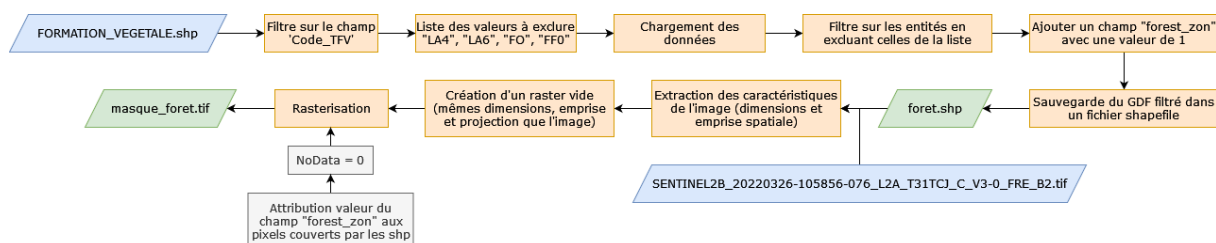


Figure 1 - Diagramme de flux de la construction d'un masque à partir de la BD Forêt

A titre d'information, nous vous donnerons le temps d'exécution des scripts afin d'avoir une idée du temps que cela prendra. Celui-ci s'est exécuté en 5 secondes.

2.2. Sélection des échantillons (5.3)

La production d'un jeu d'échantillons est essentiel pour la suite du projet car ils vont permettre de réaliser divers traitements comme l'analyse du nombre d'échantillons, l'analyse de la phénologie des peuplements purs, l'analyse de la variabilité spectrale de la BD Forêt, ainsi que la production des cartes finales.

Le script réalisant la création du jeu d'échantillons est "*sample_curation.py*". Ces échantillons seront formés à partir de la couche "*FORMATION_VEGETALE.shp*". Cette dernière contient les informations sur le type de formation végétale ainsi que l'essence de ces dernières, ce qui permet notamment de faire le lien avec les classes de la figure 2 du cours.

Les bibliothèques nécessaires à l'exécution du code sont : GeoPandas et os.

Les consignes imposent de créer un jeu d'échantillons inclus dans l'emprise des images S2, c'est pourquoi nous avons d'abord intersecté la couche des formations végétales avec l'emprise d'étude puis avec la couche "*foret.shp*". Cette dernière contient les informations du champ "Code_TFV" filtré selon les exigences de la construction du masque, mais au format vecteur.

Ensuite, nous avons initié un dictionnaire comprenant les valeurs du champ "TFV" associé aux codes du tableau de la figure 2 du cours. Cela permet de créer une seule et même table comportant les informations du champ "classif objet", avec le type de formation végétale et leur code associé.

Ces informations ont été implémentées dans le GeoData Frame "*gdf_intersect_emprise_masque_FV*" créé à la suite de l'intersection réalisée en première étape.

Pour répondre aux attentes du projet et "nettoyer" le fichier, les champs ne correspondant pas au nom et au code de la formation végétale ont été supprimés du GDF, ainsi que les entités vides (= None).

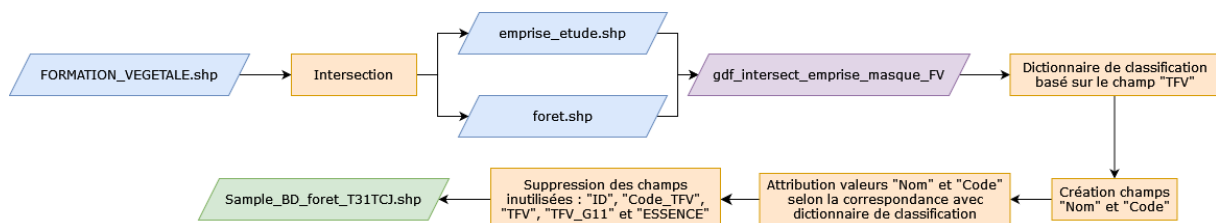


Figure 2 - Diagramme de flux de la sélection des échantillons à partir de la couche formation végétale

Le fichier en sortie se nomme "*Sample_BD_foret T31TCJ.shp*" et est stocké dans ce chemin d'accès : `/home/onyxia/work/results/data/sample'`.

Ce script a été exécuté en 34 secondes.

2.3. Analyse du nombre d'échantillons (5.4.1)

Il est important d'analyser les premier résultats que l'on a obtenus pour confirmer si toutes les données semblent correctes et théoriquement justes. Pour cela nous devons réaliser un diagramme en bâton. Ce premier diagramme est assez simple à réaliser, il suffit juste de charger la couche vecteur créer dans la partie 5.3 et vérifier que le champ "code" est bien renseigné et complet. Le diagramme est réalisé avec matplotlib et est ordonnée par nombre de polygones.

Pour réaliser le second diagramme, nous allons faire une boucle qui vient lire le raster en entrée ainsi que vecteur. Nous sélectionnons le code des polygones et nous venons tout simplement compter le

nombre de pixel qui vaut 1. Ainsi chaque code est fait un par un. Le diagramme est fait de la même manière que le précédent.

Enfin le dernier diagramme est un violin plot qui représente la distribution du nombre de pixel par polygone. Pour cela nous devons faire la somme des pixel par classe et dessiner le violin plot. Cependant le résultat n'est pas satisfaisant (partie analyse). C'est pourquoi nous avons réalisé un second diagramme avec des logarithmes. Cela atténue les trop grande variation et écart ce qui permet une meilleur visualisation des résultats.

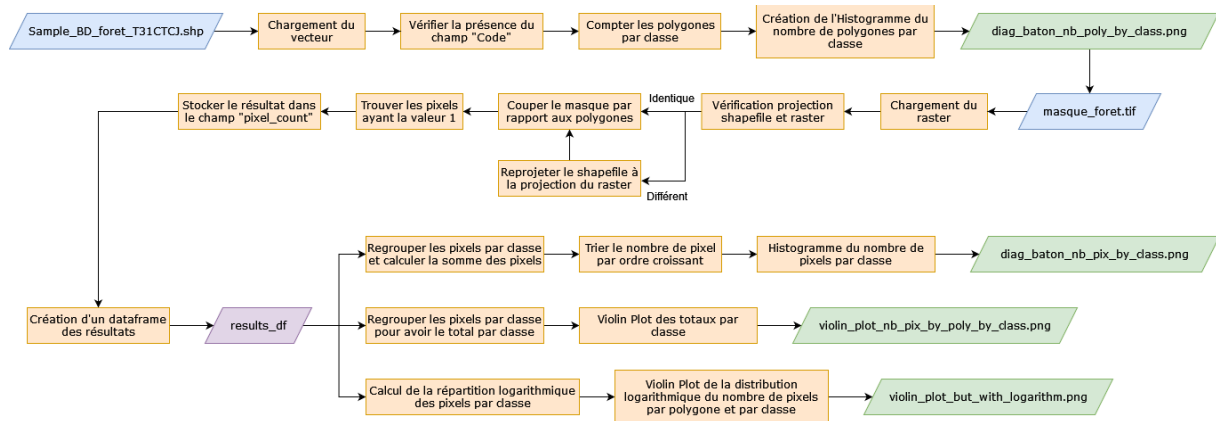


Figure 3 - Diagramme de flux de l'analyse du nombre d'échantillons

Ce script a été exécuté en 1 minute et 37 secondes.

2.4. Pré-traitement des images (5.5)

Pour cette partie il faut utiliser les 6 images téléchargées via THEIA, importées dans Onyxia donc disponibles en entrant la commande présentée dans les rappels d'objectifs (dossier images).

Les 6 images sélectionnées correspondent aux 4 saisons avec une couverture nuageuse inférieure à 15%, les voici :

- 2 images au printemps (20 mars 2022 – 21 juin 2022):
 - **26 mars 2022** : SENTINEL2B_20220326-105856-076_L2A_T31TCJ_D
→ couverture nuageuse 14%
 - **5 avril 2022** : SENTINEL2B_20220405-105855-542_L2A_T31TCJ_D
→ couverture nuageuse 0%
- 1 image en été (21 juin 2022 – 22 septembre 2022):
 - **14 juillet 2022** : SENTINEL2B_20220714-105903-040_L2A_T31TCJ_D
→ couverture nuageuse 5%
- 2 images en automne (22 septembre 2022 - 21 décembre 2022) :
 - **22 septembre 2022** : SENTINEL2B_20220922-105859-863_L2A_T31TCJ_D
→ couverture nuageuse 0%
 - **11 novembre 2022** : SENTINEL2B_20221111-105858-090_L2A_T31TCJ_D
→ couverture nuageuse 0%
- 1 image en hiver :
 - **19 février 2023** : SENTINEL2B_20230219-105857-687_L2A_T31TCJ_D
→ couverture nuageuse 11%

Les images sélectionnées sont en Flat Reflectance (corrigées et standardisées).

Nous souhaitons obtenir ici un raster contenant toutes les images satellites. Pour cela il faut harmoniser leurs résolutions et leur projection spatiale. Cette fois ci nous utiliserons rasterio plutôt que GDAL car il est plus compréhensible et facile à utiliser dans des boucles complexe de traitements. Il faut faire une triple boucle, une qui vient incrémenter un compteur, une qui sélectionne l'image et une qui réalise le travail de concaténation avec tous les paramètres demandés (masque, roi, nodata...). Cela permet de ne pas surcharger la mémoire de l'ordinateur en plus d'un visuel sur l'avancée des traitements avec des prints. La sortie est en Uint16 et fait environ 7 Go.

