

Télédétection avancée

UE 901_21

Projet de groupe



Abir BEN ABDELGHAFFAR

Manuel ESTEVEZ

Judith LACRAMPE

Gabriel ORABONA

Sommaire

Table des figures.....	3
1. Introduction.....	4
Les données utilisées.....	5
L'environnement de travail.....	5
2. Méthodologie.....	6
a. Préparation des données.....	7
b. Sélection et préparation des échantillons.....	8
c. Analyse des échantillons.....	9
d. Construction du masque forestier.....	11
e. Classification supervisée à l'échelle des pixels.....	12
2. Résultats.....	16
2.1. Analyse des échantillons.....	16
a. Distribution des échantillons.....	16
b. Phénologie des classes forestières.....	18
c. Variabilité spectrale.....	19
2.2. Carte des essences et des peuplements et qualité des classifications.....	20
3. Analyse critique.....	29
d. Discussion.....	29
e. Limites et pistes d'amélioration.....	29
Conclusion.....	31
Annexes.....	32
Annexe 1 - Utilité des scripts.....	32
Annexe 2 - Les différentes images utilisées dans les traitements.....	35
Annexe 3 - Diagrammes de flux.....	37

Table des figures

Figure 1 - Diagramme de flux global.....	7
Figure 2 - Diagramme de flux prétraitement des images.....	8
Figure 3 - Correspondance entre les classes de la BD Forêt et les nomenclatures des cartes à obtenir.....	9
Figures 4 - Divers diagrammes de flux d'analyse des échantillons sélectionnées.....	11
Figure 5 - Visualisation du masque forestier.....	12
Figure 6 - Diagramme de flux de la classification à l'échelle des pixels.....	13
Figure 7 - Arbre de décisions pour la mise en place des classes de peuplements.....	14
Figure 8 - Diagramme de flux de la classification à l'échelle des peuplements.....	15
Figure 9 - Répartition des polygones par classe.....	16
Figure 10 - Répartition des pixels par classe.....	17
Figure 11 - Répartition des pixels par polygone et par classe, sans filtrage des données relatives à la classe dominante.....	18
Figure 12 - Distribution de données par polygone et par classe, avec filtrage des données relatives à la classe dominante “Chêne”.....	18
Figure 13 - Visualisation des signatures temporelles NDVI.....	19
Figure 14 - Distance par rapport au centroïde par classe.....	20
Figure 15 - Distribution des distances par rapport au centroïde par classe.....	20
Figure 18 - Matrice de confusion classification à l'échelle des pixels.....	23
Figure 19 - Carte à l'échelle des peuplements obtenu de la reclassification (colonne “code_predit”.....	25
Figure 20 - Représentation des essences forestières original de la BD forêt (colonne “code”.....	26
Figure 21 -Estimation de la qualité de classification au niveau de peuplement.....	27
Figure 22 - Matrice de confusion classification à l'échelle des peuplements.....	28

1. Introduction

La gestion durable des forêts constitue une priorité dans le cadre des enjeux environnementaux, économiques et sociaux actuels. Les forêts, qui couvrent une part significative du territoire français, jouent un rôle fondamental dans la régulation climatique, la préservation de la biodiversité et l'approvisionnement en ressources naturelles. Pour répondre aux besoins croissants d'informations précises et actualisées sur les formations forestières, les outils de télédétection et les bases de données géographiques se révèlent essentiels.

La **BD Forêt**, une base de données de référence décrivant les essences forestières et les milieux naturels semi-ouverts, offre un panorama riche et détaillé des peuplements forestiers à l'échelle nationale. Toutefois, malgré son utilité pour de nombreuses applications (gestion des ressources, prévention des incendies, certification, etc.), cette base présente des limitations importantes, notamment en ce qui concerne l'actualisation des données et la résolution intra-peuplement. Ces contraintes réduisent son efficacité pour les analyses de précision et la cartographie à des échelles plus fines. L'émergence des technologies de télédétection, notamment les séries temporelles d'images satellites comme celles fournies par Sentinel-2, ouvre de nouvelles perspectives pour pallier ces insuffisances. Ces images permettent une analyse fine de l'évolution spatio-temporelle des formations forestières grâce à des indices tels que le NDVI (Normalized Difference Vegetation Index), offrant ainsi une vision complémentaire et potentiellement enrichissante par rapport aux bases de données statiques.

Ce projet vise à explorer la possibilité d'utiliser la **BD Forêt** comme source d'échantillons de référence pour entraîner une classification supervisée de séries temporelles d'images Sentinel-2. À travers une approche combinant données géographiques et techniques de télédétection, notre objectif est double :

- **Produire des cartes des essences forestières à l'échelle du pixel (échelle intra-peuplement) et à l'échelle des peuplements à partir des pourcentages des classes présentes**
- **Évaluer la qualité des cartes produites en termes de précision et d'utilité pour la gestion forestière et la planification territoriale.**

Le projet repose sur une méthodologie rigoureuse comprenant le prétraitement des images, la génération d'échantillons, la classification des données et l'analyse critique des résultats. À terme, il permettra de poser les bases d'une intégration plus fine entre bases de données géographiques et observations satellitaires, contribuant ainsi à la modernisation des outils de suivi et de gestion des ressources forestières.

1.1 Données et matériel

Les données utilisées

Pour ce projet, nous avons utilisé des images Sentinel-2 acquises entre janvier 2022 et février 2023. Six images, chacune comportant dix bandes, ont été téléchargées depuis la plateforme Theia-Land. Elles présentent un niveau de réflectance 2A et une couverture nuageuse inférieure à 15 %, couvrant la tuile T31TCJ, qui inclut Toulouse et ses environs.

Nous avons également exploité la base de données vectorielle **BD Forêt V2.0** pour extraire des échantillons de référence en vue d'une classification supervisée des essences forestières. Ces données ont été enrichies par d'autres sources vectorielles afin de préciser la délimitation des zones d'étude.

L'environnement de travail

Le projet a été mené sur une instance Onyxia, préconfigurée avec les bibliothèques Python indispensables au traitement des données, telles que Geopandas, OSGeo, Scikit-learn, Matplotlib et NumPy. L'utilisation de Jupyter Notebook ou de VS Code sur cette instance a simplifié le développement, la visualisation et la validation des résultats.

L'organisation du travail repose sur une arborescence structurée dans un dépôt Git, créé à l'aide de commandes exécutées depuis le terminal Onyxia. De plus, la connexion entre ce dépôt et l'environnement VS Code a grandement facilité la collaboration et la synchronisation des contributions.

2. Méthodologie

Pour l'analyse des peuplements forestiers, nous avons suivi une démarche systématique visant à garantir la reproductibilité des résultats obtenus. Cette approche inclut des étapes clés allant du pré-traitement des données à la classification supervisée, en passant par la validation et l'analyse des résultats. Voici les étapes principales, à retrouver également dans le diagramme de flux des traitements ci-dessous :

1. **Téléchargement de données satellites et création de dépôt git**
2. **Pré-traitement des données**
3. **Sélection des échantillons d'apprentissage**
4. **Analyse de la variabilité spectrale et de la signature temporelle des échantillons**
5. **Classification supervisée à l'échelle de pixel**
6. **Classification des peuplements forestiers**
7. **Analyse de la qualité des classifications**

Un diagramme de flux (*Cf. Annexe 3 - Diagrammes de flux*) est également inclus pour détailler chaque traitement effectué, les couches d'entrée utilisées, ainsi que les paramètres principaux appliqués à chaque étape. Ce diagramme permet d'avoir une vue d'ensemble claire et précise sur le flux de travail et les interactions entre les différentes étapes de traitement. L'objectif est de garantir la transparence et la reproductibilité du processus.

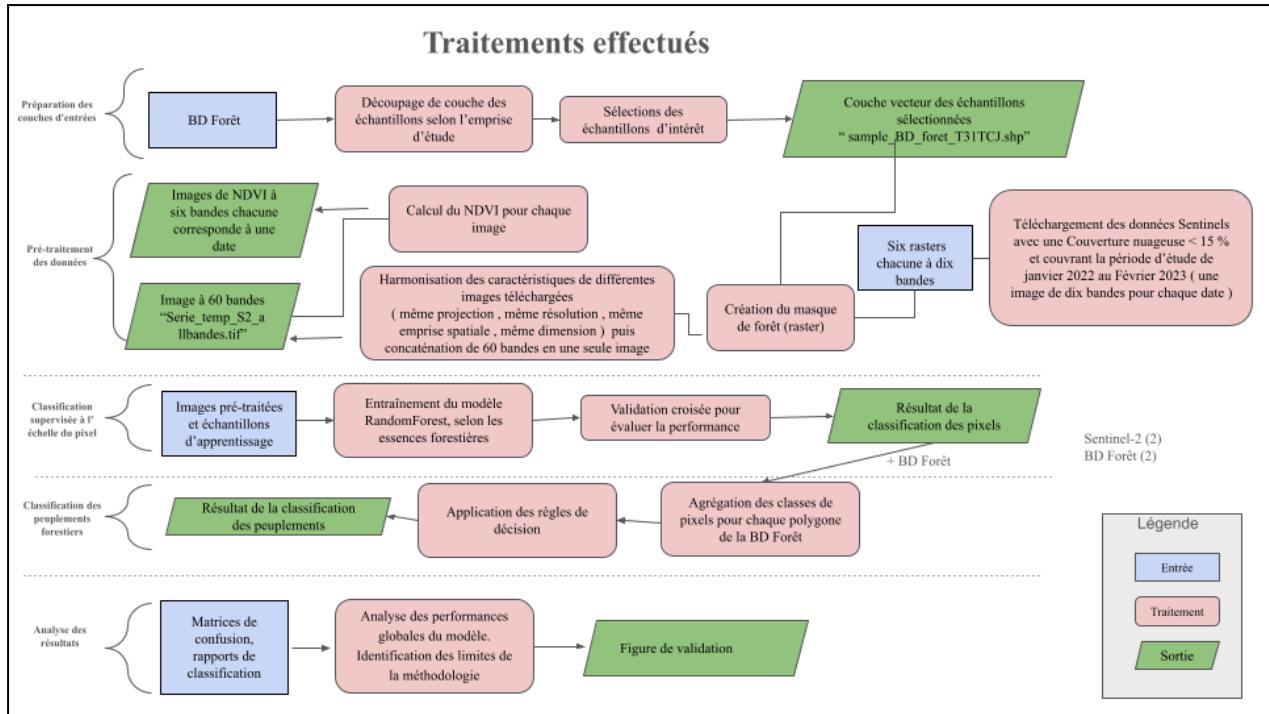


Figure 1 - Diagramme de flux global

a. Préparation des données

Les images Sentinel-2 ont été prétraitées afin d'harmoniser leur résolution spatiale, leur projection et leur découpage selon l'emprise d'étude. Ce prétraitement comprenait également la conversion des images en réflectance et la correction atmosphérique.

Le prétraitement a été réalisé pour s'assurer que toutes les images soient comparables en termes de résolution (10 m), de projection (EPSG:2154 - Lambert 93) et de type de données (Uint16). Les images ont ensuite été découpées selon l'emprise de la zone d'étude afin de se concentrer uniquement sur les zones forestières pertinentes.

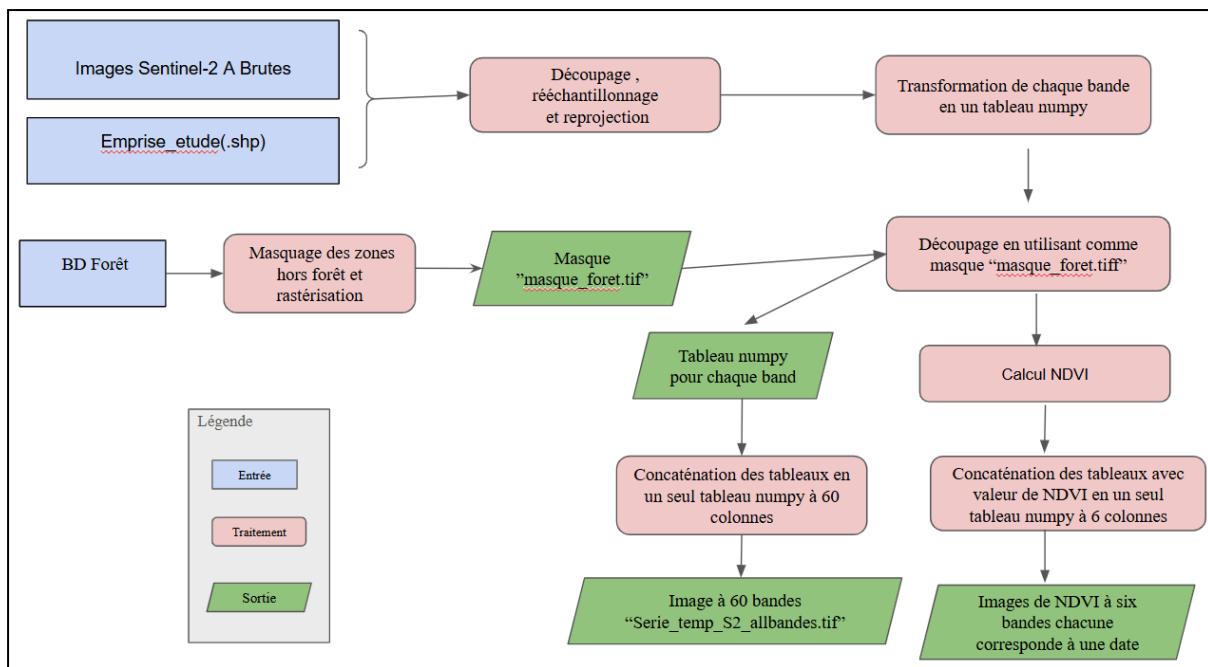


Figure 2 - Diagramme de flux prétraitement des images

b. Sélection et préparation des échantillons

À partir de la BD Forêt, des polygones représentatifs ont été extraits en respectant les classes pertinentes définies dans la nomenclature. Un fichier vecteur a été généré contenant les échantillons filtrés, destinés à l'entraînement et à la validation du modèle.

Les classes non pertinentes, telles que les landes, formations herbacées et forêts ouvertes, ont été exclues pour se concentrer sur les peuplements arborés. Les polygones retenus ont été transformés en échantillons utilisables dans le modèle de classification. Chaque polygone était lié à des informations attributaires permettant de connaître l'essence dominante, ce qui constitue la base de l'étiquetage pour la classification supervisée.