Dans une deuxième partie, nous faisons globalement la même chose. Le code calcule des indices NDVI pour chaque date en utilisant les bandes B4 et B8 des images brut du dossier. Pour récupérer ses bandes il suffit de ne pas prendre en compte les 19 premier caractère et ne pas sélectionner 2 fois la même image (date_key). Ces indices sont calculés après l'application du masque et sont empilés dans un fichier unique, où chaque bande représente le NDVI. Une dernière étape projette cette image NDVI finale au format EPSG:2154. La sortie fait 1.7 Go et est en float 32 mais la valeur de nodata est aussi de 0 et non de -9999.

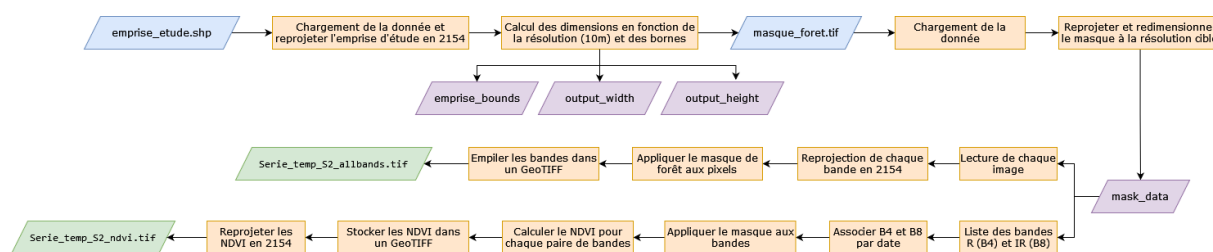


Figure 4 - Diagramme de flux du prétraitement des images

Ce script a été exécuté en 9 minutes et 11 secondes.

2.5. Etude de la phénologie des classes de la BD Forêt (5.4.2)

Ce cinquième script permet l'analyse de la phénologie des peuplements purs. Cela inclut de ne prendre en compte que les classes en gras dans la colonne classif pixel de la figure 2 et produire un graphique représentant la signature temporelle de la moyenne et l'écart-type du NDVI par classe.

Ce script sera compris dans le fichier "*sample_analysis_temp_signature.py*". Les bibliothèques permettant l'exécution de ce code sont : GeoPandas, Rasterio, os, RasterStats, Pandas et Matplotlib.

Première étape de paramétrage des données avec précision des données en entrées, qui sont "*Sample_BD_foret_T31TCJ.shp*" correspondant aux échantillons, ainsi que "*Serie_temp_S2_ndvi.tif*", le NDVI créé lors du pré-traitement des images. Pour que ces données soient accessibles il faut les charger.

Pour réaliser cette étude, il faut sélectionner les codes correspondant aux classes en gras dans la colonne "classif pixel" de la figure 2, à savoir : '12', '13', '14', '23', '24', '25'. et les filtrer dans le GDF "*gdf_classif_pixel_gras*".

Ensuite, intersecter les polygones avec le NDVI en calculant les statistiques zonales de chaque bande (moyenne et écart-type) puis stocker les résultats de ces statistiques dans un dictionnaire qui contient la classe du polygone, le numéro de bande et les statistiques.

Après cela, agréger le dataframe créé (df_ndvi) par classe et par temps pour préparer la représentation des données, l'objectif étant d'avoir une analyse temporelle.

Enfin dernière étape, la représentation sous forme graphique. Pour cela, il faut :

- Créer un dictionnaire des valeurs pour associer les classes au nom de la formation végétale correspondante,
- Créer la figure et l'axe,
- Définir les couleurs (facultatif, c'est pour personnaliser le graphique)
- Tracer la courbe moyenne ainsi que la zone d'incertitude (l'écart-type)
- Définir des étiquettes pour l'axe des x afin de faire correspondre les dates
- Configurer les axes et légendes

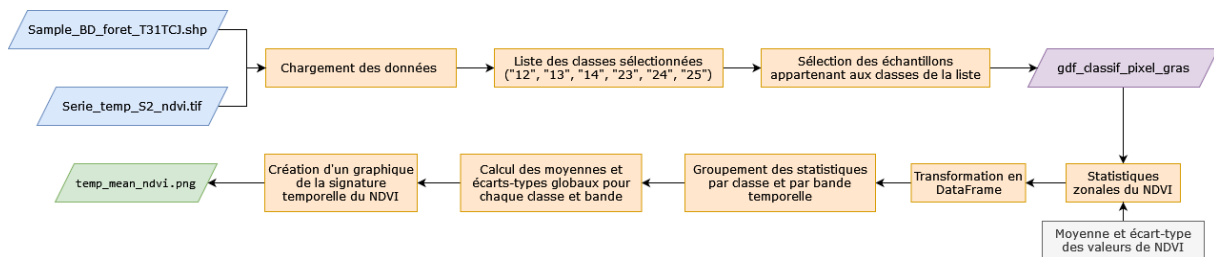


Figure 5 - Diagramme de flux de l'analyse de la phénologie des peuplements purs

La figure finale, “temp_mean_ndvi.png” a été sauvegardée dans ce chemin d'accès : /home/onyxia/work/results/figure. Elle sera présentée et analysée dans la seconde partie de ce rapport.

Ce script a été exécuté en 4 minutes et 22 secondes.

2.6. Analyse de la variabilité spectrale de la BD Forêt (5.4.3)

L'analyse de la variabilité spectrale des classes de la BD Forêt implique l'utilisation de la distance moyenne au centroïde de chaque classe. Pour réaliser cette étude nous allons séparer les classes conformément à la colonne “classif objet” de la figure 2, dont les éléments sont colorés en rouge ou en bleu. Cette analyse vise à répondre à la question suivante : “la variabilité des classes de peuplement en mélange (classes en bleu) est-elle supérieure à celles des classes de peuplements purs (classes en rouge) ?”. Cette partie étant dédiée à de la méthodologie, nous répondrons à cette question dans la partie suivante consacrée aux résultats.

Cette étude se décompose en deux parties, une première à l'échelle de l'image entière, tous polygones confondus et une seconde à l'échelle de chaque polygone.

La différence entre ces deux échelles c'est que la première analyse les classes dans le NDVI, indépendamment des polygones, permettant de comprendre la variabilité spectrale à l'échelle d'une image, et la seconde analyse chaque polygone individuellement en tenant en compte des classes associées à ces polygones (rouge ou bleu) permettant de marquer la variabilité locale et détecter les différentes répartitions spectrales dans ces polygones.

Pour réaliser les deux graphiques finaux, il faut d'abord définir les fichiers en entrée, à savoir “Sample_BD_foret_T31TCJ.shp” et “Serie_temp_S2_ndvi.tif” et les charger pour pouvoir les exploiter.

Comme précisé précédemment, cette analyse est réalisée selon des classes rouges et bleues définies comme ceci, selon la figure 2 du cours :

- classe rouge : '11', '12', '13', '14', '23', '24', '25'
- classe bleue : '15', '26', '28', '29'

La partie du script dédié à l'analyse à l'échelle de l'image entière se structure comme ceci :

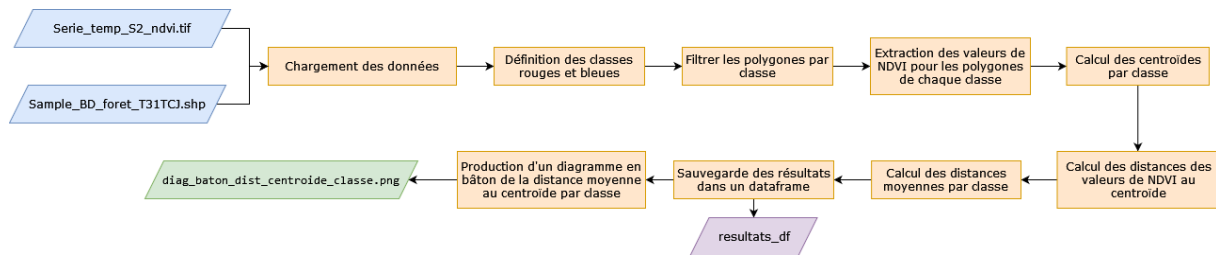


Figure 6 - Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle de l'image entière

En premier lieu, il faut filtrer les polygones par classe puis, grâce à une fonction, extraire les valeurs de NDVI pour les polygones de chaque classe. Ensuite, calculer les centroïdes et les distances moyennes en :

- Filtrant les échantillons pour chaque classe
- Extraire les valeurs de NDVI
- Calculer le centroïde correspondant à la moyenne des valeurs de NDVI
- Calculer les distances des pixels au centroïde
- Calculer la distance moyenne

Les résultats sont sauvegardés dans un dataframe “*resultats_df*”.

Ils seront ensuite présentés dans un diagramme en bâton représentant la distance moyenne au centroïde des NDVI par classe.

Ce diagramme est nommé “*diag_baton_dist_centroide_classe.png*” et enregistré dans le chemin d'accès suivant : “/home/onyxia/work/results/figure”.

La partie du script dédié à l'analyse à l'échelle du polygone se structure comme ceci :

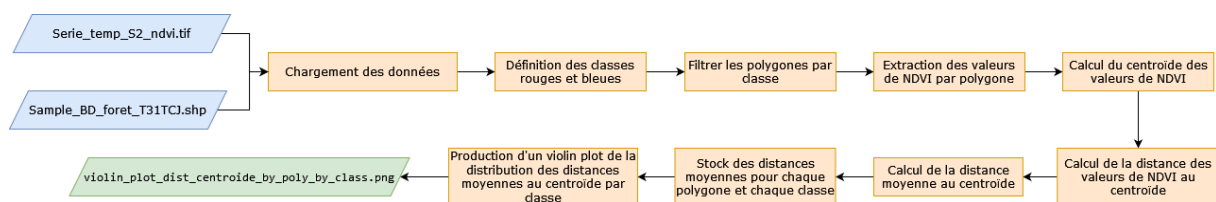


Figure 7 - Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle du polygone

Cette analyse nécessite également une fonction qui extrait les valeurs de NDVI. Après cela, il faut réaliser le traitement par classe avec une boucle qui vient :

- Filtrer les échantillons par classe
- Calculer les distances moyennes pour chaque polygone en :
 - o Extrayant les valeurs de NDVI du polygone
 - o Calculant le centroïde de moyenne des valeurs de NDVI
 - o Calculant les distances des pixels au centroïde
 - o Calculer la distance moyenne

- Ajouter la distance moyenne à la liste stockant les distances moyennes par classe initiée avant la boucle for.

Enfin, générer un violin plot des résultats avec une distribution pour chaque classe.

Le graphique en sortie représente la distribution des distances moyennes au centroïde des NDVI pour les polygones d'une classe. Il se nomme `"violin_plot_dist_centroide_by_poly_by_class.png"` et se trouve au chemin d'accès suivant : `"/home/onyxia/work/results/figure"`.

Ce script a été exécuté en 3 minutes et 16 secondes.

2.7. Production des cartes finales (5.6.3)

Dans cette partie il faut dans un premier temps réaliser un raster de ROI qui a pour valeur 0 et 1. Une fonction dans `my_fonction.py` permet de rasteriser tout vecteur avec les mêmes paramètres qu'une image en entrée. La sortie est écrite en uint8. Cette fonction aurait pu être utilisée dans les premières parties du projet, ce qui est un peu dommage. Une fois notre ROI nous alors refaire la fonction de création de centroïde vue en cour. Pour cela on vectorise le raster de NDVI et on utilise Numpy pour avoir les coordonnées x et y de chaque centroïde. On filtre votre résultat par rapport au ROI créé au début et on le transforme en geodataframe. Nous découpons ce gdf par rapport à la couche vecteur `sample_bd_foret` qui ont les codes de chaque classe. Il suffit enfin de récupérer les valeurs de ses classes ainsi que la valeur du raster de NDVI.

Nous pouvons ensuite préparer les données pour le Machine Learning. Le `test_size` doit être de 0.2 car l'on souhaite un entraînement avec 80% des données. Nous réaliserons la solution 2 du projet (stratifiée et qui prend en compte l'appartenance des pixels à un polygone).

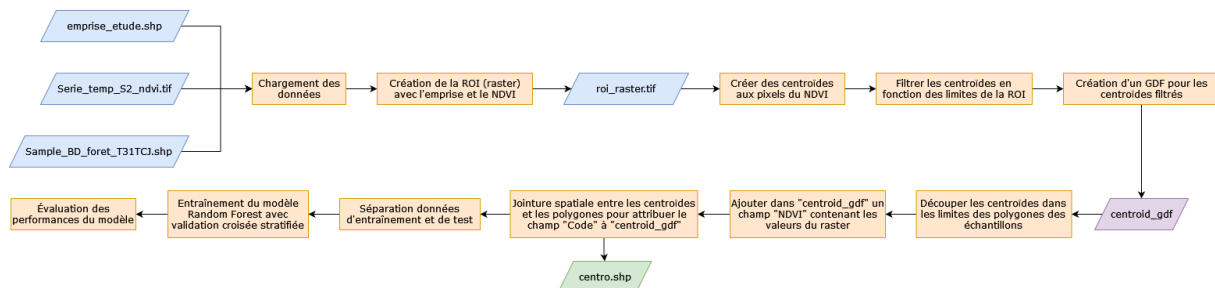


Figure 8 - Diagramme de flux de la classification par pixel

Ce script a été exécuté en 3h (premier fold en 30 min)

III. Analyse des résultats

Dès lors, les traitements principaux ont été réalisés, allant de la construction du masque à la production de cartes finales. Cette seconde partie vise à présenter les résultats, la partie discussion des résultats est réservée à la dernière partie.

3.1. Analyse des échantillons (5.4.1)

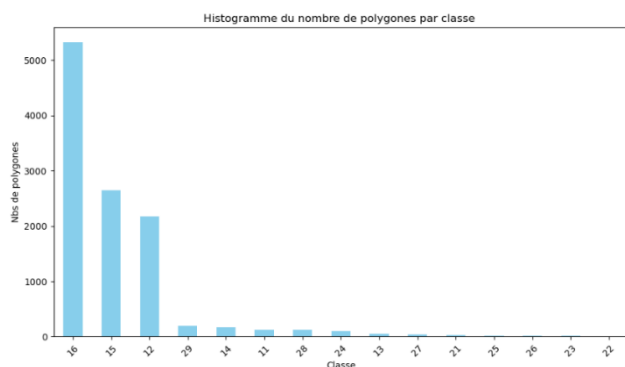


Figure 9 – Histogramme du nombre de polygone par classe

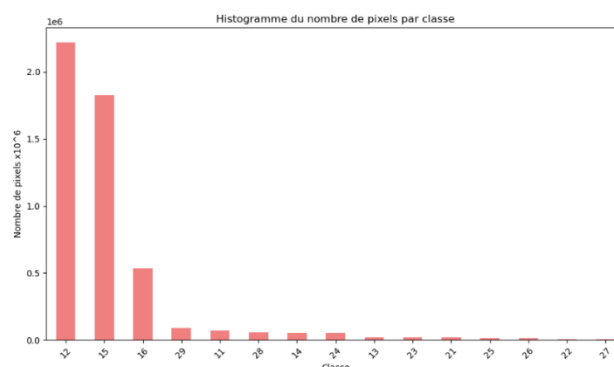


Figure 10 - Histogramme du nombre de pixels par classe

Pour l'analyse des échantillons en entrée nous pouvons constater que la classe 16 qui a beaucoup de polygone n'est pas la classe avec le plus de pixel. Cependant cette relation est légèrement corrélée. En effet les classe 12, 15 et 16 dispose du plus grand nombre de pixel et de polygone par rapport à l'autre classe. Ce qui signifie qu'il faut prendre cette importance au vu des futurs résultats et analyses.

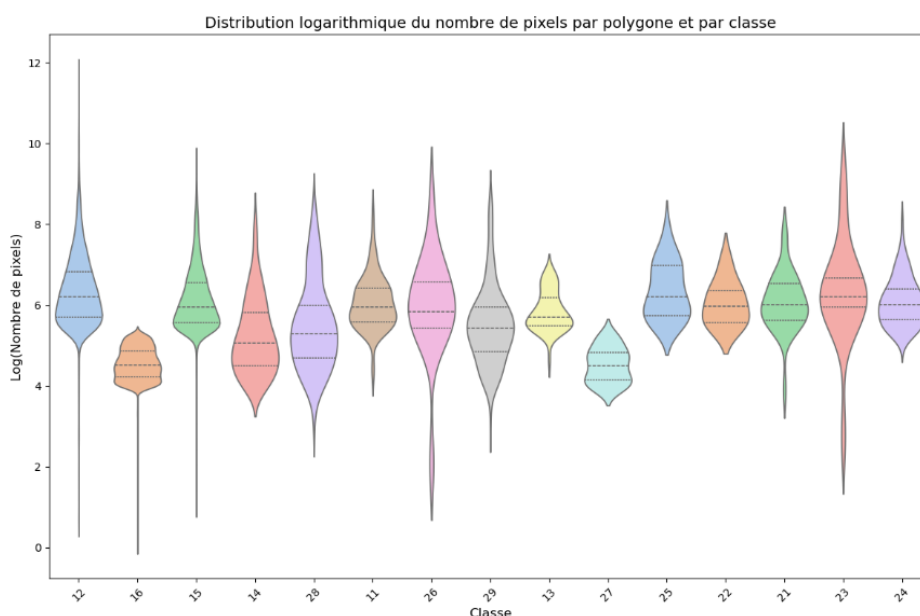


Figure 11 – Distribution logarithmique du nombre de pixels par polygone et par classe

En analysant de ce violin plot, on constate que la disparité générale entre les classes soit relativement faible. Cela indique qu'il existe des variations plus importantes dans la taille des polygones ou la densité des pixels pour ces classes spécifiques.

Certaines classes ont une distribution plus concentrée, ce qui suggère une homogénéité dans le nombre de pixels pour les polygones appartenant à ces catégories. À l'inverse, d'autres classes montrent des distributions plus larges, avec des valeurs extrêmes, signalant des polygones très différents en termes de taille ou de représentation spatiale.

3.2. Analyse de la phénologie des peuplements purs (5.4.2)

Pour rappel, la phénologie est “l’étude des variations des phénomènes périodiques de la vie animale et végétale, en fonction du climat.”. Ici l’analyse se fait donc de manière temporelle avec 6 dates représentant les 4 saisons.