Code TFV BD forêt	Classif pixel		classif objet	
	Nom	Code	Nom	Code
Forêt fermée d'un autre feuillu pur				
Forêt fermée de châtaignier pur	Autres feuillus	11	Autres feuillus	11
Forêt fermée de hêtre pur				
Forêt fermée de chênes décidus purs	Chêne	12	Chêne	12
Forêt fermée de robinier pur	Robinier	13	Robinier	13
Peupleraie	Peupleraie	14	Peupleraie	14
Forêt fermée à mélange de feuillus	Ø	.	Mélange de feuillus	15
Forêt fermée de feuillus purs en îlots	Ø	.	Feuillus en îlots	16
Forêt fermée d'un autre conifère pur autre que pin				
Forêt fermée de mélèze pur	Autres conifères autre que pin	21	Autres conifères autre que pin	21
Forêt fermée de sapin ou épicéa				
Forêt fermée à mélange d'autres conifères				
Forêt fermée d'un autre pin pur				
Forêt fermée de pin sylvestre pur	Autres Pin	22	Autres Pin	22
Forêt fermée à mélange de pins purs				
Forêt fermée de douglas pur	Douglas	23	Douglas	23
Forêt fermée de pin laricio ou pin noir pur	Pin laricio ou pin noir	24	Pin laricio ou pin noir	24
Forêt fermée de pin maritime pur	Pin maritime	25	Pin maritime	25
Forêt fermée à mélange de conifères	Ø		Mélange conifères	26
Forêt fermée de conifères purs en îlots	Ø		Conifères en îlots	27
Forêt fermée à mélange de conifères prépondérants et feuillus	Ø		Mélange de conifères prépondérants et feuillus	28
Forêt fermée à mélange de feuillus prépondérants et conifères	Ø		Mélange de feuillus prépondérants et conifères	29

Figure 3 - Correspondance entre les classes de la BD Forêt et les nomenclatures des cartes à obtenir

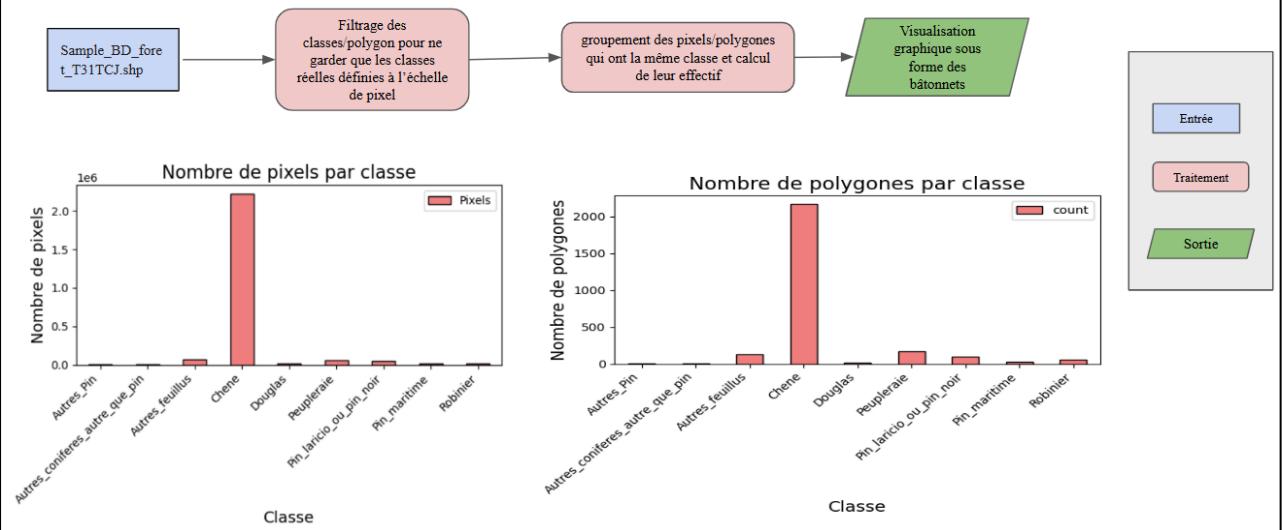
c. Analyse des échantillons

Une fois les échantillons correspondant aux classes d'intérêt extraits sous forme de jeu de données vectoriel, celui-ci conserve les attributs initiaux de la couche vecteur de la BD Forêt brute tout en ajoutant deux nouveaux attributs :

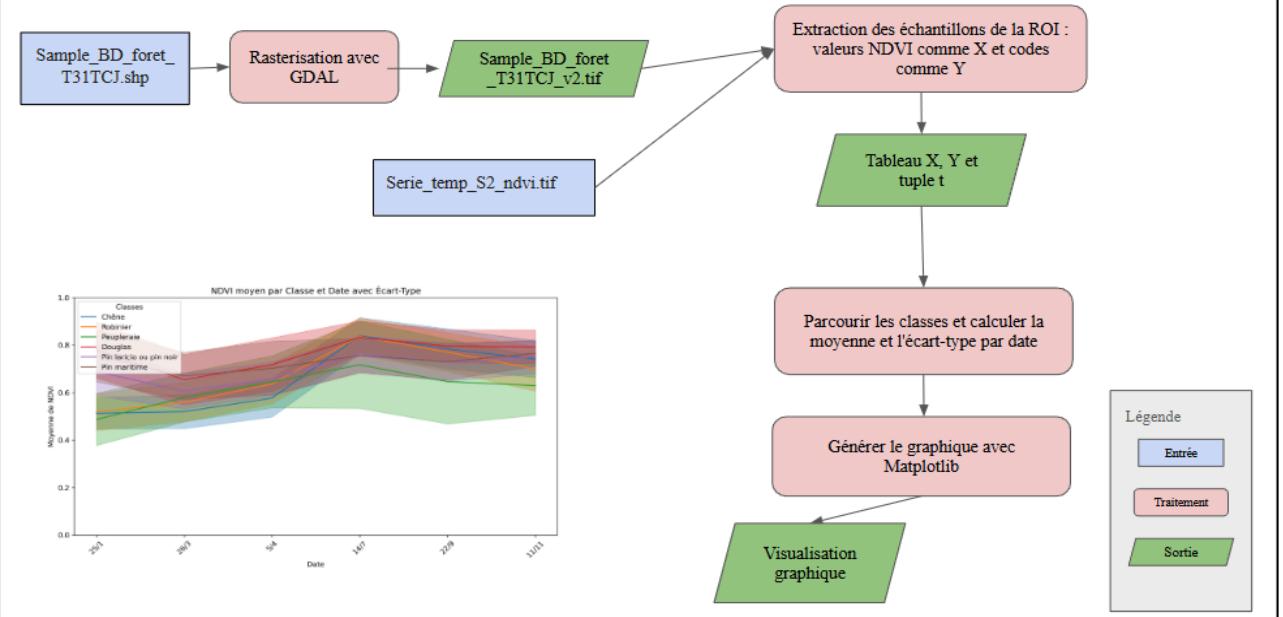
- "**Nom**" : indiquant la classe réelle de chaque échantillon.
- "**Code**" : associant à chaque classe un code unique correspondant.

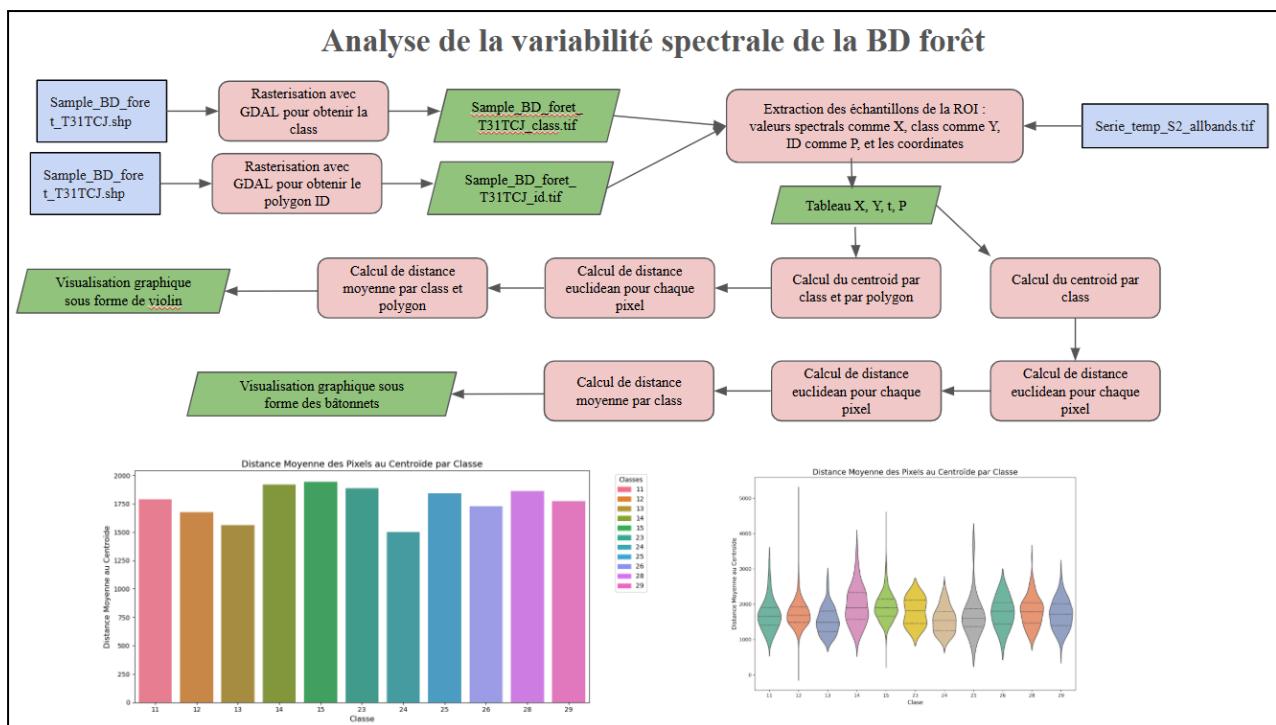
Nous analysons ensuite la répartition des polygones et des pixels dans chaque classe afin d'identifier les classes dominantes et minoritaires. Cette analyse nous permet de mieux comprendre la distribution des données et d'évaluer les résultats des classifications générées ultérieurement .

Distribution du nombre de pixels et polygons selon les classes



Phénologie des peuplements purs

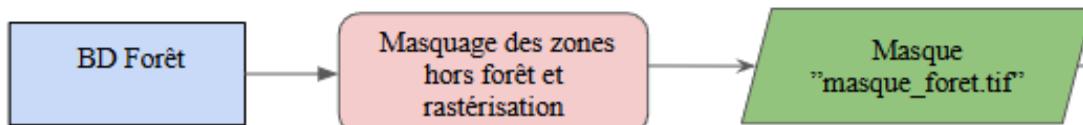




Figures 4 - Divers diagrammes de flux d'analyse des échantillons sélectionnées

d. Construction du masque forestier

Une étape clé de la méthodologie a été la construction d'un masque raster à partir de la BD Forêt afin de distinguer les zones à classer (zones forestières) des zones à ne pas classer (zones non forestières). Pour cela, nous avons exclu les polygones de type **lande**, **formation herbacée**, **forêt ouverte**, ainsi que la classe **forêt fermée sans couvert arboré**. Ces types de formations végétales ne correspondent pas aux peuplements forestiers pertinents pour notre analyse.



Le masque a été construit avec les caractéristiques suivantes :

- Format raster **GeoTIFF**, encodé en **8 bits**.
- Il possède la même emprise spatiale et la même résolution spatiale que les images Sentinel-2 prétraitées (c'est-à-dire une résolution de 10 m).
- Le fichier masque a été nommé **masque_foret.tif** et stocké dans le répertoire **results/data/img_pretraitees**.

Classe	Valeur du pixel
Zone de forêt	1
Zone hors forêt	0

- Le masque contient deux classes principales :
 - **Zone de forêt** : valeur du pixel = 1
 - **Zone hors forêt** : valeur du pixel = 0

Ce masque forestier est crucial pour garantir que seules les zones pertinentes soient incluses dans la classification supervisée. Les zones hors forêt ont été ignorées afin de limiter les erreurs liées à des classes non pertinentes, et ainsi améliorer la précision du modèle de classification. Le masque a été appliqué sur les images prétraitées afin de masquer les zones non pertinentes, ce qui a également réduit le temps de calcul nécessaire.

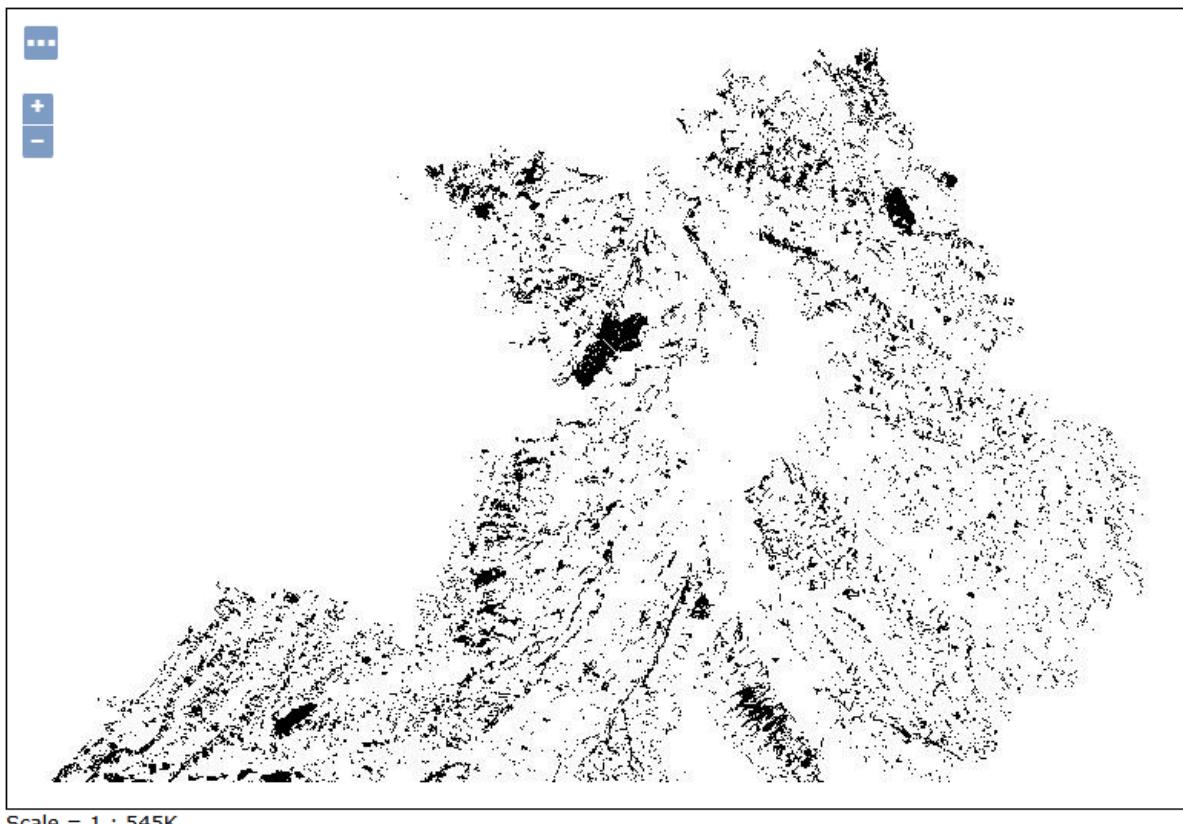


Figure 5 - Visualisation du masque forestier

e. Classification supervisée à l'échelle des pixels

Un modèle de classification supervisée a été mis en place, basé sur l'algorithme Random Forest de scikit-learn. Les échantillons d'apprentissage ont été utilisés pour entraîner le modèle, qui a ensuite été appliqué pour cartographier les essences forestières à l'échelle du pixel (fig. 6).

Le modèle Random Forest a été choisi en raison de sa capacité à gérer des données complexes, non linéaires, et sa robustesse face aux problèmes de surapprentissage. Les hyperparamètres ont été optimisés pour maximiser la précision sans compromettre la généralisation. Nous avons fixé comme valeur de nombre optimale d'itération égale à 30 et un nombre de segmentation de jeu de données d'entraînement égale à 5 (nb_folds).

Il est important de noter qu'une étape a été ajoutée afin de vérifier les rapports de classifications et les matrices de confusion produites. En effet, étant donné que certaines classes sont très peu représentées, il est arrivé que celles-ci ne soient pas présentes dans les indicateurs de qualité. Nous avons donc réalisé deux corrections supplémentaires. D'abord, dans le cas des rapports de classifications, nous avons rajouté autant de colonne de 0 qu'il manquait de classes dans chaque tableau. Ceci nous permet donc de garder les informations des autres classes pour la validation. Ensuite pour les matrices de confusions, on a décidé d'éviter le problème en supprimant chaque matrice n'ayant pas toutes les classes.

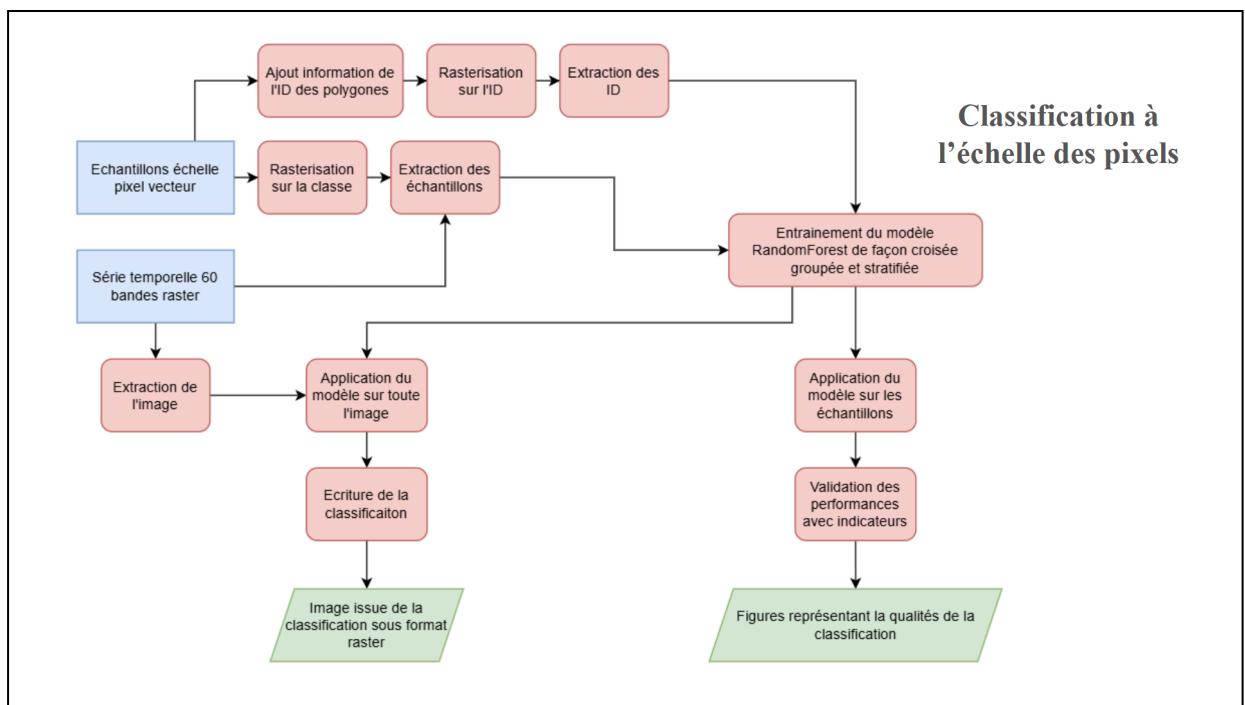


Figure 6 - Diagramme de flux de la classification à l'échelle des pixels

f. Classification à l'échelle des peuplements

La classification à l'échelle du pixel a permis de capter les différences subtiles entre les essences, tandis que l'agrégation des résultats à de données définis à l'échelle des peuplements a offert une vision simplifiée, mieux adaptée à certaines applications de gestion forestière.

Pour réaliser cette classification, on a mis en place de nouvelles classes de peuplements en respectant les règles illustrées dans l'arbre de décisions ci-dessous (fig.7).

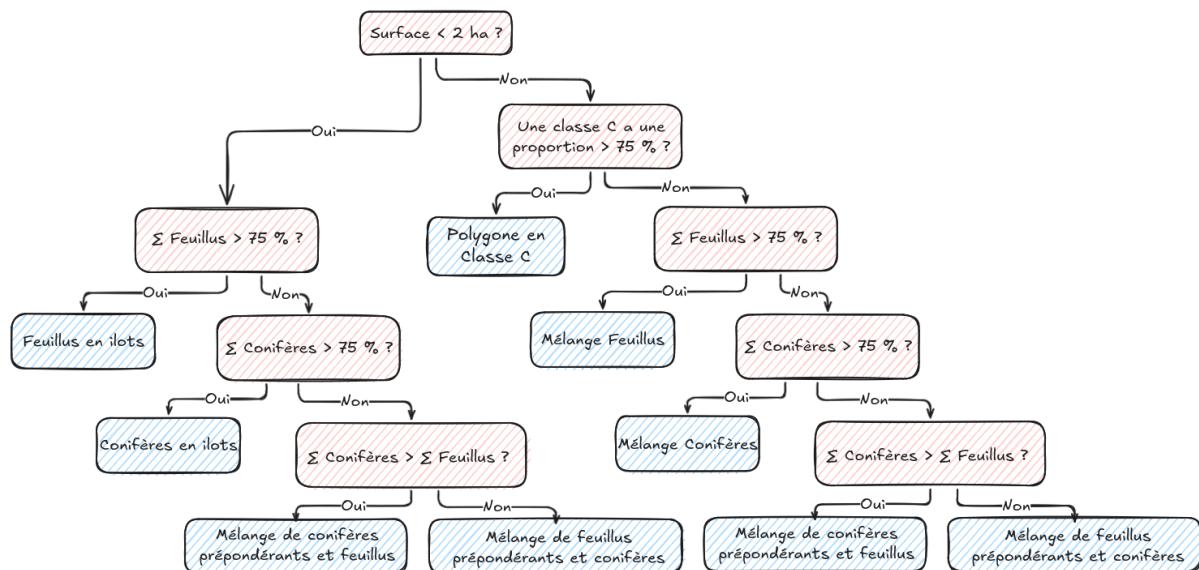


Figure 7 - Arbre de décisions pour la mise en place des classes de peuplements

Pour prédire les classes à l'échelle du polygone, nous avons d'abord calculé la superficie de chaque polygone ainsi que les pourcentages des différentes classes qui y sont présentes. Ce calcul a nécessité l'utilisation de la fonction "**zonal_stat**" de Rasterstats, qui prend en entrée la couche d'échantillons au format vecteur ainsi que le raster issu de la classification à l'échelle du pixel. La statistique "**count**" a été appliquée afin de comptabiliser le nombre de pixels par polygone.

Ensuite, en parcourant les échantillons polygone par polygone, nous avons déterminé, pour chaque classe, le pourcentage à l'échelle du polygone. Ce pourcentage a été obtenu en divisant le nombre de pixels de la classe, calculé avec **zonal_stat**, par le nombre total de pixels (fig. 8)

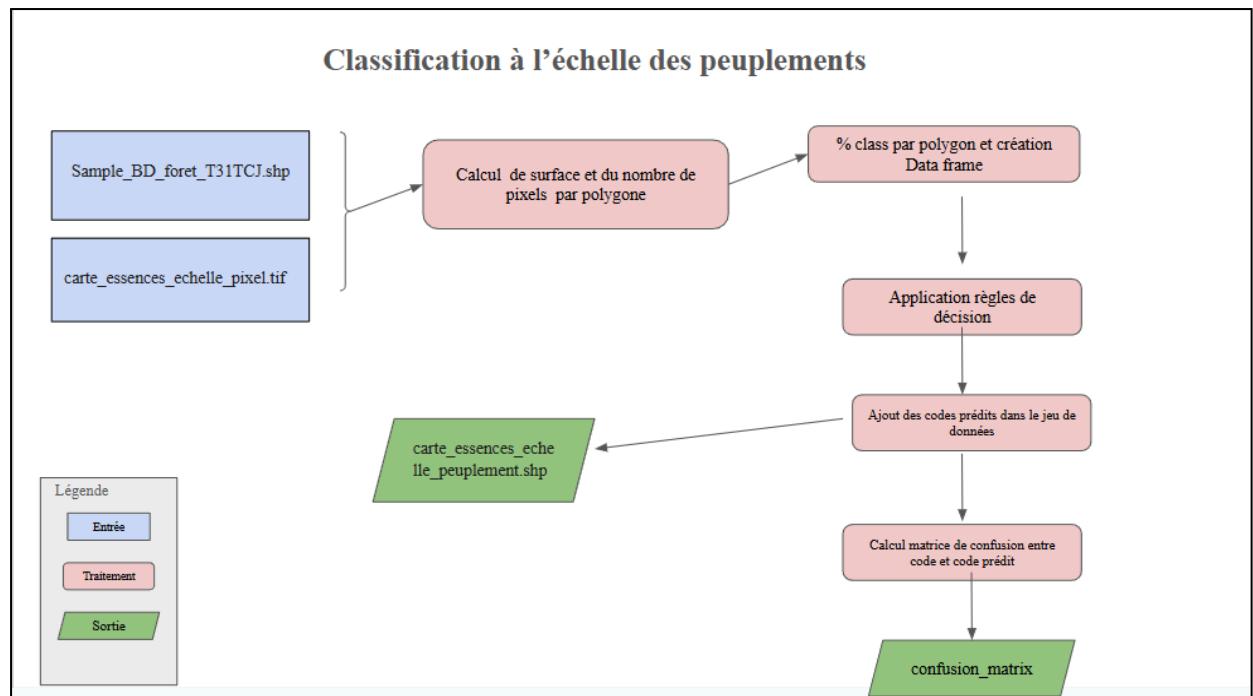


Figure 8 - Diagramme de flux de la classification à l'échelle des peuplements

3. Résultats

2.1. Analyse des échantillons

a. Distribution des échantillons

L'analyse des échantillons a permis de mettre en évidence la répartition des classes forestières et leur représentativité. Un histogramme des polygones et des pixels par classe a révélé des disparités dans leur distribution, certaines classes étant surreprésentées notamment la classe “Chêne”.

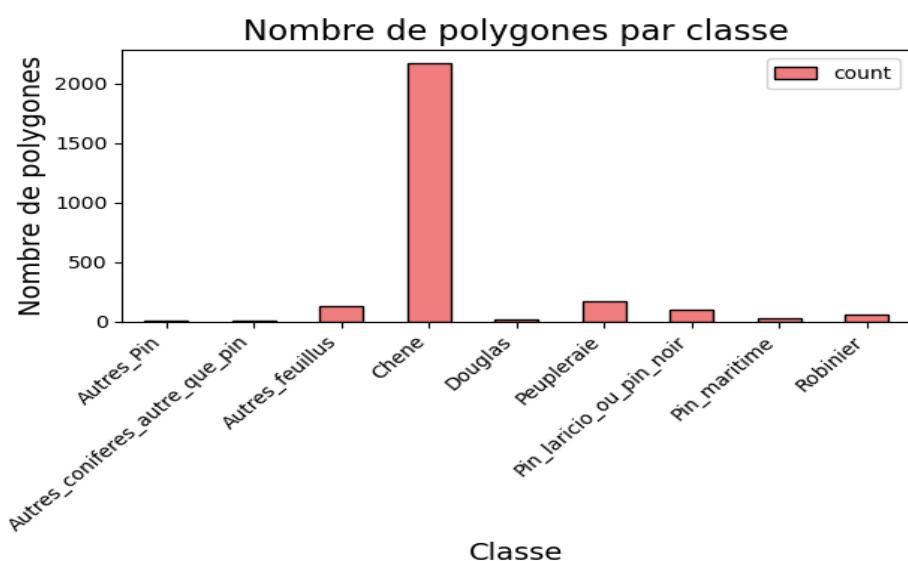


Figure 9 - Répartition des polygones par classe

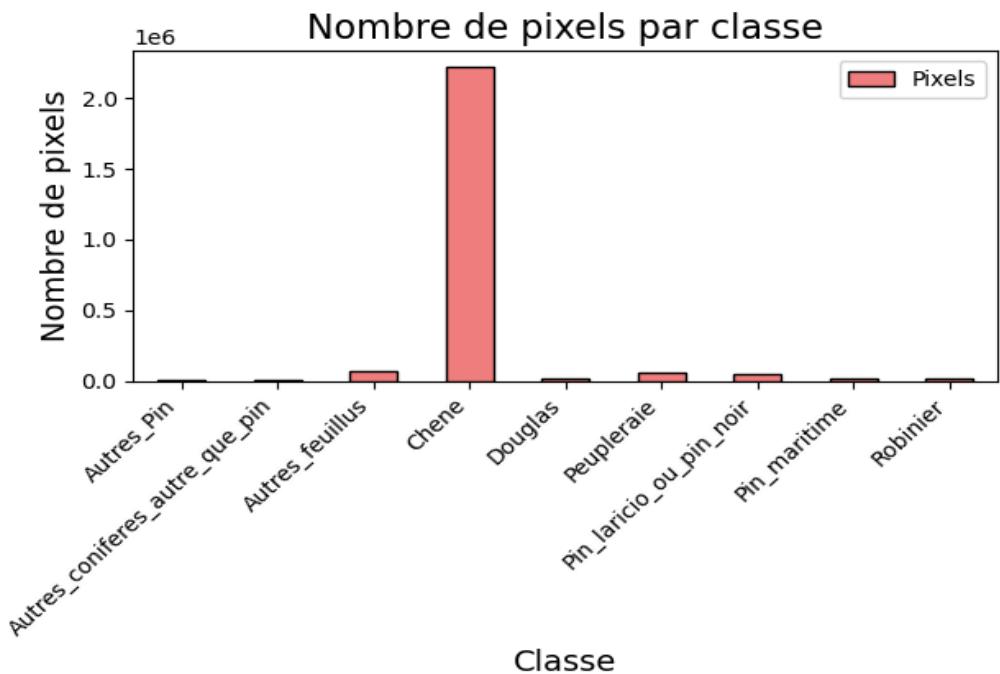


Figure 10 - Répartition des pixels par classe

Des diagrammes en violon ont également été générés afin de visualiser la répartition des classes au sein des polygones. La figure 11 montre la distribution des pixels inscrits dans un polygone pour les différentes classes sélectionnées, après application d'un filtre visant à exclure la classe "Chêne". Ce choix a été justifié par l'observation de la figure 11, où la forte prédominance de la classe "chêne" en nombre de pixels a empêché une visualisation claire de la répartition des pixels par polygone dans les autres classes.