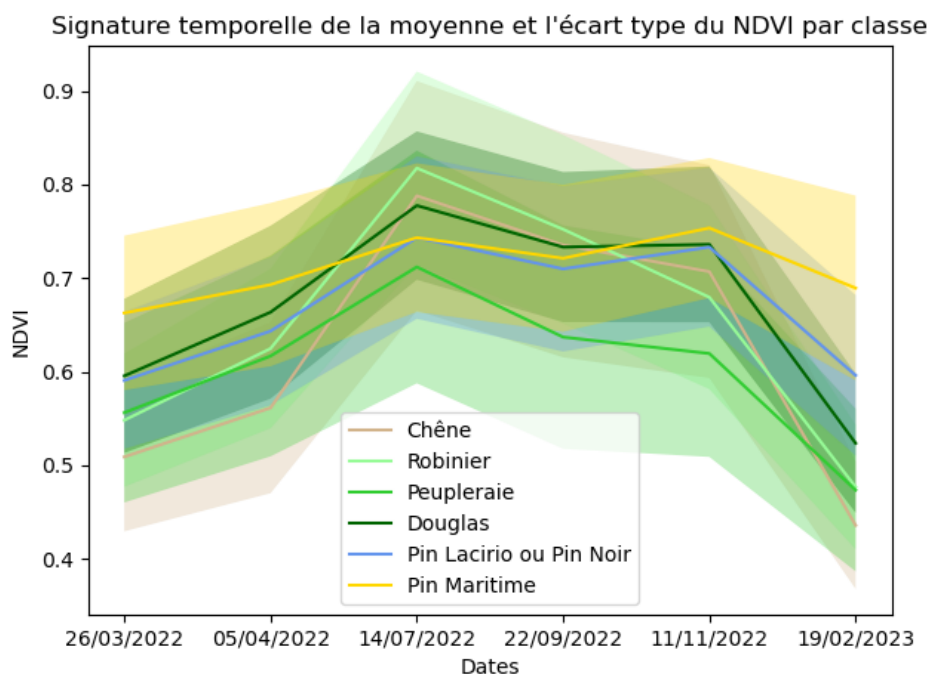


Figure 12 - Signature temporelle de la moyenne et l'écart-type du NDVI par classe

Ce graphique représente l'évolution des signatures temporelles du NDVI moyen et de l'écart-type pour 6 classes de formations végétales (chêne, robinier, peupleraie, douglas, pin laricio/noir et pin maritime).

Les courbes correspondent aux variations saisonnières du NDVI pour chaque type. De manière globale, on constate un pic de NDVI autour du mois de juillet expliqué par une forte activité chlorophyllienne, là où les végétaux reflètent le plus dans les bandes IR.

Par exemple, le robinier (feuillus) présente un pic marqué en été correspondant à une photosynthèse maximale (écart-type plus larges). Tandis que le Douglas ou le Pin noir (résineux) ont une courbe plus stable avec des variations plus faibles dans le temps reflétant ainsi leur activité photosynthétique moins saisonnière.

Les bandes d'écart-type sont généralement plus grandes pendant la période de croissance (été) et diminuent en hiver pour toutes les classes. Ces bandes montrent également une variabilité plus importante pour les feuillus (chêne, robinier, peupleraie) en particulier en automne et au printemps au moment des changements saisonniers importants (transition vers l'hiver ou l'été).

Les signatures temporelles permettent donc de faire la différence entre les comportements phénologiques des classes. Il est donc possible d'en déduire que les feuillus présentent une plus grande phénologie (variabilité temporelle) que les résineux.

3.3. Analyse de la variabilité spectrale de la BD Forêt (5.4.3)

Avant de débiter cette partie, il est important de définir ce que l'on entend par peuplement en mélange et peuplement pur. Les peuplements mélangés regroupent "*les peuplements de deux espèces ou plus*", contrairement aux peuplements purs qui sont les "*peuplements à une seule espèce*".

Concrètement dans cette partie du projet, nous souhaitons examiner la dispersion des valeurs de NDVI dans chaque classe de peuplement pour évaluer leur hétérogénéité interne.

L'analyse s'est faite en deux temps : une analyse à l'échelle globale en mesurant la dispersion des valeurs spectrales à l'ensemble des pixels d'une classe, mais aussi à l'échelle des polygones en examinant la variabilité intra-polygone pour voir si certains peuplements sont plus homogènes ou hétérogènes.

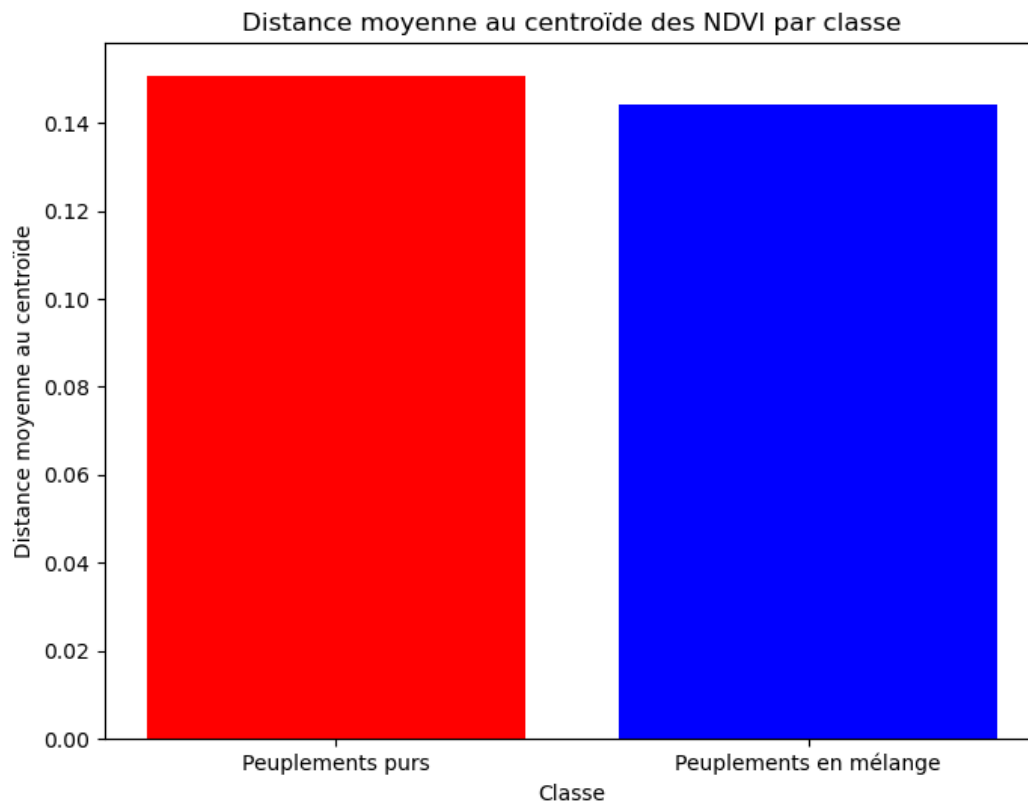


Figure 13 - Distance moyenne au centroïde des NDVI par classe à l'échelle de l'image entière

Ce diagramme en bâton présente les distances moyennes au centroïde des NDVI pour chaque classe à l'échelle de l'image entière. Chaque barre correspond à une classe : rouge pour les peuplements purs et bleu pour les peuplements en mélange.

D'après les résultats de la figure 12 la distance moyenne au centroïde est légèrement plus élevée pour les peuplements purs (0,15) que pour les peuplements en mélange (0,14). Cela reflète une hétérogénéité plus marquée pour les peuplements purs sur l'ensemble de l'image. En d'autres termes, les pixels classés comme peuplements purs présentent une plus grande dispersion autour de leur centroïde par rapport aux peuplements en mélange.

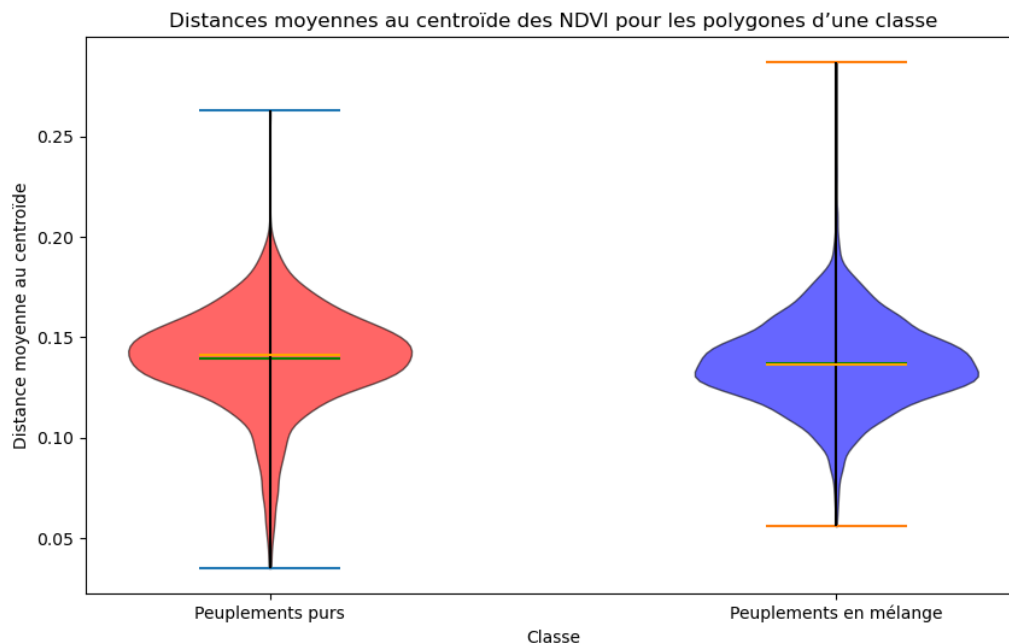


Figure 14 - Distances moyennes au centroïde des NDVI par classe à l'échelle des polygones

Enfin, ce dernier graphique montre la distribution des distances moyennes au centroïde regroupées par polygone pour deux classes : les peuplements purs (rouge) et les peuplements en mélange (bleu).

Les peuplements purs (rouge) présentent une distribution plus étendue que les peuplements en mélange (bleu), ceci indique, comme pour le diagramme en bâton, une variabilité plus importante des distances au centroïde dans les peuplements purs, confirmant leur hétérogénéité.

Pour les deux classes, l'endroit où se concentrent la majorité des distances moyennes se concentrent dans la plage centrale, la plus étendue des graphiques vers 0,15.

Pour résumer, en termes de variabilité, les peuplements purs (en rouge) montrent une distribution plus éparpillée et des distances plus importantes, ce qui indique que certains polygones présentent une hétérogénéité marquée. Quant aux peuplements en mélange (en bleu), ces derniers ont une distribution plus compacte, indiquant une variabilité spectrale moins importante à l'échelle des polygones.

IV. Discussion des résultats

Les résultats que nous obtenons tout au long de ce projet semblent corrects et pertinents. Cependant sur la partie de Machine Learning les résultats que nous avons eu sont moins encourageant et suscite une analyse plus profonde.

L'analyse des résultats montre que, pour chacun des 5 plis de la validation croisée stratifiée, le modèle Random Forest atteint une précision constante de **0.31**. Cette stabilité est confirmée par un écart type nul des scores (0.00), indiquant que le modèle performe de manière identique sur chaque sous-ensemble de données. Cependant, une précision de 0.31 est relativement faible, ce qui pourrait suggérer plusieurs points : soit les données d'entrée (NDVI en tant que seule variable explicative) ne permettent pas de distinguer efficacement les classes cibles ("Code"), soit les classes elles-mêmes sont

déséquilibrées ou difficiles à séparer. Ces résultats mettent en évidence la nécessité d'explorer d'autres caractéristiques pour améliorer la capacité de discrimination du modèle ou de vérifier la qualité et la diversité des données utilisées.

Concernant les résultats de l'analyse de la phénologie des peuplements, les résultats semblent tout à fait pertinents puisque l'on remarque une augmentation de la variabilité au moment de l'été correspondant aux périodes de plus forte activité chlorophylliennes détectées grâce aux bandes IR du NDVI.

D'après l'ONF, la phénologie des végétaux est *“l'étude de leurs phases de développements saisonniers : feuillaison, floraison, fructification, jaunissement automnal. Ces développements sont liés à certains paramètres climatiques”*. Comme en attestent nos résultats, il y a en effet une variation des développements liés à certains paramètres climatiques que l'on remarque avec les courbes des moyennes et les bandes des écart-types.

Les feuillages des espèces n'apparaissant pas au même moment. En effet, il peut y avoir *“en moyenne cinq semaines de décalage entre l'essence la plus précoce et la plus tardive”*, ce qui explique les variations inter-espèces. Les feuillus, comme leur nom l'indique, perdent leurs feuilles avec la chute des températures, c'est pourquoi leurs moyennes et écart-types chutent fortement au moment de l'hiver.

Les résineux ont des résultats plus stables. En effet, tous les résineux ne perdent pas entièrement leurs épines avec les chutes de températures, ce qui explique leur stabilité en termes de moyenne et d'écart-type, puisque leur activité chlorophyllienne est moins marquée que les feuillus.

En revanche, dans le temps, ils ont un débourrement (éclosion des bourgeons) plus tardif et leur phénologie dépend beaucoup de l'altitude. Cela pourrait représenter une des limites de notre étude, connaître l'altitude des peuplements étudiés afin d'expliquer plus précisément leurs variabilités.

Le diagramme en sortie est utilisable dans le cadre de l'étude puisqu'il permet de décrire l'évolution dans le temps des peuplements, en croisant les informations de la BD Forêt avec l'image Sentinel 2, et répondre en partie à la problématique de départ.

Cette étude pourrait être réutilisée dans une perspective temporelle plus longue en comparant la phénologie des peuplements sur plusieurs années afin de faire apparaître les différences ou ressemblances entre chaque année et chercher les causes de ces écarts.

Ensuite, pour l'analyse de la variabilité spectrale de la BD Forêt, la problématique suivante se posait : *“la variabilité des classes de peuplement en mélange (classes en bleu) est-elle supérieure à celles des classes de peuplements purs (classes en rouge) ?”*

Nous avons en partie répondu à cette question dans la partie précédente, mais nous allons détailler les résultats dans cette partie.

Pour rappel, les résultats du premier graphique montrent que la variabilité des distances moyennes au centroïde est légèrement plus marquée pour les peuplements purs que pour les peuplements en mélange avec une différence de 0,01, avec une plus grande dispersion autour du centroïde, indiquant une hétérogénéité plus élevée dans cette catégorie.

Le deuxième graphique confirme cette observation puisqu'il montre une distribution plus étendue pour les peuplements purs, tandis que les peuplements en mélange sont concertés dans une plage plus étroite.

Contrairement à l'hypothèse initiale, la variabilité des classes de peuplement en mélange est inférieure à celle des classes de peuplements purs.

La principale confusion par rapport à ces résultats est que l'on pourrait s'imaginer que les peuplements purs (tendent vers une plus grande homogénéité car les arbres sont de la même espèce) auraient une variabilité des distances moins importante que les peuplements en mélange (tendent vers une plus

grande hétérogénéité par disposent d'au moins deux espèces différentes), alors que nos résultats prouvent le contraire.

Cette hétérogénéité dans la variabilité spectrale des peuplements purs pourrait être liée à des causes internes comme l'âge des arbres, leur structure, l'état de leur environnement.

Ici les variabilités spectrales pour chaque peuplement sont modérées car les distances au centroïde sont comprises entre 0,14 et 0,15, représentant maximum 15% de la plage totale. Par conséquent, cela peut être interprété comme une dispersion modérée. Il serait intéressant de comparer ces résultats avec d'autres variabilités sur d'autres localisations pour se donner une idée de ce qu'est une dispersion extrême.

Ces graphiques peuvent être utiles pour des études environnementales parce qu'elles permettent d'identifier les zones disposant de plus de diversité ou de variabilité, ce qui pourrait aider à la gestion de ces espaces.

L'analyse est incomplète car nous ne sommes pas allés jusqu'au bout de l'étude par manque de temps et de moyens. Toutefois, pour conclure plus généralement ce projet, la BD Forêt version 2.0 pourrait être utilisée comme source d'échantillons de référence pour réaliser une classification supervisée de séries temporelles d'images Sentinel 2 mais en partie car elle reflète bien la phénologie des classes forestières, mais dispose de limites importantes notamment avec une classification supervisée qui donne des résultats peu concluants, une hétérogénéité modérée au sein de la variabilité des classes.

Pour améliorer cela, quelques perspectives d'amélioration comme l'ajout de variables supplémentaires pour affiner l'analyse, mais aussi expérimenter d'autres modèles comme les réseaux de neurones...

Bibliographie

- François Lebourgeois, Jean-Claude Pierrat, Vincent Perez, Christian Piedallu, Sébastien Cecchini, et al.. Déterminisme de la phénologie des forêts tempérées françaises : étude sur les peuplements du réseau Renecofor. *Revue forestière française*, 2008, 60 (3), pp.323-343. ff10.4267/2042/19767ff. ffhal-03449805. URL : <https://hal.science/hal-03449805/document>
- Anonyme. *La phénologie et le climat*. (2024, 7 octobre). Office National des Forêts. <https://www.onf.fr/renecofor/%2B/1772::la-phenologie-et-le-climat.html>
- Morneau, F., Vallet, P., Toïgo, M., & Dalmaso, M. (2016, janvier). *Les forêts mélangées* [Diapositives ; PDF]. La Feuille de L'inventaire Forestier IGN, France. Inventaire-forestier.ign.fr. https://inventaire-forestier.ign.fr/IMG/pdf/IF_36.pdf

Annexes :

Figure 1 - Diagramme de flux de la construction d'un masque à partir de la BD Forêt

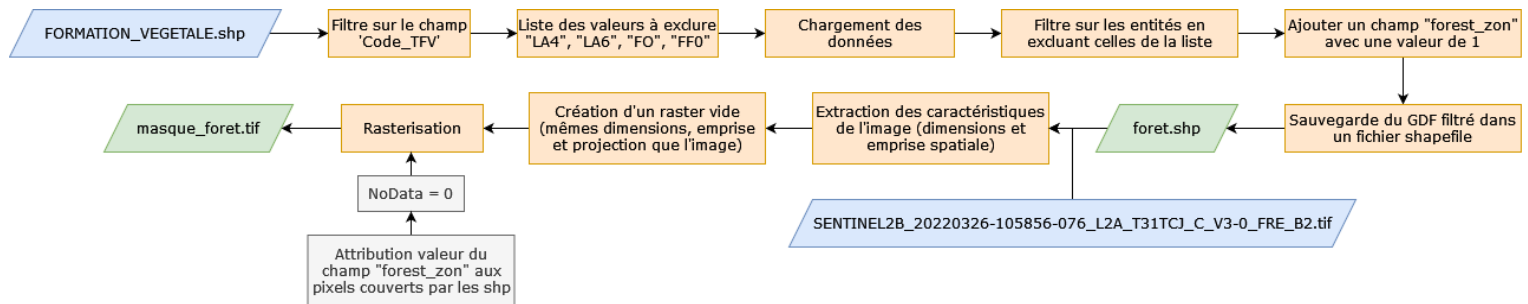


Figure 2 - Diagramme de flux de la sélection des échantillons à partir de la couche formation végétale