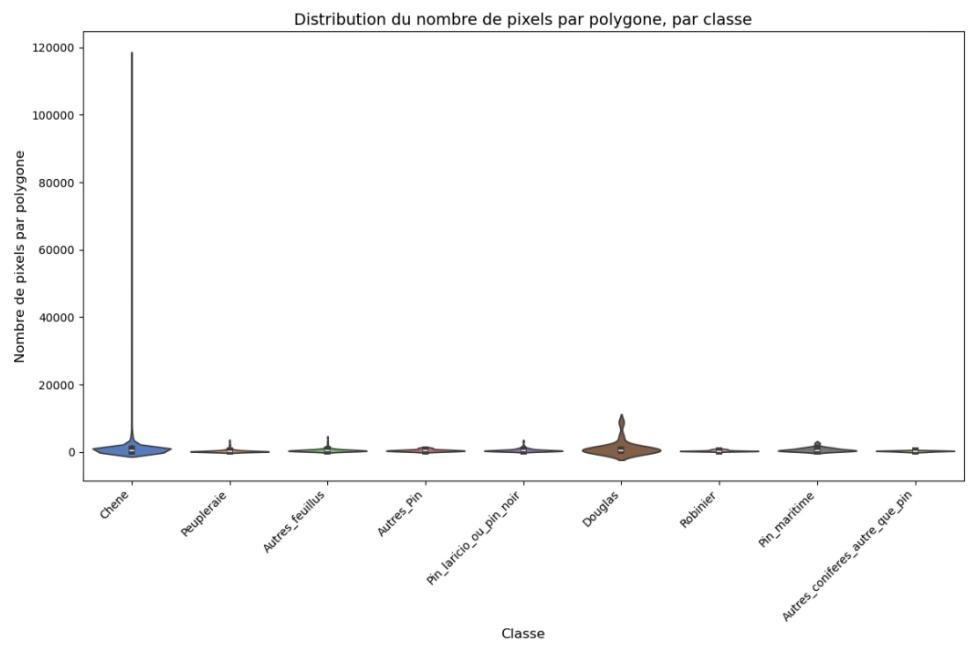


Figure 11 - Répartition des pixels par polygone et par classe, sans filtrage des données relatives à la classe dominante

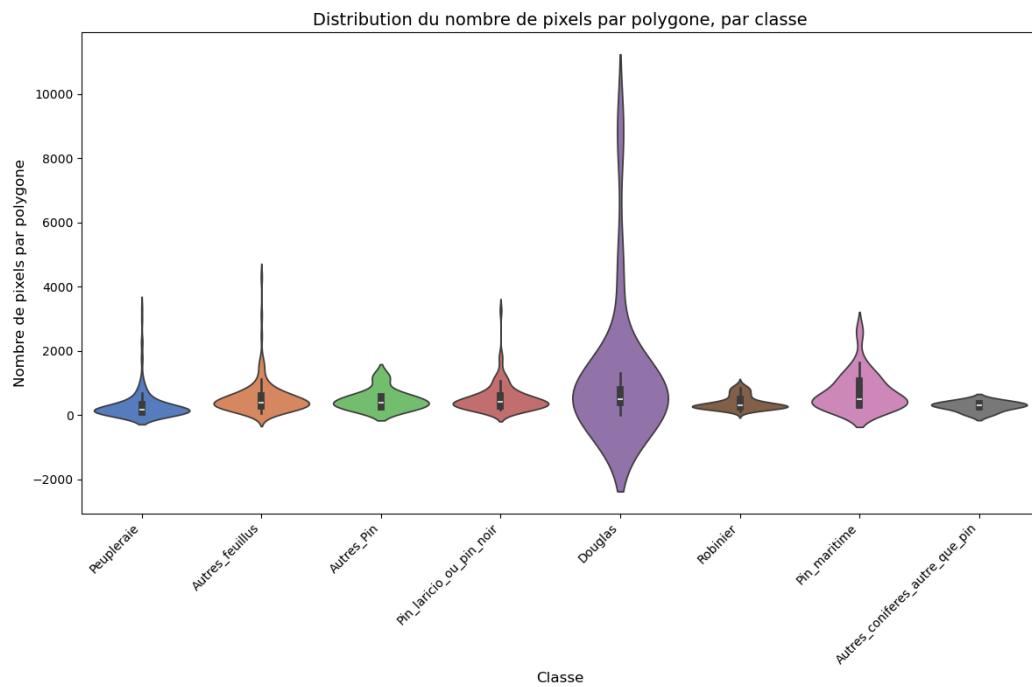


Figure 12 - Distribution de données par polygone et par classe, avec filtrage des données relatives à la classe dominante “Chêne”

b. Phénologie des classes forestières

Les signatures temporelles NDVI ont permis d'identifier des comportements phénologiques distincts, essentiels pour différencier les classes forestières. Les courbes de NDVI montrent des variations saisonnières caractéristiques pour chaque classe, facilitant leur distinction.

La phénologie des classes forestières a été analysée à travers les valeurs moyennes et les écarts types de NDVI pour chaque classe sur plusieurs dates. Ces analyses ont permis de distinguer des classes ayant des dynamiques saisonnières spécifiques, comme les feuillus qui présentent une forte variation de NDVI entre l'été et l'hiver, par rapport aux résineux qui ont un NDVI plus stable. Ces différences phénologiques ont été exploitées par le modèle de classification pour améliorer la précision de la distinction entre essences.

Peupleraie est l'espèce présentant la plus grande variabilité, avec un pic de NDVI estival particulièrement bas et l'un des NDVI les plus faibles en hiver. Chez les feuillus, Chêne montre un rebond printanier plus lent que les autres, tandis que Robinier est l'espèce au redémarrage le plus rapide.

Concernant les conifères, Douglas conserve les valeurs de NDVI les plus élevées tout au long de l'année. Pin maritime présente des valeurs similaires en hiver et au printemps mais plus faibles en été, tandis que Pin laricio ou noir affiche les NDVI les plus bas.

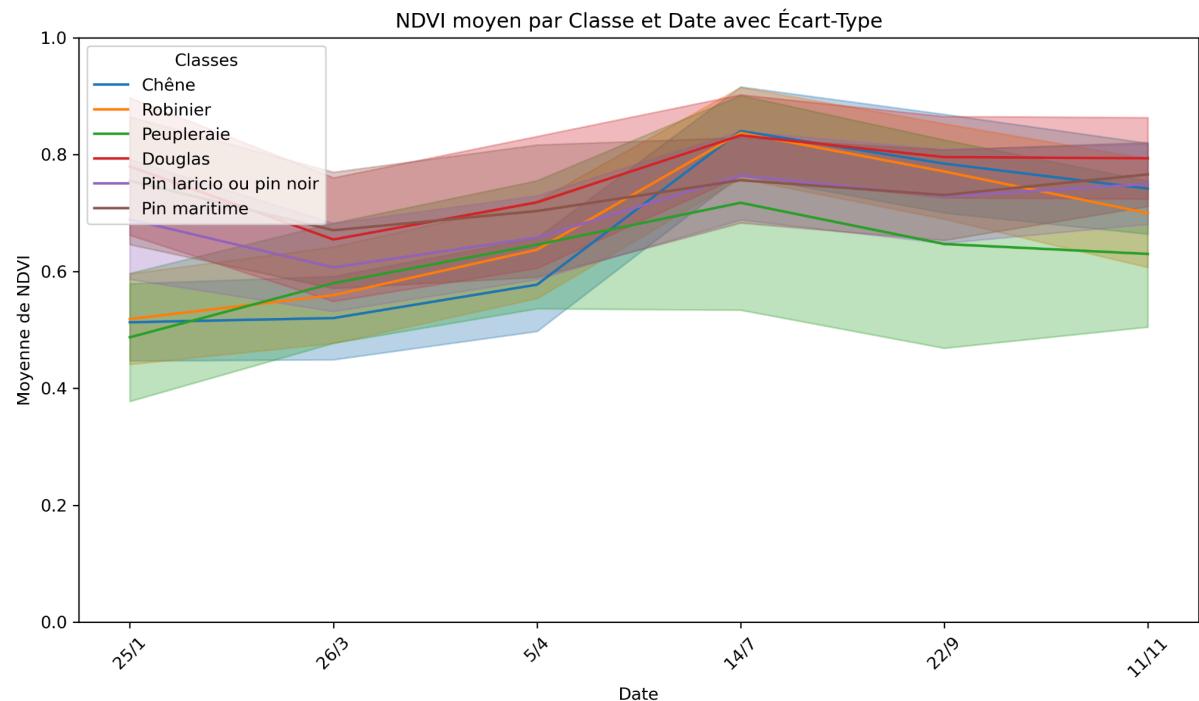


Figure 13 - Visualisation des signatures temporelles NDVI

c. Variabilité spectrale

L'étude de la variabilité spectrale a révélé que les classes en mélange (15, 26, 28 et 29) présentaient une dispersion plus moins grande que les classes pures, ce qui complique leur classification. La distance moyenne au centroïde a été utilisée pour mesurer cette variabilité, confirmant la complexité des classes mixtes.

Pour mesurer la variabilité spectrale, la distance moyenne des pixels au centroïde de la classe a été calculée. Cette mesure permet d'évaluer la cohérence interne de chaque classe : plus la distance est faible, plus les pixels d'une même classe sont spectralement similaires.

Les résultats ont révélé que les classes, qu'elles soient pures ou en mélange, présentent une variabilité intraclassse quasiment similaire. Toutefois, certaines valeurs de distances moyennes particulièrement élevées ont été observées pour la classe "Chêne" et la classe "mélange de feuillus" (fig. 14).

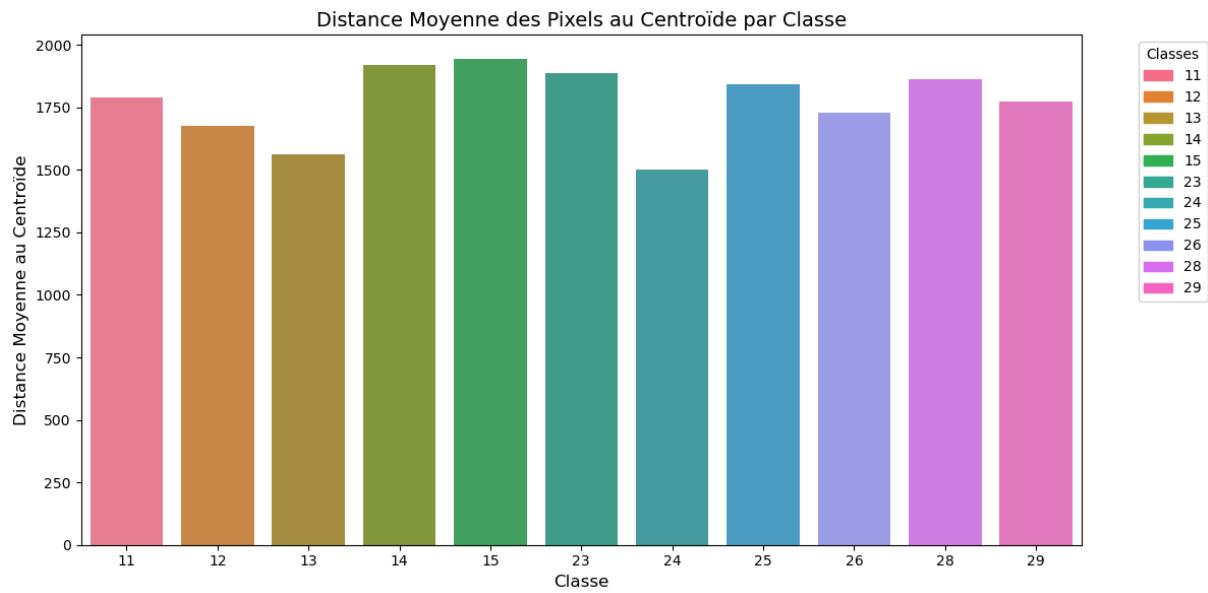


Figure 14 - Distance par rapport au centroïde par classe

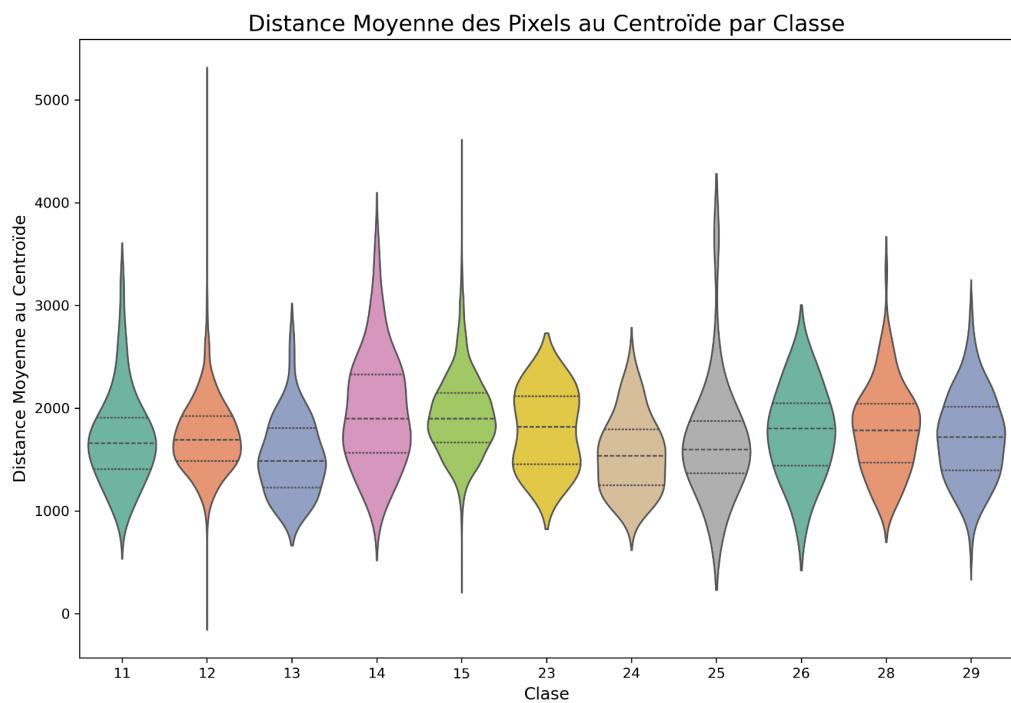
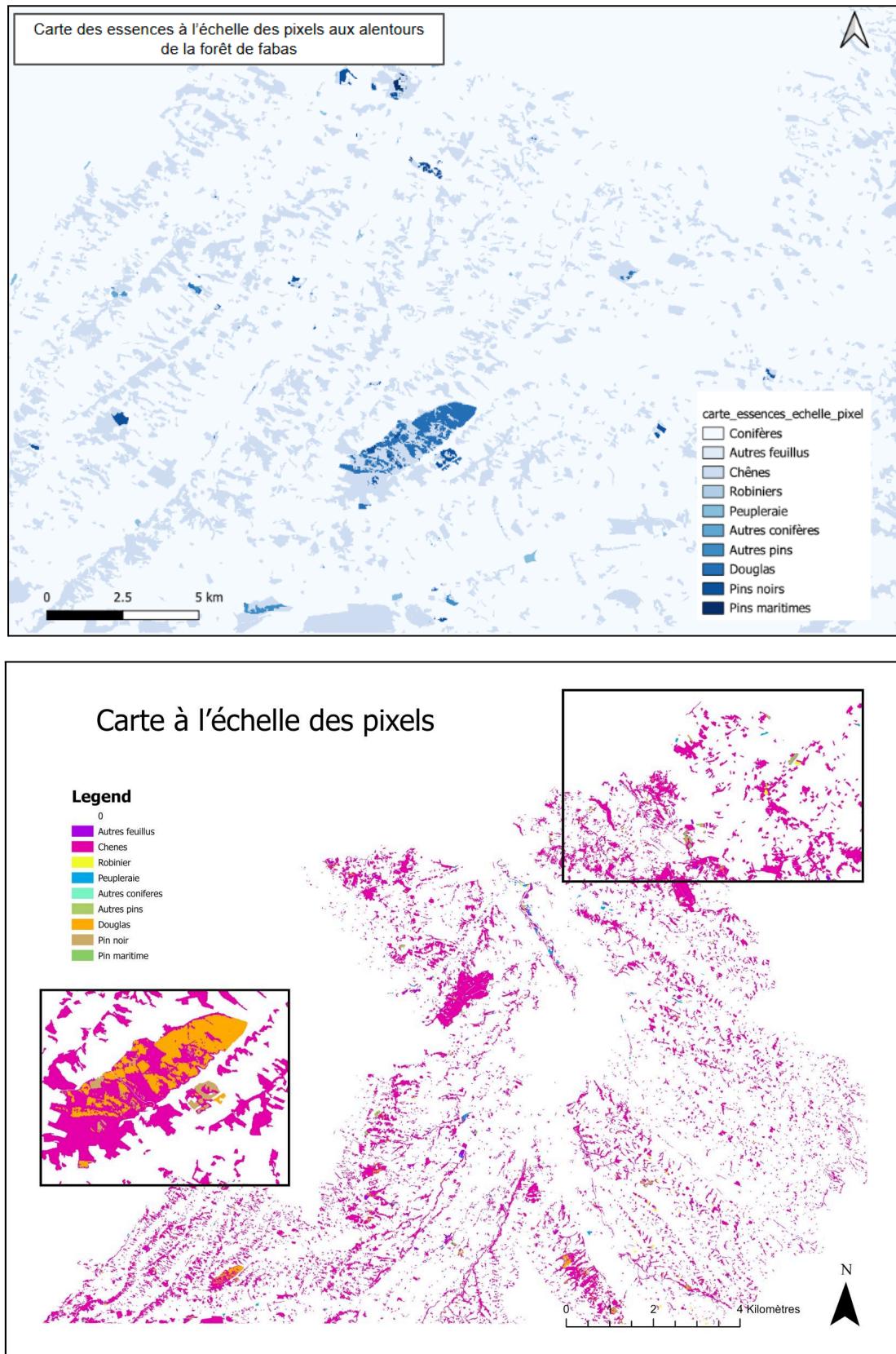


Figure 15 - Distribution des distances par rapport au centroïde par classe

2.2. Carte des essences et des peuplements et qualité des classifications

Les cartes produites offrent une vision détaillée des essences forestières à deux échelles :

À l'échelle du pixel : La classification Random Forest a permis de distinguer les différentes essences et de créer la représentation cartographique à l'échelle du pixel.