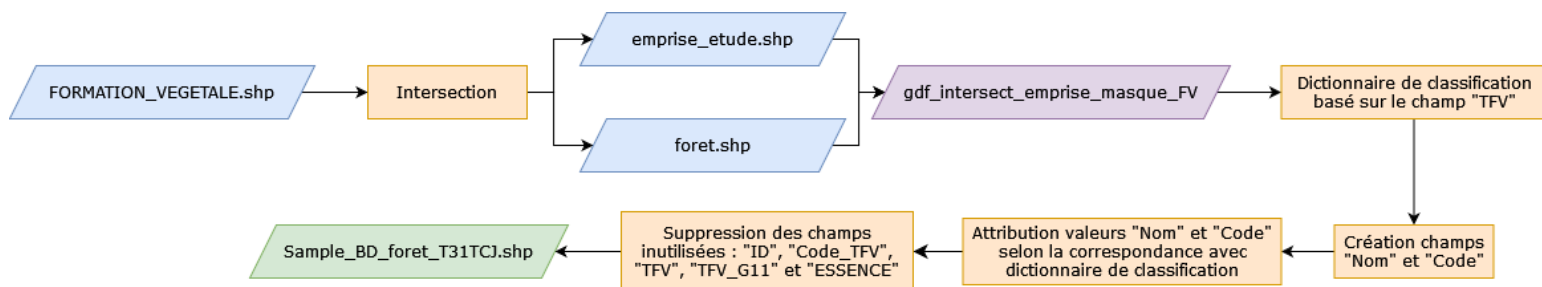


Figure 3 - Diagramme de flux de l'analyse du nombre d'échantillons

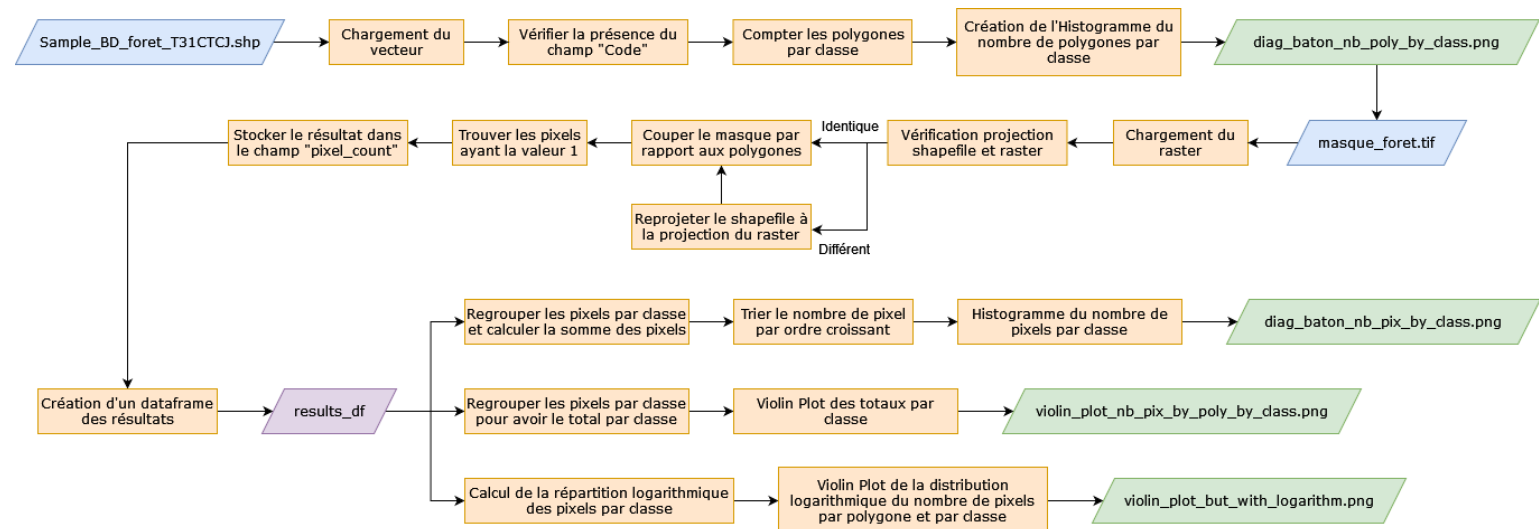


Figure 4 - Diagramme de flux du prétraitement des images

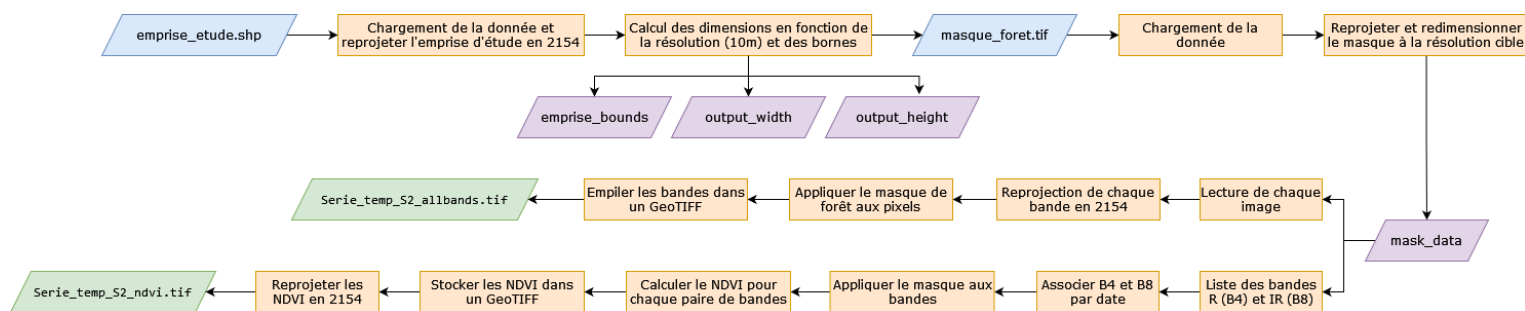


Figure 5 - Diagramme de flux de l'analyse de la phénologie des peuplements purs

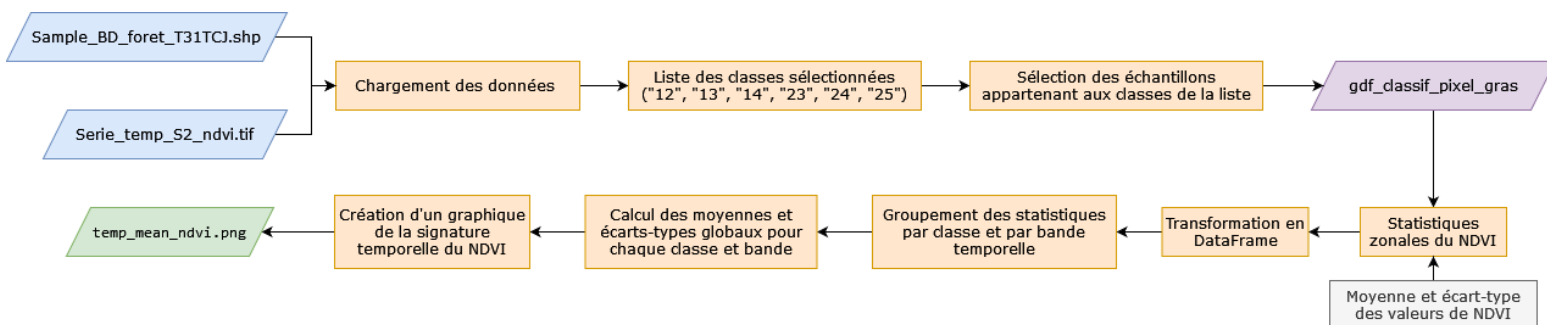


Figure 6 - Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle de l'image entière

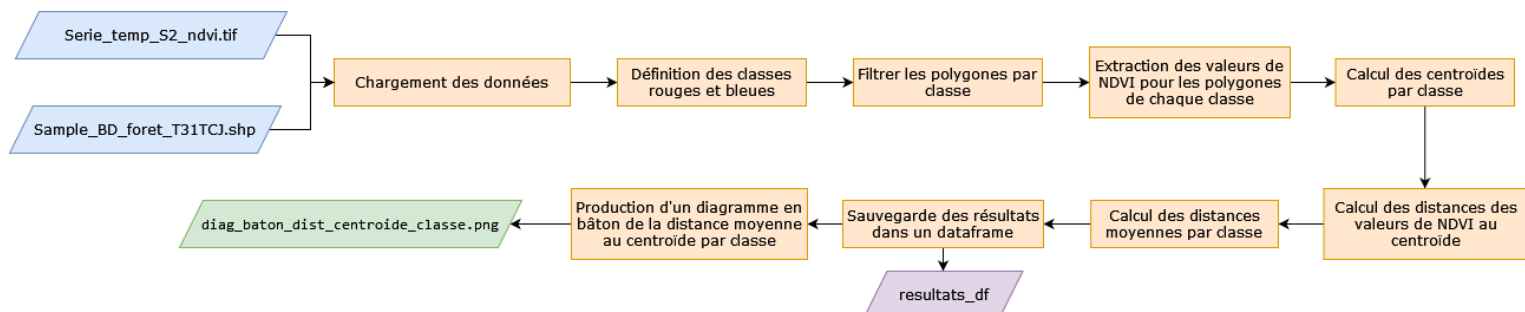


Figure 7 - Diagramme de flux de l'analyse de la variabilité spectrale de la BD Forêt à l'échelle du polygone