Figures 16 - Représentations à l'échelle de pixel avec différents contrastes de couleurs

La classification constitue une base d'information pour l'analyse des essences forestières. La classification Random Forest appliquée à la cartographie des essences forestières à l'échelle du pixel a permis d'obtenir une différenciation des types d'essences présents mais c'est nécessaire d'analyser quelle est la qualité de ces données générées.

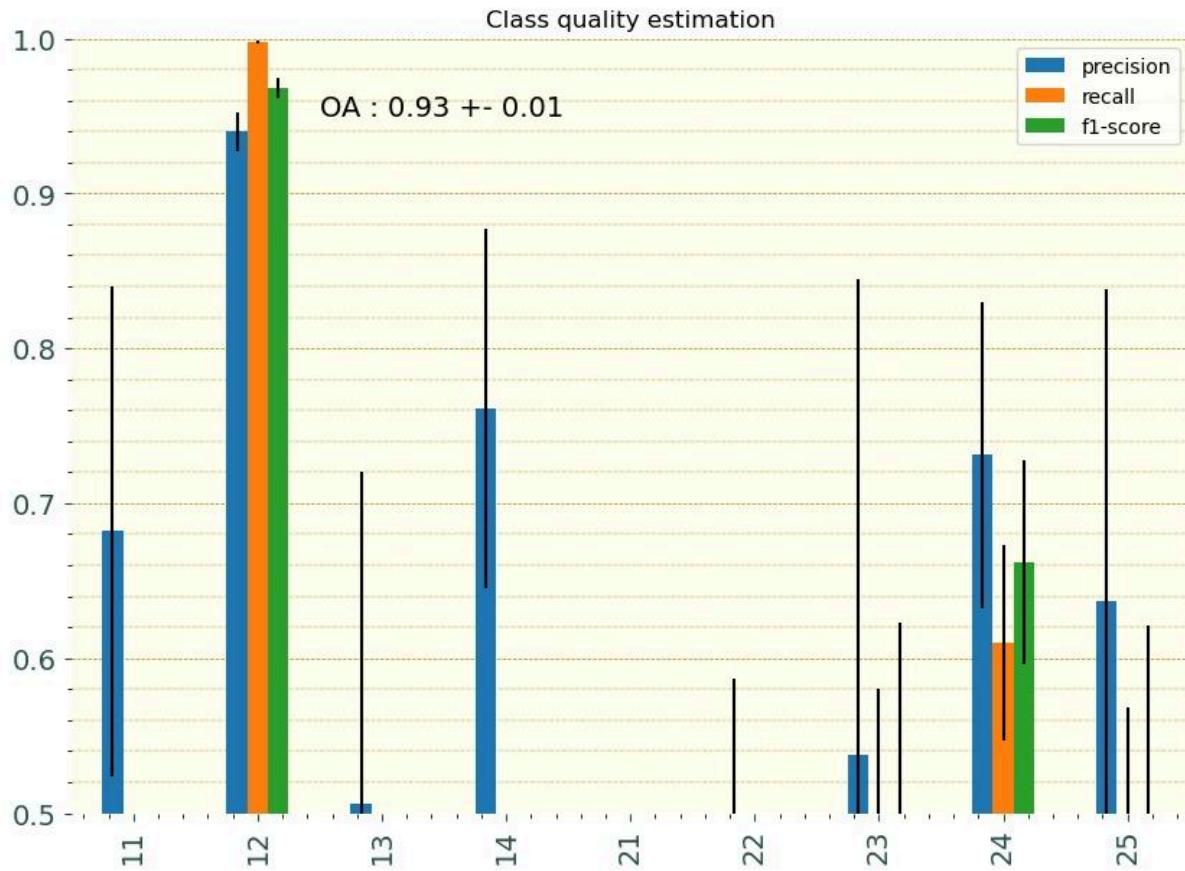


Figure 17- Estimation de la qualité de la classification à niveau de pixel

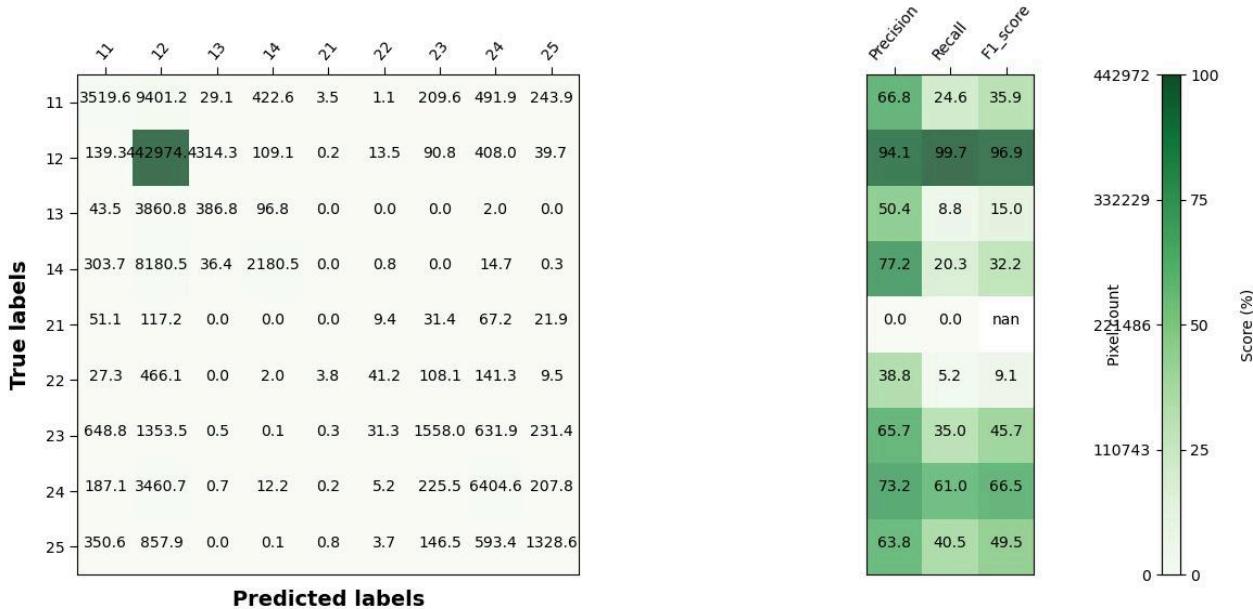


Figure 18 - Matrice de confusion classification à l'échelle des pixels

Les "True Labels" (étiquettes réelles) sont sur l'axe vertical, et les "Predicted Labels" (étiquettes prédites par le modèle) sont sur l'axe horizontal. Chaque cellule indique le nombre de pixels/classifications associés à une paire (vérité terrain et prédition).

Pour chaque classe, on a trois métriques :

- Précision : proportion de prédictions correctes parmi toutes les prédictions
- Rappel : proportion de bonnes prédictions par rapport au total des échantillons réels
- F1-score : moyenne harmonique de la précision et du rappel, indiquant un équilibre

La couleur des cellules reflète la performance : vert foncé pour des scores élevés et vert clair pour des scores plus faibles.

- La classe "12" a d'excellentes performances (Précision : 94.1 %, Rappel : 93.9 %, F1-score : 96.8 %).
- Les classes "11" et "24" ont des scores moyens.
- La classe "13" est mal classée avec des scores très faibles, probablement à cause de sa faible représentativité ou de sa proximité spectrale avec d'autres classes.

Les matrices de confusion ont révélé des performances variées selon les classes, mettant en avant des forces (classes pures bien identifiées) et des limites (confusions pour les classes en mélange):

La précision globale de la classification à l'échelle du pixel semble satisfaisante (OA = 93 %). Toutefois, ce résultat s'explique en grande partie par la capacité du modèle à identifier avec une grande précision la classe dominante **Chêne**, ce qui est logique. C'est pourquoi il est essentiel

d'examiner la précision classe par classe. Cette analyse révèle alors des confusions et des valeurs de précision bien plus faibles entre classes ayant des caractéristiques similaires, notamment entre les différentes catégories de conifères (pins noirs, pins maritimes, douglas, etc.) et entre certains feuillus. Ces confusions sont principalement dues aux similitudes spectrales des espèces en télédétection, à la variabilité intra-espèce et aux limites inhérentes à la résolution des données utilisées.

Certaines classes, comme les forêts de résineux pures (23, 24, 25), ont été bien détectées grâce à leur signature spectrale distinctive. En revanche, les peuplements mixtes, en particulier les classes "**Autres conifères autres que pin**" et "**Autres pins**" (21, 22), ont présenté des taux de confusion élevés.

Bien que les classes pures aient été globalement mieux identifiées, la classe **Peupleraie** (13) affiche une précision très faible. Cela pourrait s'expliquer par sa forte variabilité, déjà observée dans l'analyse de la signature temporelle de son NDVI, qui présente un écart-type élevé par rapport aux autres essences pures.

Ces résultats suggèrent qu'une amélioration des données d'entrée ou l'adoption d'une approche de classification plus avancée pourrait être nécessaire pour limiter ces confusions et affiner la précision de la classification.

Enfin, bien que la cartographie à l'échelle du pixel permette une représentation fine du territoire, elle peut également introduire des erreurs et un effet de fragmentation qui ne reflète pas nécessairement la réalité des peuplements forestiers. Une approche complémentaire à l'échelle des peuplements pourrait ainsi permettre de lisser ces effets et d'améliorer la cohérence spatiale des classes identifiées.

À l'échelle des peuplements : L'agrégation des résultats a simplifié la lecture des cartes tout en reflétant la composition dominante des peuplements. Cette classification à l'échelle des peuplements constitue une avancée en termes de synthèse et d'interprétation des données forestières. Elle fournit une représentation plus robuste et opérationnelle des formations forestières, bien qu'une amélioration des critères de classification des peuplements mixtes pourrait encore affiner la précision des résultats.

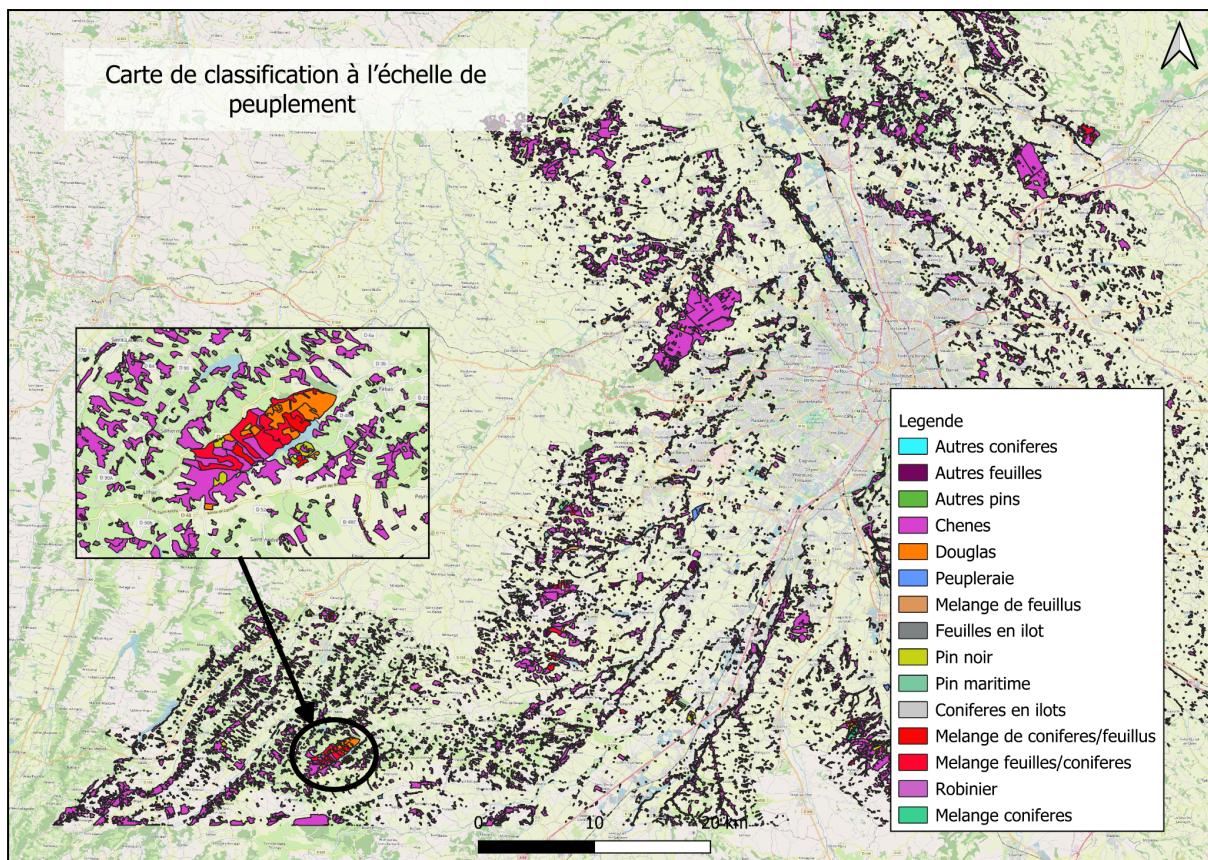


Figure 19 - Carte à l'échelle des peuplements obtenu de la reclassification (colonne “code_predit”)

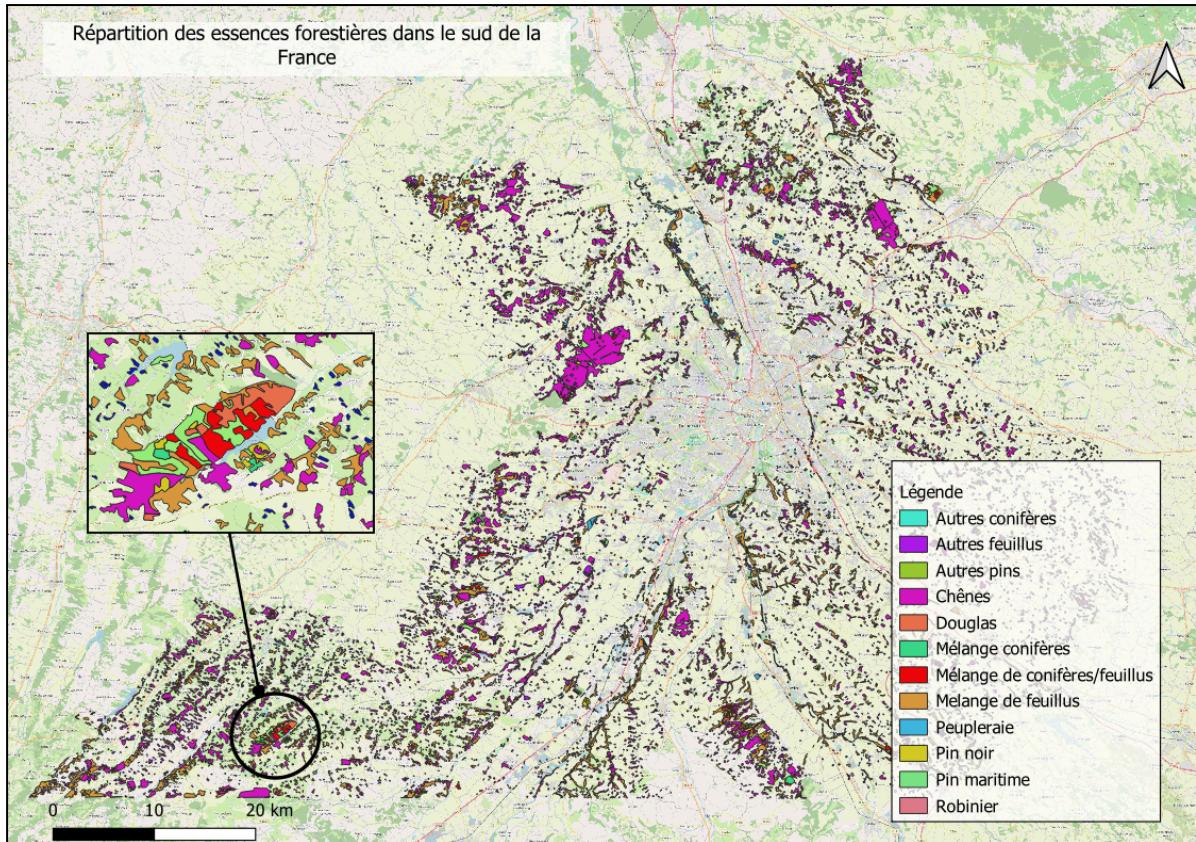


Figure 20 - Représentation des essences forestières original de la BD forêt (colonne “code”)

L’agrégation des résultats à l’échelle des peuplements apporte une simplification notable de la lecture cartographique, en réduisant la variabilité observée à l’échelle du pixel. Cette approche permet une meilleure identification des unités homogènes de végétation et reflète plus fidèlement la composition dominante des peuplements forestiers.