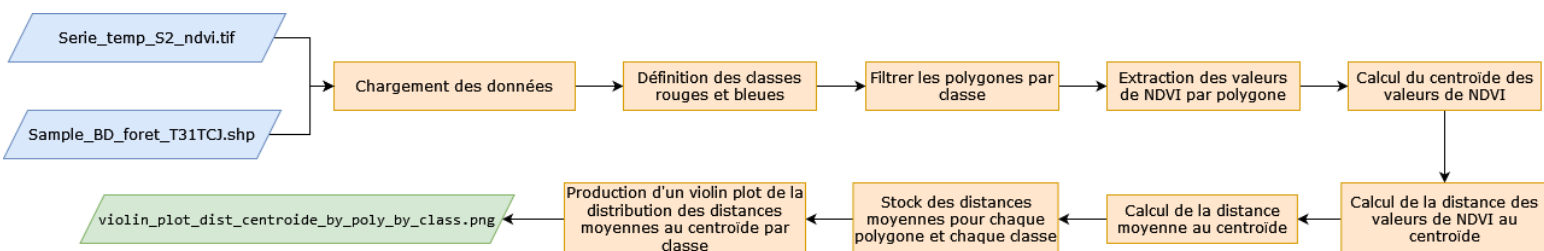


Figure 8 - Diagramme de flux de la classification par pixel

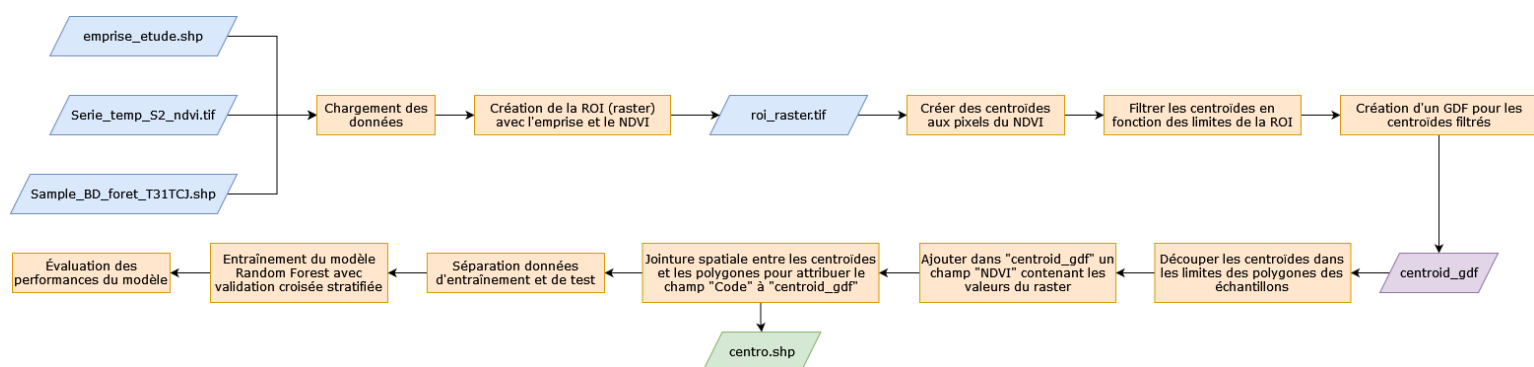


Figure 9 - Histogramme du nombre de polygone par classe

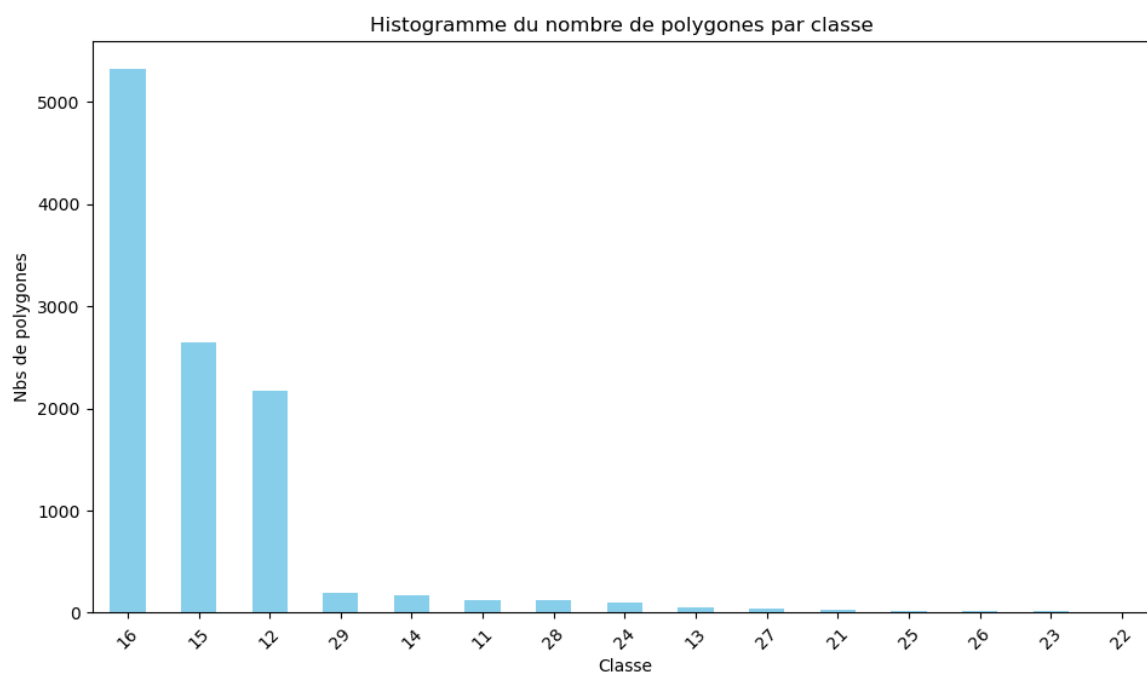


Figure 10 - Histogramme du nombre de pixels par classe

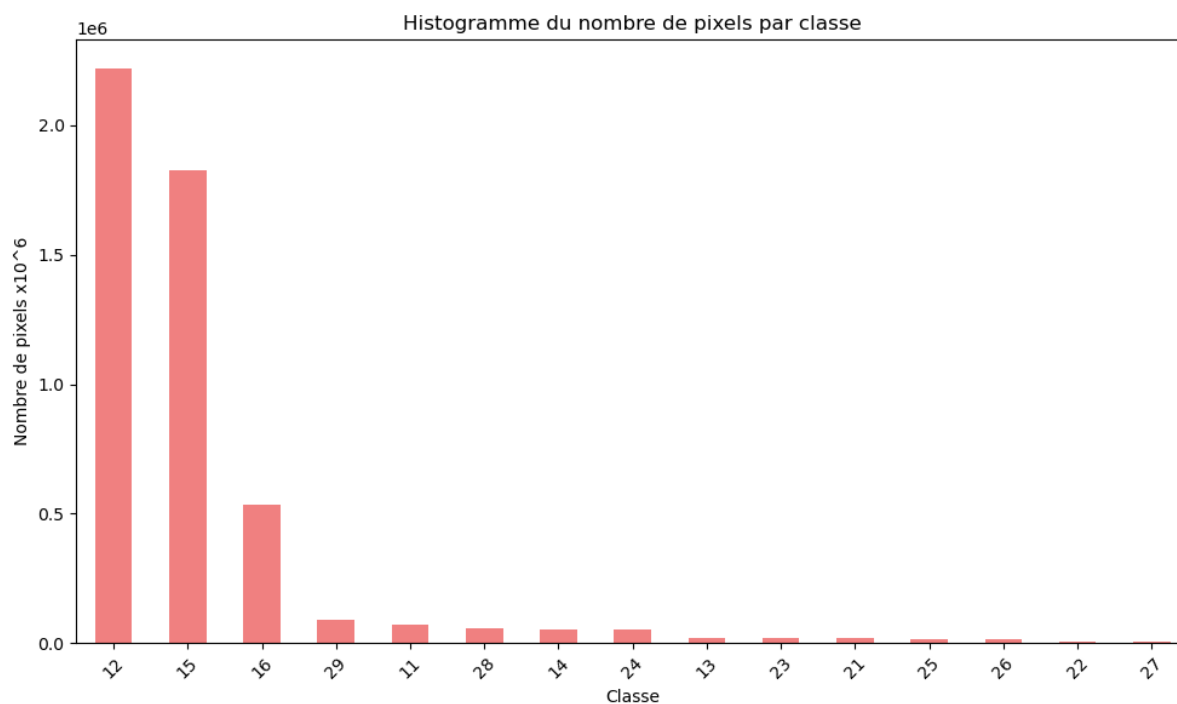


Figure 11 – Distribution logarithmique du nombre de pixels par polygone et par classe