Pour s’assurer de la qualité des résultats obtenus , une matrice de confusion a été calculée en faisant la comparaison entre le code réel associé à chaque classe et le code prédit les valeurs de précision ainsi que les histogrammes de qualité de classification ont témoigné des niveaux de précision variables entre les essences conifères et feuillus avec une forte confusion dans la détection de la classe mélange feuillus (15). De plus , pour les peuplements mixtes (28 et 29) , on constate qu’ils ont été détectés avec des faibles valeurs de précision indiquant une confusion avec d’autres classes.

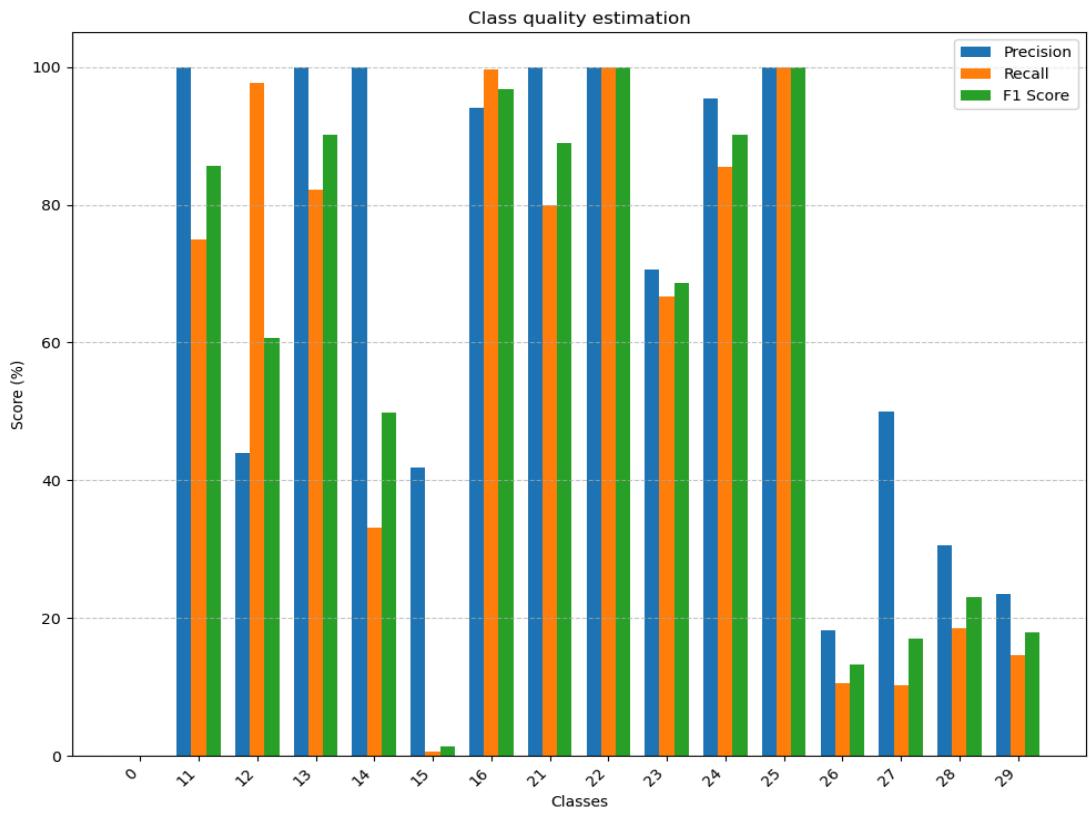


Figure 21 -Estimation de la qualité de classification au niveau de peuplement



Figure 22 - Matrice de confusion classification à l'échelle des peuplements

La matrice présentée ci-dessus montre que les échantillons ayant comme code 12, 16, 22 et 25 ont persisté avec leurs classes originelles avec des valeurs de pourcentages de précision proche ou égale à 100%.

En revanche, pour la classe 12, un nombre important de nouveaux polygones lui ont été attribués. Cela s'explique par le fait que la classification à l'échelle des pixels, combinée aux nouvelles règles de décision, a entraîné le reclassement d'une grande partie des échantillons initialement codés 15 (mélange de feuillus) vers la classe 12. Autrement dit, l'analyse a révélé que plus de 75 % des essences présentes dans ces mélanges étaient des chênes.

Par ailleurs, en observant les polygones correspondant aux mélanges de conifères et feuillus (classes 28 et 29), on constate que certaines portions ont été reclassées en classes pures, tandis que d'autres restent identifiées dans la même classe ou dans d'autres peuplements mélangés.

Cette persistance s'explique par la présence, au sein de ces polygones, de zones toujours classées comme appartenant à d'autres types de mélanges.

La présence de classes composites comme le "Mélange de feuillus prépondérants et conifères" ou le "Mélange conifères" témoigne d'une difficulté à discriminer certaines formations mixtes. Ces difficultés peuvent être dues à des limites de la télédétection pour différencier précisément les compositions forestières à partir de l'imagerie utilisée. L'évaluation de la classification, réalisée à l'aide d'échantillons de validation indépendants, a permis de mesurer les performances du modèle en termes de précision, de rappel et de score F1. Les résultats suggèrent une amélioration globale par rapport à la classification pixel, bien que des marges de progression existent pour certaines classes spécifiques.

4. Analyse critique

d. Discussion

Les résultats obtenus montrent des performances globalement satisfaisantes pour certaines classes, tandis que d'autres présentent des confusions notables. Ces confusions peuvent être attribuées à plusieurs facteurs :

1. **Similitudes spectrales** : Certaines classes d'essences, telles que les mélanges de feuillus et de conifères, présentent des signatures spectrales très similaires, ce qui rend leur différenciation difficile. Cela a entraîné des confusions, en particulier dans les zones où les proportions d'essences sont équilibrées. Par exemple, la confusion entre les mélanges de conifères et de feuillus est cohérente avec les caractéristiques spectrales proches de ces types de végétation.
2. **Hétérogénéité** : La variabilité intra-classe peut également expliquer certaines erreurs de classification. Par exemple, certaines classes comme les peuplements mixtes peuvent présenter une grande diversité de conditions locales (densité, type de sous-bois, etc.), ce qui complique la tâche du modèle.
3. **Qualité des données** : La qualité des données d'entrée, notamment la couverture nuageuse résiduelle et la résolution spatiale des images, a pu limiter la précision de la classification. Des nuages résiduels ou des ombres peuvent altérer les valeurs des indices spectrales, entraînant des erreurs de classification.

e. Limites et pistes d'amélioration

Plusieurs axes d'amélioration et de développement peuvent être envisagés :

- **Règles de décision** : Les règles de décision utilisées pour la classification des peuplements forestiers sont basées sur des seuils fixes. Ces règles pourraient être affinées pour mieux refléter la diversité des peuplements, par exemple en utilisant des approches basées sur des critères plus adaptatifs ou des méthodes d'apprentissage automatique plus avancées.
- **Enrichissement des données** : L'utilisation de données complémentaires, telles que des données LiDAR pour une meilleure estimation de la structure verticale de la végétation, pourrait permettre d'améliorer la qualité de la classification.
- **Stratégie d'échantillonnage** : La qualité des échantillons d'apprentissage a un impact direct sur la performance du modèle. Une stratégie d'échantillonnage plus représentative des différentes classes, en tenant compte de la variabilité intra-classe, pourrait améliorer la robustesse du modèle.

L'utilisation de données multispectrales supplémentaires, telles que des observations infrarouges thermiques, pourrait également améliorer la distinction entre certaines classes, notamment en période hivernale. En outre, des approches de fusion de données pourraient être explorées pour combiner des images optiques et radar, offrant une meilleure résilience aux conditions atmosphériques défavorables.

Les cartes produites sont utilisables dans le contexte de la gestion forestière, notamment pour le suivi des peuplements et la planification des interventions sylvicoles. Elles peuvent également être utiles pour des études d'impact environnemental, en fournissant une vue d'ensemble des essences présentes et de leur répartition spatiale.

Cependant, leur utilisation doit être faite avec prudence, en tenant compte des limites évoquées. Les résultats pourraient être particulièrement utiles pour une première évaluation à grande échelle, mais des vérifications sur le terrain seraient nécessaires pour valider les classifications dans des contextes spécifiques.

Enfin, ces travaux ouvrent des perspectives intéressantes pour l'amélioration des méthodes de classification des peuplements forestiers, notamment par l'intégration de nouvelles sources de données et l'utilisation de techniques d'apprentissage plus avancées, comme les réseaux de neurones convolutifs (CNN) pour une meilleure extraction des caractéristiques.

Conclusion

La gestion des forêts, face aux défis environnementaux et aux besoins croissants en ressources, nécessite des outils d'observation et d'analyse précis et actualisés. Ce projet s'est inscrit dans cette dynamique en explorant l'utilisation de la **BD Forêt® version 2.0** comme base d'échantillons de référence pour la classification supervisée de séries temporelles d'images Sentinel-2.

À travers une méthodologie rigoureuse, intégrant le prétraitement des données, la sélection des échantillons, la classification supervisée et l'analyse critique des résultats, plusieurs objectifs ont été atteints :

- **La production de cartes d'essences forestières à l'échelle du pixel et à l'échelle des peuplements.**
- **L'évaluation de la qualité des classifications, mettant en évidence les forces et limites des approches utilisées.**

Les résultats obtenus ont montré que la combinaison des données géographiques de la BD Forêt avec les capacités d'observation de Sentinel-2 permet de produire des cartes de qualité, bien qu'avec certaines limitations :

- **Les classes aux signatures spectrales proches ont entraîné des confusions logiques mais pénalisantes pour la précision des classifications.**
- **La variabilité intra-classe a parfois diminué la qualité des prédictions, notamment pour les peuplements mixtes.**

Malgré ces défis, la démarche a permis de démontrer l'intérêt de coupler bases de données géographiques et imagerie satellite pour une cartographie fine et actualisée des essences forestières. Les analyses phénologiques et spectrales menées ont également enrichi la compréhension des dynamiques végétales, ouvrant la voie à des applications futures.

Ce projet illustre bien le potentiel des outils de télédétection pour faciliter la prise de décision dans la gestion forestière et l'aménagement du territoire. Les résultats obtenus, bien que perfectibles, représentent une contribution significative à la modernisation des approches actuelles, répondant aux besoins des acteurs de la filière forêt-bois et de l'environnement.

Annexes

Annexe 1 - Utilité des scripts

my_function.py

Ce fichier regroupe toutes les fonctions réutilisables pour les autres scripts. Il est le cœur de la modularité du projet, permettant de structurer le code et de simplifier les appels dans les autres scripts.

Utilisation :

- Traiter les données géospatiales (filtrage des classes, création de masques, calculs statistiques).
- Effectuer des analyses spécifiques (calcul des distances au centroïde, NDVI, etc.).
- Gérer la classification supervisée (entraînement du modèle, validation croisée).
- Produire des graphiques et visualisations.

sample_curation.py

Ce script sert à sélectionner et préparer les échantillons d'apprentissage et de validation à partir de la BD Forêt.

Utilisation :

- Filtrer les polygones pertinents en fonction des classes d'intérêt.
- Générer un fichier vecteur contenant uniquement les échantillons valides.
- Produire le fichier Sample_BD_foret_T31TCJ.shp pour les étapes suivantes du projet.

Sortie : Sample_BD_foret_T31TCJ.shp, un fichier contenant les échantillons sélectionnés.

sample_analysis_nb_sample.py

Analyser la répartition des échantillons et produire des graphiques descriptifs.

Utilisation :

- Compter le nombre de polygones et de pixels par classe.
- Visualiser la répartition des échantillons sous forme de diagrammes en bâton.
- Identifier d'éventuelles classes surreprésentées ou sous-représentées.

Sorties :

- diag_baton_nb_poly_by_class.png : Histogramme du nombre de polygones par classe.

- diag_baton_nb_pix_by_class.png : Histogramme du nombre de pixels par classe.

sample_analysis_temp_signature.py

Étudier la phénologie des classes forestières en analysant leurs signatures temporelles NDVI.

Utilisation :

- Calculer les valeurs moyennes et les écarts types de NDVI pour chaque classe.
- Produire des graphiques montrant les évolutions temporelles du NDVI par classe.
- Identifier des tendances phénologiques permettant de distinguer les classes.

Sortie : temp_mean_ndvi.png, un graphique illustrant la variation temporelle du NDVI moyen par classe.

sample_analysis_spectral_variability.py

Analyser la variabilité spectrale des classes et comparer les classes pures et mixtes.

Utilisation :

- Calculer la distance moyenne au centroïde des pixels pour chaque classe.
- Visualiser la variabilité spectrale à l'échelle de l'image entière et des polygones individuels.
- Identifier si les classes mixtes présentent une plus grande variabilité spectrale que les classes pures.

Sorties :

- diag_baton_dist_centroide_classe.png : Histogramme de la distance moyenne au centroïde par classe.
- violin_plot_dist_centroide_by_poly_by_class.png : Violin plot de la variabilité spectrale à l'échelle des polygones.

pre_traitement.py

Préparer les données d'entrée, notamment les images Sentinel-2, pour les étapes de classification.

Utilisation :

- Harmoniser la résolution et la projection des bandes Sentinel-2.
- Découper les images selon l'emprise d'étude et appliquer un masque de forêt.
- Produire deux jeux de données :
 - Une image multi-bandes (Serie_temp_S2_allbands.tif).

- Une série temporelle d'images NDVI (Serie_temp_S2_ndvi.tif).

Sorties :

- Serie_temp_S2_allbands.tif : Image contenant toutes les bandes spectrales pour les 6 dates.
- Serie_temp_S2_ndvi.tif : Image représentant les NDVI pour les 6 dates.

classification_pixel.py

Produire une carte des essences forestières à l'échelle du pixel à partir d'une classification supervisée.