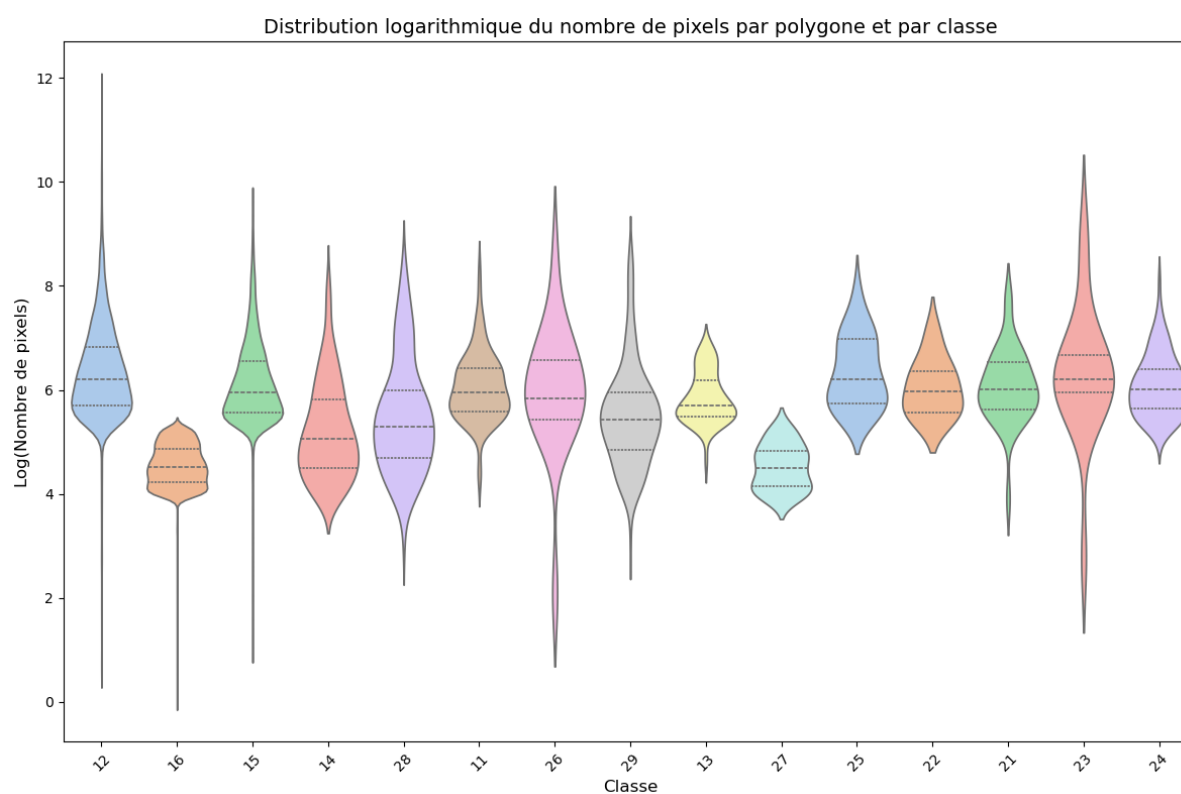


Figure 12 - Signature temporelle de la moyenne et l'écart-type du NDVI par classe

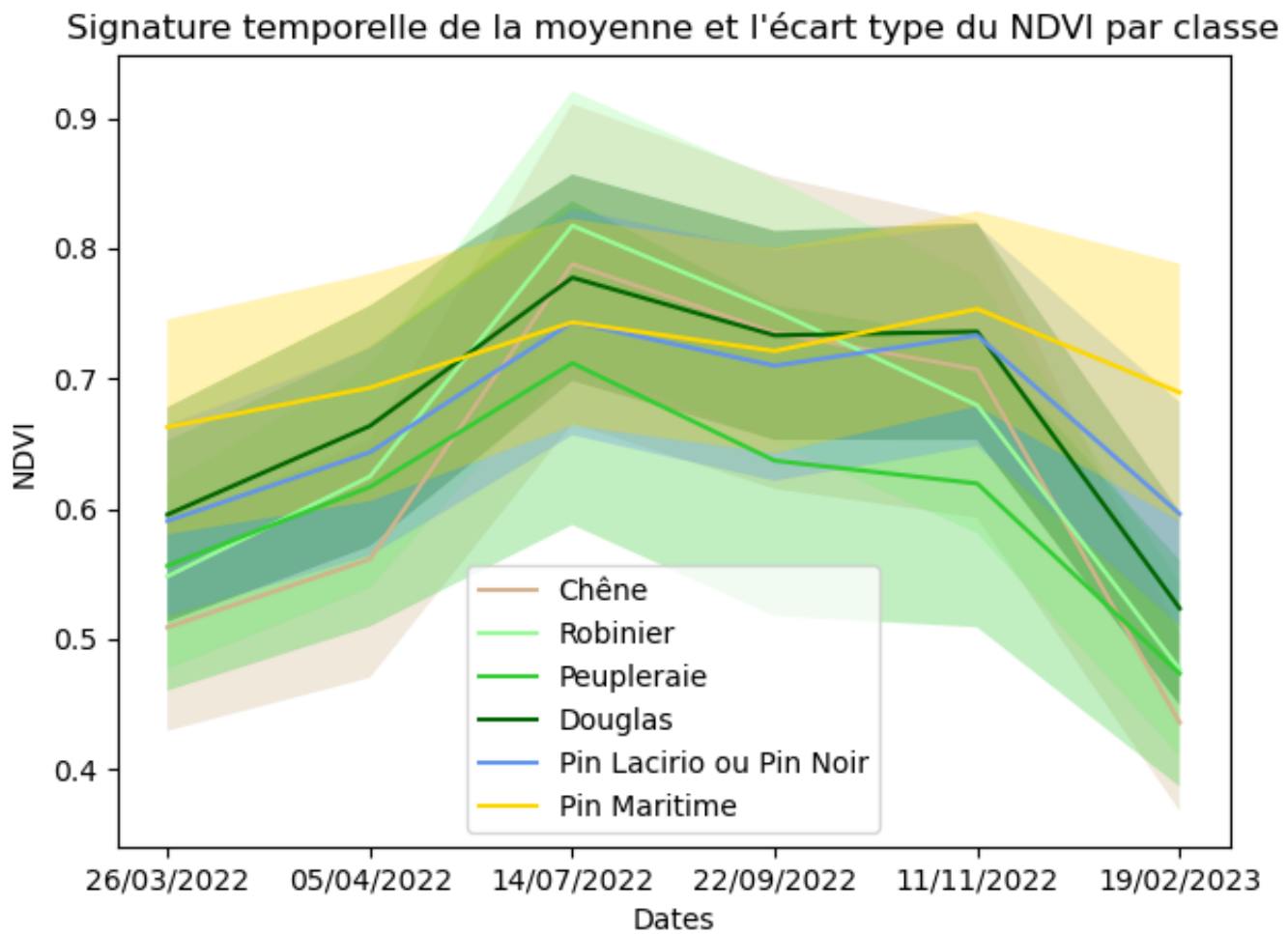


Figure 13 - Distance moyenne au centroïde des NDVI par classe à l'échelle de l'image entière

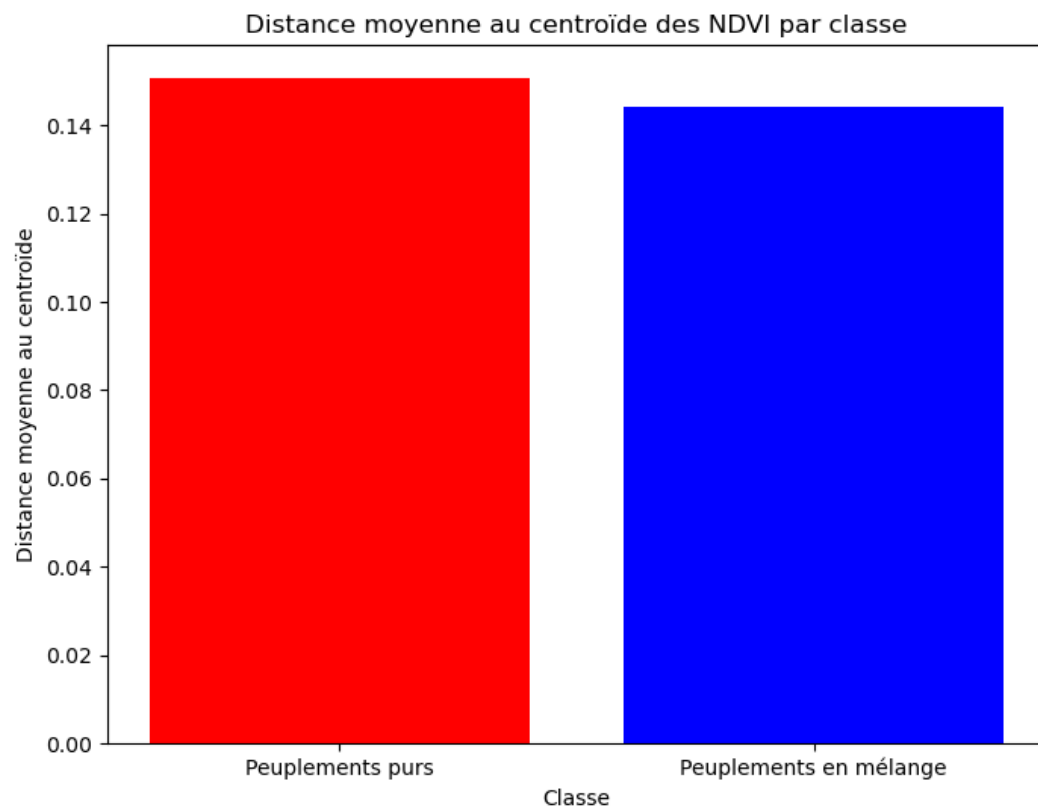


Figure 14 - Distances moyennes au centroïde des NDVI par classe à l'échelle des polygones