Utilisation :

- Entraîner un modèle Random Forest sur les échantillons sélectionnés.
- Appliquer le modèle sur l'image multi-bandes Sentinel-2.
- Générer une carte raster représentant les essences forestières à l'échelle du pixel.

Sortie : carte_essences_echelle_pixel.tif, une carte des essences forestières à l'échelle du pixel.

classification_stand.py

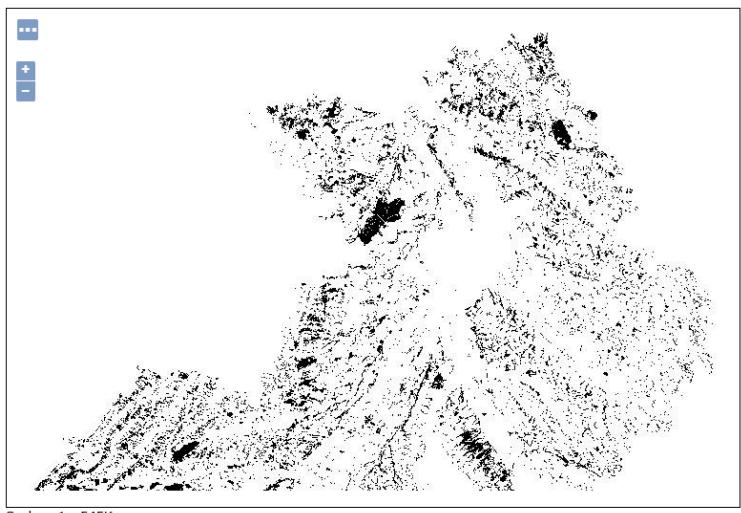
Produire une carte des peuplements forestiers en agrégeant les résultats de la classification pixel.

Utilisation :

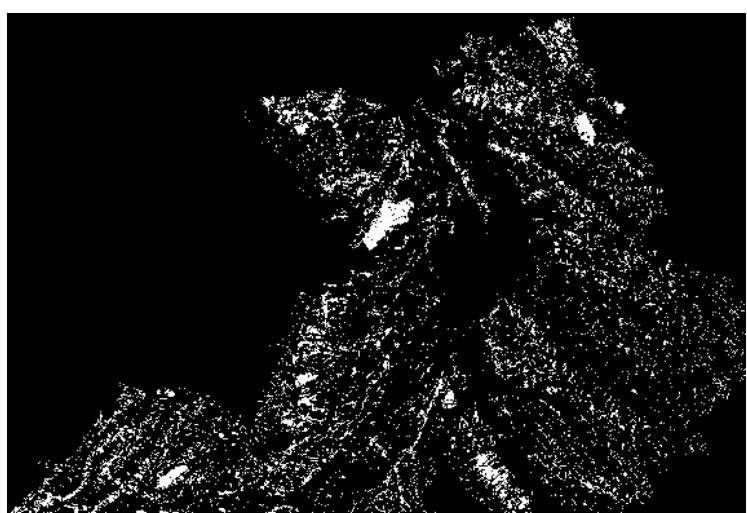
- Calculer la composition des classes de pixels pour chaque polygone de la BD Forêt.
- Appliquer des règles décisionnelles pour attribuer une classe dominante à chaque polygone.
- Mettre à jour le fichier vecteur avec les nouvelles classifications.

Sortie : Sample_BD_foret_T31TCJ_updated.shp, un fichier vecteur avec les classes prédominantes attribuées à chaque peuplement.

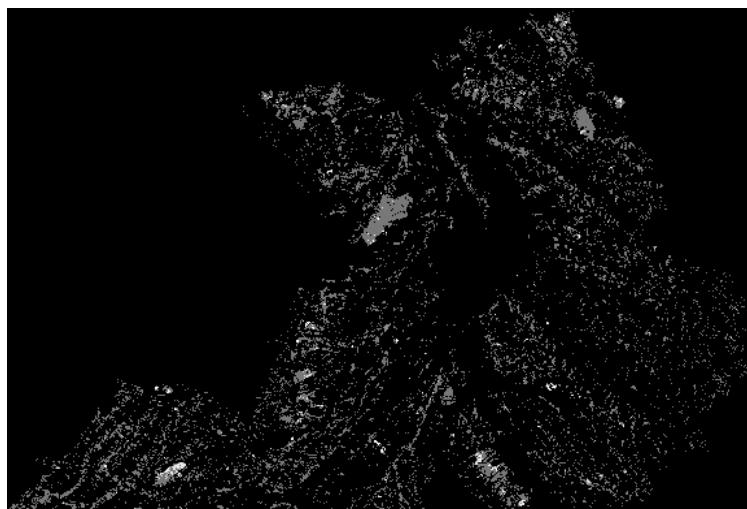
Annexe 2 - Les différentes images utilisées dans les traitements
“masque forêt”



“Serie_temp_S2_ndvi”



“carte_essences_echelle_pixel”



Annexe 3 - Diagrammes de flux

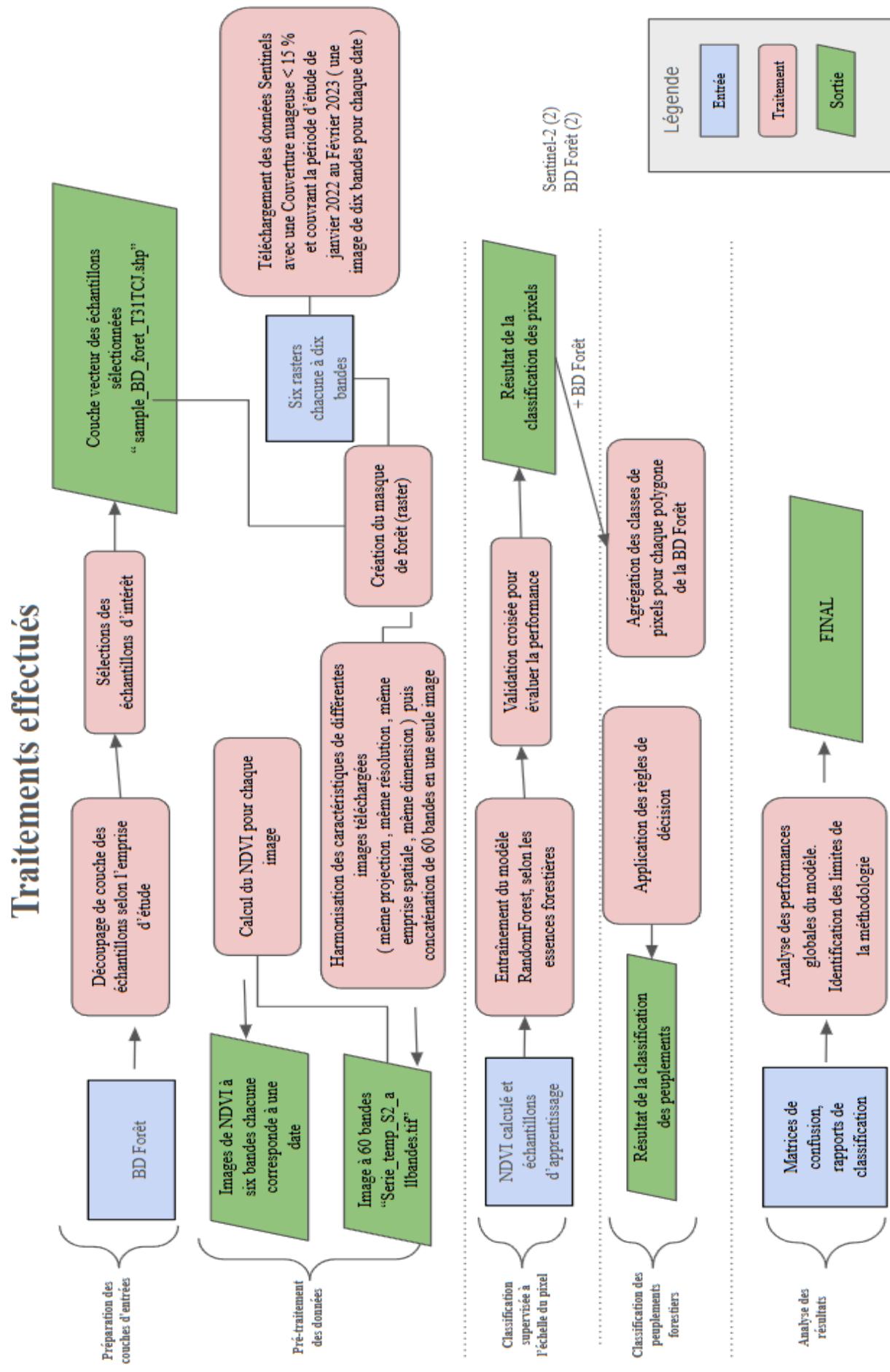
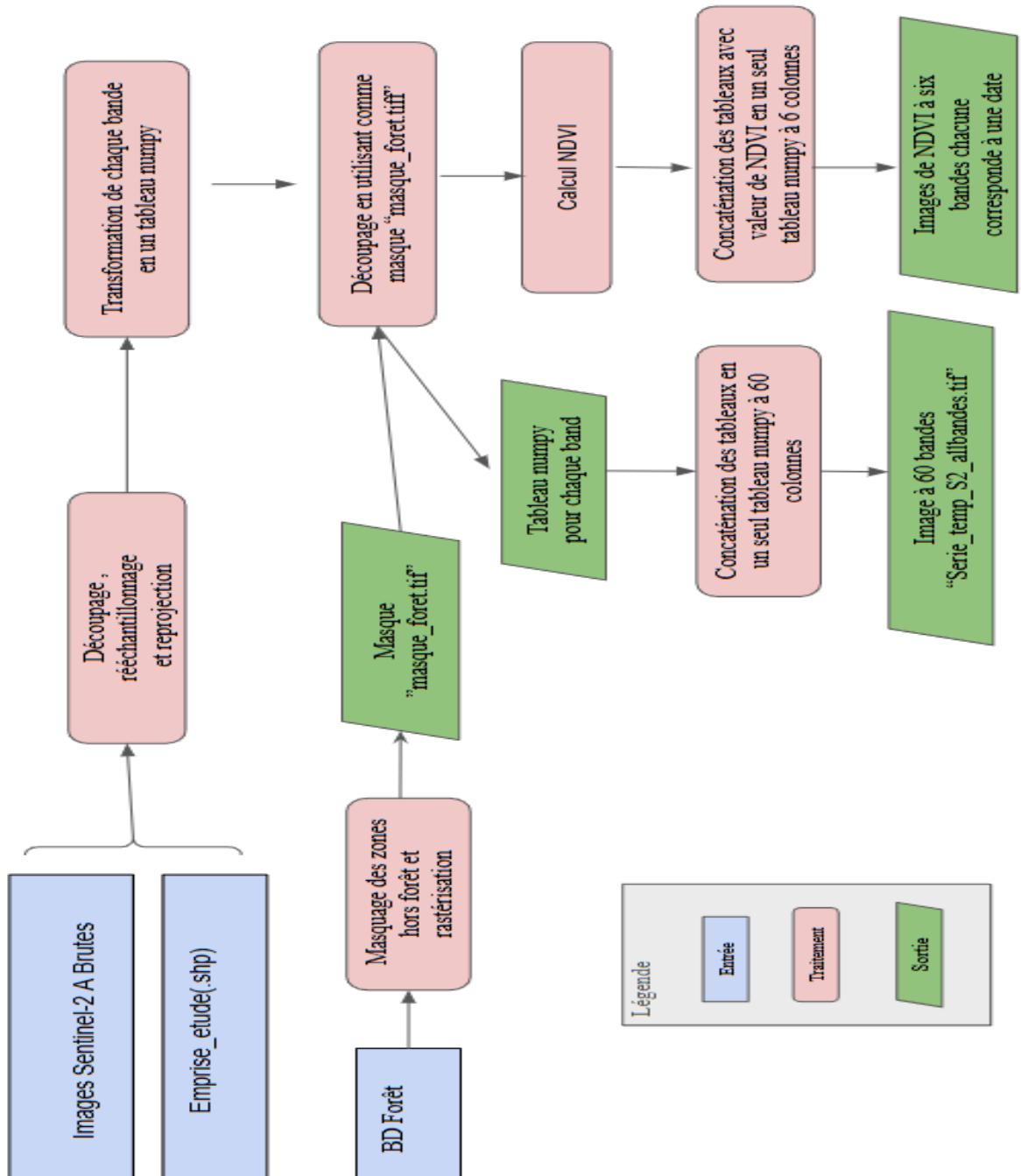
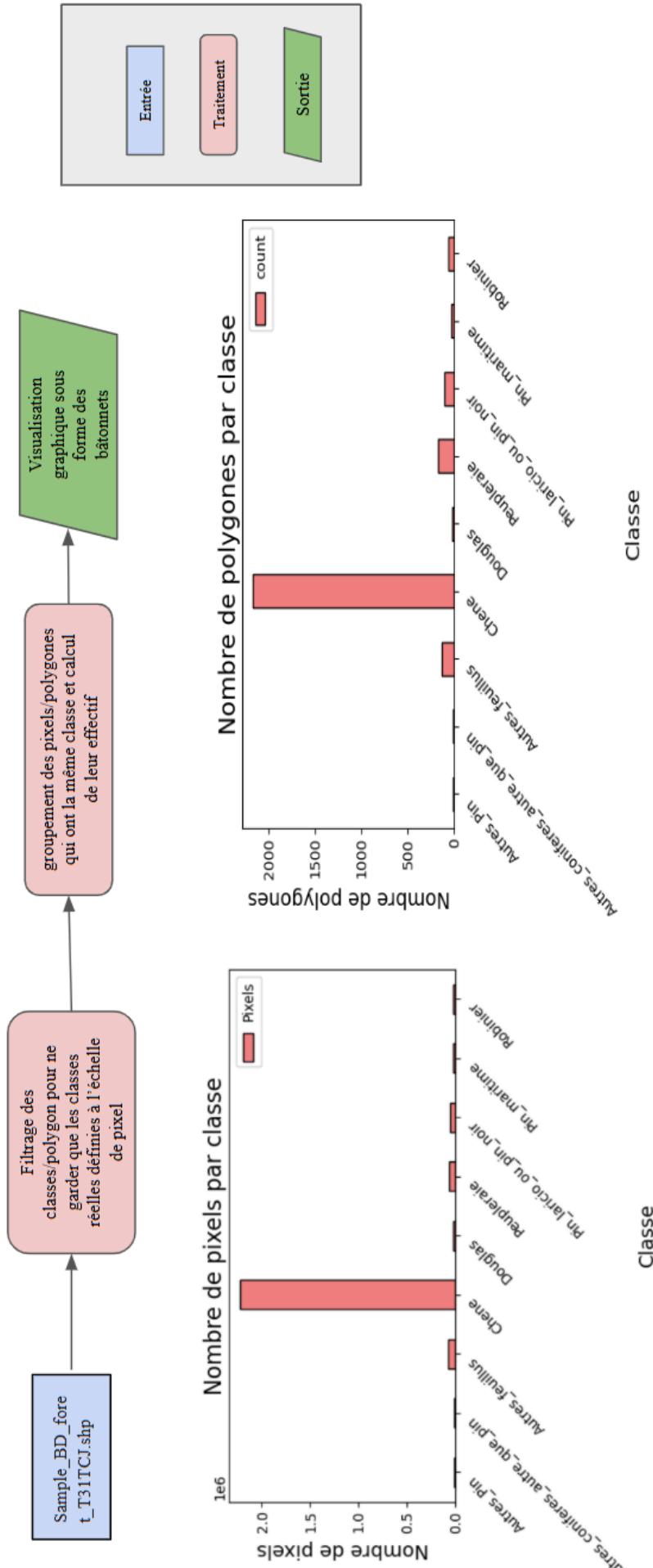


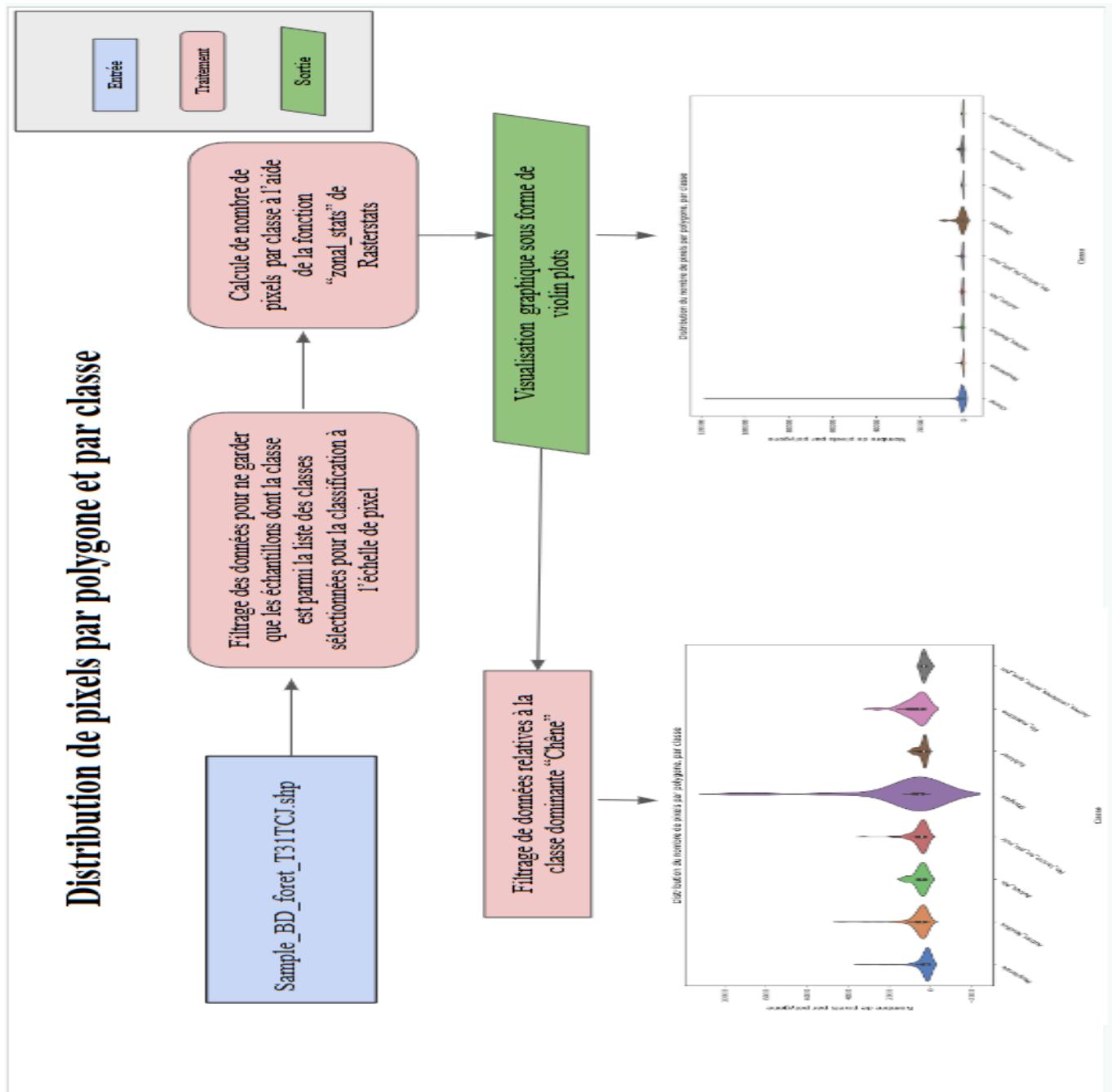
Diagramme de Flux de prétraitement



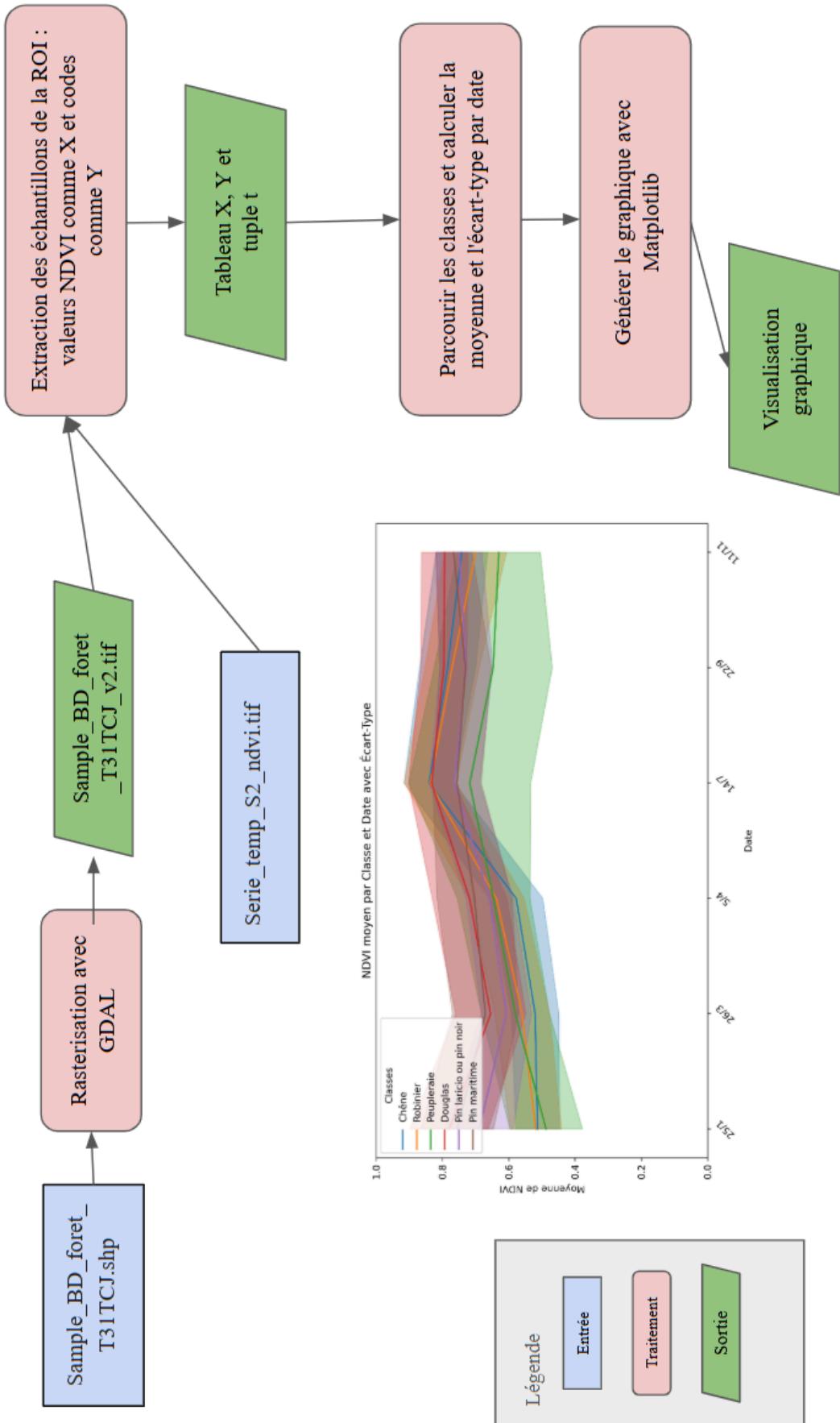
Distribution du nombre de pixels et polygons selon les classes



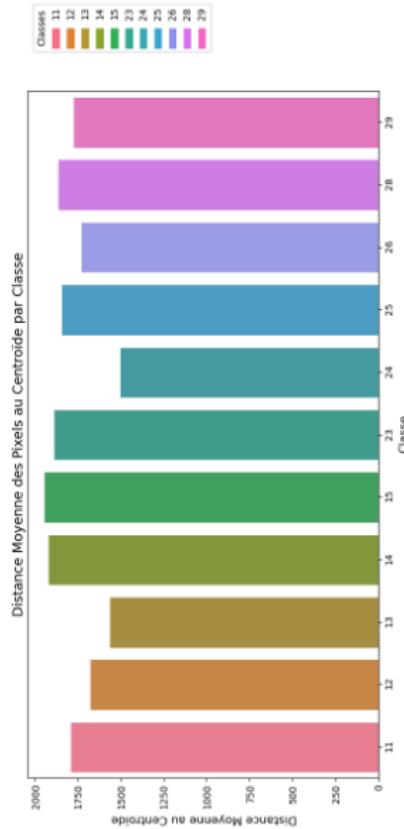
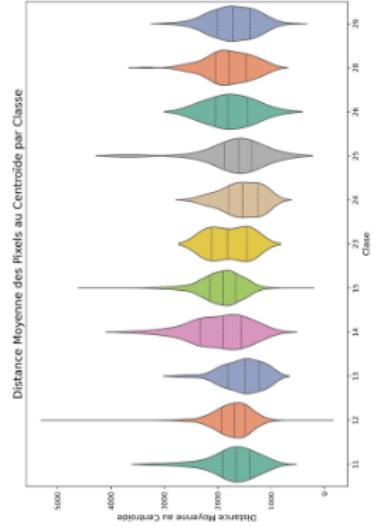
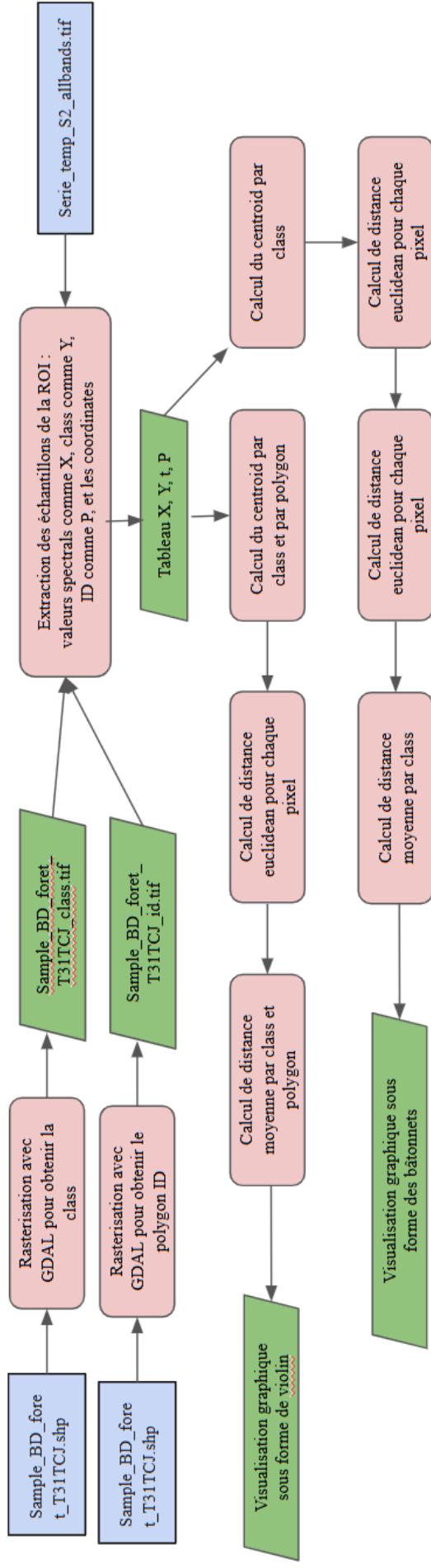
Distribution de pixels par polygone et par classe



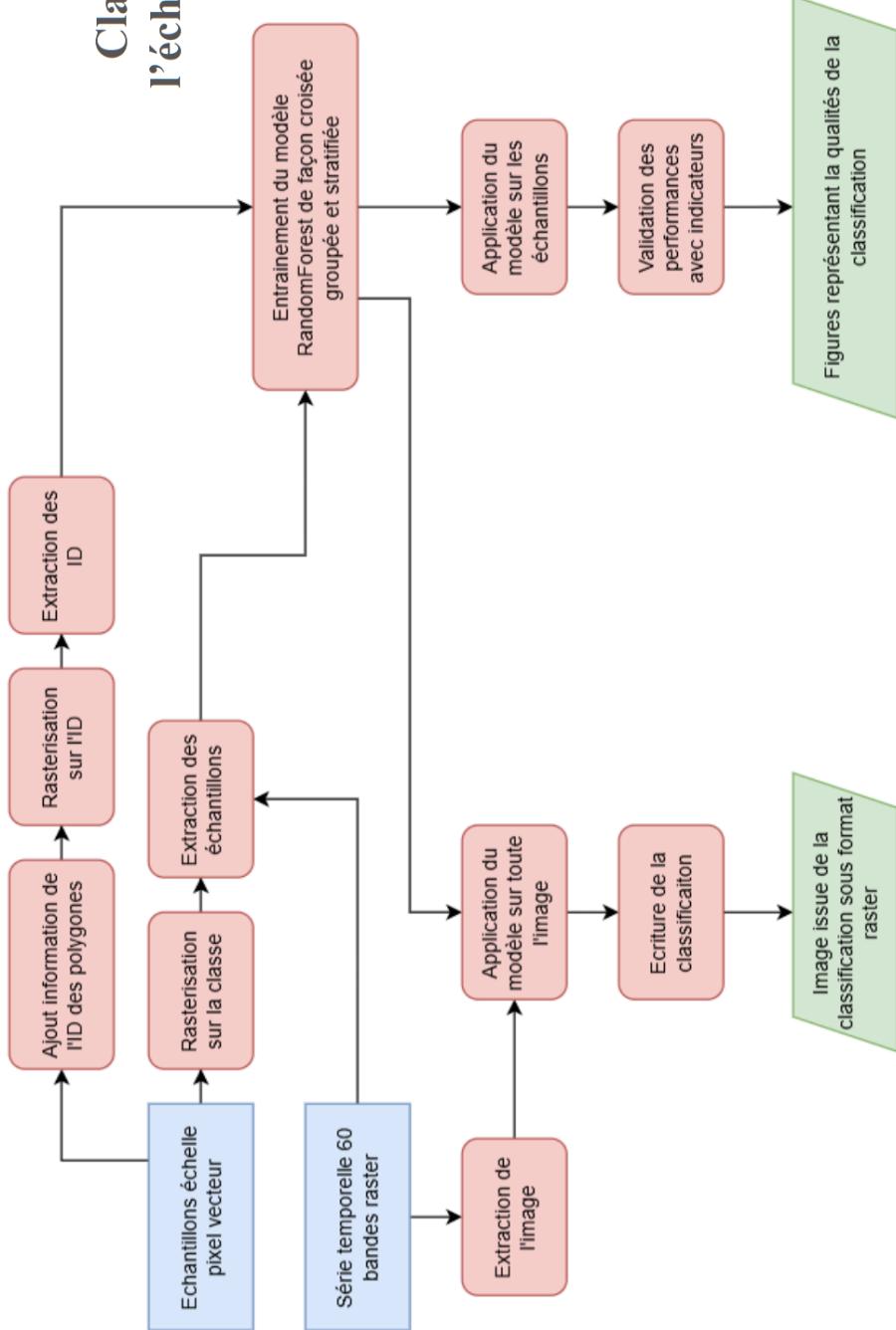
Phénologie des peuplements purs



Analyse de la variabilité spectrale de la BD forêt



Classification à l'échelle des pixels



Classification à l'échelle des peuplements

